Mach Translat (2006) 20:81–138
DOI 10.1007/s10590-006-9013-7

ORIGINAL PAPER

Syntactic mismatches in machine translation

Igor Mel'čuk · Leo Wanner

Received: 13 August 2005 / Accepted: 31 August 2006 / Published online: 14 March 2007

Abstract This paper addresses one of the central problems arising at the transfer stage in machine translation: syntactic mismatches, that is, mismatches between a source-language sentence structure and its equivalent target-language sentence structure. The level at which we assume the transfer to be carried out is the *Deep-Syntactic Structure* (DSyntS) as proposed in the Meaning-Text Theory (MTT). DSyntS is abstract enough to avoid all types of divergences that result either from restricted lexical co-occurrence or from surface-syntactic discrepancies between languages. As for the remaining types of syntactic divergences, all of them occur not only interlinguistically, but also intralinguistically; this means that establishing correspondences between semantically equivalent expressions of the source and target languages that diverge with respect to their syntactic structure is nothing else than paraphrasing. This allows us to adapt the powerful intralinguistic paraphrasing mechanism developed in MTT for purposes of interlinguistic transfer.

Keywords Transfer · Syntactic mismatch · Paraphrasing · Deep-syntactic structure · Meaning-Text Theory

1 Introduction

This paper considers the transfer stage at the level of syntactic structure in sentence-to-sentence machine translation (MT) within the transfer-based paradigm. The

I. Mel'čuk (✉)
Department of Linguistics and Translation, University of Montreal, C.P. 6128 "Centre-Ville", Montreal, QC, Canada, H3C 3J7,
e-mail: igor.melcuk@umontreal.ca

L. Wanner
Institució Catalana de Recerca i Estudis Avançats (ICREA) and Department of Technology, Pompeu Fabra University, Passeig de Circumval·lació 8, Barcelona, 08003 Spain
e-mail: leo.wanner@upf.edu

syntactic transfer stage has to deal with the transfer of linguistic information of three major types: (i) lexical units ("lexical transfer"), (ii) inflectional meanings, or grammemes ("grammemic transfer"), and (iii) syntactic constructions ("syntactic transfer").[1] We focus on one particular aspect of the syntactic transfer: *mismatches* between source and target sentence structures, a topic that has become increasingly popular during the last two decades.

In this introductory section, we will first formulate the problem of syntactic mismatches as it is known today (1.1), then indicate the interlinguistic and intralinguistic nature of syntactic mismatches (1.2), and finally sketch the relevant aspects of our theoretical framework, the Meaning-Text Theory (MTT) (1.3).

1.1 The problem stated

Syntactic mismatches pose a serious problem to MT because they require idiosyncratic transformations between the source and the target structures for each particular pair of languages involved. In order to develop a general mechanism that handles such transformations uniformly and in a systematic way we need to know all possible types of mismatches, presented in a logically derived exhaustive typology.

The major types of syntactic mismatches between different languages have been discussed—under the heading of "translation divergences"—in the influential work of Dorr (1993, 1994). Dorr distinguishes the following types of syntactic mismatches:[2]

1. Mismatches due to syntactic actant permutation, or **conversion** (Dorr: "thematic divergence"). In (1), the English syntactic subject *I* semantically corresponds to the indirect object (IndirO) мне *mne* 'to me' in Russian; and the English direct object (DirO) *picture* corresponds to the Russian subject картина *kartina* 'picture'.

   (1) I like this picture.
   Мне нравится эта картина.
   Mne navritsja èta kartina.
   TO-ME PLEASES THIS PICTURE

2. Mismatches due to dependency inversion, or **head switching** (Dorr: "demotional/promotional divergence"). The dependent adverbial modifier in the first sentence of each pair (*just* in (2), German *gern* 'with-pleasure' in (3)) semantically corresponds to the top node (finite verb) of the second sentence (French *venir* 'come' in (2), *like* in (3)), while the finite verb of the first sentence corresponds to the dependent element in the second.

   (2) I just learned that.
   Je viens de l'apprendre.
   I COME FROM THAT TO-LEARN

   (3) *Ich schwimme gern.*
   I SWIM WITH-PLEASURE
   I like swimming.

[1] In order to restrict our task reasonably, we do not consider the transfer problems related to word formation: derivation and compounding.

[2] To make Dorr's list more comparable with the types of mismatches described in this paper, we use, where it seems appropriate, our own names for types of mismatches, but we cite Dorr's terms in parentheses. We also collapsed two of Dorr's types with two others.

3. Mismatches due to lexeme-phrase substitution, or **lexical fission/fusion** (Dorr: "lexical conflational divergence"). The verbal lexeme in the first sentence of each pair in (4) and (5) corresponds to a verbal phrase in the second sentence.

   (4) I stabbed John.
   Yo le di a John una puñalada.
   I TO-HIM GAVE TO JOHN A STAB

   (5) I like Mary.
   Ich habe Mary gern.
   I HAVE MARY WITH-PLEASURE

4. Mismatches due to **part-of-speech changes** (Dorr: "categorial divergence"). In (6), the same meaning is expressed by an adjective in English and a noun in French.

   (6) I am hungry.
   J'ai faim.
   I HAVE HUNGER

5. Mismatches due to **function-word introduction/elimination** (Dorr: "structural divergence"). The same meaning is expressed by an affix in French and an auxiliary in English (7), or by a verb without a preposition in English and by a verb with a preposition in German (8).

   (7) *Je lirai.*
   I READ+fut+1st-person
   I will read.

   (8) He entered the room.
   Er trat in das Zimmer ein.
   HE STEPPED IN THE ROOM IN

Dorr's typology has served as a starting point for a number of investigations into the problem of structural mismatches in MT (see, e.g., Gawron 1999 ; Dave et al. 2001; Gupta and Chatterjee 2001). We begin with it as well, in order to develop a universal calculus of syntactic mismatches between languages and to propose a method for handling them in a uniform manner.

1.2 The intra and inter-linguistic nature of syntactic mismatches

The basic idea that underlies our work sets it aside from most other approaches in the field. We claim that the phenomenon of syntactic mismatches is as much *inter*linguistic as *intra*linguistic. In other words, semantically equivalent syntactic structures within one language (paraphrases) reveal mismatches of the same kind as those identified between equivalent syntactic structures across languages. Therefore, in MT, a source syntactic structure can be transferred by mapping it onto a mismatching target syntactic structure by the same paraphrasing model as used to map a syntactic structure onto a semantically equivalent, but syntactically mismatching structure in text generation, that is, in an intralinguistic scenario.

As was already pointed out by several researchers (see, e.g., Barnett et al. 1991; Wanner 1996), all types of syntactic mismatches listed above occur not only interlinguistically (between different languages), but also intralinguistically (inside one language) as well. For the above five interlinguistic types of mismatches, it is thus possible to quote parallel examples of paraphrases within one language; cf. (11)–(20) below, taken from German and Russian.

Before we proceed, two remarks are in order:

- Paraphrases that we use are not necessarily fully synonymous; slight semantic divergences are admitted, in the same way as translations do not always preserve absolutely the same meaning as the originals (Doherty 1999; Steiner 2001). Thus, (11)–(14) and (17)–(18) are examples of approximate paraphrases.
- Particular types of syntactic mismatches are by no means equally common in any language or possible with any lexical unit. For instance, function-word introduction/elimination is more typical interlinguistically, while lexical fission/fusion is quite common intralinguistically (at least in the languages we have considered). In German, the part-of-speech mismatch is possible with *Hunger* ‘hunger’/*hungrig* ‘hungry’, while it is impossible with the Russian equivalents голод *golod*/ голодный *golodnyj* (9).

(9) a. *Ich habe Hunger.—Ich bin hungrig.*
I HAVE HUNGER—I AM HUNGRY

b. *У меня голод.—Я голодный.
**U menja golod.—Ja golodnyj.*
TO ME HUNGER—I AM HUNGRY

With the lexical unit meaning ‘chill [illness]’ the situation is inverse (10).

(10) a. *Ich habe Schüttelfrost.* lit. ‘By me [is] a chill.’
**Es frost-schüttelt mich.* lit. ‘It chills me.’

b. У меня озноб. U menja oznob. lit. ‘By me [is] a chill.’
Меня знобит. *Menja znobit.* lit. ‘It chills me.’

Consider now examples of intralinguistic mismatches in (11)–(20).

1. Actant conversion mismatch

(11) a. *Ich mag das Bild.—Mir gefällt das Bild.*
I LIKE THE PICTURE—TO-ME PLEASES THE PICTURE

b. Я люблю эту картину.—Мне нравится эта картина.
Ja ljublju ètu kartinu.—Mne nravitsja èta kartina.
I LIKE THIS PICTURE—TO-ME PLEASES THIS PICTURE

2. Head-switching mismatch

(12) a. *Ich mag Schwimmen.—Ich schwimme gern.*
I LIKE SWIMMING—I SWIM WITH-PLEASURE

b. Я люблю плавать. – Я охотно плаваю.
Ja ljublju plavat’.—Ja oxotno plavaju.
I LIKE TO-SWIM—I WITH-PLEASURE SWIM

3. Lexical fission/fusion mismatch

(13) *Ich schoss auf John.—Ich gab auf John einen Schuss ab.*
I SHOT AT JOHN—I GAVE AT JOHN A SHOT AWAY

(14) Я ударил Джона ножом.—Я нанёс Джону ножевую рану.
Ja udaril Džona nožom.—Ja nanës Džonu noževuju ranu.
I STABBED JOHN WITH-KNIFE—I INFLICTED TO-JOHN KNIFE WOUND
'I stabbed John with a knife.'

4. Part-of-speech mismatch

(15) *Ich bin hungrig.—Ich habe Hunger.*
I AM HUNGRY—I HAVE HUNGER

(16) Меня знобит.—У меня озноб.
Menja znobit.—U menja oznob.
ME [IT]-CHILLS—BY ME [IS A] CHILL
'I have a chill.'

5. Functional word introduction/elimination mismatch[3]

(17) *Er las.—Er hat gelesen.*
HE READ-imperfect—HE HAS READ-pastpart

(18) Я буду собираться завтра.—Я соберусь завтра.
Ja budu sobirat'sja zavtra.—Ja soberus' zavtra.
I WILL PACK TOMORROW—I WILL-PACK TOMORROW

(19) *Er betrat das Zimmer.—Er trat in das Zimmer ein.*
HE ENTERED THE ROOM—HE STEPPED INTO THE ROOM IN

(20) a. Данный суффикс принадлежит множеству словообразовательных средств.
Dannyj suffiks prinadležit množestvu slovoobrazovatel'nyx sredstv.
THIS SUFFIX BELONGS TO-SET OF-DERIVATIONAL MEANS

b. Данный суффикс принадлежит к множеству словообразовательных средств.
Dannyj suffiks prinadležit k množestvu slovoobrazovatel'nyx sredstv.
THIS SUFFIX BELONGS TO SET OF-DERIVATIONAL MEANS
'This suffix belongs to the set of derivational means.'

Therefore, we may conclude that structural mismatches between semantically equivalent expressions of *different* languages constitute a particular case of a more general phenomenon:

[3] Generally speaking, the German imperfect and perfect are not synonymous, just as the Russian imperfective and perfective aspects. However, in certain contexts the sentences in (17) and (18) are nearly synonymous: cf. (i) and (ii).
(i) *Gestern las er die ganze Nacht.*
YESTERDAY READ-imperfect HE THE WHOLE NIGHT
(ii) *Gestern hat er die ganze Nacht gelesen.*
YESTERDAY HAS HE THE WHOLE NIGHT READ-pastpart
both meaning 'Yesterday, he read the whole night'.

> Establishing correspondences between semantically equivalent but structurally (= syntactically) divergent expressions is nothing else than paraphrasing. Therefore, the problem of structural mismatches in MT can be solved by using a general paraphrasing mechanism—both intra and interlinguistically.

MTT offers a general intralinguistic paraphrasing system (Žolkovskij 1967; Mel'čuk 1974, pp 149, 1988b, 1992; Milićević 2003). This paraphrasing system has already been used in MT, among others, by Sanromán Vilas et al. (1999) and Apresjan et al. (in press) intralinguistically at the source-language side to adjust source-language structures to target-language structures. We adapt this system in our approach to the interlinguistic resolution of source and target language structure mismatches.

1.3 The theoretical framework

The principle that underlies our work is as follows:

> Given the complexity of the task of MT, the stage of transfer must be alleviated as much as possible. Phenomena that are intralinguistic in nature should be treated at the source side (= in the analysis), or the target side (= in the synthesis), rather than in transfer.
> As a result, many presumed problems of transfer that concern syntactic divergences are relegated either to the analysis stage or to the synthesis stage; the transfer stage has to deal only with phenomena that essentially involve both source language $\mathcal{L}_S$ and target language $\mathcal{L}_T$, dealing with them at the deep-syntax level.

In our approach, the level at which the transfer is carried out is the deep-syntactic structure (DSyntS), as proposed in MTT. DSyntS is abstract enough to avoid all types of lexical and syntactic divergences that result from performing the transfer at the surface-syntactic level (where restricted lexical co-occurrence and language-specific constructions are represented). Therefore, as has already been pointed out for example by Sanromán Vilas et al. (1999) and Han et al. (2000), DSyntS is well suited to MT.[4]

In accordance with the stratificational character of an MTT-model, our transfer schema looks as shown in Fig. 1 (the subscript S stands for source language, and the subscript T for target language; these subscripts are used for all linguistic elements under analysis).

In (22) and Figs. 2, 3 we characterize the three types of structures implied in this schema, illustrating them for the Russian sentence (21) (already seen in (14)).

(21) Я нанёс Джону ножевую рану.
Ja nanës Džonu noževuju ranu.
I INFLICTED TO-JOHN KNIFE WOUND
'I stabbed John with a knife.'

The DMorphS of a sentence is a chain of DMorph-representations of its wordforms. The DMorph representation of a word form consists of the name of the corresponding lexeme and all necessary inflectional characteristics, "grammemes" (such as case and

[4] Actually, a syntactic representation sufficiently close to the DSyntS was introduced in MT earlier by Apresjan et al. (1989, 1992): it eliminates structural words, identifies values of lexical functions, reduces the idioms to one node, introduces explicitly the semantic grammemes, etc.

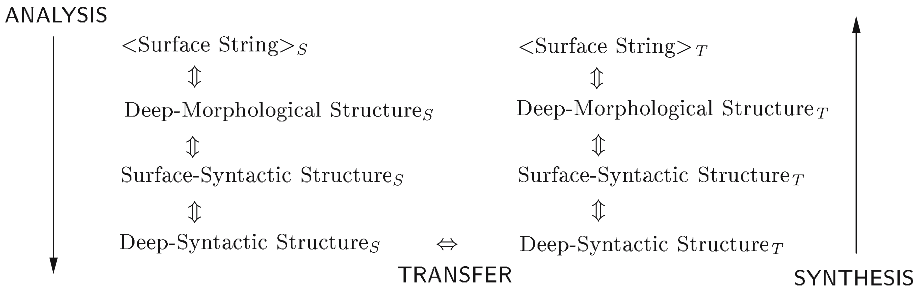
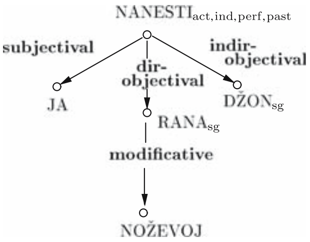

ANALYSIS
<Surface String>S
Deep-Morphological StructureS
Surface-Syntactic StructureS
Deep-Syntactic StructureS
TRANSFER
<Surface String>T
Deep-Morphological StructureT
Surface-Syntactic StructureT
Deep-Syntactic StructureT
SYNTHESIS

Fig. 1 General schema of transfer at the deep-syntactic level

Fig. 2 SSyntS for sentence (21)



number for nouns or voice, mode, aspect, tense, person and number for verbs).[5] The DMorphS of the sentence (21) appears as (22).[6]

(22) $\text{JA}_{\text{nom}} \prec \text{NANESTI}_{\text{act,ind,perf,past,masc,sg}} \prec \text{DŽON}_{\text{sg,dat}} \prec \text{NOŽEVAJA}_{\text{fem,sg,acc}} \prec \text{RANA}_{\text{sg,acc}}$

Inflectional characteristics can be absent since, as is well known, they are not present in some languages at all (for example, Vietnamese and Chinese which do not have inflectional morphology), and in other languages, many words do not inflect.

A surface-syntactic structure (SSyntS) of a sentence *S* is an unordered dependency tree whose nodes are labeled with the names of the lexemes of *S* (supplied, where necessary, with semantic grammemes), and whose arcs are labeled with names of surface-syntactic relations. The lexemes of *S* and the nodes of its SSyntS are in a one-to-one correspondence. A semantic grammeme represents a meaning; thus, the set of semantic grammemes includes number for nouns and aspect, mood and tense for verbs.[7] The set of surface-syntactic relations, which are language-specific includes such relations as subject(ival), dir(ect)-object(ival), aux(iliary), circumstantial, etc.; it is a superset of the relations used in Lexical Functional Grammar (LFG) (Bresnan 1982) in f-structure. The SSyntS for (21) is presented in Fig. 2.

Just like an SSyntS, a DSyntS is an unordered dependency tree, with however a different labeling of nodes and arcs. A detailed description of DSyntS is given in Sect. 3; here we limit ourselves to an approximate characterization. The nodes of the

[5] In our examples, we use, for the sake of a more compact presentation, subscripts instead of feature–value pairs for the representation of grammemes.

[6] The symbol "≺" indicates strict linear ordering. Note that in order not to clutter up the structures, here and henceforth, Russian lexemes in structures are given only in transcription, not in cyrillic.

[7] Syntactic grammemes, imposed by government and agreement (such as case for nouns or person and number for verbs) are not represented in SSyntS; they appear only in the DMorphS.

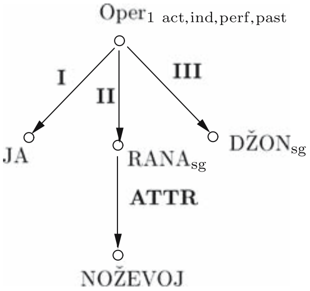


Fig. 3 DSyntS for sentence (21)

DSyntS of a sentence *S* are labeled, roughly speaking, also with the lexemes of *S*, but not in a one-to-one correspondence: lexical labels in a DSyntS are "deep representations" of the actual lexemes of *S*, or Deep LUs. They carry the same semantic grammemes as nodes of SSyntS. The arcs of a DSyntS are labeled with names of deep-syntactic relations, which are language universal and represent a generalization of surface-syntactic relations. The DSyntS for (21) is given in Fig. 3. For a detailed presentation of deep LUs and DSynt-Relations see Sect. 3.1.

The remainder of the article is organized as follows. In Sect. 2, we present the essence of our proposal ("Transfer as paraphrasing") and sketch the two main devices used by it: a bilingual lexical index and transfer paraphrasing rules. Section 3 introduces in more detail the level at which the transfer is carried out: the DSyntS. Section 4 discusses the cases of pseudo-mismatches: those that pose serious obstacles to some approaches but disappear at the level of DSyntS. Section 5 gives a formal definition of the structural mismatch, of which the syntactic mismatch is a subtype, and presents a (universal) typology of mismatches. Section 6 is the main section of the paper. After introducing MTT's paraphrasing system, it discusses the problem of how it carries over to MT. More precisely, it shows how the MTT-style paraphrasing rules can be composed of elementary transformations and how these rules can be applied in an MT scenario. In Sect. 7, we summarize the main ideas presented in the paper and draw conclusions for future work. An Appendix contains the proof of the theorem limiting the number and types of possible types of structural mismatches.

2 The proposal: transfer as paraphrasing

Our goal in this paper is twofold:

(i) to present a description of all logically possible types of syntactic mismatches;
(ii) to define and illustrate the structure of universal (paraphrasing) transfer rules necessary and sufficient for mapping between any two deep-syntactic structures that show at least one mismatch.

The nature of these rules presupposes a specific architecture of the transfer engine capable of dealing with syntactic mismatches (the Syntactic Transfer Engine, STE). A serious discussion of an STE is beyond the scope of the present paper, so we limit ourselves to a cursory characterization.

In the transition from DSyntS$_S$ to DSyntS$_T$, the STE takes care of the syntactic part of the transition, that is, of the mapping between syntactic structures, but without dealing with grammemes attached to their nodes. In other words, we leave out the problem of grammemic transfer.[8]

In order to carry out the mapping between two syntactic structures, the STE must include the following three major components:

1. A set of formalized monolingual "explanatory combinatorial dictionaries" (ECDs) for the languages involved. These lexica are independent of the pair of languages considered and direction-neutral, in the sense that each can be used either as a source- or as a target-language lexicon. They contain, among other things, the lexical co-occurrence information of the language in question specified in terms of lexical functions (LFs) (cf. Sect. 3.1.1 and Mel'čuk 1996 for a detailed introduction to LFs).
2. A set of bilingual lexical correspondence indexes for the language pairs involved. A bilingual lexical index (BLI) is specific to each pair of languages involved; it is a direction-neutral list of pairs of translationally equivalent LUs of $\mathcal{L}_S$ and $\mathcal{L}_T$.
3. A set of transfer paraphrasing rules that carry out the mapping between semantically equivalent syntactic structures of any $\mathcal{L}_S$ and $\mathcal{L}_T$.

As for the ECD, it has been described in a series of publications (Mel'čuk and Polguère 1987; Mel'čuk 1993). ECD-style dictionaries (although not yet implemented for use in NLP applications) exist for Russian (Mel'čuk and Zholkovsky 1984), French (Mel'čuk et al. 1984, 1988, 1992, 1999; Mel'čuk and Polguère, forthcoming) and Spanish (Grupo DiCE n.d.). We will presuppose that the reader has a sufficient familiarity with it. Let us briefly introduce the components 2 and 3.

2.1 Bilingual lexical index

While an ECD purports to cover the complete vocabulary of a language, i.e., to include all of its LUs, a BLI is by principle limited to deep LUs of the language pair involved (cf. Sect. 3.1.1). Thus, the English part of an English-to-$\mathcal{L}_{T_i}$ BLI does not contain PAY as in *pay attention*, but only ATTENTION: *pay* is one of the elements of the value of the LF Oper_1 of ATTENTION. Neither does it contain LAUNCH nor ATTACK_N as in *launch an attack*, but only ATTACK_V: $attack_N$ is an element of the value of the LF S_0 applied to $attack_V$, and *launch* is an element of the value of the LF $\text{Oper}_1(attack_N)$. And there is neither HEAVY nor RAIN_N as in *heavy rain*, but only RAIN_V.

As a result, a BLI includes many fewer LUs than the full vocabulary of the languages involved, which makes it easier to compile and maintain. By definition, all lexical elements admitted into a BLI are LUs in the strict sense, that is, monosemantemic (the resolution of polysemy must be done during analysis).

Lexical equivalences stored in a BLI fall into two major classes: "regular" equivalences, which do not entail structural mismatches and can be expressed in terms of LU-pairs; and "irregular" equivalences, which entail structural mismatches and require the specification of the transformation necessary to resolve these mismatches.

With respect to regular lexical equivalences, again two cases must be distinguished:

[8] The transfer of lexemes and grammemes in case of multiple correspondences is performed by a lexical or morphological transfer engine respectively; for multiple lexical transfer see Mel'čuk and Wanner (2001).

1. L_S has at least one semantically fully matching translation equivalent L_T. In this case, L_S receives only this L_T as its translation equivalent; all exact and more specific synonyms of L_T do not appear in the BLI, but are supplied in the monolingual $\mathcal{L}_T$ ECD and are selected during synthesis. This is the predominant case; formally, a translation equivalence is presented as a pair (L_S, L_T). For instance, in an English–French BLI, we find entries such as those in (23), which show two equivalents for *chair*, namely *chaise* 'furniture for sitting' and *chaire* 'professorship'.

   (23) (DEEP, PROFOND)
   (CHAIR**1**, CHAISE)
   (CHAIR**2**, CHAIRE)

   Note that even if we write the equivalences in the direction from $\mathcal{L}_S$ to $\mathcal{L}_T$, logically speaking, the BLI is non-directional, so that any of the two languages involved can be taken either as $\mathcal{L}_S$ or as $\mathcal{L}_T$.
2. L_S does not have a fully matching translation equivalent, but one or several semantically not exactly matching translation equivalents, which are intersecting synonyms of each other. In this case, L_S receives all of them as its translation equivalents. The selection between the $\mathcal{L}_T$-synonyms poses a serious problem for lexical transfer (see Iordanskaja and Mel'čuk 1997; Mel'čuk and Wanner 2001). However, it does not interfere with the resolution of structural mismatches. Formally, an entry of a BLI for this case has the form ($L_S, L_{T_1}, \ldots, L_{T_n}$). For instance, in a German–Russian BLI, we find entries such as those in (24), where *Ansprache* 'address' has multiple equivalents обращение *obraščenie* 'address', призыв *prizyv* 'appeal'/'call' and воззвание *vozzvanie* 'emotional appeal to do something difficult and/or dangerous'.

   (24) (ANSPRACHE, OBRAŠČENIE,
   PRIZYV,
   VOZZVANIE)

As far as irregular lexical equivalences are concerned, each translation equivalence is of the form (L_S, L_T, Ψ), where Ψ is an LF such that, when applied to L_S, it returns L_T: $\Psi(L_S) = L_T$. In this way, Ψ unambiguously specifies the type of the mismatch provoked by translating L_S as L_T and the operation needed for its resolution. Consider the examples in (25). (Anti, Conv$_{21}$, and ^{II}Adv$_1$ are LFs whose values are found in the corresponding monolingual ECDs.)

(25) a. (SHALLOW, PROFOND, Anti)
b. (LIKE, PLAIRE, Conv$_{21}$)
c. (SOLER, HABITUELLEMENT, ^{II}Adv$_1$)

The example (25a) means that in order to translate *shallow* into French, we need to use PROFOND 'deep', which is an interlinguistic antonym of SHALLOW: French has no adjective meaning 'shallow'. At the same time, Ψ = Anti triggers a lexical equivalence rule (see below) that adds a negative expression to PROFOND, producing *peu profond*, which is the standard translation of *shallow*. This lexical equivalence can also be considered in the opposite direction: to translate *profond* into English, PROFOND is specified as Anti(SHALLOW), therefore, PROFOND $\equiv$ Anti(SHALLOW) = *deep*, which is the correct translation.[9] Along the same lines, *peu profond* is Anti(PROFOND) ≡ Anti(Anti(SHALLOW)) = *shallow*.

The example (25b) means that when translating LIKE as PLAIRE ‘to please’, which is semantically perfect, we need to make a 2-to-1 syntactic conversion (26).

(26) Julie likes Paul. ≡ *Paul plaît à Julie*.

In (25c) the Spanish verb SOLER (roughly ‘to usually do’) is translated by the French adverb HABITUELLEMENT ‘usually’, which is an element of the value of the LF ${}^{II}\mathrm{Adv}_1$ (given by Ψ). Ψ calls in the corresponding lexical equivalence rule, which, in its turn, ensures the structural transformations that lead to (27). For the reverse passage (from French to Spanish) the same equivalence rules are applied from right to left.

(27) *Maria suele leer.* ≡ *Maria lit habituellement.* ‘Maria usually reads.’

In addition to purely lexical equivalences presented above, languages feature translation equivalences between an LU and a grammeme/derivative, such as English WANT ≡ Japanese -TAI, a verb suffix ‘desiderative’ (see (69) below). The BLI sketched here does not consider such equivalences.

2.2 Transfer paraphrasing rules

Three different sets of transfer paraphrasing rules are distinguished:

1. A set of elementary lexical equivalences expressed in terms of LFs; they are universal in the sense that they do not depend on the pair of languages considered. These equivalences are needed because syntactic mismatches essentially depend on lexical co-occurrence behavior of translation equivalents: specific lexical units admit or reject specific syntactic constructions. Each lexical equivalence is supplied with references to elementary syntactic operations that perform syntactic transformations of the target DSyntS required by the target LU, see immediately below; these operations form elementary paraphrasing rules.
2. A set of elementary syntactic operations that carry out the transformation of the target deep-syntactic tree S_{DSynt_T} that is triggered by the application of a lexical equivalence and that is needed to ensure the well-formedness of S_{DSynt_T} and its semantic equivalence to the source deep-syntactic tree S_{DSynt_S}.[10]
3. A set of syntactic adjustment operations that take care of the context in which the paraphrastic substitution occurs: they reattach the incoming and outgoing branches that remained dangling as a result of the lexical paraphrastic substitution. Together with the elementary paraphrasing rules, they constitute the set of transfer rules proper.

[9] Obviously, PROFOND ≡ DEEP.

[10] In point of fact, there exist transformations of S_{DSynt_T} that are not lexically triggered, that is, purely syntactic transformations; consider, for instance, the difference between a participial phrase in (iii) and a relative clause in (iv).

(iii) Мальчик, уже прочитавший книгу, ушел.
Mal’čik, uže pročitavšij knigu, ušel.
BOY, ALREADY HAVING-READ BOOK, LEFT

(iv) The boy who had already read the book left.

Our model accounts for this type of transformation as well. However, in order not to make the presentation even heavier we do not consider them in this paper.

3 The level of transfer: deep-syntactic structure

A well-chosen and well-defined formalism for the representation of syntactic structure to be used at the stage of transfer is crucial for a successful treatment of structural mismatches. This structure must allow for relegating all intralinguistically treatable phenomena to analysis or synthesis, and thus admitting into transfer only phenomena that are essential for translation. Such a structure is the DSyntS of MTT (Mel'čuk 1974, 1988a, pp 59–66, 2004a).

3.1 The notion of Deep-Syntactic Structure

We begin with the formal definition of DSyntS.

Definition 1 (DSyntS). Let L_d, G_{sem} and R_{dsynt} be three disjunct alphabets, where L_d is the set of deep lexical units (LUs) of $\mathcal{L}$, G_{sem} is the set of semantic grammemes, and R_{dsynt} is the set of names of deep-syntactic relations.

A DSyntS of $\mathcal{L}$, S_{DSynt}, is a quintuple over $L_d \cup G_{sem} \cup R_{dsynt}$ of the form

$$S_{DSynt} = \langle N, A, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g} \rangle,$$

where the set N of nodes and the set A of directed arcs (or branches) form a dependency tree (with a source node n^s and a target node n^t defined for each arc), $\lambda_{l_s \to n}$ is a function that assigns to each $n \in N$ an $l_s \in L_d$, $\rho_{r_s \to a}$ is a function that assigns to each $a \in A$ an $r_s \in R_{dsynt}$, and $\gamma_{n \to g}$ is a function that assigns to the name of each LU associated with a node $n_i \in N$, $l_i \in \lambda_{n \to g}(N)$, a set of corresponding grammemes $G_t \in G_{sem}$.

According to Definition 1, a DSyntS is defined over three alphabets of a natural language $\mathcal{L}$: L_d (the set of deep LUs), G_{sem} (the set of semantic grammemes) and R_{DSynt} (the set of deep-syntactic relation names). Let us discuss L_d, G_{sem} and R_{DSynt} in turn.

3.1.1 L_d: Deep lexical units

The set of deep LUs of $\mathcal{L}$ contains all LUs (i.e., lexemes and idioms) of $\mathcal{L}$, with the following additions and eliminations. Two types of artificial LUs are added to the stock of deep LUs of $\mathcal{L}$: (i) symbols of LFs (see immediately below); (ii) fictitious lexemes, which represent idiosyncratic syntactic constructions of $\mathcal{L}$ (cf. Sect. 3.2.1). Three types of LUs of $\mathcal{L}$ are excluded from the stock of $\mathcal{L}$'s deep LUs: (i) "structural words", which include, on the one hand, analytical realizations of grammemes (such as articles and auxiliaries), and on the other hand, governed prepositions and conjunctions; (ii) substitute pronouns, that is, pronouns of the third person, which replace nouns (*he*, *she*, *it*, *they*, *these*); (iii) values of LFs.

The first two types of emiminations do not need special explanations; the third, however, calls for a few comments, although LFs have already been introduced in great detail in a series of publications (Mel'čuk 1974, 1995a, 1996). LFs are a formal means to encode lexico-semantic derivation and restricted lexical co-occurrence relations. More specifically, an LF is a function that associates with an LU L (argument, or "keyword", of the LF) a set of lexical units (the "value" of the LF) that are approximately synonymous to each other (28).

Table 1 Examples of paradigmatic LFs

LF	keyword (argument)	value
S_0	discover accept offer despise steal	discovery acceptance [an] offer contempt theft
S_1	smoke walk beauty steal sue	smoker walker [a] beauty thief plaintiff
A_1	silence importance duty look [for] win	in [silence], //silent of [importance], //important obliged in search [of] victorious
Conv_{21}	follow send husband higher behind	precede receive wife lower in front of

(28) $\mathbf{f}\colon L_d \to \mathbb{P}^L$ or $\mathbf{f}(L_k) = \{L_{v_1}, L_{v_2}, \ldots, L_{v_n}\}$

An element *e* of an LF's value expresses a particular meaning such that the choice of *e* depends on the keyword L_k and *e* appears in a particular deep-syntactic role with respect to L_k. Note that, as a conventient abbreviation, a less accurate notation is currently used (29).

(29) $\mathbf{f}(L_k) = L_{v_1}, L_{v_2}, \ldots, L_{v_n}$

LFs fall into two major types: paradigmatic LFs and syntagmatic LFs. Paradigmatic LFs represent lexico-semantic derivations such as the name of the action, state, process, etc.: a deverbal noun (S_0); the name of the *i*th actant of an action, state, process, etc. (S_i); the name of the characteristic property of the *i*th actant of an action, state, process, etc. (A_i); the name of a conversive of *L* (Conv_{ij}). Some examples of each type of derivation are given in Table 1.

Syntagmatic LFs represent collocations, among them: an intensifying modifier (Magn); a light verb that takes the *i*th DSynt-actant of its keyword *L* as Subject and *L* itself as DirO (Oper_i); a verb denoting the realization of a "requirement" contained in the definition of *L* by the referent of *i*th DSynt-actant of *L* (Real_i); a verb denoting a typical sound produced by the referent of its keyword *L* (Son). Examples are shown in Table 2.

The value of a syntagmatic LF is in most cases expressed along with *L* as a syntactic dependent or the governor of *L*. But in some cases, an element of the value of **f**(*L*) expresses the meaning of the LF **f** together with the meaning of *L*, as a consequence, replacing the latter in the sentence. Such an element is called "fused" and is indicated by the symbol '//', as for example with *heavy rain* versus *downpour* (30) or *to spread butter* versus *to butter* (31). The notion of fused element of an LF value is also relevant for the paradigmatic LFs (see examples under A_1 in Table 1).

Table 2 Examples of syntagmatic LFs

LF	keyword	value
Magn	rain	heavy, hard
	accent	heavy, thick
	alike	as two peas in a pot
	basis	firm, solid
	attention	assiduous, close
Oper$_1$	action	take
	cartwheel	turn
	cough	give
	danger	pose
	activity	engage [in]
Real$_1$	brake	slam [on]
	treaty	comply [with]
	duty	discharge
	promise	keep
	obligation	meet
Son	door	creaks
	brake	screeches
	insect	buzzes, hums
	machine gun	chatters
	thunder	claps, rolls

(30) *heavy* = Magn(RAIN)
downpour = //Magn(RAIN)

(31) *spread* = PreparReal$_1$(BUTTER)
butter = //PreparLabreal$_{12}$(BUTTER)

In the DSyntS, an LU L_1 that is an element of the value of a syntagmatic LF **f** of the keyword L_2 is replaced with the functional notation **f**(L_2) as in (32).

(32) *heavy rain* = Magn(RAIN)←**ATTR**-*rain*

An LU L_1 that is a fused element of the value of an **f**, L_1 = //**f**(L_2), is represented in the DSyntS by the symbol of the corresponding LF. Thus, German *Platzregen* 'heavy rain' is written in a DSyntS as //Magn(REGEN), and Russian завтракать *zavtrakat'* 'have breakfast' appears as //Real$_1$(ZAVTRAK)[11] For each LF, a rule establishes the equivalence between its fused and non-fused value elements (33).

(33) //Magn(L) ≡ Magn←**ATTR**-L, //Real$_1$(L) ≡ Real$_1$(L)-**II** → L, etc.

As a result, a fused element of an LF-value in a DSyntS can be always rewritten as a non-fused variety, which may be required for paraphrasing.

As far as paradigmatic LFs are concerned, a lexeme L_1 that represents an element of the value of a paradigmatic LF **f** of the keyword L_2 is replaced in the DSyntS by the functional notation only if the following three conditions simultaneously apply:

[11] The LUs are *Regen* 'rain' and завтрак *zavtrak* 'breakfast'.

- **f** does not stand for a synonym, an antonym or a conversive.[12]
- L_1 is derived from L_2 such that the meaning of L_1 is the exact composition of the meanings of L_2 and **f** (the corresponding information is supplied by the lexicon). In other words, if the semantic difference between L_1 and L_2 is not fully expressed by **f** (i.e., there are some semantic "additions"), L_1 is not replaced by $\mathbf{f}(L_2)$. Thus, German *erhältlich* 'obtainable', which is fully covered by Able$_2$(ERHALTEN), appears in the DSyntS as Able$_2$(ERHALTEN), where *erhalten* is the verb 'obtain'. On the contrary, German GESCHWISTER ≈ 'siblings' is, roughly speaking, Mult('sibling'), but there is no exact equivalence: GESCHWISTER is not simply the set of all siblings, it is the set of all siblings who are children of the same parents or who are siblings of the same person; cf. *die Geschwister Mayer* lit. 'the Mayer siblings' vs. *meine Geschwister* 'my siblings'. As a result, the LU GESCHWISTER has to appear in the DSyntS$_S$ as such (rather than to be replaced by Mult(BRUDER oder SCHWESTER) 'brother or sister'). On the other hand, Russian куча [вопросов] *kuča [voprosov]* 'heap [of questions]' can be encoded in the DSyntS as Mult(VOPROS), with LF вопрос *vopros* 'question'. The information on the possibility of such a replacement is supplied in the lexical entry for the corresponding keyword.
- L_1 is derived from L_2 such that the meanings of L_1 and L_2 are equal and L_2 is semantically basic with respect to L_1.[13] Thus, for example, the verb *attack*, which is V$_0$(ATTACK$_N$), does not appear in the DSyntS as such: the noun *attack* is not semantically basic for the verb *attack*. Rather the inverse is true: the verb is basic for the noun and therefore the noun *attack* appears in the DSyntS as S$_0$(ATTACK$_V$).

LFs that are applicable to a specific LU are listed in the entry for this LU in the Explanatory Combinatorial Dictionary of the corresponding language.

3.1.2 G_{sem}: deep grammemes

Deep (or semantic) grammemes of LUs are directly linked to meaning. Such are grammemes of number and determination for nouns, and of tense, aspect, mood, and voice for verbs. In contrast, nominal case, adjectival number, gender and case and verbal person and number are syntactic grammemes and therefore do not appear in the DSyntS, nor in SSyntS: they are determined by government and agreement and are introduced closer to the surface (in the deep-morphological structure of the sentence).

3.1.3 R_{DSynt}: deep-syntactic relations

R_{DSynt} includes nine universal dependency relations:[14]

[12] As a matter of fact, this statement is approximate: certain types of antonyms and conversives *should* be given in the DSyntS in LF-notation. This concerns, in the first place, those antonyms and conversives that are derived morphologically, i.e., following a regular pattern. Such antonyms are, for instance, reversives of Bantu languages: Swahili *fung*(*-a*) 'fasten' vs. *fung+u*(*-a*) 'unfasten', *kunj*(*-a*) 'fold' vs. *kunj+u*(*-a*) 'unfold', *tat*(*-a*) 'tangle' - *tat+u*(*-a*) 'untangle', *fumb*(*-a*) 'close' vs. *fumb+u*(*-a*) 'open'. Corresponding English examples include verbal pairs of the type *delete* vs. *undelete*, *button* vs. *unbutton*, *bolden* vs. *unbolden*. Conversives that must be encoded in the DSyntS by the symbol of the corresponding LF include passive verbal forms.

[13] Semantic basicness is indicated in the lexicon. In most cases, a verb is semantically basic with respect to an equisignificant noun, and an adjective with respect to an equisignificant adverb.

[14] For a detailed presentation of the dependency relations in DSyntS, see Mel'čuk (2004a, b).

(i) six actantial DSyntRels (**I**, **II**, …, **VI**), which represent the relations between a predicate LU and the syntactic implementations of its arguments, i.e., its deep-syntactic actants (DSyntAs), which are, roughly speaking, generalizations of syntactic complements in the broad sense of the term (subject, direct object, indirect object, etc.);
(ii) attributive DSyntRel (ATTR), which represents the relation between a modified LU and its modifier;
(iii) coordinative DSyntRel (COORD), which represents the relation between the coordinated elements;
(iv) appenditive DSyntRel (APPEND), which represents the relation between the top node of a clause and any of its "extrastructural" elements (an address, a prolepsis, an interjection, a sentence adverb, or a parenthetical).

3.2 Description and examples of DSyntS

Let us make explicit the characteristics of DSyntS and give a few sample structures.

3.2.1 Description of DSyntS

The definition of the DSyntS (Definition 1, p 12) implies the following.

- An analytical form is represented by one node with corresponding grammemes; thus, *have been paid* appears as PAY$_{\text{ind,pass,pres,perf,non-progr}}$.
- All governed prepositions and conjunctions are omitted as in (34).

(34) insist on departure: INSIST-**II** → DEPARTURE
quarrel between friends: QUARREL-**I** → FRIENDS
know that she is sleeping: KNOW-**II** → SLEEP$_{\text{ind,act,pres,non-perf,progr}}$-**I** → SHE

- All substitute (3rd person) pronouns are replaced by their antecedents, for example (35a) appears as in (35b).

(35) a. Taking the book, John put it on the table.
b. Taking the book, John put THE BOOK on the table.

- An idiom is represented by one node, as exemplified in (36).

(36) a. John got his second wind: JOHN←**I**-⌜GET-ONE'S-SECOND-WIND⌝
b. John barks up the wrong tree: JOHN←**I**-⌜BARK-UP-THE-WRONG-TREE⌝

Note that such expressions as *of course*, *with respect to*, *by the way*, *give up*, *passer-by* are considered to be idioms and thus also represented each by a single node (cf. Sect. 4.1.3).
- Generally speaking, an LU L_1 that is an element of the value of an LF of the LU L_2 is replaced with the name of the LF; thus, *heavy rain* appears as MAGN ← ATTR-RAIN, *turn* [*a*] *cartwheel* as Oper$_{\mathit{1}}$ **II** → CARTWHEEL, and *keep* [*one's*] *promise* as Real$_1$-**II** → PROMISE.

These five conventions mean that a number of lexical elements that appear in the sentence are not represented in its DSyntS. On the other hand, the DSyntS contains four types of lexical elements that do not appear on the surface.

- Zero LUs and zero forms of non-zero LUs, that is, linguistic signs that lack an overt signifier; for example, in Russian, the indefinite human agent Ø$^{\text{PEOPLE}}_{\text{pl}}$ or the zero form of the copula быть *byt'* 'be' in the present tense of the indicative (37).[15]

(37) a. Ивана спасли. *Ivana spasli.* IVAN SAVED-3pl
'Ivan was saved [by some people]' :
Ø$^{\text{PEOPLE}}{}_{\text{pl}}$ ←**I**-SPASTI$_{\text{ind,past}}$-**II**→IVAN

b. Иван студент. *Ivan student.* IVAN STUDENT 'Ivan is a student' :
IVAN←**I**-BYT'$_{\text{ind,pres}}$-**II**→STUDENT

- LUs that are elided on the surface in a particular context; for example, Spanish *duermo* 'I sleep' appears in DSyntS as YO←**I**-DORMIR$_{\text{ind,pres,non-perf,non-progr}}$.
- Fictitious LUs, which represent meanings expressed by syntactic constructions. A syntactic construction that carries a lexical meaning 'L' is utterly idiosyncratic. Therefore, in order to preserve the abstract nature and the universal character of DSyntRels, the meaning 'L' has to be encoded in the DSyntS by a fictitious lexeme &L.
For instance, the meaning 'approximately' expressed by the Russian numeral inversion construction is represented in the DSyntS by the fictitious lexeme &приблизительно (*priblizitel'no* 'approximately').[16] Thus, phrases of the type exemplified in (38) are represented as shown.

(38) a. метров двести *metrov dvesti* METERS TWO-HUNDRED
'about two hundred meters' :
METR-**ATTR**→200-**ATTR**→&PRIBLIZITEL'NO

b. человек пятнадцать *čelovek pjatnadcat'* PEOPLE FIFTEEN
'about fifteen people' :
CELOVEK-**ATTR**→15-**ATTR**→&PRIBLIZITEL'NO

Similarly, the meaning 'have to', expressed in Russian by the construction N$_{\text{dat}}$+INF), is represented in the DSyntS by the fictitious lexeme &надо (*nado* ≈ 'necessary'); thus, example (39) has the DSyntS as in Fig. 4.

(39) Мне есхчё работать *Mne eščё rabotat'.* TO-ME STILL TO-WORK
'I still have to work.'

As indicated, the copula BYT' appears on the surface as a zero word form. Note that in order to make our syntactic structures more surveyable, we allow ourselves, here and below, to omit all grammemes that are not necessary for the understanding of a given representation.
- LF symbols, with the specification of the keyword (in surface-syntactic structures LF-symbols are replaced by actual lexemes that are values of the LFs in question).

[15] The Russian zero lexeme Ø$^{\text{PEOPLE}}_{\text{pl}}$ roughly corresponds semantically to the impersonal pronoun such as German *man* or French *on*, although their use is far from identical (Mel'čuk 1988a, pp 314–320, 1995b).

[16] The symbol '&' indicates the fictitious character of this LU.

Fig. 4 DSynts for example (39)

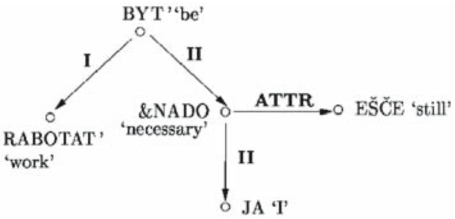


Fig. 5 DSyntS of German equivalent in example (40)

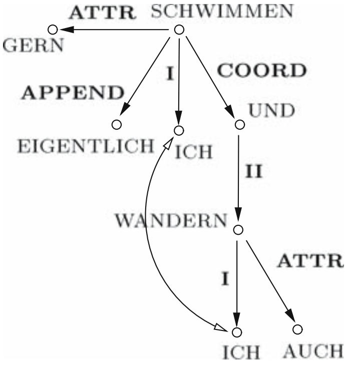
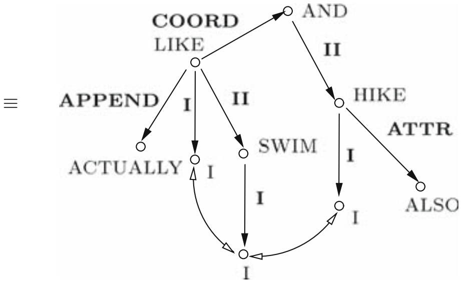


Fig. 6 DSyntS of English equivalent in example (40)



3.2.2 *Examples of DSyntS*

DSyntSs of translationally equivalent sentences in (40) are shown in Figs. 5 and 6. The bidirectional arrow linking the two occurrences of ICH 'I' stands for the coreferentiality relation.

(40) a. *Eigentlich schwimme ich gern und ich wandere auch.*
ACTUALLY SWIM I WITH-PLEASURE AND I HIKE ALSO
b. Actually, I like swimming and I hike also.

Figures 7–9 show DSyntSs that include LFs for the three examples in (41).

Fig. 7 DSyntSs for example (41a)

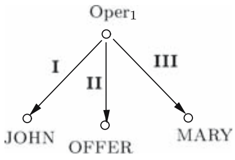


Fig. 8 DSyntSs for example (41b)

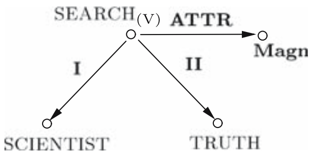
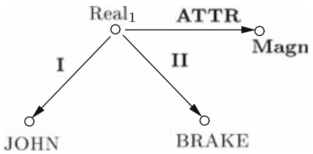


Fig. 9 DSyntSs for example (41c)



(41) a. John made Mary an offer.
b. Scientists frantically search for truth.
c. John vigorously slammed on the brake.

3.3 DSyntS in contrast with other transfer representations

Compared to other common transfer representations such as Jackendoff's (1990) Lexical Conceptual Structures (LCSs), LFG's f-structures (Bresnan 1982), and semantic structures inspired for example by Discourse Representation Theory (DRT) (Kamp and Reyle 1993; Reyle 1993) or Situation Semantics (Barwise and Perry 1983), DSyntS is more general and thus allows better abstraction from lexical and surface-syntactic idiosyncrasies reflected in these representations. For instance, according to Dorr (1999, pp 612), the LCS for *enter* as in (42a) is as in (43a), while that of its Spanish equivalent *entrar*, whose second actant subcategorizes for a PP with *en* as in (42b), is shown in (43b).

(42) a. Maria entered the house.
b. *Maria entró en la casa.*
Maria entered into the house

(43) a. [Event GO$_{\text{Loc}}$([Thing W],
[Path TO$_{\text{Loc}}$([Position IN$_{\text{loc}}$([Thing W], [$_{\text{Location}}$ *Z])])])]
b. [Event GO$_{\text{Loc}}$([Thing W],
[Path *TO$_{\text{Loc}}$([Position IN$_{\text{loc}}$([Thing W], [$_{\text{Location}}$ Z])])])]

With (44) as the interlingua representation for (42), both source-language-to-interlingua and target-language-to-interlingua mismatches occur. In contrast, the DSyntSs of (42a, b) are the same (45) (cf. also Sect. 4.1.2).

(44) [Event GO$_{\text{Loc}}$([Thing MARIA],
[Path TO$_{\text{Loc}}$([Position IN$_{\text{loc}}$([Thing MARIA],
[$_{\text{Location}}$ HOUSE])])])]

(45) a. MARIA←**I**-ENTER$_{\text{past}}$-**II**→HOUSE

b. MARIA←**I**-ENTRAR$_{\text{past}}$-**II**→CASA

With LFG's f-structure as transfer representation, the representations of (42a, b) are as in (46) (see Kaplan et al. 1996) for an equivalent example). Thus, as in the case of LCSs, adjustments during the transfer stage are necessary.

(46) a.

$$\left(\begin{array}{ll}
\text{PRED} & \text{`enter}\langle(\uparrow \text{SUBJ}, \uparrow \text{OBJ}\rangle\text{'}\\
\text{TENSE} & \text{PAST}\\
\text{SUBJ}f_2 & \left(\begin{array}{ll}\text{PRED} & \text{`Maria'}\\ \text{NUM} & \text{SG}\\ \text{GEND} & \text{FEM}\end{array}\right)\\
\text{OBJ}f_3 & \left(\begin{array}{ll}\text{PRED} & \text{`house'}\\ \text{NUM} & \text{SG}\\ \text{SPEC} & \left(\begin{array}{ll}\text{DEF} & +\\ \text{PRED} & \text{`the'}\end{array}\right)\end{array}\right)
\end{array}\right)$$

b.

$$\left(\begin{array}{ll}
\text{PRED} & \text{`entrar}\langle(\uparrow \text{SUBJ}, \uparrow \text{AOBJ}\rangle\text{'}\\
\text{TENSE} & \text{PAST}\\
\text{SUBJ}t_2 & \left(\begin{array}{ll}\text{PRED} & \text{`Maria'}\\ \text{NUM} & \text{SG}\\ \text{GEND} & \text{FEM}\end{array}\right)\\
\text{AOBJ} & \left(\begin{array}{ll}\text{PRED} & \text{`en}\langle(\uparrow \text{OBJ}\rangle\text{'}\\ \text{PCASE} & \text{AOBJ}\\ \text{OBJ}t_3 & \left(\begin{array}{ll}\text{PRED} & \text{`casa'}\\ \text{NUM} & \text{SG}\\ \text{GEND} & \text{FEM}\\ \text{SPEC} & \left(\begin{array}{ll}\text{DEF} & +\\ \text{PRED} & \text{`la'}\end{array}\right)\end{array}\right)\end{array}\right)
\end{array}\right)$$

Dorna and Emele (1996) use a variant of the DRT-representation (Underspecified Discourse Representation Structures, (UDRSs)) as a transfer representation. As UDRSs, the corresponding structures for our example (42) are approximately as in (47).

(47) a. `[ L̄1:enter(E), L2:arg1(E,i1), L3:arg2(E,i2), L4:Maria(i1),`
`L5:house(i2), L6:def(i2)]`

b. `[L1:entrar(E), L2:arg1(E,i1), L3:arg2(E,i2), L4:Maria(i1),`
`L5:en(i2,i3), L6:casa(i3), L7:def(i3)]`

Similarly to the f-structures, the UDRSs reveal (as would all Davidsonian formalisms) a divergence here. Moreover, none of the transfer representations known to us deals systematically with phraseology, that is, with idioms and collocations (cf. also Sects. 4.1.3 and 4.1.4).

In the next section, we investigate in more detail the types of idiosyncrasies that can be eliminated when DSyntS is used as the transfer representation.

4 Reduction of divergences by the use of DSyntS

The use of the DSyntS in transfer allows for the elimination of two types of structural divergences: (i) divergences caused by surface-syntactic phenomena and

(ii) divergences caused by restricted lexical co-occurrence (if the co-occurrence is described in terms of LFs). In our approach, these are in point of fact pseudo-mismatches.

4.1 Pseudo-mismatches due to surface-syntactic phenomena

Some of the structural divergences discussed in the literature show up only if transfer is carried out at the surface-syntactic or even deep-morphological level. However, since we assume that transfer must be done at the deep-syntactic level, the corresponding phenomena are processed intralinguistically: in the analysis or synthesis stage as appropriate. As a result, divergences of this kind disappear, turning out to be pseudo-mismatches. These phenomena include: (i) auxiliaries of all kinds, (ii) governed prepositions and conjunctions, (iii) (parts of) idioms, (iv) syntactic idiosyncrasies.

4.1.1 Pseudo-mismatches due to auxiliaries

Inflectional meanings, independently of whether they are expressed analytically (by auxiliaries or other analytical markers) or synthetically (by bound morphemes), are grammemes, which are specified in the DSyntS as feature–value pairs of the respective lexical units.[17] Therefore, although for example the French and English sentences in (7), Sect. 1.1, show an obvious structural divergence, there is no mismatch between their DSyntSs (48).

(48) $\textsc{lire}_{\text{fut}}$-**I**→MOI ≡ $\textsc{read}_{\text{fut}}$-**I**→I

The linguistic signs that express inflectional meanings are specified *intra*linguistically either in the SSynts (analytical markers, i.e., word forms) or in the deep-morphological structure (morphemes, i.e., parts of word-forms). This guarantees that not only verbal auxiliaries, but also analytical vs. synthetic articles (49a), analytical vs. synthetic adjectival degrees (49b) are "leveled out" in the DSyntS and thus do not entail structural mismatches.

(49) a. (French) *le loup* THE WOLF ⇔ (Romanian) *lupul* WOLF-def 'the wolf'
b. (English) *more beautiful* ⇔ (German) *schöner* BEAUTIFUL-comp

4.1.2 Pseudo-mismatches due to governed prepositions and conjunctions

The specification of governed prepositions and conjunctions is part of idiosyncratic surface-syntactic characteristics of particular LUs: they are specified in the Government Pattern (GP ≈ subcategorization frame) of their governor. Therefore, they do not need to be represented in the DSyntS.[18] As a result, the DSyntSs of the English and German sentences in (8) show no mismatch (50).

[17] As already mentioned, for the sake of a more compact presentation, we use simple subscripts (names of grammemes) instead of feature–value pairs.

[18] During analysis, all elements governed by an LU L_S are identified and then eliminated in the transition $\text{SSyntS}_S \Rightarrow \text{DSyntS}_S$ in accordance with the GP of L_S. During synthesis, the corresponding elements are introduced into the SSyntS_T in the transition $\text{DSyntS}_T \Rightarrow \text{SSyntS}_T$ in accordance with the GP of the translation equivalent L_T of L_S.

(50) HE←**I**–ENTER$_{\text{past}}$–**II**→ROOM ER←**I**-EINTRETEN$_{\text{past}}$-**II**→ZIMMER

Similarly, the three further English–French examples (51)–(53) show the elimination of surface-syntactic divergences (the diverging elements are set in bold face) at the level of DSyntS (the DSyntSs of each pair of sentences in (4.1.2) are structurally identical).

(51) *J'attends Anne*. I AWAIT ANNE
I am waiting **for** Anne.
MOI←**I**-ATTENDRE-**II**→ANNE
I←**I**-WAIT-**II**→ANNE

(52) His being [here] bothers me.
Qu*'il soit* [*là*] *me dérange*. THAT HE BE [HERE] ME BOTHERS
HE←**I**-BE←**I**-BOTHER-**II**→I
LUI←**I**-ÊTRE←**I**-DÉRANGER-**II**→MOI

(53) *Je veux partir*.
I want **to** leave.
MOI←**I**-VOULOIR-**II**→PARTIR
I←**I**-WANT-**II**→LEAVE

4.1.3 Pseudo-mismatches due to idioms

We call an "idiom" a multiword expression that constitutes one lexical unit: *kick the bucket*, *in order to*, *give up*, *of course* etc. An idiom is represented in the DSyntS by one single node, which allows for the elimination of structural divergences that occur when an idiom in L_S corresponds to a single lexeme in L_T as exemplified by *curry favor* [*with N*] equivalent to Russian подлизываться [к N] *podlizyvat'sja* [k N], *pass water* to French *uriner*, *in order to* to French *pour*. An important class of idioms in English is phrasal verbs;[19] for example *give up*, equivalent to French *abandonner*, *get away* to *échapper*, *get out* to *sortir*, *keel over* to *tourner de l'oeil* [lit. 'turn the eye']. All such idioms show no structural difference between languages as far as their representation at the level of DSyntS is concerned: each is represented by a single node.

4.1.4 Pseudo-mismatches due to syntactic idiosyncrasies

Numerous structural divergences are due to surface-syntactic idiosyncrasies as in the French–German–Russian example (54).

(54) a. *Les rhinocéros tournaient leur tête horrible et cornue.*
THE RHINOCEROSES TURNED THEIR HEAD HORRIBLE AND HORNED
b. *Die Nashörner drehten ihre hässlichen (*und) behörnten Köpfe.*
THE RHINOCEROSES TURNED THEIR HORRIBLE (*AND) HORNED HEADS
c. Носороги ворочали своими безобразными (*и) рогатыми головами.
*Nosorogi voročali svoimi bezobraznymi (*i) rogatymi golovami.*
RHINOCEROSES TURNED THEIR HORRIBLE (*AND) HORNED HEADS

[19] In the current literature, phrasal verbs are often considered as a phenomenon apart, while in our approach they are most adequately treated as a subclass of idioms.

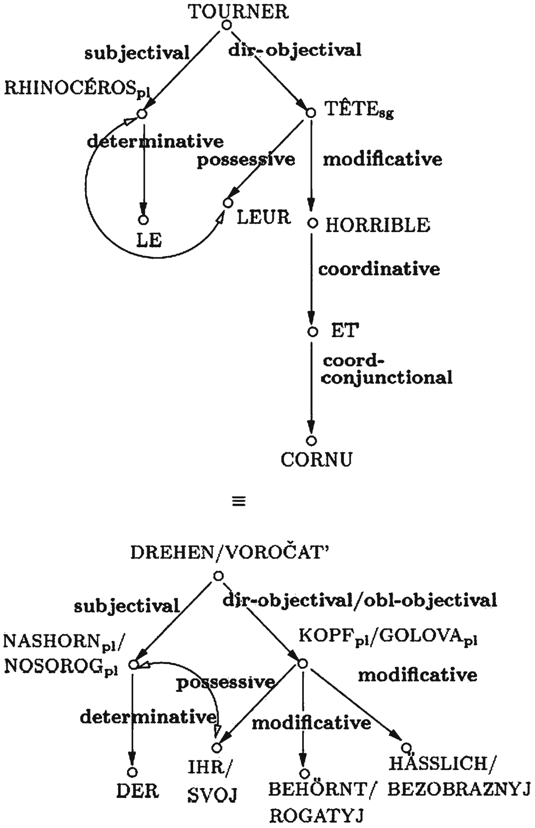


Fig. 10 Fragments of the SSyntSs of the sentences in (54)

The SSynts of (54a) shows a structural divergence with the structures of (54b,c): French *horrible et cornue* (with conjunction) versus German *hässlichen behörnten* and Russian безобразными рогатыми *bezobraznymi rogatymi* (without conjunction); cf. the corresponding fragments of the SSyntSs in question in Fig. 10.

However, in the DSyntS, the French conjunction *et* is not represented: its presence in (54a) is a surface-syntactic idiosyncrasy of French. French does not easily tolerate asyndetic conjunction of adjectives and requires *et* independently of the meaning of the conjoined adjectives; that is, in (54a) *et* is a typical surface-syntactic structural element. DSyntSs of the sentences in question are structurally identical; cf. the DSyntS for all three given in Fig. 11.

This example shows three further divergences, namely a syntactic, a morphological and a lexical one.

First, in French and German (as well as in English), the noun for 'rhinoceros' has a definite article, which is absent in Russian. However, this article is an analytical marker of the grammeme of definiteness, so that in the DSyntS it will be represented as grammemic feature–value symbol. Therefore, there will be no structural mismatch. Note however that a morphological mismatch remains, the resolution of which requires a much deeper processing.

Second, the singular of French TÊTE vs. the plural of German KOPF and Russian GOLOVA is again an idiosyncrasy of French syntax: a single "distributed" object in a construction with a plural subject that refers to the "possessors" of the object must

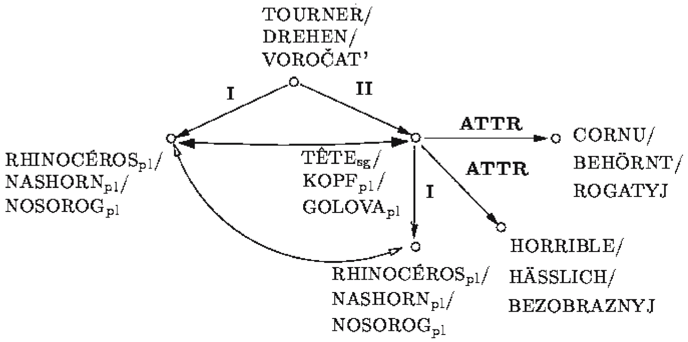


Fig. 11 DSyntS for the sentences in (54)

be in the singular. However, French admits the contrast between the singular and plural in this position (in the case of a multiple "distributed" object as in (55a) if everyone had just one horse, versus (55b), if everyone had several horses). Therefore, the difference must be preserved in the French DSyntS, which creates a morphological mismatch. Still, no structural mismatch results.

(55) a. *Les cowboys se sont arrêtés pour faire boire leur cheval.*
lit. 'The cowboys stopped to make drink their horse'
b. *Les cowboys se sont arrêtés pour faire boire leurs chevaux.*
lit. 'The cowboys stopped to make drink their horses'

Third, where French and German have a "normal" possessive adjective (LEUR, IHR), Russian imposes the use of the reflexive possessive adjective СВОЙ *svoj*, which follows from an idiosyncratic syntactic rule of Russian: if the possessor of a direct object is coreferential with the subject, it must be expressed by SVOJ. Like the above-mentioned surface-syntactic divergences, this divergence also disappears in the DSyntS (cf. Fig. 11).

4.2 Pseudo-mismatches due to restricted lexical co-occurrence

Words and phrases that are values of LFs often diverge between $\mathcal{L}_S$ and $\mathcal{L}_T$. However, when values of LFs are replaced with the corresponding LFs in the DSyntS, these divergences disappear and thus do not cause structural mismatches. Consider the examples in (56) and (57), where (a, b) show translationally equivalent sentences with significant surface divergences and (c) the DSyntSs of the same sentences, which show no mismatch.

(56) a. It rains heavily/hard.
b. *Il pleut comme vache qui pisse.*
IT RAINS AS COW THAT PISSES
c. RAIN/PLEVOIR-**ATTR**→Magn

(57) a. They are glued to the TV set.
b. Они не отходят от телевизора.
Oni ne otxodjat ot televizora.
THEY NOT STEP-BACK FROM TV-SET
c. THEY/ONI←**I**-MagnReal$_I$-**II**→TV-SET/TELEVIZOR

In (56), HEAVILY/HARD and COMME VACHE QUI PISSE are values of the LF Magn, applied to the verbs RAIN and PLEUVOIR. In (57) the expressions BE GLUED and NE OTXODIT' are values of the complex LF MagnReal$_1$, which roughly means 'use intensely', applied to the nouns TV-SET and TELEVSIOR.

5 Structural mismatches: definition and typology

As shown in the preceding section, the use of DSyntS allows for the elimination of divergences between the source and the target sentences in many cases. However, it does not of course eliminate all mismatches. In order to be able to deal with the mismatches remaining at the DSynt level, we need a rigorous definition of the general notion of "structural mismatch". We provide such a definition in Sect. 5.1, then we sketch a general typology of structural mismatches (Sect. 5.2), and finally, we examine the syntactic mismatches to be handled at the DSynt level (Sect. 5.3).

5.1 The definition of structural mismatch

A structural mismatch between two DSyntSs S_1 and S_2 is defined as a violation of the isomorphism between them; therefore, let us first introduce the concept of DSyntS-isomorphism.

Definition 2 (Isomorphism of DSyntSs). Let there be two DSyntSs
$S_1 := \langle N_1, A_1, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g} \rangle$ and
$S_2 := \langle N_2, A_2, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g} \rangle$
as defined in Definition 1.
In S_1 and S_2, $G_{sem_1} := \gamma_{n \to g}(N_1)$ and $G_{sem_2} := \gamma_{n \to g}(N_2)$; $\boldsymbol{n}^s_{a_i}$, $\boldsymbol{n}^t_{a_i}$ are the source and the target node of the arc a_i, respectively. Furthermore, let there be:

(1) a node mapping function ν which maps a node $n_i \in N_1$ onto the node $n_j \in N_2$ [which ensures the one-to-one correspondence between the nodes of S_1 and S_2],
(2) a lexical label translation function τ which maps a lexical node label l_i of the node $n_i \in N_1$ onto the lexical node label l_j of the corresponding node $n_j \in N_2$ [which ensures the translation equivalence of the LUs labeling the corresponding nodes; τ is realized as the bilingual lexical index introduced in Sect. 2.1].

S_1 and S_2 are isomorphic iff there exist three one-to-one functions g_N: $N_1 \to N_2$, g_A: $A_1 \to A_2$, g_G: $G_{sem_1} \to G_{sem_2}$, and for all nodes $n_{1,i}$ and arcs $a_{1,k}$ of S_1 and all nodes $n_{2,j}$ and arcs $a_{2,l}$ of S_2 the following restrictions are satisfied:

1. $g_N(n^s_{a_{1,k}}) = n^s_{g_A(a_{1,k})}$
2. $g_N(n^t_{a_{1,k}}) = n^t_{g_A(a_{1,k})}$
3. $\lambda_{l_s \to n}(n_{1,i}) = \tau(\lambda_{l_s \to n}(g_N(n_{1,i})))$,
4. $\rho_{r_s \to a}(a_{1,k}) = \rho_{r_s \to a}(g_A(a_{1,k}))$ [the corresponding arcs in S_1 and S_2 must have the same label],
5. $\gamma_{n \to g}(\lambda_{n \to g}(n_{1,i})) = g_G(\gamma_{n \to g}(\tau(\lambda_{n \to g}(g_N(n_{1,i})))))$

In 3–5, the inverse is also true because the corresponding functions are one-to-one functions.

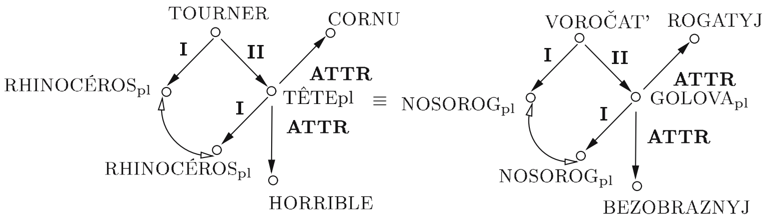


Fig. 12 Isomorphic semantically equivalent DSynt-structures

This definiton means that two DSyntSs are isomorphic if and only if (i) each LU in one of them corresponds to an LU in the other (and vice versa), (ii) all grammemes of two corresponding LUs correspond as well, and (iii) if two LUs in the first structure are linked by a DSyntRel, then their counterparts in the second are linked by the same DSyntRel, directed in the same way.

Definition 2 specifies DSyntS-isomorphism in general. However, in the context of MT, we are interested in the isomorphism of the DSyntSs of semantically equivalent sentences of $\mathcal{L}_S$ and $\mathcal{L}_T$. (Two syntactic structures S_1 and S_2 are semantically equivalent if they can be mapped onto the same semantic structure; the definition of semantic structure is beyond the scope of the paper.) Therefore, the functions τ and g_G are language translation functions and represent semantic equivalence between LUs and between grammemes of the source and target languages. Figure 12 shows an example of two isomorphic semantically equivalent DSynt-structures.

Now, we can readily define the concept of structural mismatch.

Definition 3 (Structural Mismatch). Two semantically equivalent DSyntSs S_1 and S_2 show a structural mismatch iff they are not isomorphic.

A structural mismatch is thus a symmetrical relation between two DSyntSs. Therefore, we consider a mismatch between DSyntS$_S$ and DSyntS$_T$ to be the same as the mismatch between DSyntS$_T$ and DSyntS$_S$.

5.2 Structural mismatches between languages: major types

From Definitions 2 and 3 it follows that the violation of the isomorphism between two DSyntSs can be due to the absence of one-to-one correspondence:

- between their LUs,
- between grammemes of the corresponding LUs,
- between DSyntRels that link the corresponding LUs.

The absence of one-to-one correspondence between particular elements of two DSyntSs may entail the absence of one-to-one correspondence between elements of other types. Thus, if one LU in DSyntS S_1 corresponds to several LUs in S_2, there will unavoidably be a divergence between DSyntRels.

Theoretically, the absence of one-to-one correspondence between LUs and grammemes of two DSyntSs can appear as (i) correspondence of one element to a configuration of elements of the same kind, (ii) of one element to nothing, and (iii) of one element to one or several elements of another kind. For DSyntRels, the absence of one-to-one correspondence means that the DSyntRel$_S$ corresponds to a DSyntRel$_T$

with (iv) a different name, (v) a different direction, or (vi) a different governor. The combination of all these possibilities gives rise to a considerable number of structural mismatches. However, these can be reduced to five major types, by imposing the following linguistic constraints:

– No LU$_S$ and no DSyntRel$_S$ can be eliminated/introduced.[20]
– No LU$_S$ and no source grammeme can correspond to a DSyntR$_T$ (and vice versa).

As a result, we consider the following five major types of mismatches:

(a) Mismatches with respect to LU$_S$:
  1. an LU$_S$ corresponds to a configuration of LU$_T$s,
  2. an LU$_S$ corresponds to a target grammeme or a configuration of target grammemes.

(b) Mismatches with respect to grammemes:
  3. a source grammeme corresponds to a configuration of target grammemes,
  4. a source grammeme does not correspond to any element in DSyntS$_T$.[21]

(c) Mismatches with respect to DSyntRel$_s$S:
  5. a DSyntRel$_S$ corresponds to a DSyntRel$_T$ with a different label or a different direction.

Let us illustrate these five types of mismatches.

5.2.1 $LU_S \equiv LU_{1_T} + LU_{2_T} + LU_{3_T} + \cdots$

The mismatch of this type, which can be called "node fission/fusion", is a correspondence between an LU$_S$ and a multilexemic target structure, i.e., a phrase. It is the most common type of mismatch in MT. It divides into three further subtypes:

(i) A source monolexemic LU$_S$ corresponds to a target (binary) collocation; for example, Russian простудиться *prostudit'sja* and German *sich erkälten* [= ERKÄLTEN$_{\text{pron}}$] correspond to the English phrase *to catch a cold*. Similarly, French *se suicider* [= SUICIDER$_{\text{pron}}$] and English *commit suicide*. Another example is the English verb *rain* and Russian идёт дождь *idёt dožd'* 'is-going [the] rain'.

(ii) A source monolexemic LU$_S$ that is a derived or compound word corresponds to a target free phrase; cf. (58).

(58) a. Russian вы+ползти *vy+polzti* OUT-CRAWL 'crawl out [of]'
French *sortir en rampant* 'exit by crawling'

[20] Actually, the elimination/introduction of LUs in translation is possible (v) and (vi).
(v) This allows **us/one** to consider the following case.
Ceci permet de considérer le cas suivant. lit. 'This allows to consider …'
(vi) I take the red **one**.
Ich nehme die Rote. lit. 'I take the red'
In all such cases, here we are looking at the omission (or introduction, as the case may be) of "semi-auxiliary" LUs required by idiosyncratic rules and restrictions of a particlar language. We chose not to tackle these cases here in order to simplify our task. However, this does not distort the proposed general picture of syntactic mismatches.

[21] As, for example, the Japanese politeness grammemes are not rendered grammatically in European languages, while the European verbal tense and nominal number are omitted in Vietnamese.

b. Russian пере+плыть *pere+plyt'* CROSS-SWIM 'swim across'
French *traverser à la nage* 'cross at the swim'

c. German *Berlin+reise* BERLIN-TRIP
English *trip to Berlin*

d. German *Arbeit+s+methode* WORK-link-METHOD
Russian рабочий метод *rabočij metod* 'work method'

A compound LU—including free nominal compounds such as *Berlinreise* 'Berlin-trip' and *Arbeitsmethode* 'work-method' and verbs with incorporated actants as in (59)—is represented in the DSyntS as one LU-node, with its internal structure explicitly specified in terms of component lexemes and relations between them.[22] Thus, a Chukchee verbal form in (59) with incorporated elements is represented in the DSyntS by one single node, while its translation in European languages is a whole sentence, represented by a multiple-node tree.

(59) a. *Tə+ngəran+otkočj+əntə+vat+ərkən*
1sg FOUR TRAP SET$_V$ pres
'I set four traps'

b. *Mən+nəki+ure+qəpl+uvičven+mək*
1pl-imper NIGHT LONG-TIME BALL PLAY 1-pl-imper
'Let us play ball at night for a long time'

c. A source monolexemic LU$_S$, which is a fused element of the value of an LF, corresponds to a combination of the non-fused element of the value of the same LF with the argument of the LF. Examples are given in (60).

(60) a. German *duschen* '[to] shower' [= //Real$_1$(DUSCHE)]
Russian принять душ *prinjat' duš* 'receive shower'
[= Real-II→DUŠ]

b. Serbian *zaposliti se* '[to] start-work oneself'
[= //Caus$_1$Oper$_1$(POSAO)]
English *land a job* [= Caus$_1$Oper$_1$(JOB)-**II**→JOB]

c. German *schütten* lit. '[to] downpour' [= //Magn(REGNEN)]
English *rain heavily* [= RAIN-**ATTR**→Magn(RAIN$_V$)]

In (60a) German *duschen* is the fused value of the LF Real$_1$ applied to the noun *Dusche* 'shower': //Real$_1$(DUSCHE) = *duschen*, while the Russian принять душ *prinjat' duš* is a combination of the value of Real$_1$ applied to the noun душ *duš* 'shower' with DUŠ itself: Real$_1$(DUŠ) = *prinjat'*⊕DUŠ.[23] In (60b) the Serbian ZAPOSLITI_SE is the fused value of the LF Caus$_1$Oper$_1$ applied to the noun POSAO 'job': //Caus$_1$Oper$_1$(POSAO) = *zaposliti se*, while *land a job* is a combination of Caus$_1$Oper$_1$(JOB) and *job*. In (60c), the German *schütten* is a fused value of the LF Magn applied to the verb REGNEN 'rain': //Magn(REGNEN) = *schütten*, while *rain heavily* is a combination of Magn(RAIN$_V$) = *heavily* and *rain*.

[22] However, we do not enter here into the problem of compounds and their representation in DSyntS.

[23] "⊕" stands here and below for the operation of linguistic union.

5.2.2 $LU_S \equiv \{g1_T, g2_T, \ldots\}$

The case of a source monolexemic LU$_S$, possibly introduced by an auxiliary, corresponding to one or several target grammemes is also rather frequent. Some examples are shown in (61)–(65).

(61) *Als ich kam, war er dabei, eine Flasche Wein zu entkorken.*
WHEN I CAME, WAS HE AT-THIS, A BOTTLE WINE TO UNCORK
Когда я пришел, он открывал бутылку вина.
Kogda ja prišel, on otkryval butylku vina.
WHEN I CAME, HE WAS-OPENING BOTTLE OF-WINE
'When I came, he was opening a bottle of wine.'

(62) *Il est en train d'écrire une lettre.*
HE IS IN COURSE OF TO-WRITE A LETTER
He is writing a letter.

(63) I can read this book.
(Hungarian) *Ezt a könyvet olvás+hat+om.*
THIS THE BOOK READ-CAN-1sg

(64) my house
(Hungarian) *ház+am* HOUSE-MY

(65) under the table
(Lezgi) СТОЛ+ДИК *stol+dik* TABLE-UNDER

In (61), the German lexeme DABEI in the construction [*zu* V_{inf}] *dabei (sein)* '(to be) V-ing' corresponds to the Russian grammeme 'imperf(ective aspect)' on $\tau(V_{\text{inf}})$ where открывал *otkryval* is OTKRYVAT'$_{\text{imperf,past}}$. In (62), the French deep lexical unit EN_TRAIN in the construction *(être) en train* [*de* V_{inf}] '(to be) V-ing' corresponds to the English grammeme 'progr(essive)' on $\tau(V_{\text{inf}})$: *writing* is WRITE$_{\text{progr,pres}}$. In (63), the English modal verb *can* corresponds to the grammeme 'potentialis' in Hungarian, so *olváshat-* is represented as OLVÁS$_{\text{potentialis}}$. Similarly in (64), the English possessive adjective corresponds in Hungarian to the grammeme of belonging: *házam* = HÁZ$_{\text{1sg}}$. Example (65) shows how a meaningful preposition in an Indo-European language corresponds, as a rule, to a grammatical case in a Daghestanian language such as Lezgi (and, often, in Finno-Ugric and other languages): столдик *stoldik* = [STOL$_{\text{subessive}}$].

5.2.3 $g1_S \equiv \{g1_T, g2_T, \ldots\}$

An example of this type of mismatch is the translation of the German 'imperfect' tense grammeme into Russian, where it corresponds to two grammemes: 'imperf' (aspect) and 'past' (tense), as exemplified in (66). German *kam* = KOMMEN$_{\text{imperf}}$ corresponds to Russian приходил *prixodil* 'came' = PRIXODIT'$_{\text{imperfective,past}}$.

(66) *Ich kam oft.* I CAME OFTEN
Я приходил часто. *Ja prixodil často.* I CAME OFTEN.

5.2.4 $g1_S \Leftrightarrow -$

The elimination/introduction of grammemes in the DSyntS$_T$ occurs when the target language lacks an inflectional category present in the source language and has no lexical or grammatical equivalence. For instance, nominal number and verbal tense are absent from Vietnamese and Chinese, which do not express at all the corresponding information. The same happens with Japanese and Korean politeness: in Indo-European languages, the information on the social hierarchy between the participants of a speech act is not systematically expressed in each verbal form. As a result, in the process of translation from Vietnamese into English, for example, the grammemes of nominal number and verbal tense must be introduced into the target DSyntS. Inversely, English-to-Vietnamese translation entails, generally speaking, the elimination of these grammemes.[24]

5.2.5 DSyntRel$_S$ ⇔ DSyntRel$_T$ (with DSyntRel$_S$ ≠ DSyntRel$_T$)

The absence of a one-to-one correspondence between source and target DSyntRels can be manifested as one of the three following subcases: branch relabeling, branch inversion, and branch transposition (raising/lowering).[25] All of them are triggered either by a specific pair of lexical translation equivalents or by a specific pair of syntactic construction translation equivalents.

Branch relabeling: LU$_S \xrightarrow{r_1} LU'_S \equiv LU_T \xrightarrow{r_2} LU'_T$. Branch relabeling subsumes all kinds of DSyntRel name modification. This includes, as a major particular case, all the varieties of syntactic conversion as exemplified in (67)–(69).

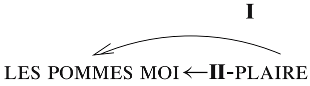
(67) a. He vomits.
HE←**I**-VOMIT

b. он рвёт. *Ego rvët.* HIM VOMITS
ON←**II**-RVAT.

(68) a. I like apples.
I←**I**-LIKE-**II**→APPLES

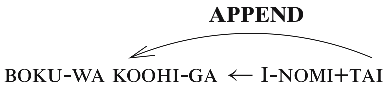
b. Мне нравяатся тижение яблоки. *Mne navratsja jabloki.*
TO-ME PLEASE APPLES
JA←**II**-NRAVIT’ SJA-**I**→JABLOKI

c. *Les pommes me plaisent.* THE APPLES TO-ME PLEASE



(69) a. I want to drink coffee.
I←**I**-WANT-DRINK-**II**→COFFEE

b. 僕はコーヒーが飲みたい。*Boku-wa koohi-ga nomi-tai.*
I-topic COFFEE-nom DRINK-IS-DESIRABLE



[24] Actually, the state of affairs is much more complex: the source grammemes that do not have a direct correspondence in the target language can still be expressed by a combination of existing target grammemes and lexical means. One such grammeme is for example Russian aspect. The literature on the expression of Russian aspect in aspectless languages is very rich; cf., among others, Sacker (1983), Abraham (2003), Paslawska and von Stechow (2003).

[25] As for branch introduction/elimination, these operations presuppose the introduction/elimination of nodes, and, therefore, are not considered here.

In example (69), the Engish lexeme WANT corresponds to the Japanese grammeme 'desiderative' (the case is discussed in 5.2.2: the correspondence of a lexeme to a grammeme). In other words, WANT together with its second actant *V* (the verb DRINK) is translated as a deverbal adjective with the suffix [*V*+]-TAI. At the same time, two branch relabelings occur: (i) the equivalent of the second actant of the verb DRINK (i.e., COFFEE) becomes the first actant of the adjective NOMI+TAI 'drink is desirable'; (ii) the equivalent of the first actant of the verb WANT (*I*) becomes an **APPEND**-dependent of the top node (a prolepsis on the surface),[26] expressing the theme of the Japanese sentence.

Branch inversion: $\mathrm{LU}_S \xrightarrow{r_1} \mathrm{LU}'_S \equiv \mathrm{LU}'_T \xrightarrow{r_2} \mathrm{LU}_T$. Branch inversion (head-switching) most often involves branch relabeling (and a change of the part of speech of one of the corresponding nodes), as exemplified in (70)–(72).

(70) a. *dès la majorité atteinte* SINCE THE MAJORITY REACHED
DÈS LA MAJORITÉ-**ATTR**→ATTEINTE

b. с момента достижения совершеннолетия
s momenta dostiženija soveršennoletija
FROM MOMENT OF-REACHING OF-MAJORITY
S_MOMENTA DOSTIŽENIJA-**II**→SOVERŠENNOLETIJA

(71) a. in early/mid/late January
IN EARLY/MID/LATE←**ATTR**-JANUARY

b. в начале/середине/конце января
v načale/seredine/konce janvarja
V NAČALE/SEREDINE/KONCE-**II**→JANVARJA

(72) a. *Ich schwimme gern.* I SWIM WITH-PLEASURE
ICH←**I**-SCHWIMME-**ATTR**→GERN

b. I like swimming.
I←**I**-LIKE-**II**→SWIMMING

Note, however, that branch inversion may also not involve branch relabeling (73).

(73) a. He hobbled away.
HE←**I**-HOBBLED–**ATTR**→AWAY

b. *Il partit en clopinant.* HE LEFT WHILE HOBBLING.
IL←**I**-PARTIT–**ATTR** →EN_CLOPINANT

Branch transposition:
$\mathrm{LU}'_S \xleftarrow{r_1} \mathrm{LU}_S \xrightarrow{r_2} \mathrm{LU}''_S \xrightarrow{r_3} \mathrm{LU}'''_S \equiv \mathrm{LU}'''_T \xleftarrow{r'_3} \mathrm{LU}'_T \xleftarrow{r'_1} \mathrm{LU}_T \xrightarrow{r'_2} \mathrm{LU}''_T$. Branch transposition (raising/lowering) also involves branch relabeling, for example (74) and (75).

(74) a. I wash his hands.
II
I WASH HIS←**I**-HANDS

[26] A prolepsis is a particular sentence element: the leftmost nominal phrase that expresses a fronted topic (cf. Mel'čuk 2001, pp 130f).

b. *Je lui lave les mains.* I TO-HIM WASH THE HANDS

II

JE LUI←**III**-LAVE LES MAINS

(75) a. *Pierre semble dormir.* PIERRE SEEMS TO-SLEEP
PIERRE←**I**-SEMBLE-**II**→DORMIR

b. Кажеця что Пвер спит. *Kažetsja čto P'er spit.*
SEEMS THAT SLEEPS PIERRE
KAŽETSJA-**I**→[*čto*] SPIT-**I**→P'ER

5.3 Syntactic structural mismatches

The five types of structural mismatches fall naturally in two groups: mismatches that involve a grammeme-to-grammeme translation (inflectional structural mismatches, Sects. 5.2.2–5.2.4) and those that do not involve a grammeme-to-grammeme translation (lexico-syntactic structural mismatches, Sects. 5.2.1 and 5.2.5). The two groups are of quite a different linguistic nature. In order to facilitate our task, in what follows, we consider only lexico-syntactic structural mismatches, referring to them for short as "syntactic mismatches".

The syntactic mismatches at the DSynt-level discussed in the preceeding section are of four types (in the order of increasing complexity):

1. *Branch relabeling mismatch*: the source and the target structures diverge with respect to the names of two corresponding DSynt-relations.
2. *Head-switching mismatch*: the source and the target structures diverge in that in one of them an LU L_{1_S} syntactically depends on LU L_{2_S}, while in the other one the dependency relation is inverted, namely L_{2_T} depends on L_{1_T} (where $L_{2_T} = \tau(L_{2_S})$ and $L_{1_T} = \tau(L_{1_S})$).
3. *Node fission/fusion mismatch*: the source and the target structures diverge in that a node in one of them corresponds to a non-unit subtree.
4. *Branch transposition mismatch*: the source and the target structures diverge in that a subtree in one of them is moved to another governor in the other.

Figure 13 presents the above four types of mismatches.[27]

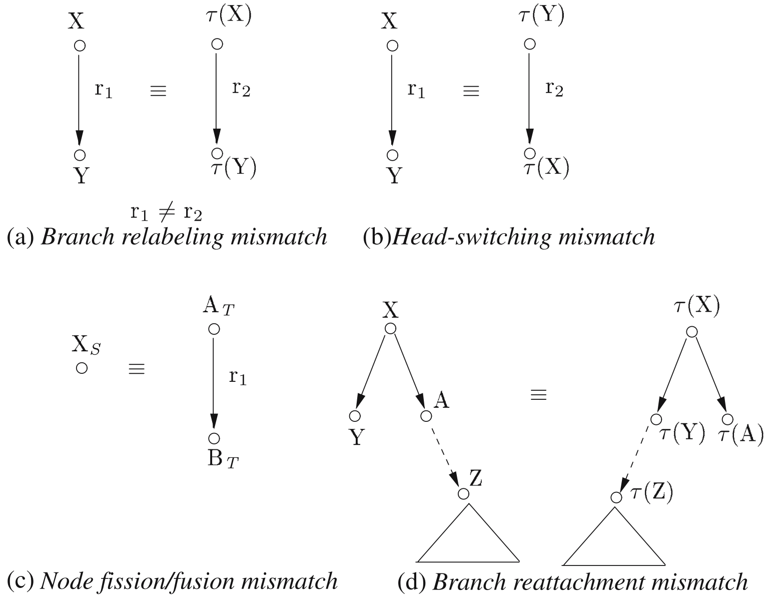
Interestingly, these four types cover all possible syntactic mismatches that can and should be treated at the DSynt-level.[28] This can be stated as Theorem 1 and its proof is given in the Appendix.[29]

Theorem 1 (Types of syntactic mismatches) *Let S_S and S_T be two translationally (= semantically) equivalent deep-syntactic structures such that between S_S and S_T one or several syntactic mismatches occur. Then, these mismatches can be only of the above four types.*

[27] In order to avoid cluttering the presentation, we use henceforth—where it seems appropriate—subscripts to indicate "source" or "target" language lexeme (instead of using the translation function τ).

[28] As we will see immediately below, there are syntactic mismatches that should not be dealt with at the DSynt-level.

[29] The elimination/introduction of semi-auxiliary LUs (see footnote 20) is covered by Type 3; formally, it corresponds to node fission/fusion.



(a) *Branch relabeling mismatch* (b)*Head-switching mismatch*

(c) *Node fission/fusion mismatch* (d) *Branch reattachment mismatch*

Fig. 13 Four possible types of syntactic mismatches

Note that the admitted types of syntactic mismatches do not include the many-to-many ($n : m, n \neq m$) mismatch. This is due to the following two considerations (of a very different nature):

- On the one hand, a non-compositional expression of n elements (i.e., an n-word idiom) of $\mathcal{L}_S$ can correspond to a non-compositional expression of m (i.e., an m-word idiom) of $\mathcal{L}_T$. For instance, the French idiom (76a), meaning 'to hate each other,' has seven surface words, while its Russian equivalent (76b) has six; the structures of these idioms are also very different.

(76) a. *ne pas pouvoir se voir en peinture*
neg. NOT BE-ABLE refl. TO-SEE IN PAINTING
'to be unable to see each other in a painting'

b. друг друга на дух не переносить
drug druga na dux ne perenosit'
EACH OTHER IN SPIRIT NOT TOLERATE

However, such cases, that is idioms of all types, do not create syntactic mismatches since in our approach any idiom is represented in the DSyntS by one single node (cf. Sect. 3).[30]

[30] Idioms are one of the three major types of phraseological expressions: idioms, collocations and pragmatemes (Mel'čuk 1995a). Given, as just pointed out, that an idiom is represented as one node, it cannot lead to an $n : m$ mismatch. Collocations are semantically and syntactically compositional and can, therefore, be transferred using the proposed types of rules. Pragmatemes such as *Hold the line*, *No parking*, *Best before*, i.e., expressions that are in many cases semantically and syntactically compositional, being restricted only by the situation of the corresponding speech act, behave from the viewpoint of transfer as one unit and thus should be treated formally, just like idioms, as one node in the DSyntS.

Table 3 Comparison of the major types of syntactic mismatches with Dorr's types of divergences

Syntactic mismatches	Dorr's divergences
Branch relabeling	Thematic divergence
Head-switching	Demotional divergence/ Promotional divergence
Fission/fusion	Lexical divergence/ Conflational divergence
Branch transposition	— [not considered]
—	Categorial divergence

– On the other hand, it is possible that an $\mathcal{L}_S$ sentence and its $\mathcal{L}_T$ equivalent differ "globally" in that the underlying common meaning is distributed between lexical items and syntactic constructions of both sentences in a completely different way, as in example (77).

(77) a. *Ceux qui persévèrent aboutissent.*
'Those who persevere succeed.'

b. Если упорно идти вперёд, то своего добьёшься.
Esli uporno idti vperëd, to svoego dob'ëš'sja.
IF PERSISTENTLY TO-GO FORWARD, THEN YOUR-DUE [YOU]-OBTAIN

Pairs of sentences of this type do manifest syntactic mismatches of the *n* : *m*-type, but the extent of mismatching is such that the sentences in question are simply incommensurable at the DSynt-level; they also reveal a semantic mismatch. In such a case, the source sentence must be analyzed down to its semantic representation and transfer must take place at the semantic level.

The four major types of mismatches correspond to divergences proposed in Dorr (1993) as shown in Table 3. Demotional divergence and promotional divergence are inverse with respect to each other, as are lexical and conflational divergence. Therefore, they constitute in fact a single mismatch type: in our terms, head-switching and fission/fusion, respectively. Dorr's classification does not contain the branch transposition type of mismatch. Her categorial divergence and structural divergence disappear from our typology due to the use of the DSyntS: they must be treated within a monolingual linguistic model.

The statistical distribution of the individual types of mismatches, which is interesting from the viewpoint of practical applications, is, of course, specific to each pair of languages involved. To ascertain these would require extensive study which we could not undertake in the scope of our current work.

6 Transfer in the paraphrasing paradigm

In this section, we introduce MTT's paraphrasing system and show how it can be applied in MT to handle syntactic mismatches.

6.1 The starting point: MTT's paraphrasing system

The main three features of the MTT paraphrasing system (Žolkovskij and Mel'čuk 1967; Mel'čuk 1974, pp 141–176, 1988a, pp 77–79, 1993; Milićević 2003), which was developed for intralinguistic paraphrasing, are: (i) lexical and structural paraphrasing means are strictly separated, (ii) lexical paraphrasing is based on LFs, and (iii) structural paraphrasing is based on the DSyntS.

Two major types of paraphrasing rules are distinguished: (a) lexical paraphrasing rules and (b) structural paraphrasing rules.

6.1.1 Lexical paraphrasing rules

A lexical paraphrasing rule expresses semantic (quasi-)equivalence between two configurations of LUs (in a DSyntS) whose syntactic properties might differ. Notationally, $L_1 \equiv L_2, L_1 \equiv L_2 - \mathbf{r} \rightarrow L_3$ (where **r** is a DSyntRel), etc. Each of these LU-configurations can be used instead of another in paraphrasing, provided that the necessary syntactic readjustments (in the DSynt-structures being processed) are carried out and all contextual constraints are respected.

All lexical equivalences at the level of DSyntS can be expressed in terms of LFs, as in examples (78)–(80).

(78) $R_{\mathbf{lex1}}$: $L_{(V)} \equiv \mathrm{Conv}_{21}(L_{(V)})$

In (78), $L_{(V)}$ stands for "an LU which is a verb" or a verbal expression of the form 'BE + Adv/Adj' (such as *be afraid*). For instance, *be afraid* ≡ *frighten*, *belong* ≡ *include*, *follow* ≡ *precede*. The verb *frighten* is a conversive of *be afraid*, with the permutation of deep-syntactic actants **I** and **II**; this relation is symbolized as FRIGHTEN = Conv_{21}(BE_AFRAID). The same holds for the other pairs above.

The rule in (79a) covers, for instance, the equalities be *quickly* ≡ *quickly* (79b), *usually* to ≡ *usually* (79c), *hurry* ≡ *in a hurry*. *Quickly* is an adverb equivalent of *be quick* that characterizes the action of deep-syntactic actant **I** of BE_QUICK (this is expressed by the right subscript "1" to Adv) and has the deep-syntactic actant **II** of BE_QUICK as its governor (this is expressed by the left superscript "II"). The other pairs show the same relationship.

(79) a. $R_{\mathbf{lex2}}$: $L_{(V)} \equiv {}^{\mathrm{II}}\mathrm{Adv}_1(L_{(V)})$

b. John was quick to react. ≡ John reacted quickly.

c. John used to swim. ≡ John usually swam.

The rule in (80) describes such cases as *be responsible for* ≡ *the responsibility lies with*, *be stressed* ≡ *the stress falls on*, *respond* ≡ *the response comes from*. The verb *lie*, which is the value of the LF Func_1 applied to the noun RESPONSIBILITY, which is a nominalization of BE_RESPONSIBLE; Func_1 is a light verb that takes RESPONSIBILITY as its subject and the deep-syntactic actant **I** of RESPONSIBILITY as its Obl(ique)Obj. As a result, the expression *X is responsible for Y* is semantically equivalent to the expression *the responsibility for Y lies with X*.[31]

(80) $\mathbf{R}_{\mathbf{lex}3}$: $\mathbf{L}_{(V)} \equiv \mathrm{S}_0(\mathrm{L}_{(V)}) \oplus \mathrm{Func}_1(\mathrm{S}_0(\mathbf{L}_{(V)}))$

[31] It has to be emphasized that all semantic equivalences between DSyntSs cited in this paper are considered up to the "Communicative" (or "Information") Structure. As a result, in two semantically equivalent DSyntSs, their Themes and their Rhemes may diverge. The communicative equivalence is ensured by *communicative* rules, which are not discussed in this paper.

To sum up: If two (configurations of) lexical units L_1 and L_2 are related by a lexical paraphrasing rule (i.e., if they are semantically equivalent), one can replace the other in any DSyntS, provided all relevant structural adjustments are carried out. These structural adjustments are expressed in terms of structural paraphrasing rules.

6.1.2 Structural paraphrasing rules

A structural paraphrasing rule expresses the correspondence between two subtrees of non-lexicalized deep-syntactic structures: $S_1 \leftrightarrow S_2$. This means that S_1 and S_2 can be mutually substituted when required by a lexical replacement. In other words, a structural paraphrasing rule is triggered by the application of a lexical paraphrasing rule. Consider the examples (81)–(83).

(81) a. R**struct1**: A←**I**-X-**II**→B ↔ B←**I**-Y-**II**→A
b. John is afraid of the consequences. ≡ The consequences frighten John.
c. JOHN [A]←**I**-IS AFRAID [X]-**II**→OF THE CONSEQUENCES [B]
d. THE CONSEQUENCES [B]←**I**-FRIGHTEN [Y]-**II**→JOHN [A]

(82) a. R**struct2**: X-**II**→B ↔ B-**ATTR**→ Y
b. John was quick to react. ≡ John reacted quickly.
c. JOHN WAS QUICK [X]-**II**→TO REACT [B].
d. JOHN REACTED [B]-**ATTR**→QUICKLY [Y]

(83) a. R**struct3**: A←**I**-X ↔ Y←**I**-Z-**II**→A
b. Terrorists are responsible for this attack. ≡ The responsibility for this attack lies with terrorists.
c. TERRORISTS [A]←**I**-ARE RESPONSIBLE [X] [FOR THIS ATTACK]
d. THE RESPONSIBILITY [Y] [FOR THIS ATTACK]←**I**-LIES [Z]-**II**→WITH TERRORISTS [A]

As mentioned above, lexical and structural paraphrasing rules are related. Each lexical rule has one or several structural rules associated with it; for each element of a lexical rule, we specify its correspondence with an element of the structural rule in question. For instance, the full specification of R**lex1** is as in (84).

(84) R**lex1**: R**struct1**$(X := L_{(V)};\, Y := \mathrm{Conv}_{21}(L_{(V)}))$

This is to be read as follows: **R**$_{\textbf{lex1}}$ is served by the structural rule **R**$_{\textbf{struct1}}$; the variable X in **R**$_{\textbf{struct1}}$ is instantiated by $L_{(V)}$ and Y by a 2/1-conversive of $L_{(V)}$.

The associated structural rules ensure the syntactic adjustments which are required by the application of the lexical rule.

The important advantages of such a paraphrasing system are:

(a) Simplicity: the rules are maximally simple from a linguistic viewpoint (although not necessarily elementary in the strict sense).
(b) Linguistic validity: each rule is linguistically valid in the sense that it corresponds to an intuitively perceived linguistic phenomenon.

(c) Universality: the universal nature of the DSyntS leads to an exhaustive calculus of logically possible types of lexical and structural paraphrasing rules. Lexical paraphrasing rules, formulated in terms of LFs, are expected to be sufficient for the description of all regular lexical equivalences in human languages. Structural paraphrasing rules, formulated in terms of elementary transformations of DSyntSs, are equally expected to be sufficient for the description of all possible syntactic transformations.

(d) Small number of rules: the paraphrasing system contains no more than 100 lexical and no more than 50 syntactic rules (in Mel'čuk 1992, less than 60 lexical rules and less than 40 structural rules are presented).

(e) Modularity and portability: individual rules can be modified, replaced or introduced without affecting the rest of the system.

(f) Precise formulation: the description of a rule in formal terms makes the system suitable for implementation.

We propose to use the paraphrasing system of MTT for transfer in MT in order to solve the problem of structural mismatches. More precisely, we propose to extend the technique of monolingual paraphrasing to the transfer of mismatching syntactic structures between the source and the target languages. This requires:

- additional standardization of rules;
- compiling a set of elementary syntactic operations and specifying a mechanism for the composition of structural paraphrasing rules from these operations;
- elaboration of a formal paraphrasing grammar.

6.2 Paraphrasing rules in transfer

In the monolingual scenario, paraphrasing rules are applied to the DSyntS S of a sentence Se in order to produce an equivalent DSyntS S', which will give rise to a paraphrase of Se. In the transfer scenario, paraphrasing rules are applied to the DSyntS S_S of the source sentence Se_S in order to produce an equivalent DSyntS S_T of the target language, which gives rise to the sentence Se_T, a translation of Se_S (in a more compact notation: $Se_T = \tau(Se_S)$).

Formally, paraphrasing of DSyntS S in the monolingual scenario may be considered as tree rewriting, which implies that the resulting DSyntS S' is, in a sense, the original S: lexico-syntactically modified but semantically the same. In other words, the paraphrasing is carried out on the original DSyntS. This is justified since the paraphrasing modifications most often affect only a small part of S, the rest being left untouched. Tree rewriting has been central to transfer-oriented MT for a long time, recently, especially in connection with the transfer of packed linguistic representations (Emele and Dorna 1998; Dymetman and Tendeau 2000). On the other hand, tree rewriting is also a specific case of graph rewriting. In theoretical computer science, graph transformation is commonly described in terms of graph grammars. We may thus interpret the DSynt-transfer paraphrasing grammar as a graph rewriting grammar. An early implementation of such a grammar for the German–Russian language pair is presented in Specht (2002).

However, in the interlingual scenario, paraphrasing is more appropriately viewed as tree transduction (more generally: graph transduction), which implies that the original DSyntS S_S is mapped onto the resulting DSyntS S_T such that S_T is a translational equivalent of S_S, with S_S being left untouched. Tree transduction is more appropriate for DSynt-based transfer in MT because of the following three observations:

(a) The left-hand side and the right-hand side of a transfer rule are defined over disjunct alphabets. Tree rewriting would thus lead to intermediate structures that are, strictly speaking, incorrect since they contain node labels from both $\mathcal{L}_S$ and $\mathcal{L}_T$.

(b) The application of transfer rules is unordered. Therefore, in order to avoid cases where a structural transformation required by a lexical equivalence is not applicable because (some of) the elements of its left-hand side image in S_S have already been substituted by elements from $\mathcal{L}_T$, tree rewriting would require that syntactic transformations be applied to "hybrid" structure fragments. This is, again, to be avoided.

(c) The substitution of a lexical configuration by its translational equivalent in S_S may lead to a linguistically invalid intermediate structure, which requires the application of further lexical translation equivalences in order to "get repaired". As a consequence, tree rewriting would require that lexical equivalences trigger not only structural transformations, but also further lexical equivalence substitutions. This would make transfer an unnecessarily complex procedure.

Therefore, we interpret the DSynt-transfer paraphrasing procedure as tree transduction. In formal terms, our DSynt-transfer paraphrasing is a graph transduction grammar whose theoretical foundations represent an extension of the graph rewriting grammars known from computer science.

6.2.1 Paraphrasing rule application in transfer

As we have seen above, in the traditional intralinguistic paraphrasing setting of MTT, an individual lexical paraphrasing rule triggers the application of one or several structural rules, with each of the latter implying one or several concomitant syntactic operations. In order to avoid redundancy, we suggest a fully compositional approach to the definition of a paraphrasing rule in the transfer scenario, such that a transfer rule π consists of a lexical equivalence statement $\mathbf{L_{lex}} \equiv \mathbf{R_{lex}}$ and a number of syntactic transformations of the form $\mathbf{L_{synt}} \leftrightarrow \mathbf{R_{synt}}$. Before we continue with the formalization of the notion of paraphrasing rules, let us sketch the application of a paraphrasing rule π to a given DSyntS S_S. The application is performed in six steps:

1. Identify an image of $\mathbf{L_{lex}}$ in S_S.
2. Check the conditions that must be fulfilled for π to be applicable.
3. Introduce the image of $\mathbf{R_{lex}}$ into the fragment of the target DSyntS already obtained by previous transfer rule applications; let the resulting forest be named $\bar{S}_T$.
4. Identify the image of $\mathbf{L_{synt}}$ in S_S and the images of the nodes of $\mathbf{R_{synt}}$ in $\bar{S}_T$ for each syntactic transformation associated with the given lexical equivalence $\mathbf{L_{lex}} \equiv \mathrm{R}_{\mathbf{lex}}$.
5. Connect the images of the nodes of $\mathbf{R_{synt}}$ in $\bar{S}_T$ for each syntactic transformation associated with the given lexical equivalence in accordance with the DSyntRels in $\mathbf{R_{synt}}$; let the resulting graph (which may still be a forest) be named S'_T.
6. Mark the image of $\mathbf{L_{lex}}$ in S_S as transduced.

In a realistic application, any paraphrasing transfer rule π is applied to a genuine *substructure* of a source DSyntS S_S, along with structural transformations. Therefore, another type of rule—of a more general nature—proves necessary for DSyntS-paraphrasing: rules that specify how to deal with arcs contiguous to the nodes affected by a paraphrasing rule. They may need to be "reattached" to different nodes of the image of $\mathbf{R_{synt}}$ than before. We call the procedure of reattaching "syntactic tree adjustment".

Taking the syntactic tree adjustment into account, the application of a paraphrasing rule π to a given DSyntS S_S must be extended by a seventh step:

7. Adjust the image of $\mathbf{R_{synt}}$ in S'_T.

As will become clear later, the application of paraphrasing transfer rules of an STE is order-independent. However, we must take into account that

- the $\mathbf{L_{lex}}$s of several rules may overlap or even coincide;
- it cannot be taken for granted that the application of a subset of rules leads to a well-formed translation of a given DSyntS$_S$.

In order to avoid erroneous or incomplete translations, the STE uses a "clustering" mechanism that has extensively been tested with a graph transduction grammar implemented for text generation (Bohnet and Wanner 2001). Before any rule is applied to the source DSyntS$_S$, the clustering mechanism first groups rules that are compatible (i.e., whose $\mathbf{L_{lex}}$s do not overlap) and that cover a connected fragment of the DSyntS$_S$; then it selects those compatible rule groups that fully cover the DSyntS$_S$. Applying the selected rule groups in sequence, the STE obtains alternative well-formed and valid translations of DSyntS$_S$.

In what follows, we focus on individual paraphrasing rules. We define the lexical equivalences and the syntactic transformations triggered by them in terms of "elementary paraphrasing rules." Elementary paraphrasing rules are extended to paraphrasing rules proper in order to accomodate for the necessary syntactic tree adjustments.

6.2.2 Elementary paraphrasing rules

Let us start with the formal notion of elementary paraphrasing rule and then illustrate its central elements.

Notion of elementary paraphrasing rule: An elementary paraphrasing rule is defined over the bilingual lexical index (BLI) as a quadruple that consists of a lexical equivalence rule, the set of syntactic operations triggered by the lexical equivalence, a set of correspondences between the variables of the lexical equivalence and syntactic operations, and a (potentially empty) set of conditions that restrict the application of the rule. The conditions may concern grammatical or semantic features of the elements of $\mathbf{L_{lex}}$ or $\mathbf{R_{lex}}$ or the context of $\mathbf{L_{lex}}$ or $\mathbf{R_{lex}}$ within DSyntS$_S$, or DSyntS$_T$ respectively

Definition 4 (Elementary paraphrasing rule). Let lexeq$_{\pi_e}$:= ($\mathbf{L_{lex}} \equiv \mathbf{R_{lex}}$) be a lexical equivalence defined over the BLI, Sy$_{\pi_e}$:= {$\mathbf{L_{synt}} \leftrightarrow \mathbf{R_{synt}}$} the set of syntactic operations triggered by lexeq$_{\pi_e}$, and Φ a pair of correspondences (ϕ_L, ϕ_R), where ϕ_L establishes the correspondences between the elements of all $\mathbf{L_{synt}} \in$ Sy$_{\pi_e}$ and the elements of $\mathbf{L_{lex}}$, while ϕ_R establishes the correspondences between the elements of all

$\mathbf{R}_{\mathbf{synt}} \in \mathrm{Sy}_{\pi_e}$ and the elements of $\mathbf{R}_{\mathbf{lex}}$. Then, $\pi_e := (\mathrm{lexeq}_{\pi_e}, \mathrm{Sy}_{\pi_e}, \Phi)$ is an *elementary paraphrasing rule* without application conditions. If there are any conditions Co that must be fulfilled for a π_e to be applicable, then $\pi_e := (\mathrm{lexeq}_{\pi_e}, \mathrm{Sy}_{\pi_e}, \Phi, \mathrm{Co})$.

Let us now discuss first the sets of lexical equivalences and universal syntactic operations over which elementary paraphrasing rules are defined and then the syntactic adjustments.

Interlinguistic Lexical Equivalences: An elementary lexical equivalence is a generalization over the BLI; it specifies—in terms of LFs—which target LU(s) can be substituted for a given source LU L_i, in any given sentence *Se* in order to obtain a translation of *Se*. More precisely, it establishes a semantic equivalence either

A. between an LU L_S and an LU L_T given by a lexical-functional expression of the form $f(L'_T)$ (or of the form $f_1(L'_T) \oplus f_2(L'_T)$), or
B. between LUs specified as LF-expression or (LF-configuration expressions) in $\mathcal{L}_S$ and $\mathcal{L}_T$, respectively.

Lexical equivalences of type A are further subdivided into five different subclasses:[32]

[A1] **Synonymic equivalences:** a lexical unit L_S is replaced by its "synonym" in $\mathcal{L}_T$, which can be written as (85).

(85) $L_S \equiv \mathrm{Syn}(L_S, \mathcal{L}_T)$

This lexical equivalence rule captures all fully matching translation equivalence LU pairs in the BLI (cf. case 1, in the description of the BLI, Sect. 2, p 10), for example French–German (AMOUR, LIEBE) 'love,' German–English (PFLANZE, PLANT).

[A2] **Antonymic equivalences:** a lexical unit L_S is replaced in $\mathcal{L}_T$ by its antonym (expressed in terms of the LF 'Anti') plus a negative expression; cf. (86).

(86) $L_S \equiv \mathrm{Anti}(L_S, \mathcal{L}_T)$-$\mathbf{ATTR}\!\rightarrow$NEG

In English, typical instances of NEG are: *not, far from*, *without*, *less than*; in French, one of the instances of NEG is *peu* 'little'.

This lexical equivalence rule covers the Anti-translations in the BLI; cf. English–German–Russian–French (SHALLOW, SEICHT, МЕЛКИЙ [*melkij*], PROFOND, Anti), the correspondence already cited above, and French–English (IGNORER ['not-know'], KNOW, Anti) as in (87), with L_S instantiated as French IGNORER and Anti(L_S, English) as KNOW (Fig. 14).

(87) *J'ignore ces données.*
I don't know these data.

[A3] **Conversive equivalences:** a lexical unit L_S is replaced by its conversive Conv_{ijkl}, the most common being the 2/1-conversion (88).

(88) $L_S \equiv \mathrm{Conv}_{21}(L_S, \mathcal{L}_T)$

[32] To adjust the notation of LFs to the interlingual scenario, we extend the domain of LFs by a further dimension: LF: $V_S \times \mathcal{L} \rightarrow V_T$, where V_S is the source language vocabulary, $\mathcal{L}$ is the set of target languages, and V_T is the target language vocabulary.

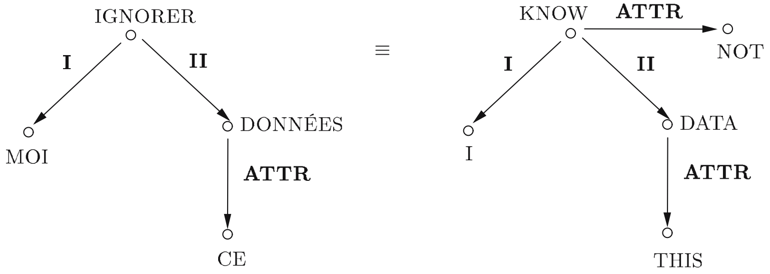


Fig. 14 Equivalent DSyntSs of sentences in (87)

Obviously, this lexical equivalence rule accounts for the entries in the BLI whose translation equivalents are conversives:[33] for example, Spanish–English (PARECER, THINK, Conv_{21}), (COBRAR, PAY, Conv_{21}), as in (89).

(89) a. *¿Qué te parece este libro?* WHAT TO-YOU SEEMS THIS BOOK?
What do you think of this book?

b. *Nos cobras.* FROM-US YOU-RECEIVE.
We pay (you).

[A4] **Derivative equivalences:** a lexical unit L is replaced by one of its derivatives: $L_S \equiv \mathrm{DER}(L_S, \mathcal{L}_T)$, with DER $\in$ {Sing, Mult, S_i, Adv_i, ...} (i = 1,2,...), that is, as in (90).

(90)
$$
\begin{aligned}
&L_S && = \equiv = \mathrm{Sing}_{\mathrm{pl}}(L_S, \mathcal{L}_T)\\
&L_{S_{(V)}} && \equiv \ \mathrm{Oper}_1(\mathrm{S}_1(L_S, \mathcal{L}_T), \mathcal{L}_T) \ \textbf{-II}\!\to \mathrm{S}_1(L_S, \mathcal{L}_T)\\
&L_{S_{(Adv)}} && \equiv \mathrm{V}_0(L_S, \mathcal{L}_T)\\
&L_{S_{(V)}} && \equiv \mathrm{Adv}_1(L_S, \mathcal{L}_T)\\
&\ldots
\end{aligned}
$$

Samples from the BLI that are covered by the lexical equivalence rules in (90) include German–English (GESCHWISTER, BROTHERS-AND-SISTERS, $\mathrm{Sing}_{\mathrm{pl}}$), French–Russian (SOIGNER ['treat'], лечащий врач [*lečaščij vrač* 'treating doctor'], S_1), and English–French (GO OUT, SORTIR ['go-out'], V_0), as in (91).

(91) a. *Die Geschwister trafen sich regelmäßig.*
THE BROTHERS-AND-SISTERS MET refl. REGULARLY
The brothers and sisters met regularly.

b. *Qui la soigne?* WHO HER(obj.) TREATS
Кто её лечащий врач? *Kto eë lečaščij vrač?*
WHO-IS HER(gen.) TREATING DOCTOR?

c. He crawled out of the den.
Il est sorti de la tannière en rampant.
HE IS GONE-OUT OF THE DEN BY CRAWLING

[33] The operation of conversion implies the permutation of DSyntAs with respect to the corresponding SemAs; elimination and introduction of DSyntAs may also occur. If we consider only LUs with no more than four actants (which is the most current case) and disregard elimination/introduction of DSyntAs, the maximal number of theoretically possible elementary conversive equivalences is thus 4! = 24. However, most of these equivalences do not occur in natural languages.

Crawl out is translated into French as *sortir en rampant* by the application of two lexical equivalence rules: $L_S \equiv \text{Syn}(L_S, \mathcal{L}_T)$ (to account for the translation of *crawl*), and $L_{S_{(Adv)}} \equiv \text{V}_0(L_S, \mathcal{L}_T)$ and its accompanying syntactic operations (to account for the translation of *out*).[34] In general, these two lexical equivalences suffice to ensure the translation of all English constructions of the type $L_{(V)}\textbf{-II}\rightarrow\text{Prep}_{\text{dir}}$, where $L_{S_{(\text{V})}}$ is a verb denoting the manner in which a movement is performed and Prep_{dir} is a preposition expressing the direction of the movement, into French, as in (92) for example.

(92) swim across ≡ *traverser en nageant* ‘cross by swimming’
fly into ≡ *entrer en volant* ‘enter by flying’
run out of ≡ *sortir en courant* ‘go-out by running’

The English directional adverb is translated as a French main verb;[35] this entails the structural head-switching operation and the transformation of the English main verb into Adv_1 (realized in French by an *en V-ant* construction).

[A5] **Collocational equivalences:** a lexical unit L_S is replaced by a collocation, i.e., semantically equivalent to L_S and consists of $\text{DER}(L_S)$ and an LF $(\text{DER}(L_S))$. A finite set of collocational equivalences is available. The most frequent equivalence is the one that involves such LFs as Oper_I and Labor_{32} (93).

(93)
$$L_{S_{(\text{V})}} \equiv \text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T) \leftarrow\textbf{II}\text{-Oper}_1(\text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T))$$
$$L_{S_{(\text{V})}} \equiv \text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T) \leftarrow\textbf{IV}\text{-Labor}_{32}(\text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T))$$

This equivalence accounts for such entries in the BLI as French–English (SE_SUICIDER$_V$, SUICIDE$_N$, S_0), where *se suicider* is lit. ‘to suicide oneself’; and German–Russian (MIETEN, PROKAT, S_0), where *mieten* is the verb ‘to rent’, and прокат *prokat* the noun ‘rent’. Oper_1(SUICIDE) = *commit* [∼] and Labor_{32}(PROKAT) = взять на ∼ (with взять на *vzjat’ na* ‘take on’) are added as support verbs by structural rules. Other collocation equivalences include Labor_{12}, Real_1, and Labreal_{12} as in (94).

(94)
$$L_{S_{(\text{V})}} \equiv \text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T)\leftarrow\textbf{III}\text{-Labor}_{12}(\text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T))$$
$$L_{S_{(\text{V})}} \equiv \text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T)\leftarrow\textbf{II}\text{-Real}_1(\text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T))$$
$$L_{S_{(\text{V})}} \equiv \text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T)\leftarrow\textbf{III}\text{-Labreal}_{12}(\text{S}_0(L_{S_{(\text{V})}}, \mathcal{L}_T))$$

For the first of the three equivalences in (94), a sample BLI-entry is Russian–German (выразить, AUSDRUCK, $\text{S}_0\oplus\text{Labor}_{12}$), where выразить *vyrazit’* is the verb ‘to express’, while *Ausdruck* is the noun ‘expression’ as exemplified in (95).

(95) a. В этой речи, он выразил свои сомнения в успешном исходе дела.
V ètoj reči, on vyrazil svoi somnenija v uspešnom isxode dela.
IN THIS SPEECH HE EXPRESSED HIS DOUBTS IN SUCCESSFUL OUTCOME OF-ENTERPRISE

[34] The conditions defined for the corresponding paraphrasing rules ensure their correct sequential application.

[35] In accordance with MTT, we consider ACROSS, INTO, OUT_OF, etc. to be deep adverbs.

b. *In dieser Rede brachte er seine Zweifel bezüglich eines positiven Ausgangs des Unternehmens zum Ausdruck.*
IN THIS SPEECH BROUGH HE HIS DOUBTS REGARDING A POSITIVE OUTCOME OF-THE ENTERPRISE TO EXPRESSION
'In this speech he expressed his doubts about the successful outcome of the enterprise.'

Labor$_{12}$(AUSDRUCK) = *bringen* [*zum* ∼] lit. 'bring [to ∼]' is introduced as support verb by auxiliary syntactic rules.

The second and third equivalences in (94) are instantiations of the lexical mismatch type discussed as case (iii) in Sect. 5.2.1: a fused value of an LF **f**, //**f**(L_S), corresponds to a non-fused value of **f**, **f**(L_T)⊕L_T, when this is applied to the translation equivalent of L_S, that is L_T. Thus the verb *to bottle* [= //Labreal$_{12}$(BOTTLE$_N$)] is equivalent to French *mettre en bouteilles* lit. 'put into bottles' [= Labreal$_{12}$(BOUTEILLE)-**III**→BOUTEILLE$_{\mathrm{pl}}$] as in (96).[36]

(96) Bottled in Château Blaignan
Mis en bouteilles au Château Blaignan

Given that the corresponding DSyntSs contain fused LF-expressions as node labels (//Labreal$_{12}$(BOTTLE$_N$), //Real$_1$(DUSCHE) 'shower', etc.), the BLI-entries for these equivalences are standard entries that do not show any divergence between $\mathcal{L}_S$ and $\mathcal{L}_T$, as in English–French (BOTTLE$_N$,BOUTEILLE), German–Russian (DUSCHE, DUŠ) (where *Dusche* and душ *duš* both correspond to the noun 'shower').[37]
The lexical equivalences of type **B** (see above), i.e., of equivalences between LFs, include, among others, the two equivalences in (97) (see Mel'čuk 1992 for a relatively complete list).

(97)
$$\begin{aligned} &\mathrm{Func}_2(L_S) &&\equiv \mathrm{Oper}_1(L_S,\mathcal{L}_T)\\ &\mathrm{Labor}_{12}(L_S) &&\equiv \mathrm{Oper}_2(L_S,\mathcal{L}_T)\end{aligned}$$

To account for these equivalences, the translation of the keywords in the BLI suffices, as in English–French (THANKS, REMERCIEMENTS) or Russian–German (BOMBARDIROVKA, BOMBARDEMENT) exemplified in (98).

(98) a. My thanks go to John.
J'exprime mes remerciements à John.
I-EXPRESS MY THANKS TO JOHN.

b. Американская авиация подвергла город трёхдневной массированной бомбардировке.
Amerikanskaja aviacija podvergla gorod trëxdnevnoj massirovannoj bombardirovke.
AMERICAN AVIATION SUBJECTED CITY TO-THREE-DAY MASSIVE BOMBARDMENT
Die Stadt lag drei Tage lang unter einem massiven Bombardement der amerikanischen Luftwaffe.
THE CITY LAY THREE DAYS LONG UNDER A MASSIVE BOMBARDMENT OF-THE AMERICAN AIRFORCE

[36] See also example (60) above, for illustration.

[37] The Labreal$_{12}$(BOUTEILLE) 'bottle' construction requires BOUTEILLE to be in the plural; this is indicated in the monolingual French ECD.

The transformation of the syntactic constructions that accompany the substitution of one LF by an equivalent LF (such as $\mathrm{Func}_2(L_S)$-**I**→ L_S⇒ $\mathrm{Oper}_1(L_S, \mathcal{L}_T)$-**II**→ $(L_S, \mathcal{L}_T)$) is taken care of by the associated structural rules.

Further common lexical collocational equivalences of type **B** include those in (99).

(99)
$$\begin{aligned}
&\mathrm{Oper}_1(L_S) &&\equiv \mathrm{Oper}_2(L_S, \mathcal{L}_T)\\
&\mathrm{Func}_1(L_S) &&\equiv \mathrm{Func}_2(L_S, \mathcal{L}_T)\\
&\mathrm{Labor}_{12}(L_S) &&\equiv \mathrm{Labor}_{32}(L_S, \mathcal{L}_T)\\
&\mathrm{Oper}_1(L_S) &&\equiv \mathrm{Func}_1(L_S, \mathcal{L}_T)\\
&\mathrm{Oper}_1(L_S) &&\equiv \mathrm{Labor}_{12}(L_S, \mathcal{L}_T)\\
&\mathrm{Real}_1(L_S) &&\equiv {}^{II}\mathrm{Adv}_1\mathrm{Real}_1(L_S, \mathcal{L}_T)
\end{aligned}$$

All these LF-equivalences can be reduced to different types of conversion: $\mathrm{Oper}_2(L_S) \equiv \mathrm{Conv}_{321}(\mathrm{Oper}_1(L_S, \mathcal{L}_T))$, etc. However, for simplicity's sake, we spell out the direct LF-equivalences explicitly.

Universal Elementary Syntactic Operations: Elementary syntactic operations triggered by individual lexical equivalence rules fall into three classes: (i) branch relabeling, (ii) branch inversion, and (iii) branch transposition.[38] Branch transposition subdivides further into two subcases: (iii.a) branch raising and (iii.b) branch lowering.

(i) Branch relabeling is defined as in (100),

(100) $X_S \xrightarrow{r_i} Y_S \leftrightarrow X_T \xrightarrow{r_j} Y_T, \quad r_i \neq r_j$

where X_S/Y_S are source language lexical unit variables, X_T/Y_T the translation equivalent variables of X_S/Y_S, and $r_i, r_j \in$ {**I**, **II**, ..., **VI**, **ATTR**, **COORD**, **APPEND**}.

The most common of the relabeling operations are the two relabelings that express the **I-II** and **II-I** conversions, i.e., those that are associated with the lexical equivalence rule $L \equiv \mathrm{CONV}_{21}(L, \mathcal{L}_T)$, see (101).

(101) X_S-**I**→ Y_S ↔ X_T- **II**→Y_T
X_S-**II**→ Y_S ↔ X_T-**I**→Y_T

A clear case of this kind of relabeling is observed for BLI-equivalences between personal intransitive verbs in English and impersonal transitive verbs in Russian and German such as *be sick* тошнить *tošnit'* 'vomit' expressed as (BE_SICK, TOŠNIT', $\mathrm{Conv}_{\Delta 1}$); and English–German (SHAKE, SCHÜTTELN, $\mathrm{Conv}_{\Delta 1}$) (here, Δ stands for "dummy actant": English IT, German ES, Russian Ø_{3SG}), as in (102).

(102) a. I am sick.
Меня тошнит. *Menja tošnit.* TO-ME IT-VOMITS
I←**I**-BE_SICK ↔ JA←**II**-TOŠNIT'

b. I am shaking.
Mich schüttelt es. ME SHAKES IT
I←**I**-SHAKE ↔ ICH←**II**-SCHÜTTELN

[38] Recall that these syntactic operations are used together with lexical equivalence rules of the type $L_S \equiv L_T$ to ensure the well-formedness of the resulting DSyntS and, most importantly, its semantic equivalence with the starting DSyntS (which could be affected by the syntactic divergences between L_S and L_T).

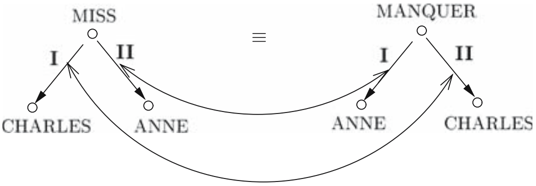


Fig. 15 Equivalent DsyntSs of sentences in (103)

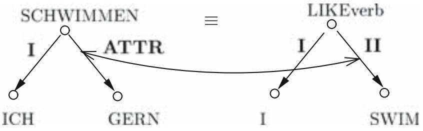


Fig. 16 Equivalent DsyntSs of sentences in (106a)

A more general case of conversion is the widely cited equivalence (MISS, MANQUER, Conv_{21}) as in (103) and Fig. 15.

(103) Charles misses Anne.
Anne manque à Charles. ANNE LACKS TO CHARLES

The structural operations involved here are shown in (104).

(104) $X_S\text{-}\mathbf{I}\!\rightarrow Y_S = \leftrightarrow = X_T\text{-}\mathbf{II}\!\rightarrow Y_T$
$X_S\text{-}\mathbf{II}\!\rightarrow Y_S \leftrightarrow X_T\text{-}\mathbf{I}\!\rightarrow Y_T.$

(ii) **Branch inversion** is defined as in (105a). The most frequent branch inversions include those in (105b, c).

(105) a. $Y_S \xleftarrow{r_i} X_S \leftrightarrow Y_T \xrightarrow{r_j} X_T,$
b. $X_S\text{-}\mathbf{ATTR}\!\rightarrow Y_S \leftrightarrow Y_T\text{-}\mathbf{II}\!\rightarrow X_T,$
c. $X_S\text{-}\mathbf{ATTR}\!\rightarrow Y_S \leftrightarrow Y_T\text{-}\mathbf{III}\!\rightarrow X_T.$

Transformations of this type are triggered by such derivative lexical equivalence rules as $L_{S_{(V)}} \equiv {}^{II}\text{Adv}_1(L_S,\mathcal{L}_T)$ and $L_{S_{(Adv)}} \equiv \text{V}_0(L_S,\mathcal{L}_T)$, exemplified by English–German *like* vs. *gern* 'with pleasure' example (3), repeated here as (106a) for convenience; and English–Russian *used to* vs. обычно *obyčno* 'usually' (106b), shown in Figs. 16 and 17, respectively.

(106) a. (LIKE, GERN, ${}^{II}\text{Adv}_1$)
I like swimming.
Ich schwimme gern. I SWIM WITH-PLEASURE
b. (USED_TO, OBYČNO, ${}^{II}\text{Adv}_1$)
He used to take the train
Он обычно ездил на поезде.
On obyčno ezdil na poezde.
HE USUALLY RODE ON TRAIN

Strictly speaking, branch inversion as presented above subsumes two different operations: (a) branch inversion *per se* and (b) branch relabeling.

Note that derivative lexical equivalence does not necessarily involve succeeding structural operations; consider, for instance, the equivalence $L_S \equiv \text{Mult}(L_{S_{sg}},\mathcal{L}_T)$ that

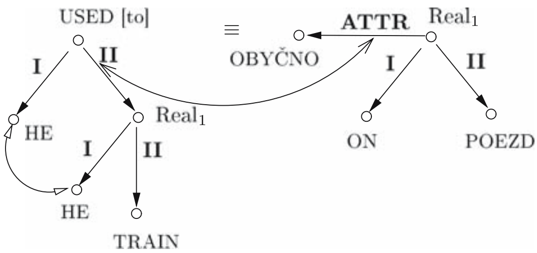


Fig. 17 Equivalent DsyntSs of sentences in (106b)

applies for example, to English–French (VOTER, ÉLECTORAT, Mult), where *électorat* corresponds to 'electorate', as in (107). No structural operations are involved in the realization of this equivalence.

(107) The voters decided differently.
L'électorat a décidé autrement.
THE ELECTORATE HAS DECIDED DIFFERENTLY

(iii) **Branch transposition** is defined as in (108),

(108) $Y_S \xrightarrow{q_i} X_S \xrightarrow{r_i} T_S \leftrightarrow X_T \xleftarrow{q_j} Y_T \xrightarrow{r_j} T_T$

where T_S, T_T are subtrees and q is a deep-syntactic relation not affected by the transposition. Branch transposition moves a subtree from its governor either one level up—to this governor's governor—or one level down—to one of its governor's dependents. In other words, if the operation is applied from left to right it corresponds to branch raising; applied in the opposite direction, it describes branch lowering.

Frequent transpositions include, for instance those listed in (109).

(109) Y_S-**II**→ X_S-**I**→T_S ↔ X_T ←**II**-Y_T-**III**→T_T
Y_S-**II**→ X_S-**I**→T_S ↔ X_T ←**II**-Y_T-**I**→T_T
Y_S-**I**→ X_S-**I**→T_S. ↔ X_T ←**II**-Y_T-**I**→T_T

They are triggered, for instance, by some special synonymous lexeme equivalences or LF-equivalences. For instance, the first can be triggered by synonymous lexical equivalences of the type (WASH, LAVER), (TEAR, DÉCHIRER).

More precisely, this type of equivalence is triggered by the synonymous lexical equivalence between an English verb that denotes physical impact (such as WASH, TEAR) and the corresponding verb, for example, in French, when its DirO is the name of a bodypart.[39] Consider for illustration the pair (110), Fig. 18 (cf. also (74) in Sect. 5.2.5).

(110) I wash Paul's hands.
Je lave les mains à Paul. I WASH THE HANDS TO PAUL

[39] Strictly speaking, such a generalized synonymic equivalence is better described in terms of an elementary syntactic equivalence rather than in terms of a lexical equivalence. The introduction of such elementary syntactic equivalences whose application can trigger structural transformations (in the same way lexical equivalences do) will be necessary in a more comprehensive presentation of the STE.

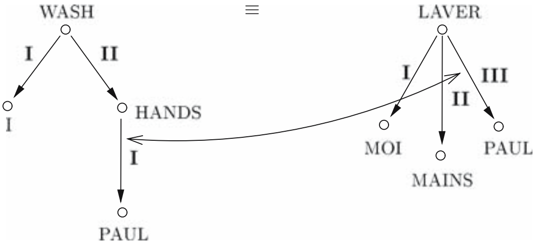


Fig. 18 Equivalent DsyntSs of sentences in (110)

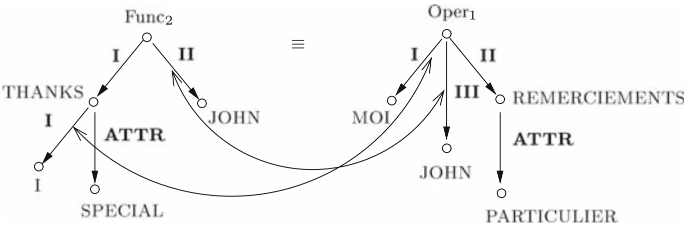


Fig. 19 Equivalent DsyntSs of sentences in (111)

The second transposition is applied in the context of synonymous lexical equivalences of the type French *sembler* to Russian казаться *kazat'sja* 'seem', as illustrated in (75).

The third of the transpositions in (109) comes to bear after the LF-equivalence Func$_2$ ≡ Oper$_1$ has been applied as in (111a, b). The structural transformation applied here is shown in (111c). Fig. 19 shows the corresponding DsyntSs.

(111) a. My special thanks go to John.

b. *J'exprime des remerciements particuliers à John*
I-EXPRESS SOME THANKS SPECIAL TO JOHN.

c. Func$_2$-**I**→THANKS-**I**→I ↔ MOI←**I**-Oper$_1$-**II**→REMERCIEMENTS

Consider finally an example of both reversed and transposed branch equivalence, in (112) and Fig. 20.

(112) Он закончил свою речь призывом к молодым.
On zakončil svoju reč' prizyvom k molodym.
HE ENDED HIS SPEECH BY-APPEAL TO YOUNG-ONES
At the end of his speech, he appealed to youth.

Note that the pair of structures in Fig 20 manifests another non-trivial lexical equivalence, covered by the rule $L_S \equiv \mathrm{Mult}(L_{S_{\mathrm{sg}}}, \mathcal{L}_T)$: MOLODOJ$_{\mathrm{pl}}$ ≡ YOUTH.

6.2.3 Syntactic tree adjustments

Syntactic tree adjustment transformations ("elementary adjustments", for short) are connection maintenance transformations triggered by equivalence transformations. Adjustments ensure the well-formedness of the resulting structure after the

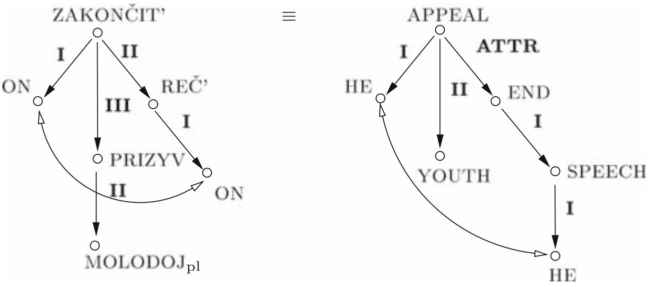


Fig. 20 Equivalent DzyntSs of sentences in (112)

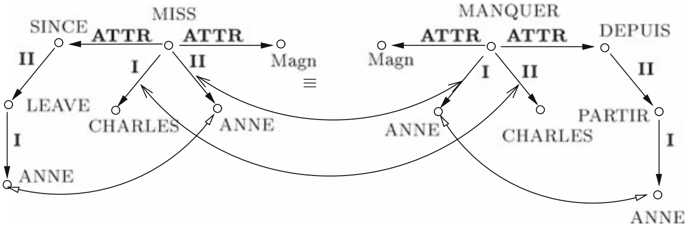


Fig. 21 Equivalent DsyntSs of sentences in (113)

application of an elementary lexical equivalence transformation. Three major cases of syntactic adjustments must be distinguished.

[1] A lexical unit L_S (i.e., an element of $\mathbf{L}_{lex}$) is replaced by an LU L_T (i.e., $\mathbf{R}_{lex}$) without inversion of syntactic dependency between L_S and L_T; this case covers synonymic and conversive lexical equivalences: simple or involving transpositions of subtrees, but not involving head-switching. It does not require any adjustment apart from the straightforward reattachment.

Type 1 adjustment rule: All contiguous branches of L_S are reattached to L_T without any change.

For illustration of this type of adjustment, let us modify example (103) as (113) and Fig. 21.

(113) Charles misses Anne terribly [since she left].
Anne manque à Charles énormément [*depuis* …].
ANNE LACKS TO CHARLES ENORMOUSLY [SINCE …].

As Fig. 21 shows, the $\mathcal{L}_T$ DSynt-substructures that correspond to -**ATTR**→Magn (with Magn(MISS) = *terribly*) and -**ATTR**→SINCE-**II**→LEAVE-**I**→ANNE are reattached without any change to MANQUER.

[2] A lexical unit L_S is replaced by an LU L_T with the inversion of syntactic dependency between L_S and L_T; this case covers lexical equivalences that entail head-switching. The following adjustments are in order:

Type 2 adjustment rule

1. If L_S has a DSyntA **I**, it is reattached to the new head.
2. As far as modifiers are concerned, two types of modifiers must be distinguished: modifiers of the first type characterize its governor, so to speak, internally,

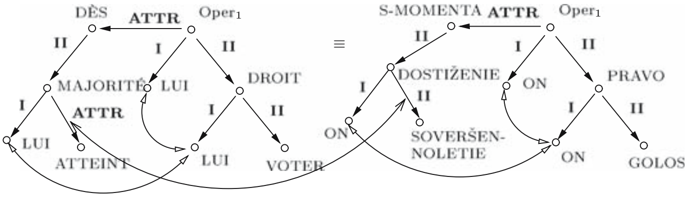


Fig. 22 Equivalent DsyntSs of sentences in (114)

qualifying or quantifying it; modifiers of the second type characterize it externally, specifying its coordinates in space and time, its evaluation by the speaker, etc. Accordingly, two cases must be dealt with:

(a) Outgoing **ATTR**-branches of DSyntA **II** of L_S (if L_S is a verb) or the governor of L_S (if it is an adverb) that characterize the event externally. They are reattached to the governor of L_T (if L_S is a adverb) or to DSyntA **II** of L_T (if L_T is a verb).

(b) Outgoing **ATTR**-branches of DSyntA **II** of L_S/the governor of L_S that characterize the event internally as well as the actantial branches. They are not affected by the adjustment.

The French–Russian pair in (114) and Fig. 22 shows this type of adjustment.[40]

(114) a. *Il a le droit de voter dès sa majorité atteinte.*
HE HAS THE RIGHT OF TO-VOTE SINCE HIS MAJORITY REACHED

b. Он имеет право голоса с момента достижения им совершеннолетия.
On imeet pravo golosa s momenta dostiženija im soveršennoletija.
HE HAS RIGHT OF-VOTE FROM MOMENT OF-REACHING BY-HIM MAJORITY

'He has the right to vote since reaching majority.'

The translation of *majorité* as совершеннолетие *soveršennoletie* requires the inversion and relabeling of the dependency relation between MAJORITÉ and ATTEINT. As a consequence, the first actant of the equivalent of MAJORITÉ is reattached to достижение *dostiženie* 'reaching' (the new head), as is the equivalent of the subtree modifying MAJORITÉ.

[3] A lexical unit L_S is replaced by a binary subtree $L_{T_1}\text{-}\mathbf{r}\rightarrow L_{T_2}$. This is the most complex case of adjustment. All dependents of L_S have to be distributed between L_{T_1} and L_{T_2}, the distribution being controlled by semantics. We cannot delve here into all semantic subtleties, but, roughly speaking, the adjustment rules for this case are as follows:

Type 3 adjustment rule

1. actant **I** of L_S is reattached to L_{T_1}.
2. Actants **II–VI** of L_S are reattached to L_{T_2} if $L_{T_2} = \mathrm{DER}(L_{T_1})$, otherwise they are reattached to L_{T_1}.

[40] In this example, ATTEINDRE 'reach' (as a past participle *atteinte*) and достичь *dostič'* 'reach' (as the noun достижения *dostiženija*) are elements of the value of the LF IncepOper$_1$(MAJORITÉ/совершеннолетие) *soveršennoletie* 'majority'. To simplify the presentation of the example, we did not make this explicit in the corresponding figure.

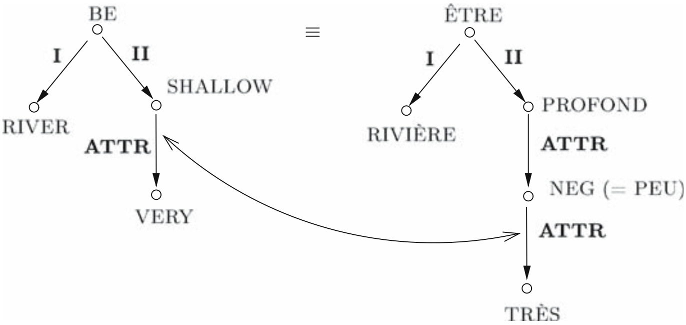


Fig. 23 Equivalent DsyntSs of (115)

3. **ATTR**-nodes that characterize L_S internally are reattached to L_{T_2}; those that characterize L_S externally, are reattached to L_{T_1}.

Adjustments of this type must be performed, for instance, in the case of antonymic equivalences. If a source adjectival lexeme L_{A_S} is paraphrased by its antonymic equivalence "Anti(L_{A_S})-**ATTR**→ NEG", any modifier of L_{A_S} becomes a modifier of the negation NEG. Let us give an example (115) with the equivalence (SHALLOW, PROFOND, Anti) already referred to above (see example (25a) and discussion thereof), as seen in Fig. 23.

(115) The river is very shallow.
La rivière est très peu profond.
THE RIVER IS VERY LITTLE DEEP

The equivalent of VERY, i.e., TRÈS, is reattached to the negation lexeme (PEU 'little'). This adjustment can be generalized as follows:

if an adjectival L_{A_S} corresponds to an adjectival or a nominal phrase: L'_{Adv_T}←**ATTR**-L_{A_T} or L'_{A_T}←**ATTR**-L_{N_T}
then the translation equivalent of any modifier of L_{A_S} is a modifier of L'_{Adv_T}

Adjustments of type 3 are also necessary in the case of collocational equivalences. Let us assume that a source verbal lexeme L_{V_S} corresponds to a verbal phrase $\tilde{L}_{V_T}$-$r \rightarrow L_{N_T}$, where $\tilde{L}_{V_T}$ is an LF of the type Oper or Real. Then:

- the translation equivalent of DSyntA **I** of L_{V_S} is the DSyntA **I** of $\tilde{L}_{V_T}$;
- the translation equivalents of DSyntAs **II**, **III**, … are the DSyntAs **II**, **III**, … of L_{N_T} if the government pattern (GP) of L_{N_T} admits them, or of $\tilde{L}_{V_T}$, again, if the GP of $\tilde{L}_{V_T}$ admits them; if both GPs admit these actants, optional variants of actant distribution are available;
- the translation equivalents of any qualifying modifiers are reattached to L_{N_T};
- the translation equivalents of any circumstantial modifiers are reattached to $\tilde{L}_{V_T}$.

The corresponding adjustments are illustrated, for example, by (116)–(118) and Figs. 24–25.

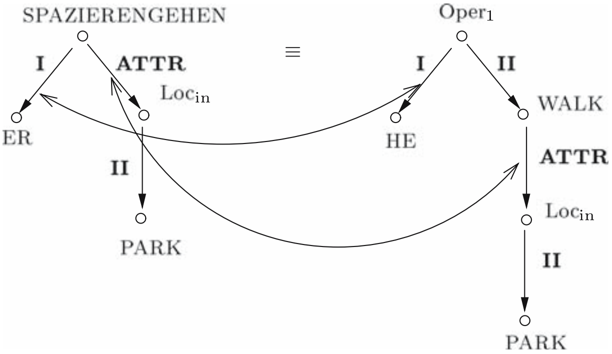


Fig. 24 Equivalent DsyntSs of sentences in (117)

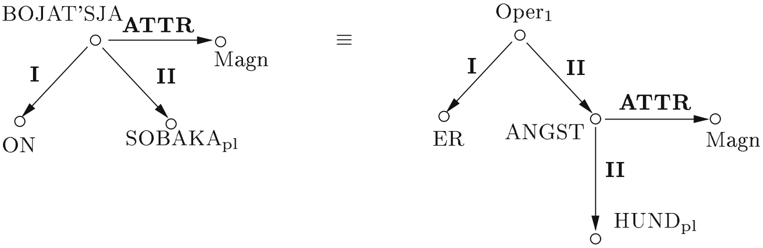


Fig. 25 Equivalent DsyntSs of sentences in (118)

(116) a. *Und nun dusche ich.* AND NOW SHOWER I.
b. А сейчас я приму душ. *A sejčas ja primu duš.*
AND NOW I WILL-RECEIVE SHOWER
'And now I'll take a shower.'

(117) a. *Er ging im Park spazieren.* HE WENT IN-THE PARK TO-WALK
b. He went for a walk in the park.

(118) a. Он очень боится собак. *On očen' boitsja sobak.*
HE VERY IS-AFRAID OF-DOGS.
b. *Er hat große Angst vor Hunden.* HE HAS GREAT FEAR BEFORE DOGS
'He is very afraid of dogs.'

As shown in Fig. 24, Loc$_{in}$ is an LF standing for a preposition with locative meaning, as in *at the station*, *at the conference*, *on the island*, *in the office*, etc.: Loc$_{in}$ and its dependents are lowered and reattached to WALK.

As seen in Fig. 25, the Magn that depends in Russian on the verb бояться *bojat'sja* 'be afraid' is lowered in German to the noun ANGST 'fear' (which is a dependent of the new head—an empty verb Oper$_1$).

A more complex variant of type 3 adjustment is illusrated by example (119) and Fig. 26.

(119) a. Вчера был мороз 30 градусов.
Včera byl moroz 30 gradusov.
YESTERDAY WAS FROST 30 DEGREES
b. Yesterday, it was freezing–30 degrees.

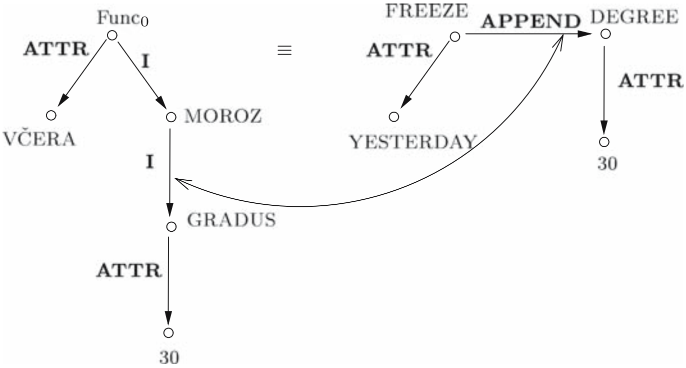


Fig. 26 Equivalent DsyntSs of sentences in (119)

The Func_0-**I**→MOROZ 'frost' branch is substituted in English by the lexeme FREEZE. As a consequence, the subtree with the first actant of *moroz* as head must be re-attached to the new head, FREEZE as an APPEND-dependent. The modifier of Func_0 is also reattached to the new head.

6.2.4 Notion of paraphrasing transfer rule proper

Now we are in a position to define genuine paraphrasing rules as they are supposed to be used in transfer.

Definition 5 (Paraphrasing transfer rule) Let Π be the set of elementary paraphrasing rules and TA the set of syntactic tree adjustments as outlined above. A paraphrasing transfer rule p is defined over Π and TA as a triple $p := \langle \pi_e, \text{TA}_p, \Phi_{tr} \rangle$ where:

- $\pi_e \in \Pi$ is a quadruple $\langle \text{lexeq}_{\pi_e}, \text{Sy}\pi_e, \Phi, \text{Co} \rangle$ as defined in Definition 4;
- TA_p is the set of tree adjustments required for π_e; $\text{ta}_p \in \text{TA}_p$ is defined as a pair of subtrees, the left-hand subtree $\mathbf{L_t}$ and the right-hand subtree $\mathbf{R_t}$, such that ta_p is given by a bidirectional transformation $\mathbf{L_t} \leftrightarrow \mathbf{R_t}$;
- Φ_{tr} is a pair of mappings $(\phi_{tr_L}, \phi_{tr_L})$, where ϕ_{tr_L} maps the instantiated elements of $\mathbf{L_{synt}}$ of all operations $\text{sy}_{\pi_e} \in \text{Sy}\pi_e$ onto $\mathbf{L_t}$, and ϕ_{tr_R} maps the instantiated elements of $\mathbf{R_{synt}}$ of all $\text{sy}_{\pi_e} \in \text{Sy}\pi_e$ onto $\mathbf{R_t}$.

For purposes of illustration, the three paraphrasing rules that handle the examples (113)–(115) are shown as (120)–(122). Note that in order not to make the illustration more complex than necessary, we do not provide the general variants of these rules but, rather, the variants tuned to the examples. The examples are repeated for convenience.

(120) Charles misses Anne terribly. ≡ *Anne manque à Charles énormément.*

π_e: =lexeq_{π_e}: =$L_S \equiv \text{Conv}_{21}(L_S, \mathcal{L}_T)$
$\text{Sy}\pi_e$: X_S-**I**→ $Y_S \leftrightarrow X_T$-**II**→ Y_T
X_S-**II**→ $Y_S \leftrightarrow X_T$-**I**→ Y_T
TA_p: X_S-**ATTR**→ $Y_S \leftrightarrow X_T$-**ATTR**→ Y_T

(121) *Il a le droit de voter dès sa majorité atteinte.* ≡*On imeet pravo golosa s momenta dostiženija im soveršennoletija.*

$$
\begin{aligned}
&\pi_e\text{: =lexeq}_{\pi_e}\text{: =}L_S \equiv \mathrm{Syn}(L_S, \mathcal{L}_T)\\
&\quad \mathrm{Sy}_{\pi_e}\text{:}\qquad X_S\text{-}\mathbf{ATTR}\rightarrow Y_S \leftrightarrow Y_T\text{-}\mathbf{II}\rightarrow X_T\\
&\quad \mathrm{Co}: Y_S \text{ is a verb}\\
&\mathrm{TA}_p\text{: =}X_S\text{-}\mathbf{I}\rightarrow Z_S \leftrightarrow Y_T\text{-}\mathbf{I}\rightarrow Z_T\\
&\qquad A_S\text{-}\mathbf{r}\rightarrow \mathrm{X}_S \leftrightarrow \mathrm{A}_T\text{-}\mathbf{r}\rightarrow \mathrm{Y}_T
\end{aligned}
$$

(122) The river is very shallow. ≡ *La rivière est très peu profond.*

$$
\begin{aligned}
&\pi_e\text{: =lexeq}_{\pi_e}\text{: =}L_S \equiv \mathrm{Anti}(L_S, \mathcal{L}_T)\text{-}\mathbf{ATTR}\rightarrow\mathrm{NEG},\\
&\quad \mathrm{Sy}_{\pi_e}\text{:}\qquad -\\
&\mathrm{TA}_p\text{: =}X_S\text{-}\mathbf{ATTR}\rightarrow Z_S \leftrightarrow \mathrm{NEG}_T\text{-}\mathbf{ATTR}\rightarrow Z_T,\\
&\qquad (\text{if } Z_S \text{ is an intensifier})
\end{aligned}
$$

Analogous rules can be readily given for the other translation examples cited in this article.

7 Summary

Mismatches between source and target syntactic structures are a challenge for MT. As mentioned in the Introduction, a prominent strand of research in MT takes up Dorr's proposal to address this challenge by using an interlingua representation (see Weinstein et al. 1997; Dave et al. 2001; Gupta and Chatterjee 2001, among others). However, as is well known, an approach in the interlingua paradigm faces two major difficulties: the definition of a truly language-independent representation that can serve as an interlingua, and the development of a deep analysis module for mapping the $\mathcal{L}_S$-sentences onto the interlingua representation. Therefore, the interlingua is often biased—usually towards English, as for example the LCS used by Dorr (1993) and in subsequent works. This makes it necessary to address the mismatches in the same way as in the transfer-based paradigm. On the other hand, an approach in the transfer paradigm runs the risk of using a transfer structure that is not abstract enough to avoid what we called "pseudo-mismatches," leading to a high number of idiosyncratic transfer rules.

A recent trend in corpus-based MT is based on the alignment of (potentially non-isomorphic) syntactic structures in parallel corpora and learning of translation rules; (see, e.g., Gildea 2003; Čmejrek et al. 2003; Cicekli and Güvenir 2003). Being relevant for the future, approaches of this type require large and rich corpora to contain enough training material for the derivation of stable translation mappings. Such corpora are and will still be scarce in the near future. Therefore, a transfer-based model that adequately handles the problem of syntactic mismatches continues to be in great demand. In this paper, we argued that:

(a) Transfer in MT is best handled at the level of DSyntS, as proposed by MTT; DSyntS shows an appropriate degree of abstraction to serve as a convenient transfer structure; the small number of DSynt-relations (nine in total) leads to a restricted set of lexical equivalence and syntactic transformation rules (for instance, 72 branch relabelings and 72 branch inversions are logically possible, of which only a small subset is found in language pairs).

(b) Transfer at the DSynt-level is best handled as interlingual paraphrasing; to develop the corresponding mechanisms, the MTT intralingual paraphrasing system can be used as a starting point.

(c) From a formal viewpoint, transfer as paraphrasing is best handled in terms of a graph rewriting grammar.

Our argumentation is based on the rigorous definitions of a few key concepts: (i) structural mismatch as a particular case of non-isomorphism of two DSyntSs; (ii) paraphrasing transfer rules; (iii) DSynt-paraphrasing graph rewriting grammar and its rules.

To facilitate the understanding of the presentation, we used illustrative data from English, French, German, and Russian, examples of mismatches in translation supplied with detailed explanations; materials from other languages are also used occasionally.

Large-scale implementations of transfer mechanisms that are based on a DSynt-like structure and on MTT's paraphrasing apparatus (cf. first of all Apresjan et al. 1989, 1992; in press) and the operational implementation of a graph transduction grammar formalism (Bohnet and Wanner 2001; Bohnet 2005) demonstrate that our proposal can be readily implemented for use in practical MT.

Acknowledgements This paper has been read and commented upon by M. Alonso Ramos, I. Boguslavskij, S. Kahane, J. Milićević, and V. Specht. We would like to express our heartfelt gratitude to all of them as well as to the three anonymous reviewers. Special thanks to S. Kahane, whose sharp criticisms and constructive suggestions helped make this paper much better than it would be otherwise.

Appendix: proof of theorem 1

Let us repeat here, for the convenience of the reader, the theorem from p 31.

Theorem 1 (Types of syntactic mismatches) *Let S_S and S_T be two translationally (= semantically) equivalent deep-syntactic structures such that between S_S and S_T one or several syntactic mismatches occur. Then, these mismatches can be only of the above four types.*

We show by induction that there are no other types of syntactic mismatches.

Proof Let us assume that S_S and S_T show a mismatch that is different from the four in Fig. 13. If we can show that this assumption is false or that it leads to a contradiction with our other assumptions with respect to DSyntS, we prove that only mismatches 1–4 exist.

1. Let S_S be a one-node structure that consists of n_{s_1}. From our assumption that between S_S and S_T no fission/fusion mismatch occurs, we can conclude that S_T consists of one node n_{t_1} that is the translation equivalent of n_{s_1}. Therefore, S_S and S_T are isomorphic and there is no syntactic mismatch between S_S and S_T.
2. Now, let S_S be a two-node structure that consists of n_{s_1} and n_{s_2}. Wellformedness criteria for DSyntS require that n_{s_1} and n_{s_2} be connected via a single relation r_s: $n_{s_1} \xrightarrow{r_s} n_{s_2}$. Due to our assumption of no fission/fusion mismatch between S_S and S_T, S_T consists of two nodes n_{t_1} and n_{t_2} (we assume, without loss of generality,

that n_{t_1} is the translation equivalent of n_{s_1} and n_{t_2} the translation equivalent of n_{s_1}). n_{t_1} and n_{t_2} are connected by a relation r_t. Due to our assumption that there is no head-switching mismatch between S_S and S_T, $S_T = n_{t_1} \xrightarrow{r_t} n_{t_2}$, and due to the assumption that there is no branch relabeling mismatch, $r_s = r_t$. This means that S_S and S_T are isomorphic.

3. Let us now assume that S_S is a three-node structure that consists of n_{s_1}, n_{s_2}, and n_{s_3}. In order to be well-formed, S_S must contain two relations r_{s_1} and r_{s_2}.
   They are defined as follows: $n_{s_1} \xleftarrow{r_{s_1}} n_{s_2}$ and $n_{s_2} \xrightarrow{r_{s_1}} n_{s_3}$.
   As above, due to our assumption of no fission/fusion mismatch between S_S and S_T, S_T contains the same number of nodes as S_S: the nodes n_{t_1}, n_{t_2}, and n_{t_3}. In accordance with the well-formedness criteria, S_T also contains two relations r_{t_1} and r_{t_2}.
   According to our assumption, no branch relabeling and no head-switching mismatch occur. Therefore, we can conclude that $r_{t_1} = r_{s_1}$ and $r_{t_2} = r_{s_2}$. In other words, it is not the case that the relations r_{t_1} and r_{t_1} changed their source or target nodes: the head of r_{t_1} and r_{t_2} is thus n_{t_2}: the translation equivalent of n_{s_1}, the dependent of r_{t_1} is $n_{t_1} = n_{s_1}$ and the dependent of r_{t_2} is $n_{t_3} = \tau(n_{s3})$. That is, S_S and S_T must again be isomorphic.

$n-1$. In accordance with our induction strategy, let us hypothesize that $S_{S_{n-1}}$ is a DSyntS with $n-1$ nodes (and therefore with $n-2$ relations), and that the corresponding target structure $S_{T_{n-1}}$ is either isomorphic with $S_{S_{n-1}}$ or reveals one or several mismatchs of types 1–4.

n. We have to prove now that a source structure S_{S_n} with n nodes is also either isomorphic with its corresponding target structure or show a mismatch of a pattern as predetermined by 1–4.

A DSyntS with $n-1$ nodes can be extended to a DSyntS with n nodes: an additional node is introduced and connected via a relation with the root node, with a leaf node, or with an internal node. In the first case, the root node can be either the head or the tail of the relation; to fulfil the well-formedness criterion of the DSyntS as a tree, in the second and third case, the leaf or the internal node, as the casse may be, must be the tail of the relation.
Let us consider these four cases in turn.

(a) Let the relation r_α hold between the new node n_{s_n} and the root of S_{S_n}, n_{root}, given as follows: $n_{s_n} \xrightarrow{r_\alpha} n_{\text{root}}$.
    Due to our assumption that there is no fission/fusion mismatch, a single node n_{t_n} in the target structure corresponds to n_{s_n}. Furthermore, due to the assumption that neither a branch relabeling nor a head-switching mismatch occurs, the relation r_α appears at the target side as $n_{t_n} \xrightarrow{r_\alpha} \tau(n_{\text{root}})$. The only new mismatch not yet covered could be that one or several subtrees under the root is/are moved to another governor. However, this would be a branch reattachment mismatch, which implies a violation of our premise that no branch reattachment mismatch occurs.

(b) Let the relation r_α hold between the new node n_{s_n} and the root of S_{S_n}, n_{root}, given as follows: $n_{\text{root}} \xrightarrow{r_\alpha} n_{s_n}$. An analogous argumentation as above shows that in $\tau(n_{\text{root}}) \xrightarrow{r_\alpha} n_{t_n} \subset S_{\mathrm{T}_n}$ corresponds to $n_{\text{root}} \xrightarrow{r_\alpha} n_{s_n}$ and both relations are isomorphic if no mismatch of the type 1–4 occurs. Again, only a reattachment of the subtrees under the root can lead to a mismatch not considered yet in the $(n-1)$th step of our induction (that is, in the discussion of the mismatches between $S_{S_{n-1}}$ and $S_{T_{n-1}}$). But since we excluded a reattachment mismatch, we can maintain that no new mismatch occurred at all.

(c) Let the relation r_α hold between the new node n_{s_n} and a leaf node n_α defined as $n_\alpha \xrightarrow{r_\alpha} n_{s_n}$. Then (by analogy to case (b)), no further mismatch occurs between S_{S_n} and S_{T_n}.

(d) Let the relation r_α hold between the new node n_{s_n} and a an internal node of $S_{S_{n-1}}$ defined. Then (by analogy to case (a)), no further mismatch occurs between S_{S_n} and S_{T_n}.

We have thus proven that any mismatch that may occur between any S_{S_n} and its translation equivalent structure S_{T_n} belongs to one of types 1–4. □

References

Abraham W (2003) Canonic and non-canonic deliberations about epistemic modality: Its emergence out of where? In Koster J, van Riemsdijk H (eds) Germania et alia: A linguistic webschrift for Hans den Besten. Published online at www.let.rug.nl/k̃oster/DenBesten/Abraham.pdf [Last accessed 27 July 2006]

Apresjan JD, Boguslavskij IM, Gecelevič E, Iomdin LL, Lazurskij AV, Percov NV, Sannikov VZ, Cinman LL [Апресян ЮД, Богуславский ИМ, Гецелевич Е, Иомдин ЛЛ, Лазурский АВ, Перцов НВ, Санников ВЗ, Цинман ЛЛ] (1992) Лингвистический процессор для сложных систем [Linguistic processor for complex information systems]. Nauka, Moskva

Apresjan JD, Boguslavskij IM, Iomdin LL, Lazurskij AV, Percov NV, Sannikov VZ, Cinman LL [Апресян ЮД, Богуславский ИМ, Иомдин ЛЛ, Лазурский АВ, Перцов НВ, Санников ВЗ, Цинман ЛЛ] (1989) Лингвистическое обеспечение системы ЭТАП-2 [Linguistic software of the ÈTAP-2 system]. Nauka, Moskva

Apresjan JD, Boguslavsky IM, Iomdin LL, Tsinman LL (in press) Lexical functions in actual NLP-applications. To appear in Wanner L (ed) Selected lexical and grammatical topics in the Meaning-Text Theory: In honour of Igor Mel'čuk, John Benjamins, Amsterdam pp 199–228

Barnett J, Mani I, Rich E, Aone C, Knight K, Martinez JC (1991) Capturing language-specific semantic distinctions in interlingua-based MT. In: Machine translation summit III proceedings, Washington, DC, pp 25–32

Barwise J, Perry J (1983) Situations and attitudes. The MIT Press, Cambridge, MA

Bohnet B (2005) Textgenerierung als Transduktion linguistischer Strukturen [Text generation as the transduction of linguistic structures]. Dissertation, Universität Stuttgart, Germany

Bohnet B, Wanner L (2001) On using a parallel graph rewriting grammar formalism in generation. In: Proceedings ACL 2001 workshop, eighth European workshop on natural language generation (EWNLG), Toulouse, France, pp 47–56

Bresnan J (ed) (1982) The mental representation of grammatical relations. MIT Press, Cambridge, MA

Cicekli I, Güvenir HA (2003) Learning translation templates from bilingual translation examples. In: Carl M, Way A (eds) Recent advances in example-based machine translation. Kluwer Academic Publishers, Boston, pp 247–278

Čmejrek M, Cuřín J, Havelka J (2003) Treebanks in machine translation. In: Proceedings of the second workshop on treebanks and linguistic theories, Växjö, Sweden, 209–212

Dave S, Parikh J, Bhattacharyya P (2001) Interlingua-based English–Hindi machine translation. Mach Translat 16:251–304

Doherty M (1999) Sprachspezifische Aspekte der Informationsverteilung [Language-specific aspects of the distribution of information]. Akademie Verlag, Berlin

Dorr BJ (1993) Machine translation: A view from the lexicon. The MIT Press, Cambridge, MA

Dorr BJ (1994) Machine translation divergences: A formal description and proposed solution. Comput Ling 20:579–634

Dorna M, Emele MC (1996) Efficient implementation of a semantic-based transfer approach. In: Proceedings of the 12th European conference on artificial intelligence, Budapest, 567–571

Dymetman M, Tendeau F (2000) Context-free grammar rewriting and the transfer of packed linguistic representations. In: Proceedings of the 18th international conference on computational linguistics, COLING 2000 in Europe, Saarbrücken, Germany, pp 1016–1020

Emele MC, Dorna M (1998) Ambiguity preserving machine translation using packed representations. In: COLING-ACL '98: 36th annual meeting of the Association for Computational Linguistics and 17th international conference on computational linguistics, Montreal, Que, 365–371

Gawron JM (1999) Abduction and mismatch in machine translation. Technical report, SRI International, Stanford, CA

Gildea D (2003) Loosely tree-based alignment for machine translation. In: ACL-03: 41st annual meeting of the Association for Computational Linguistics, Sapporo, Japan, pp 80–87

Grupo DiCE (n.d.) DICE: Diccionario de colocaciones del español [DICE: Dictionary of Spanish collocations]. Website http://www.dicesp.com [Last accessed 7 July 2006]

Gupta D, Chatterjee N (2001) Study of divergence for example based English-Hindi machine translation. In: STRANS–2001, Symposium on translation support systems, Kanpur, India

Han C, Lavoie B, Palmer M, Rambow O, Kittredge R, Korelsky T, Kim N, Kim M (2000) Handling structural divergences and recovering dropped arguments in a Korean/English machine translation system. In: White JS (ed) Envisioning machine translation in the information future: 4th conference of the Association for Machine Translation in the Americas, AMTA 2000, Cuernavaca, Mexico, October 2000, Springer Verlag, Berlin, pp 40–53

Iordanskaja L, Mel'čuk, IA (1997) Le corps humain en russe et en français: Vers un dictionaire explicatif et combinatoire bilingue. [The human body in French and Russian: Towards an explanatory and combinatorial dictionary]. Cah Lexicol 70:249–281

Jackendoff R (1990) Semantic structures. The MIT Press, Cambridge, MA

Kamp H, Reyle U (1993) From discourse to logic. Reidel, Dordrecht

Kaplan RM, Netter K, Wedekind J, Zaenen A (1996) Translation by structural correspondences. In: Dalrymple M, Kaplan RM, Maxwell JT, Zaenen A (eds) Formal issues in lexical-functional grammar, CSLI, Stanford, CA, pp 311–329

Mel'čuk IA [Мельчук ИА] (1974) Опыт теории лингвистических моделей "Смысл ⇔ Текст" [Outline of the "Meaning⇔Text" theory of linguistic models]. Nauka, Moskva

Mel'čuk IA (1988a) Dependency syntax: theory and practice. State University of New York Press, Albany, NY

Mel'čuk IA (1988b) Paraphrase et lexique dans la théorie linguistique Sens-Texte. [Paraphrase and lexicon in the Meaning-Text linguistic theory]. Cah Lexicol 52.1:5–50, 52.2:5–53

Mel'čuk IA (1992) Paraphrase et lexique:la théorie Sens-Texte et le dictionnaire explicatif et combinatoire [Paraphrase and the lexicon:the Meaning-Text Theory and the explanatory and combinatorial dictionary]. In: Mel'čuk et al (1992), pp 9–59

Mel'čuk IA (1993) The future of the lexicon in linguistic description and the explanatory combinatorial dictionary. In: Lee I-H (ed) Linguistics in the morning calm 3. Linguistic Society of Korea, Seoul, pp 181–270

Mel'čuk IA (1995a) Phrasemes in language and phraseology in linguistics. In: Everaert M, van der Linden EJ, Schenk A, Schreuder R (eds) Idioms: structural and psychological perspectives, Lawrence Erlbaum Associates, Hillsdale, NJ, pp 167–232

Mel'čuk IA (1995b) Syntactic or lexical zero. In: Mel'čuk IA [Мельчук ИА] Русский язык в модели Смысл ⇔ Текст [Russian language in the Meaning ⇔ Text model]. Wiener Slawistischer Almanach, Wien, pp 169–205

Mel'čuk I (1996) Lexical functions: A tool for the description of lexical relations in a lexicon. In: Wanner L (ed) Lexical functions in lexicography and natural language processing. John Benjamins, Amsterdam, pp 37–102

Mel'čuk IA (2001) Communicative organization in natural language (The semantic-communicative structure of sentences). John Benjamins, Amsterdam

Mel'čuk IA (2004a) Actants in semantics and syntax I: Actants in semantics. Ling 42:1–66

Mel'čuk IA (2004b) Actants in semantics and syntax II: Actants in syntax. Ling 42:247–291

Mel'čuk I, Arbatchewsky-Jumarie N, Dagenais L, Elnitsky L, Iordanskaja L, Lefebvre M-N, Mantha S (1988) Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II [Explanatory combinatorial dictionary of modern French. Lexical-semantic studies II]. Montréal, Les Presses de l'Université de Montréal

Mel'čuk I, Arbatchewsky-Jumarie N, Elnitsky L, Iordanskaja L, Lessard A (1984) Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I [Explanatory combinatorial dictionary of modern French. Lexical-semantic studies I]. Les Presses de l'Université de Montréal, Montréal

Mel'čuk I, Arbatchewsky-Jumarie N, Iordanskaja L, Mantha S (1992) Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III [Explanatory and combinatorial dictionary of modern French. Lexico-semantic research III]. Les Presses de l'Université de Montréal, Montréal

Mel'čuk I, Arbatchewsky-Jumarie N, Iordanskaja L, Mantha S, Polguère A (1999) Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques IV [Explanatory combinatorial dictionary of modern French. Lexical-semantic studies IV]. Les Presses de l'Université de Montréal, Montréal

Mel'čuk IA, Polguère A (1987) A formal lexicon in the Meaning-Text theory (or how to do lexica with words). Comput Ling 13:276–289

Mel'čuk IA, Polguère A (forthcoming) Lexique actif du français: Dictionnaire de dérivations sémantiques et de collocations [Active French lexicon: Dictionary of semantic derivations and collocations]. Duculot, Louvain-la-Neuve

Mel'čuk I, Wanner L (2001) Towards a lexicographic approach to lexical transfer in machine translation (illustrated by the German–Russian language pair). Mach Translat 16:21–87

Mel'čuk IA, Zholkovsy AK (1984) Explanatory combinatorial dictionary of modern Russian. Wiener Slawistischer Almanach, Wien

Milićević J (2003) Modélisation sémantique, lexicale et syntaxique de la paraphrase [Semantic, lexical and syntactic models of paraphrase]. Thèse de doctorat, Département de linguistique, Université de Montréal, Montréal

Paslawska A, von Stechow A (2003) Perfect readings in Russian. In: Rathert M, Alexiadou A, von Stechow A (eds) Perfect explorations. Mouton de Gruyter, Berlin, pp 307–363

Reyle U (1993) Dealing with ambiguities by underspecification: Construction, representation and deduction. Jnl Semant 10:123–179

Sacker U (1983) Aspektueller und resultativer Verbalausdruck [Aspectual and resultative verbal expression]. Gunter Narr Verlag, Tübingen

Sanromán Vilas B, Lareo Martín I, Alonso Ramos M (1999) Transferencia léxica y reglas de paráfrasis: verbos denominales de SP cognado [Lexical transfer and rules of paraphrase: denominal verbs with a cognate object]. Rev Soc Españ Proc Leng Nat 25:183–189

Specht V (2002) Entwurf und Implementierung einer Paraphrasierungsgrammatik für maschinelle Übersetzung [Design and implementation of a paraphrase grammar for machine translation]. Institut für Informatik, Universität Stuttgart, Germany

Steiner E (2001) Translations English – German: Investigating the relative importance of systemic contrasts and of the text-type "translation". In: Proceedings of the symposium "Information structure in a cross-linguistic perspective", SPRIKreports 7, Oslo

Wanner L (1996) Lexical choice in text generation and machine translation. Mach Translat 11:3–35

Weinstein CJ, Lee YS, Seneff S, Tummala DR, Carlson B, Lynch JT, Hwang JT, Kukolich LC (1997) Automated English/Korean translation for enhanced coalition communication. Lincoln Lab Jnl 10:35–60

Žolkovskij AK, Mel'čuk IA [Жолковский АК, Мельчук ИА] (1967) О семантическом синтезе [On semantic synthesis]. Probl Kybernet 19:177–238