Morphological mismatches in machine translation

Igor Mel'čuk · Leo Wanner

Abstract   This paper addresses one of the least studied, although very important, problems of machine translation—the problem of morphological mismatches between languages and their handling during transfer. The level at which we assume transfer to be carried out is the *Deep-Syntactic Structure* (DSyntS) as proposed in the Meaning-Text Theory. DSyntS is abstract enough to avoid all types of surface morphological divergences. For the remaining 'genuine' divergences between grammatical significations, we propose a morphological transfer model. To illustrate this model, we apply it to the transfer of grammemes of definiteness and aspect for the language pair Russian–German and German–Russian, respectively.

Keywords   Machine translation · Transfer · Grammatical signification · Morphological mismatch · Deep-syntactic structure · Meaning-Text Theory · Definiteness · Aspect

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

L. Wanner
Department of Information and Communication Technologies, Pompeu Fabra University,
C. Roc Boronat, 138, 08018 Barcelona, Spain
e-mail: leo.wanner@upf.edu

L. Wanner
Institució Catalana de Recerca i Estudis Avançats (ICREA),
08010 Barcelona, Spain

Writing Conventions, Abbreviations, and Notations

BI	Bilingual index
DSyntA	Deep-syntactic actant
DSynt-CommS	Deep-syntactic communicative structure
DSyntS	Deep-syntactic structure
d	A particular derivateme
ECD	Explanatory combinatorial dictionary
g	A particular grammeme
GS	Grammatical signification
L	A particular lexeme
$\mathcal{L}$	A particular language
LF	Lexical function
LU	Lexical unit
MT	Machine translation
MTT	Meaning-Text Theory
SAE	Standard Average European (language)
SSyntS	Surface-syntactic structure
X_S/X_T	Source/target X (with X being an LU, $\mathcal{L}$, …)

Lexical units are in CAPITALS; in the text, grammatical significations are in SMALL CAPITALS and 'single quotes'; (quasi-)grammemes, when encoded as subscripts to LUs, are in small capitals without quotes. Linguistic examples are in *ordinary italics*; meanings, including glosses, are in 'single quotes'.

1 Introduction

Translation, in particular, Machine Translation (MT), can be considered from three different angles: lexical translation, syntactic translation, and morphological translation. The main challenge within each of these tasks are *mismatches*: roughly speaking, situations where a source language element does not correspond in a one-to-one way to a target language element. Some aspects of lexical and syntactic mismatches were considered by the authors in Mel'čuk and Wanner (2001, 2006); in the present article, we focus on mismatches between what we call *morphological*, or *grammatical*, *significations*: *grammemes*, *quasi-grammemes* and *derivatemes* in the source language $\mathcal{L}_S$ and the target language $\mathcal{L}_T$. As in our previous work, we assume that the transfer is done at the level of the Deep-Syntactic Structure (DSyntS) as defined in the Meaning-Text Theory (MTT); see Sect. 4.1, Definition 5.

In order for our task to be better presented and our proposal to be better understood, let us start by situating our approach with respect to the field (Subsect. 1.1), then give the definitions of grammatical significations (Subsect. 1.2), contrast morphological translation with lexical and syntactic translation by briefly presenting all three of them (Subsect. 1.3), and, finally, formulate the problem of morphological translation (Subsect. 1.4).

1.1 Situating our approach

As is well known, there are two major methodologies to address MT (and computational linguistics in general): empirical versus rational. From the practical standpoint, the empirical methodology, which attempts to automatically derive translation models from corpora, proves to be most effective for obtaining cost-effective, robust and good quality translations in the shortest delay possible. *Statistical* MT (see, e.g., Marcu et al. 2006, Ney et al. 2007) and *example-based* MT (Nagao 1984, Somers 1999) are instantiations of this methodology. However, from the theoretical standpoint, which is essentially interested in formalizing and modelling the intuition of human translators, the rational methodology seems to be more adequate. One way to follow this methodology is for a linguist to study parallel corpora and formulate translation rules; this is what can be called the "orthodox linguistic" approach to MT. The present paper follows this approach. Our goal is to work out regularities observed with respect to morphological mismatches, propose useful generalizations, and develop an exhaustive typology of morphological phenomena underlying mismatches. We believe that a rule-based transfer model, which makes explicit the criteria for the correspondence between grammatical significations in two different languages $\mathcal{L}_1$ and $\mathcal{L}_2$, is useful for a better formal understanding of natural language and can be exploited for different applications in NLP, in particular, for example-based MT. Thus, our handcrafted rules can serve, along with linguistic data, as training material.

1.2 Grammatical significations

The definitions of grammatical notions presented below do not claim to be final. However, they are precise enough for our task.

For the definitions of grammatical significations, the notion of *inflectional category* of lexical units (LUs) is central.

Definition 1 (*Inflectional category* (*of LUs*))

An *inflectional category* of a particular class of LUs is a finite set of mutually exclusive significations that are

1. obligatory, in the sense that any LU of the given class must be accompanied by one of these significations
   and
2. regular, in the sense that any LU of the given class combines with the corresponding signification in an entirely compositional way.

Definition 2 (*Grammeme*)

A *grammeme* is a signification that belongs to an inflectional category.

For instance, the English noun has the inflectional category of number, which means that every English noun in the text must have one of the two possible number grammemes: 'SINGULAR' or 'PLURAL'; a noun without number is unthinkable in English. The meaning of the singular and the plural is joined to the meaning of the LU by (quite) regular general rules.

Definition 3 (*Quasi-grammeme*)

A *quasi-grammeme* is a signification that does not belong to an inflectional category, but is formally expressed by the same linguistic means as grammemes (namely, by morphological means: mainly, affixes). It is

1. not obligatory
   but
2. regular (in the same sense as a grammeme).

For instance, the English possessive form in *-'s* (as in *John's*) expresses the quasi-grammeme 'POSSESSIVE'. Unlike genuine grammemes, it is not obligatory: many English nouns do not have it at least according to many style books (cf. WATER: $^?$*the water's color*, ATTACK: $^?$*the attack's success*, etc.), and those which can take an *-'s* do not require it (cf. *John's book = the book of John*). However, it is absolutely regular.

Grammemes and quasi-grammemes distinguish wordforms of the same LU.

Definition 4 (*Derivateme*)

A *derivateme* is a signification that does not belong to an inflectional category, but is formally expressed by the same linguistic means as grammemes. It is

1. not obligatory,
   and
2. not necessarily regular, in contrast to (quasi-)grammemes

For instance, the English derivateme 'ONE WHO/WHICH X- ES' (SMOKE∼SMOK+ER, RUN ∼ RUNN+ER, etc.) is expressed by a suffix, the way the most English grammemes are.

Derivatemes distinguish different LUs, in contrast to (quasi-)grammemes. The property of derivatemes to produce new LUs underlies their distinct presentation in the DSyntS. (Quasi-)grammemes appear as feature-value pairs attached to LU labels; for simplicity's sake we write only the values, attaching them as subscripts to LUs:

HOUSE$_{PL}$ ⇔ *houses*
WRITE$_{PASS, PERF, PROGR, PAST}$ ⇔ *had been being written*

Derivatemes, on the contrary, form part of derived LU labels: HELP+'ONE WHO …' (=HELPER), Sp. CASA+'DIMIN(utive)' (=CASITA) 'nice little house', Rus. PIT' 'drink'+'DO-…-SJA' (=DOPIT'SJA 'suffer something bad because of excessive drinking'). While for (quasi-)grammemes standard labels are readily available, for derivatemes this is not the case; some of them are conveniently referred to by their signified (as 'ONE WHO …'), for some others standard labels can be used (as 'DIMIN'), and still for others it is preferable to use their signifier (as 'DO-…-SJA').

In this paper, we limit ourselves to fully regular and productive derivation. The derivatemes considered here thus form different LUs that need not be stored in the lexicon.

Grammatical significations are subdivided into semantic and syntactic. Semantic grammatical significations are those that are in principle directly related to meaning, i.e. to elements of the Semantic Representation; the examples above illustrate semantic

(quasi-)grammemes and derivatemes. Note that a semantic (quasi-)grammeme may be controlled by an element of the context. For instance, the plural of BOOK in *five books* is imposed by the numeral and thus is not significative. Nevertheless, since the grammeme 'PL' belongs to a semantic inflectional category, it is still considered semantic and must be present in the DSyntS.

Syntactic grammatical significations are those that are not directly related to meaning, but serve only to mark syntactic roles. Examples of syntactic grammemes, which are imposed by government and agreement, are case for nouns and person and number for verbs. Deverbal nouns (*nomina facti* such as FIGHT$_{(N)}$, MOVEMENT, and CONTINUATION) feature a syntactic derivateme.

1.3 Three aspects of translation

We illustrate the three-pronged distinction of the translation process in MT by an example of bidirectional translation between two equivalent German/Russian sentences in (1).

(1) a. Ger. *Gestern zog der Minister die Reisewarnung für Rom zurück*
       'Yesterday, the minister withdrew the warning against traveling to Rome'.
    b. Rus. *Včera ministr otmenil predostereženie protiv poezdok v Rim* [idem].

Lexical translation deals with the mapping of LUs of the source language $\mathcal{L}_S$ to the corresponding LUs of the target language $\mathcal{L}_T$. In (1), there are, first of all, such uncontroversial lexical correspondences as VČERA ≡ GESTERN 'yesterday', MINISTR ≡ MINISTER 'minister' and RIM ≡ ROM 'Rome'. These correspondences can be easily accounted for in a lexical *bilingual* (*transfer*) *index*. More complex correspondences such as OTMENIT′ ≡ ZURÜCKZIEHEN 'withdraw' can be dealt with by using *lexical functions* (LFs), as presented in Mel'čuk (1996), Apresjan et al. (2007): *otmenit′* = LiquFunc$_0$(PREDOSTEREŽENIE 'warning') ≡ LiquFunc$_0$(WARNUNG 'warning') = *zurückziehen*. The values of LFs are specified in the monolingual $\mathcal{L}_S$ and $\mathcal{L}_T$ lexica. And finally, correspondences such as *predostereženie protiv poezdok* lit. 'warning against travels' ≡ *Reisewarnung* lit. 'travel.warning' are specified in two different lexical resources: in the lexical bilingual index (POEZDKA ≡ REISE, PREDOSTEREŽENIE ≡ WARNUNG) and in the corresponding entries of the monolingual lexicon of $\mathcal{L}_T$: thus, the entry for Ger. WARNUNG contains the information that WARNUNG takes, under specific conditions, its third actant as the dependent element of a compound in which WARNUNG is the main component.[1] In point of fact, the above lexical equivalences POEZDKA ≡ REISE and PREDOSTEREŽENIE ≡ WARNUNG are not one-to-one, and therefore, imply lexical mismatches. Thus, POEZDKA translates also as FAHRT; similarly, WARNUNG can be PREDUPREŽDENIE or

---

[1] The actantial structure of Ger. WARNUNG can be assumed to be as follows: X [⇔ I] *warns* Y [⇔ II] *against* Z [⇔ III]. The variables X, Y and Z refer to the semantic slots that correspond to the arguments 1, 2, and 3 of the predicate 'warn'. I, II, and III refer to the corresponding deep-syntactic actantial slots. The same symbols (I, II, III) are used as names of deep-syntactic relations (see the examples immediately below). This practice is common in syntax: the name of a dependent is often used as a label for the corresponding syntactic relation; cf. "subject" ∼ "subject relation", "object" ∼ "object relation", etc.

OPOVEŠČENIE. The problem of lexical mismatches is discussed in detail in Mel'čuk and Wanner (2006), and we can ignore it here.

Syntactic translation deals with the mapping of DSyntSs, more precisely, of subtrees of DSyntSs of $\mathcal{L}_S$ to subtrees of DSyntSs of $\mathcal{L}_T$ (see Mel'čuk and Wanner 2006).[2] In (1), several substructures of (a) and (b) are isomorphic and, therefore, allow for a one-to-one translation:

(2) a. Rus. $\mathrm{Adv_{temp}} \leftarrow \mathbf{ATTR} - \mathrm{V_{fin}(Adv_{temp}} = \textit{včera})$
     $\equiv$
     Ger. $\mathrm{Adv_{temp}} \leftarrow \mathbf{ATTR} - \mathrm{V_{fin}(Adv_{temp}} = \textit{gestern})$
   b. Rus. $\mathrm{N} \leftarrow \mathbf{I} - \mathrm{V_{fin}(N} = \textit{ministr}$ and $\mathrm{V_{fin}} = \textit{otmenit}')$
     $\equiv$
     Ger. $\mathrm{N} \leftarrow \mathbf{I} - \mathrm{V_{fin}(N} = \textit{Minister}$ and $\mathrm{V_{fin}} = \textit{zurückziehen})$[3]
   c. Rus. $\mathrm{V}-\mathbf{II} \rightarrow \mathrm{N(V} = \textit{otmenit}'$ and $\mathrm{N} = \textit{predostereženie})$
     $\equiv$
     Ger. $\mathrm{V}-\mathbf{II} \rightarrow \mathrm{N(V} = \textit{zurückziehen}$ and $\mathrm{N} = \textit{Warnung})$

There is also a more complex correspondence, namely between two constructions (already discussed above from the lexical angle) that are superficially very similar, but, in point of fact, far from isomorphic:

(3)   Rus. $\mathrm{N_1} - \mathbf{II} \rightarrow \mathrm{N_2} - \mathbf{II} \rightarrow \mathrm{N_3}$
       ($\mathrm{N_1} = \textit{predostereženie}, \mathrm{N_2} = \textit{poezdka}$, and $\mathrm{N_3} = \textit{Rim}$; the prepositions *protiv* 'against' and *v* 'to' are governed and thus do not appear in the DSyntS)
       $\equiv$
       Ger. $[\mathrm{N_2} + \mathrm{N_1}] - \mathbf{ATTR} \rightarrow \mathrm{Prep} - \mathbf{II} \rightarrow \mathrm{N_3}$
       ($\mathrm{N_1} = \textit{Warnung}, \mathrm{N_2} = \textit{Reise}, \mathrm{Prep} = \textit{für}, \mathrm{N_3} = \textit{Rom}$)

The sentences in (1) feature two syntactic mismatches. First, the Russian NP *predostereženie protiv poezdok* lit. 'warning against travels' corresponds to the German compound *Reisewarnung* lit. 'travel.warning'. Second, while the governor of the Russian PP *v Rim* 'to Rome' is *poezdka* 'travel', that of its German equivalent cannot be *Reise* since it is the dependent element of a compound; therefore, the German PP *für Rom* lit. 'for Rome' is subordinated to *Reisewarnung*. Third, the dependency relation between *poezdka* 'travel' and *v Rim* 'to Rome' is actantial (so that the preposition *v* 'to' is empty), while the relation between *Reisewarnung* 'travel.warning' and *für Rom* 'for Rome' is modifying (so that the preposition *für* 'for' is meaningful: $\approx$ 'concerning'). As mentioned above, we ignore syntactic mismatches in this paper and refer the interested reader to Mel'čuk and Wanner (2006).

Morphological translation deals with the mapping between semantic grammatical significations in $\mathcal{L}_S$ and $\mathcal{L}_T$, i.e. semantic derivatemes or semantic (quasi-) grammemes.[4] The DSyntSs of the sentences in (1) show three morphological mis-

---

[2] For more details on DSyntS in this paper, see the last paragraph of Sect. 2 and Subsect. 4.1.

[3] The movement of the German separable verbal prefix *zurück-* in (1a) to the end of the clause is a surface-structure phenomenon and, therefore, not relevant at the deep-syntactic level, where the verb *zurückziehen* is always represented by one node.

[4] Syntactic derivatemes and syntactic grammemes do not appear at the deep-syntactic level and, therefore, are not an object of transfer.

matches. First, the German radical *Reise-* 'travel' within the compound *Reisewarnung* 'travel.warning' does not carry any number grammeme (it is not a node in the DSyntS), while its Russian equivalent, a separate lexeme *poezdka*, must be in the plural. Second, the nominal grammeme 'DEF(inite)' in Ger. *der Minister* 'the minister' and *die (Reise)warnung* 'the (travel.)warning' is not reflected in Russian due to the absence of the inflectional category of definiteness in this language. It is precisely because of this absence that the translation from Russian into German presents a specific challenge: how to compute the proper article in German. Third, the German verbal grammeme 'IMPERF' (in the traditional grammar; see, however, Appendix A.1.1) corresponds in this context to the pair of Russian grammemes 'PAST'+'PERFECTIVE' and vice versa; Russian has the inflectional category of aspect, which German lacks altogether.

1.4 The problem of morphological translation

While lexical and syntactic problems of MT have received considerable attention in the literature, its morphological aspects are underexplored; we know only a few systematic (mostly philological) studies of the problem.[5] At the same time, the translation of meaningful grammatical significations (such as the grammemes of nominal number or of verbal tense and aspect) requires access to complex abstract semantics, which still defies a rigorous description. Therefore, the interlingual correspondences between grammatical significations and their equivalents are difficult to formulate.

In this article, we address the problem of the transfer of meaningful morphological significations and, in particular, morphological mismatches. More precisely, in Sect. 6 we focus on the translation of the grammemes of two prominent inflectional categories of German and Russian: nominal definiteness in German (absent from Russian) and verbal aspect in Russian (absent from German). Both categories are notoriously difficult to handle. Despite extensive work on the category of definiteness in many article languages (Bisle-Müller 1991; Hansen 1995; Vater 1997, 2002; Leiss 2000; Heim 2001; Himmelmann 2001; Kovtun 2003), there is still no clear-cut picture of German article semantics; the same applies to Slavic, in particular Russian, aspect (see, among others, Wierzbicka 1967; Glovinskaja 1982, 2001; Padučeva 1996, 1998). Obviously, we cannot hope to provide a rigorous and exhaustive semantic description of the corresponding grammemes; we will limit ourselves to formulating some approximate rules for "generating" definiteness grammemes under Russian–German translation and aspect grammemes under German–Russian translation. In so doing, we attempt to place our description within the framework of a simplified but universal typology of morphological semantic (quasi-)grammeme and derivateme mismatches.

Let us emphasize that we consider the translation of grammatical significations based on the following assumption. While in the DSyntS a lexical label corresponds to a fully disambiguated word, i.e. a word taken in one particular sense, a grammatical

---

[5] Cf., e.g. Sacker (1983) for the translation of Russian aspect into German; Hansen (1995) for the translation of German articles into Turkish; Kovtun (2003) for the translation of German articles into Ukranian and Russian; and Vater (1997) for a contrastive study of definiteness in German and Polish.

signification is not disambiguated; it can have several meanings. This assumption determines to a large extent the content and outline of our article.[6]

The remainder of the article is organized as follows. In Sect. 2, we briefly introduce our theoretical framework—the Meaning-Text Theory. Section 3 presents a typology of correspondences between semantic grammatical significations. Section 4 discusses the morphological discrepancies between the source DSyntS and target DSyntS: first the mismatches that need to be accounted for (Sect. 4.1) and then deviances that can be avoided when the transfer is done between DSyntSs (Sect. 4.2). Section 5 contains a sketch of our model of morphological translation. In Sect. 6, we apply this model to the translation of the categories of definiteness and aspect between German and Russian to illustrate the handling of such difficult mismatches. Finally, Sect. 7 summarizes the central aspects of our approach and draws some conclusions. Two appendices complement the article. In Appendix 1, linguistic exposés that are relevant to the topic, but not indispensable for the understanding of our approach, are presented. Appendix 2 contains the textual material which is used in Sect. 6 to illustrate our translation model: two original texts in German and their Russian translations, plus the list of translated grammemes with the indication of morphological transfer rules that were used during the translation.

2 Theoretical framework

Given the complexity of the task of MT, the stage of transfer must be simplified as much as possible. Phenomena that are intralinguistic in nature should be treated at the source side (= during the analysis), respectively the target side (= during the synthesis), rather than in the transfer. In our approach, the level at which the transfer is carried out is the DSyntS, as proposed in the MTT. As already pointed out in the literature, the DSyntS is well-suited to serve as a transfer level in MT (Apresjan et al. 1989; Sanromán Vilas et al. 1999; Han et al. 2000; Mel'čuk and Wanner 2006). This is due to the fact that the DSyntS is a generalization of a family of equivalent surface structures of a given language $\mathcal{L}$ such that it "levels out" contextually determined syntactic phenomena of $\mathcal{L}$. As a consequence, DSyntS is, generally speaking, simpler and more homogeneous than the corresponding surface structures.

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).

The three types of structure implied in this schema and illustrated for the Russian sentence (4) are characterized below:

---

[6] This assumption is due to the essentially different nature of lexical and grammatical entities. The linguistic properties of an LU are related to its meaning; lexical units of the same form but with different meanings tend to show different distribution, co-occurrence, and/or morphology. The linguistic properties of a grammatical signification, on the contrary, are not related to its meanings. For instance, no matter what a plural means in a given language, it always behaves in the same way; it controls the same agreement and has the same expression. This opposition underlies the different treatment of lexical and grammatical significations in the linguistic tradition.
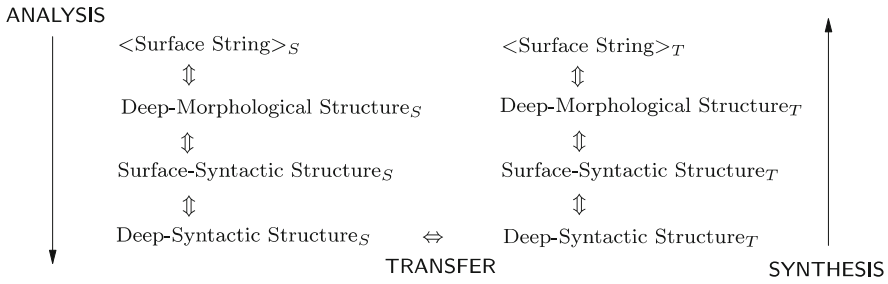
ANALYSIS

$\langle$Surface String$\rangle_S$ $\langle$Surface String$\rangle_T$

$\Updownarrow$ $\Updownarrow$

Deep-Morphological Structure$_S$ Deep-Morphological Structure$_T$

$\Updownarrow$ $\Updownarrow$

Surface-Syntactic Structure$_S$ Surface-Syntactic Structure$_T$

$\Updownarrow$ $\Updownarrow$

Deep-Syntactic Structure$_S$ $\Leftrightarrow$ Deep-Syntactic Structure$_T$

TRANSFER   SYNTHESIS

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

(4)   Rus. *V aprele Džon posetil Rim* 'In April John visited Rome'.

The Deep-Morphological Structure (DMorphS) of a sentence *S* 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 number and case for nouns or voice, mood, aspect, tense, person and number for verbs, etc.), and quasi-grammemes, if any. The DMorphS of sentence (4) appears as (5):[7]

(5)   V $\prec$ APREL$'_{SG,LOC}$ $\prec$ DŽON$_{SG,NOM}$ $\prec$
        In   April   John

        POSETIT$'_{ACT,IND,PERF,PAST,MASC,SG}$ $\prec$ RIM$_{SG,ACC}$
          visited   Rome

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

The Surface-Syntactic Structure (SSyntS) of a sentence *S* is an unordered depen-dency tree whose nodes are labeled with the names of the lexemes of *S* (supplied, where necessary, with semantic grammemes,[8] i.e. number for nouns, and voice, mood, aspect and tense for verbs), and whose arcs are labeled with names of surface-syn-tactic relations. The lexemes of *S* and the nodes of its SSyntS are in a one-to-one correspondence. The set of surface-syntactic relations, which are language-specific, includes such relations as *subject(ival)*, *dir-object(ival)*, *aux(iliary)*, *circumstantial*, etc. The SSyntS of (4) is presented in (6):[9]

_____

[7] The symbol '$\prec$' indicates the strict linear ordering.

[8] Syntactic (quasi-)grammemes are not represented in the SSyntS (just as they are not in the DSyntS); they appear only in the Deep-Morphological Structure.

[9] Given that the nodes of a syntactic tree are labeled with specific LUs, we should, in principle, supply the node labels with distinguishing lexicographic numbers; we dispense with this for the sake of simplic-ity. (Quasi-)grammemes and derivatemes are not "disambiguated" because their surface behavior does not depend on their meaning as stated above in footnote 6.

(6)

$$POSETIT'_{ACT,IND,PERF,PAST}$$
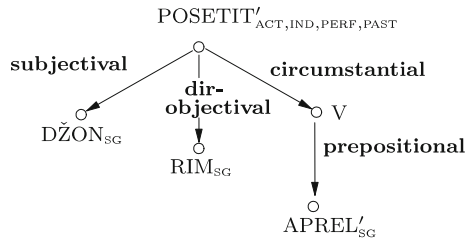
Just as an SSyntS, a DSyntS is an unordered dependency tree, but with a different labeling of nodes and arcs. The nodes of the DSyntS of a sentence S are labeled, roughly speaking, also with the LUs of S, but not in a one-to-one correspondence; lexical labels in a DSyntS are "deep representations" of the actual LUs of S, or *deep* LUs. A deep LU is, roughly speaking, a full LU; it can be a simple LU, a derived LU, a compound, or an LF. The deep LUs carry the same semantic grammemes as the LUs at the nodes of SSyntS. The arcs of a DSyntS are labeled with names of deep-syntactic relations, which are cross-linguistically universal and represent a generalization of surface-syntactic relations. The DSyntS of (4) is given in (7):

(7)
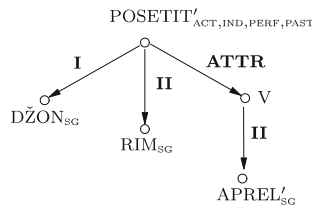
All grammatical significations in the remainder of this article are to be viewed in the context of DSyntSs as illustrated in (7).

3 Typology of correspondences between grammatical significations

As stated in Subsect. 1.2, three families of grammatical significations (GS) are distinguished: (i) grammemes, (ii) quasi-grammemes, and (iii) derivatemes (Mel'čuk 2006, pp. 22ff). The following typology covers all of them; it presents all possible correspondences between a GS of $\mathcal{L}_S$ and its equivalents in $\mathcal{L}_T$. Once again, since we deal with morphological translation at the DSynt-level, only semantic GSs are considered. These GSs and their target equivalents appear in the DSyntS$_S$ and DSyntS$_T$, respectively. To the best of our knowledge, a GS is never expressed by a syntactic relation; therefore, it can correspond only to one of the two kinds of linguistic items: either a GS$_T$ or an LU$_T$, or else to nothing (i.e. remained unexpressed). The typology consists thus of the following three major types of interlingual grammatical signification correspondences:

   (i) $GS_S \Leftrightarrow GS_T$,
  (ii) $GS_S \Leftrightarrow LU_T$,
 (iii) $GS_S \Leftrightarrow$ — [nothing].

Let us discuss each of them in turn, separating (quasi-)grammemes (denoted as g_S and g_T, respectively) from derivatemes (denoted as d_S/d_T).

The above typology does not cover the theoretically possible syntagmatic one-to-many correspondences between grammatical significations, as observed, for instance, in the transfer between Spanish and Russian; the Spanish preterite must be translated by the Russian past tense of the perfective aspect; cf. Sp. *murió* '[he/she] died' ≡ Rus. *on umer* 'he died' ('PRET' ≡ 'PERF', 'PAST'); for paradigmatic GS correspondences, see Subsect. 3.1.2.

3.1 $GS_S \Leftrightarrow GS_T$

$GS_S \Leftrightarrow GS_T$ subsumes two subtypes of correspondence: a GS_S corresponds to a GS_T and vice versa (one-to-one correspondence); a GS_S corresponds to several alternative GS_Ts (one-to-many paradigmatic correspondence) and thus requires selection conditions.

3.1.1 One-to-one correspondence between a GS_S and a GS_T

Two subcases can be distinguished: $g_S \Leftrightarrow g_T$ and $d_S \Leftrightarrow d_T$.

3.1.1.1 *Grammemes*: $g_S \Leftrightarrow g_T$ A grammeme-to-grammeme correspondence presupposes that both $\mathcal{L}_S$ and $\mathcal{L}_T$ have the respective inflectional categories and these categories are semantically sufficiently close.

For the time being, we do not know of any convincing example of a one-to-one correspondence between two genuine grammemes of a language pair. Even such grammemes as these of nominal number ('SG' and 'PL') or of verbal tense ('PRES', 'PAST', 'FUT') in relatively close languages show multiple violations of a one-to-one correspondence (cf. next subsection). However, if we agree to speak of near-one-to-one correspondence, then, for instance, in SAE languages[10] in most cases 'SG' corresponds to 'SG' and 'PL' to 'PL'.

As for quasi-grammemes, one-to-one correspondences are more common. For instance, the quasi-grammeme 'CAUS(ative)' in Turkic languages is very likely to correspond, at least in most cases, to the same quasi-grammeme in Hungarian or Japanese. Consider an example for Turkish (8) and Hungarian (9):[11]

(8) a. *Et*   +Ø   *piş*   +*iyor*   +Ø
        meat NOM cook_V   IND.PRES   3SG
        'Meat is cooking'.

    b. *Aşçi* +Ø   *et*   +*i*   *piş*   +***ir***   +*iyor*   +Ø
        cook_N   NOM meat ACC cook_V   **CAUS**   IND.PRES   3SG
        '[The] cook is cooking [the] meat'.

---

[10] SAE = Standard Average European, in terms of B. Whorf.

[11] In Turkish: *ç* = /č/, *ş* = /š/; in Hungarian: *cs* = /č/, *s* = /š/, *sz* = /s/.

    c.   *Anna*   +Ø   *et*   +*i*   *aşçi*   +*ya*
           mother   NOM meat   ACC cook$_N$ DAT
           piş   +*ir*   +*t*   +*iyor*   +Ø
           cook$_V$ **CAUS**   **CAUS**   IND.PRES   3SG
           'Mother makes [the] cook cook [the] meat'.

(9) a.   *Hús* +Ø   *fő*   +Ø   +Ø
           meat   NOM cook$_V$ IND.PRES   3SG
           'Meat is cooking'.

    b.   *A szakács* +Ø   *a hús* +*t*   *fő*   +*z*   +Ø   +*i*
           the cook$_N$ NOM the meat ACC cook$_V$ **CAUS** IND.PRES 3SG
           'The cook is cooking the meat'.

    c.   *Anya* +Ø   *a hús* +*t*   *a szakács* +*csal*
           mother   NOM the meat   ACC the cook$_N$ INSTR
           fő   +*z*   +*et*   +Ø   +*i*
           cook$_V$ **CAUS**   **CAUS**   IND.PRES   3SG
           'Mother makes the cook cook the meat'.

The parallelism of the causatives in both languages is obvious.[12]

3.1.1.2 *Derivatemes: $d_S \Leftrightarrow d_T$*   The German derivateme 'DIMIN(utive)', expressed by
the suffix -*chen*, regularly corresponds to the Spanish derivateme 'DIMIN', expressed
by a series of morphonologically distributed suffixes: -*ito*, -*cito*, -*illo*, -*cillo*; cf. two
German examples from the Internet and their translation into Spanish:

(10) a. *Schon bald sah er ein kleines Weg+**chen**, das zu einem Bauernhof führte*
        lit. 'Soon he saw a small nice.little.road that led to a farm'.
        ≡
        *Ya pronto vio un pequeño camin+**ito** que conducía a una granja* [idem].
    b. *Ein merkwürdiges Düft+**chen** kommt mit einem Lüft+**chen** aus der Küche*
        lit. 'A strange nice.little.smell comes with a nice.little.draft from the kitchen'.
        ≡
        lit. *De la cocina llega con un air+**ito** un olor+**cito** estraño* [idem].

Further instantiations of this kind of correspondence include international prefixes
such as *anti*-, *pro*-, *mini*-, *super*-, and so on.

3.1.2 One-to-many paradigmatic correspondence between GS$_S$ and GS$_T$

Again, two subcases can be distinguished: $g_S \Leftrightarrow g_{1_T}|g_{2_T}|g_{3_T}|\ldots$ and $d_S \Leftrightarrow d_{1_T}|d_{2_T}|$
$d_{3_T}|\ldots$

---

[12] Of course, due to lexicalized causative forms, there are numerous exceptions. Thus, in Turkish, 'kill' is
a causative verb formally derived from **öl**+*mek* 'die': **öl**+***dür***+*mek*, while in Hungarian, **hal**+*ni* 'die' and
gyilkol+*ni* 'kill' have two unrelated stems.

3.1.2.1 *Grammemes:* $g_S \Leftrightarrow g_{1_T} | g_{2_T} | g_{3_T} | \ldots$ A classical example of this subcase is the nominal number: 'SG_S' corresponds most frequently to 'SG_T', but also—in several special cases—to 'PL_T'; 'PL_S' corresponds most frequently to 'PL_T', but can also correspond to 'SG_T'. Consider five examples of 'SG' $\Leftrightarrow$ 'PL' correspondences:

1. A mass noun in $\mathcal{L}_S$, which does not have a plural, can correspond to a plural count noun in $\mathcal{L}_T$.

    (11) a. Eng. *fight against graft*$_{\text{SG}}$

    $\equiv$

    Fr. *lutte contre les pots-de-vin*$_{\text{PL}}$
    lit. 'fight against the bribes'

    b. Rus. *Ja ljublju gorox*$_{\text{SG}}$ 'I like peas'.

    $\equiv$

    Ger. *Ich mag Erbsen*$_{\text{PL}}$ [idem]

    Slightly different and more complex is the correspondence between 'SG' and 'PL' if the singular mass noun has a "lexical pluralization;" cf.:
    Eng. *information*$_{\text{SG}}$ (no plural) $\equiv$ Fr. *les informations*$_{\textbf{PL}}$, but also
    pieces of information $\equiv$ *les informations*
    Eng. *advice*$_{\text{SG}}$ (no plural) $\equiv$ Fr. *les conseils*$_{\textbf{PL}}$, but also
    pieces of advice $\equiv$ *les conseils*

2. A singular collective noun in $\mathcal{L}_S$ corresponds to a plural count noun in $\mathcal{L}_T$ (a case which is very close to the preceding one): Ar. *naḥl+u* 'bee' (collective) $\equiv$ *bees*.[13]

    (12) Ar. *Jaʔiišu al+naḥl+ u*$_{\textbf{SG}}$ *fii džamiiʔ+i anḥa'+ i uruubaa*
    lit. 'Lives the.bee in all corners of.Europe'.

    $\equiv$

    Bees$_{\text{PL}}$ *live all over Europe.*

3. The singular definite noun in the generic reading in $\mathcal{L}_S$ corresponds to a plural definite noun in $\mathcal{L}_T$:[14]

    (13) Ger. *Warum trinkt der Russe*$_{\text{SG}}$ *Wodka und der Schotte*$_{\text{SG}}$ *Whisky?*
    lit. 'Why does the Russian drink vodka and the Scot whisky?'

    $\equiv$

    Rus. *Počemu russkie*$_{\textbf{PL}}$ *p'jut vodku, a šotlandcy*$_{\textbf{PL}}$ *viski?*
    lit. 'Why do Russians drink vodka and Scots whisky?'

---

[13] Arabic derives singulatives from collectives; thus, from *naḥl+u* we get *naḥl+at+u* 'bee', which forms its own plural *naḥlaatu*: *Hunaaka naḥl+at+u+n madžit+at+u+n ʕalaa al+minḍadat+i* 'There is a dead bee on the table' vs. *Hunaaka ʕašarat+u naḥl+aat+u+n madžitat+u+n ʕalaa al+minḍadat+i* 'There are 10 dead bees on the table'.

[14] Depending on the context, the correspondence may be more complex. Thus, for Ger. *Der Durchschnittsrusse trinkt im Jahr 230 Liter Milch* lit. 'The average.Russian drinks in.the year 230 liters milk', the translation of *der Durchschnittsrusse* must be in the singular: Rus. *Srednestatističeskij russkij*$_{\text{SG}}$ *vypivaet za god 230 litrov moloka* lit. 'Average Russian drinks.out per year 230 liters milk'; here, the singular of *russkij* 'Russian' is imposed by *srednestatističeskij*. However, another translation, with the plural, is also possible: *Russkie*$_{\text{PL}}$ *vypivajut v srednem za god po 230 litrov moloka* lit. 'Russians drink.out on average per year each 230 liters milk'. We can obtain the second variant by using DSynt-paraphrasing rules (see Mel'čuk and Wanner 2006).

4. A dual noun in a language with singular $\sim$ dual $\sim$ plural distinctions (such as Arabic or Slovene) corresponds to a plural noun in a language without the dual: Ar. *ʕajn+aani*$_\text{DU}$ lit. 'pair of eyes' $\equiv$ *eyes*$_\text{PL}$ or *saaq+aani*$_\text{DU}$ lit. 'pair of legs' $\equiv$ *legs*$_\text{PL}$. If the noun does not refer to one of the two parts of a "paired" object, in a language without a dual the numeral 'two' has to be added: Ar. *kitaabu* 'book' versus *kitaab+aani* $\equiv$ *two books*.[15]

5. French (and to a lesser extent Spanish) has the so-called distributive singular noun (i.e. a nominal direct object in the singular referring to one of several entities, each of which is associated with each of the many referents of the plural subject). Such a singular noun corresponds to a plural noun in English, German, Russian, etc.:

   (14) a. Fr. *Les étudiants ont levé leur main*$_\text{SG}$
            lit. 'The students have raised their hand' (each of them raised only one hand)
            and
         b. Fr. *Les étudiants ont levé leurs mains*$_\text{PL}$
            lit. 'The students have raised their hands' (each of them raised both hands)
            $\equiv$
         c. *The students raised their hands*$_\text{PL}$;
            Ger. *Die Studenten hoben ihre Hände*$_\text{PL}$;
            …

Another well-known case of one-to-many GS-correspondence is the verbal tense in SAE languages. For instance, both Russian and German have similar tenses: 'PRES(ent)', 'PAST', and 'FUT(ure)'.[16] Nonetheless, there is no one-to-one correspondence between these grammemes. Thus, in German, the grammeme 'PRES' is much more often used to refer to future events than in Russian or English:

(15) *Dann kommt*$_\text{PRES}$ *er morgen/*…wird er morgen kommen*$_\text{FUT}$
     lit. 'Then comes he tomorrow/will he tomorrow come'.
     'Then he will come tomorrow'/*'…he comes tomorrow'.
     $\equiv$
     Togda on pridët$_\text{FUT}$ *zavtra/*…prixodit*$_\text{PRES}$ *zavtra*
     [idem].

A similar case occurs between German and Spanish. The German imperfect (in our notation: 'PAST', 'NON-PERF') corresponds in Spanish to the imperfect or to the preterit:

(16) a. *Als ich im Zimmer war*$_\text{PAST,NON-PERF}$, *kam Vater herein*
        lit. 'When I was in the room, father came in'.
        $\equiv$
        Cuando estaba$_\text{IMPF}$ *yo en la habitación, entró el padre.*

---

[15] This is, in fact, a complex case of correspondence, with the addition of a lexeme; see Subsect. 3.2 below.

[16] This becomes explicit if we organize German verb forms along two independent axes: TENSE = {'PRES', 'PAST', 'FUT'} and PERFECTIVITY = {'NON-PERF(ect)', 'PERF'}; see Appendix 1, A1.1.

Table 1 German versus Ewenki actor derivateme equivalents

$V - en_{INF}$		V-*er*			
		'who is X-ing now'/'who X-es professionally'			
German					
jag + en	'hunt'	*Jäg + er*	'hunter'		
fisch + en	'fish'	*Fisch + er*	'fisher'		
führ + en	'guide'	*Führ + er*	'guide'		
V		X-*d′e*		X-*mnī*	
		'who is X-ing now'		'who X-es professionally'	
Ewenki					
beju	'hunt'	*beju + d′e*	'hunting man'	*beju + mnī*	'hunter'
olomī	'fish'	*olomī + d′e*	'fishing man'	*olomī + mnī*	'fisher'
elge	'guide reindeer'	*elge + d′e*	'guiding man'	*elge + mnī*	'guide'

      b. *Vor zwei Jahren war*$_{PAST,NON-PERF}$ *ich in Paris*
        'Two years ago I was in Paris'.
        ≡
        Hace dos años estuve$_{PRET}$ *in París.*

3.1.2.2 *Derivatemes: $d_S \Leftrightarrow d_{1_T}|d_{2_T}|d_{2_T}|\ldots$* Multiple paradigmatic derivateme correspondence is less common. A clear-cut example is the correspondence of the actor derivateme in German or English with the actual actor or usual actor derivateme in Ewenki (a Tungus language of Siberia); cf. some German–Ewenki equivalents in Table 1.

    Another example is the correspondence between diminutives in German, Spanish, Russian, on the one hand, and Italian, on the other.[17] For instance, German has a diminutive suffix -*chen*, which has a rather positive meaning: '[X] perceived as little and nice', but can imply something negative, because 'not big enough'. In contrast, Italian has two series of contrasting diminutive suffixes: positive -*ino*, -*etto*, etc., and negative -*uccio*; cf. (17):

(17)   a. *Kürzlich hatte er ein vielversprechendes Idee+**chen**.*
        ≡
       *Poco tempo fa ha avuto una ide+**ina** molto promettente*
       lit. 'Recently he had a very promising nice.little.idea'.
      b. *Kürzlich hatte er ein müdes Idee+**chen**.*
        ≡
       *Poco tempo fa ha avuto una fiacca ide+**uccia***
       lit. 'Recently he had a shallow little.idea'.

---

[17] For a very detailed contrastive discussion of diminutives in Italian, German and English (see Dressler and Merlini 1994, pp. 84–414).

3.2 $GS_S \Leftrightarrow LU_T$

As is the case with $GS_S \Leftrightarrow GS_T$, the $GS_S \Leftrightarrow LU_T$ correspondences subsume logically the same two subtypes: a GS_S corresponds to a LU_T and vice versa (one-to-one correspondence); a GS_S corresponds to several alternative LU_Ts (one-to-many paradigmatic correspondence).

3.2.1 One-to-one correspondence between GS_S and LU_T

As before, GS_S may be a (quasi-)grammeme or a derivateme.

3.2.1.1 *(Quasi-)grammemes: $g_S \Leftrightarrow LU_T$*   Let us first give five examples of grammemes.

1. The nominal dual number corresponds in most cases to the lexeme meaning 'two', with the exception of paired objects; see Subsect. 3.1.2.1, Item 4 above.

   (18)   a. Ar. *Aʕṭat+nii Diina tuffaaḥat+ini*_{DUAL,ACC}

   $\equiv$

   b. *Dina gave me **two** apples.*

2. The nominal possession grammemes of languages such as Hungarian or Turkish correspond in an SAE language to possessive pronominal adjectives: Hung. *könyv+**em*** '**my** book', *ház+**ad*** '**your** (sing.) house', *szóbá+**nk*** '**our** room'; Turk. *kitab+**ım*** '**my** book', *ev+**in*** '**your** (sing.) house', *oda+**mız*** '**our** room'.

3. A semantic locative case grammeme of such a language as Hungarian, Finnish or any Daghestanian language corresponds to a meaningful preposition in SAE languages:

   (19)   Hungarian

   a. *A könyv az asztal+ **on** *_{SUPERESSIVE} *van*

   lit. 'The book on the table is'.

   $\equiv$

   *The book is **on** the table.*

   b. *A könyv+et az asztal+**ra***_{SUPERLATIVE} *tet+t+em.*

   $\equiv$

   *I put the book **on** the table.*

4. The verbal 'CIT(ative)' grammeme in Bulgarian has to be rendered—if $\mathcal{L}_T$ does not have the corresponding grammatical means—by lexical means.[18] A rough translation of the citative into English is *It is said that …*

   (20)   Bulgarian

   a.

Kupil	*si*	*e*_{IND}	*novo palto.*
bought	himself	is	new coat
versus			
Kupil	*si*	*Ø*_{CIT}	*novo palto.*
bought	himself	Ø	new coat

   $\equiv$

---

[18] In Bulgarian, the citative exists only in the past and future, unlike, for instance, Estonian, which has also the present of the citative.

He bought himself a new coat.
versus
It is said that he bought himself a new coat.

b.

$Šte_{IND}$	si	kupi	novo palto.
will	himself	buys	new coat

versus

$Štjal_{CIT}$ da	si	kupi	novo palto.
would	himself	buys	new coat

≡

He will buy himself a new coat.
versus
It is said that he will buy himself a new coat.

Unlike the previous examples, in (20) the English equivalent of the Bulgarian citative is not unique; in different contexts, different translation equivalents might be more appropriate (such as *He is said to …*, *Perhaps, he …*, *I have heard that he …*). However, it is enough to indicate in the bilingual index, BI (see Sect. 5), the above equivalent only. The other must be computed within English rather than established during the transfer procedure.[19]

5. The grammeme 'COND(itional)' in French can be used to express the Speaker's uncertainty, very similar to the citative in Bulgarian. In Russian, this meaning can be rendered, among other things, by the LU JAKOBY 'presumably' or by a phrase of the type KAK SOOBŠČAETSJA lit. 'as is.communicated':

(21)   Fr. *Il aurait_{COND} reçu une offre de 2 millions de dollars*
       'He is said to have received an offer of two million dollars'.
       ≡
       Rus. *Emu **jakoby** <**kak soobščaetsja**> bylo predloženo dva milliona dollarov*
       lit. 'To.him presumably <as is.communicated> were offered two million dollars'.

As for quasi-grammemes, we will consider three examples: 'CAUS(ative)', 'PO-TENT (ialis)' and 'DESIDER(ative)'.

1. The causative quasi-grammeme often corresponds to a special verbal lexeme in a language without morphological causative.

(22)   Hung. *Ír+at_{CAUS}+om a level+et János+sal*
       lit. '[I] make.write the letter by.Janosh'.
       ≡
       Ger. *Ich **lasse** Janosch den Brief schreiben*
       lit. 'I let Janosh the letter write'.
       ≡
       *I **let/make** Janosh write the letter.*

---

[19] In practical MT, the proposed translation equivalent *It is said that* must be marked for possible human postediting.

Similar correspondences are observed, of course, with English and French as $\mathcal{L}_T$, but not, for instance, with Slavic languages, which do not have a standard expression for causative meaning.

2. Hungarian does not have a verb for the meaning 'can' = 'be allowed to', 'have the possibility of'. This meaning is always expressed by the quasi-grammeme of potentialis. When translating from Hungarian into a language such as English, French, etc., the corresponding modal verb must be introduced; cf.:

(23) a. *Jár+ok gyalog* lit. '[I] walk on.foot'.
          versus
      b. *Jár+**hat**+ok gyalog* lit. '[I] may.walk on.foot' =
          'I am allowed to walk [for instance, by my doctor]'.

3. The Japanese quasi-grammeme 'DESIDER(ative)' corresponds in SAE languages to a lexeme meaning 'want, desire'; cf.:

(24)   *Bokuwa   kohiga   nomi   +**tai*** [NOMU DESIDERAT].
       I   coffee   drink   is.desirable
       ≡
       *I **want** to drink coffee.*

This example implies also a deep-syntactic mismatch; for this reason it has already been considered under a different angle in Mel'čuk and Wanner (2006, pp. 110–111).

3.2.1.2 *Derivatemes: $d_S \Leftrightarrow LU_T$* $d_S \Leftrightarrow LU_T$ correspondences fall into three major classes:

(i)   a derivational means[20] is translated by a separate LU:
      $[L_S+\text{'DERIVATEME}_S\text{'}] \Leftrightarrow [L_T\ L'_T]$,
      where $L_T = \tau_l(L_S)$ and $L'_T = \tau_l(\text{'DERIVATEME}_S\text{'})$[21]
(ii) a derivational means is translated by an LU which makes part of a compound:
      $[L_S+\text{'DERIVATEME}_S\text{'}] \Leftrightarrow [L_T + L'_T]$
(iii) a source LU combined with a derivational means is translated by an LU which replaces both:
      $[L_S+\text{'DERIVATEME}_S\text{'}] \Leftrightarrow [L'_T]$,
      where $L'_T = \tau_l(L_S+\text{'DERIVATEME}_S\text{'})$

The first class can be illustrated by the following four examples.

1. The Czech derivateme 'LANGUAGE', expressed by the suffix *-ština/ -č(t)ina*, corresponds to the English lexeme LANGUAGE:[22]

čeština	'Czech language'	*angličtina*	'English language'
mad'arština	'Hungarian language'	*němčina*	'German language'
norština	'Norvegian language'	*turečtina*	'Turkish language'
ruština	'Russian language'	*řečtina*	'Greek language'
etiopština	'Ethiopian language'	*baskičtina*	'Basque language'

---

[20] Derivational means include affixes, conversions and reduplications.

[21] For translation functions τ, see 5.1.

[22] Along with this suffix, Czech also has expressions with the noun *jazyk* 'language': *český jazyk*, *ruský jazyk*, etc. The derivatives with the suffix *-ština* are slightly more colloquial.

2. The German complex verbal prefixes of movement verbs *hinauf-/herauf-* versus *hinab-/herab-*, *hinein-/herein-* versus *hinaus-/heraus-*, etc. encode two types of orientation: (i) deictic orientation (with respect to the Speaker) expressed by *hin-*, if the position of the Speaker is not the endpoint of the movement or is irrelevant, and by *her-*, if the position of the Speaker is at the endpoint of the movement; (ii) spatial orientation expressed by *auf-* 'up' versus *ab-* 'down', *ein-* 'in' versus *aus-* 'out', etc. For instance, *hinaufsteigen* lit. 'climb up from me' versus *heraufsteigen* lit. 'climb up to me' and *hinabsteigen* lit. 'climb down from me' versus *herabsteigen* lit. 'climb down to me'.

    In Russian, the different spatial orientations are expressed by different verbs, while deictic orientation need not (and sometimes cannot) be expressed: *hinaufsteigen* ≡ *podnimat′sja* 'climb up' (with loss of deictic information), *heraufsteigen* ≡ *podnimat′sja* (*ko mne*) 'climb up (to me)', *hinabsteigen* ≡ *spuskat′sja* 'climb down' (with loss of deictic information), *herabsteigen* ≡ *spuskat′sja* (*ko mne*) 'climb down (to me)'. As the glosses show, the Russian verbs do not indicate the position of the Speaker with respect to the movement. Therefore, during the translation from Russian into German, the position of the Speaker must be determined based on the situational information.

3. The Spanish derivateme 'IMPACT/BLOW OF X' is translated into French by the noun COUP 'blow'.

    (25)   Sp. *Recibió un botell+**azo** en la cabeza.*
           ≡
           Fr. *Il a reçu **un coup de bouteille** sur la tête*
           lit. 'He received a blow of bottle on the head'.

4. The Russian derivateme 'BEGIN' denoted by the inchoative prefix *za-* is expressed in English by an LU: verbs START or BEGIN.

    (26)   ***za**+begat′* ≡ ***start** running*, ***za**+pet′* ≡ ***start** singing*,
           ***za**+goret′sja* ≡ ***start** burning*, ***za**+govorit′* ≡ ***start** talking*, etc.

The second class contains, for example, the following correspondence. The Russian suffix -ONOK 'young of …' is translated into German by the LU [*das*] JUNGE 'young of …', which builds a compound together with the name of the animal:

(27) a. Rus. MEDVED′ 'bear' ∼
          MEDVED′+ONOK = MEDVEŽONOK 'bear cub'
          ≡
          Ger. BÄR 'bear' ∼
          BÄR+EN+JUNGE = BÄRENJUNGE 'bear cub'
      b. Rus. DEL′FIN 'dolphin' ∼
          DEL′FIN+ONOK = DEL′FINËNOK 'dolphin calf'
          ≡
          Ger. DELPHIN 'dolphin' ∼
          DELPHIN+JUNGE = DELPHINJUNGE 'dolphin calf'

The third class can be illustrated by a slightly more complex phenomenon. The Russian suffix -ATINA 'meat of …' is rendered into French by converting the name of the animal (a count noun) into the name of its meat (a mass noun), i.e. by a separate

lexeme (another lexicographic sense of the given animal noun that need not appear in the dictionary since it corresponds to a productive regular polysemy schema, typical of French):

(28) a. Rus. MEDVED′ 'bear' ∼
          MEDVED′+ATINA = MEDVEŽATINA 'bear meat'
          ≡
          Fr. [un] OURS 'bear' ∼ [de l']OURS 'bear meat'
      b. Rus. ZAJAC 'hare' ∼
          ZAJAC+ATINA = ZAJČATINA 'hare meat'
          ≡
          Fr. LIÈVRE 'hare' ∼ [du] LIÈVRE 'hare meat'

3.2.2 One-to-many paradigmatic correspondence between GS_S and LU_T

Although the correspondence of a (quasi-)grammeme to several LUs is less frequent, it can be encountered. Again two subcases can be distinguished:

$$g_S \Leftrightarrow LU_{T_1} \mid LU_{T_2} \mid LU_{T_3} \mid \ldots \text{and } d_S \Leftrightarrow LU_{T_1} \mid LU_{T_2} \mid LU_{T_3} \mid \ldots$$

3.2.2.1 *Grammemes: $g_S \Leftrightarrow LU_{T_1} \mid LU_{T_2} \mid LU_{T_3} \mid \ldots$* For instance, the Spanish grammeme 'PL' has the following two alternative lexical equivalences:

1. In combination with certain human nouns of masculine gender, the 'PL' grammeme creates the name of a couple:

mis tios lit. 'my uncles'	≡	*my uncle and his wife* or
		my aunt and her husband
los reyes lit. 'the kings'	≡	*the king and his wife* or
		the queen and her husband
los príncipes lit. 'the princes'	≡	*the prince and his wife* or
		the princess and her husband

    $[X]_{pl}$ is translated as '$[\tau(X)]$ and $[\tau(X)$'s] wife' or '$[\tau(X)]$ and $[\tau(X)$'s] husband'. This case is lexically highly restricted, but we choose to present it because it illustrates a phenomenon regular in certain languages (e.g. in Turkic languages, $[X]_{pl}$, where X is a proper name, means '[X] and [X's] people').

2. The 'PL' grammeme on a common noun combined with a human proper noun in apposition is used to create the name of an enterprise: *Harinas Alonso* lit. 'flours Alonso', *Flores Ramos* 'flowers Ramos', *Cervezas August Damm* lit. 'beers August Damm', etc. It is rendered into English as *supplies*:

Harinas Alonso	≡	*Alonso Flour Supplies*
Flores Ramos	≡	*Ramos Flower Supplies*
Cervezas August Damm	≡	*August Damm Beer Supplies*

    $[X]_{PL}[N_{(hum\text{-}prop)}]$ is translated as $[N_{(hum\text{-}prop)}\tau(X_{SG})]$ 'supplies'.

3.2.2.2 *Derivatemes: $d_S \Leftrightarrow LU_{T_1} \mid LU_{T_2} \mid LU_{T_3} \mid \ldots$* The following two examples from Russian and Spanish illustrate this case.

1. Russian has a highly productive deverbal derivateme [X] 'DO+L+SJA' with the meaning 'something bad happens to X because of X's doing L excessively', as in

X do+katat′+sja [*na lyžax*] 'something bad happens to X because of his excessive skiing', *X do+guljat′+sja* 'something bad happens to X because of his excessive going for a walk', *do+čitat′+sja* '…because of his excessive reading', etc. As far as we know, other languages do not have such a derivateme. For instance, in German, the translation equivalent can be one of the two following idioms: in the past tense, *X L, das hat X*$_{pron}$ *nun davon* lit. 'X L, this has X now there.from'; in the future tense, *X L, dann wird X*$_{pron}$ *schon sehen* lit. 'X L, then will X already see'. Cf. the following Russian–German sentence pairs:

(29) a. *On katalsja na velosipede v centre goroda – i dokatalsja*
          lit. 'He rode his bicycle downtown—and something bad happened to him because of that'.
          ≡
          Er ist im Stadtzentrum Fahrrad gefahren – das hat er nun davon [idem].

      b. *Esli budeš' čitat' tak mnogo detektivov, ty dočitaeš'sja*
          lit. 'If you read so many mystery novels, something bad will happen to you because of that'.
          ≡
          Wenn du weiterhin so viele Krimis liest, wirst du schon sehen.

Obviously, the translations proposed are not unique; in particular contexts, other translations can be preferable or even necessary. For instance, in (29a), another valid (more narrow) translation is possible, namely, when the bad thing that happened to X is a traffic accident: *Er ist im Stadtzentrum Fahrrad gefahren – bis es ihn erwischt hat* lit. 'He rode the bicycle downtown—until it caught him' [= 'He rode the bicycle downtown until he had an accident'].

Furthermore, the [X] *do+L+sja* verbs take an actant of the form *do* 'till' + N$_{GEN}$, which expresses what exactly happened to X. In these cases, yet a different, more specific translation is required:

(30) *On dorabotalsja do ser′eznogo nervnogo sryva*
      lit. 'He worked himself until a serious nervous breakdown'.
      ≡
      Er hat gearbeitet, bis er einen Nervenzusammenbruch erlitt
      lit. 'He has worked until he suffered a nervous breakdown'.

However, these complications do not change the general picture.

2. A Spanish diminutive derivateme is usually translated into English by the adjective LITTLE or SMALL: *hotel+**ito** ≡ **little** hotel, mes+**ita** ≡ **small** table*.

3.3 $GS_S \Leftrightarrow —$

In this case, GS_S can be only a grammeme: a source (semantic) derivateme absent from the $\mathcal{L}_T$ can always have an LU$_T$ as a standard equivalent (see Subsect. 3.2). Since derivatemes are not obligatory and thus not very frequent, the systematic use of a corresponding LU$_T$ is warranted even if it does not fit well in all contexts; in the

worst case, it can lead to a text which is less than ideally idiomatic, but not ungrammatical.[23]

In the case of grammemes, such a strategy would inevitably result in a severe violation of the style conventions in $\mathcal{L}_T$. For instance, in the case of translation from Korean and Japanese, the full reflection of the politeness markers leads to an unreadable, almost humoristic text; cf. a Korean example:[24]

(31) Korean

Kyosu $+$***nim*** $+\emptyset$,   ***cinci*** $+\emptyset$
professor   HON   VOC_{HON}   honorable.meal   NOM

tu $+$***si*** $+$***p*** $+$***si*** $+$***o!***
honorably.eat   HON_{SBJ}   HON_{ADDR}   REQ   IMPER_{DEFER}

lit. 'O honorable.professor, deign.honorably.eat honorable.meal!'
versus
Cingu $+\emptyset$ $+$***ya***,   ***pap*** $+\emptyset$ ***mek*** $+$***e*** $+$***la!***
friend   PLAIN   VOC_{PLAIN}   meal   NOM   eat   INF   IMPER_{PLAIN}

lit. 'Friend, eat meal!'

Other well-known examples of source language grammemes that disappear in $\mathcal{L}_T$ include:

–   The grammemes of the nominal number in SAE languages: the category of nominal number is absent from Far East languages such as Chinese, Vietnamese and Japanese. When translating towards one of these languages, the corresponding grammemes are to be omitted, while under inverse translation they need to be computed, based on contextual information retrieved from the source text.
     A similar case is related to the translation of compounds. As a general rule, a dependent element of a compound does not show inflection. Thus, the first element of a German compound of the type *Reisewarnung* lit. 'travel.warning', i.e. *Reise-* 'travel' (see above), does not carry a grammeme of otherwise obligatory category of number. In languages where such a compound is translated by two nouns, the equivalent of *Reise* must have a number grammeme: Rus. *predostereženie protiv poezdok*_{PL} lit. 'warning against travels'.
–   Definiteness in article languages vs. absence of definiteness in so many other languages; aspect of Slavic languages vs. the absence of this type of aspect in most other languages. Consider an example of the translation from French to Russian, where the French grammeme 'DEF' disappears without a trace:

---

[23] Cases of type Fr. *cable métallique* [with *métallique* = A$_0$(MÉTAL)] ≡ *metal cable*, where the derivateme A$_0$(MÉTAL) disappears, do not contradict our statement; this is a purely syntactic derivateme (covered by the LF A$_0$), while we consider here only semantic derivatemes.

[24] The following abbreviations are used in this example: 'HON': honorific, 'HON_{SBJ}': subject honorific, 'HON_{ADDR}': addressee honorific, 'DEFER': deferential, 'IMPER_{DEFER}': deferential imperative, 'REQ': requestive mood. CINCI- 'meal' and TU- 'eat' are special "deferential" stems. The signs related to the expression of the politeness level are boldfaced. See Appendix 1, A1.2, for some details on the Korean politeness markers.

(32)   Fr. ***L'incorporation*_{DEF}** *devient obligatoire sous* **les** *conditions*_{DEF} *correspondantes*

'Incorporation becomes obligatory under the corresponding conditions'.

≡

Rus. *Inkorporacia stanovitsja objazatel′noj pri sootvestvujuščix uslovijax.*

For a presentation of definiteness and aspect under translation, see Sect. 6.

– The grammemes of tense and aspect are absent from Chinese and Vietnamese. Under translation from these languages into an SAE language, they have to be computed from context clues (see, e.g. Olsen et al. 2001).

4 Morphological divergences in DSyntSs

Some of the grammatical signification correspondences in the above typology constitute a mismatch at the DSynt-level and need to be handled as morphological mismatches during morphological transfer. Others turn out to be "pseudo-mismatches," which disappear when the source sentence is parsed down to the DSyntS. Let us discuss, in what follows, both types of divergences.

4.1 Morphological mismatches in DSyntSs

In order to determine which of the grammatical signification correspondences constitute a morphological mismatch, we need a formal definition of DSyntS as well as the definitions of isomorphism of DSyntSs and that of a linguistic mismatch at the DSynt-level in general.[25]

Definition 5 (*Deep-Syntactic Structure, D Synt S*)

Let L_d, G_{sem}, D_{sem} and R_{dsynt} be four disjunct alphabets for language $\mathcal{L}$, where: L_d is the set of deep lexical units (LUs), G_{sem} is the set of semantic (quasi-)grammemes, D_{sem} is the set of semantic derivatemes,[26] and R_{dsynt} is the set of deep-syntactic relations.

A DSyntS of sentence S in $\mathcal{L}$, denoted as S_{DSynt}, is an 8-tuple over $L_d \cup G_{sem} \cup D_{sem} \cup R_{dsynt}$ of the following form:

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

where:

---

[25] For a more detailed presentation of DSyntS in the context of MT (see Mel'čuk and Wanner 2001, 2006).

[26] MTT distinguishes two types of semantic derivatemes: (1) derivatemes that are represented in the DSyntS indirectly by an LF, as in S₁(SMOKE_V)[= *smoker*], Fr. Mult(ÉLECTEUR 'voter') = *électorat* 'electorate', etc., and (2) derivatemes that are represented directly by a derivateme symbol, as in Sp. CEREZA 'cherry'⊕'TREE' [= *cerezo* 'cherry tree']. In the context of DSyntS-transfer, only the second type of derivatemes needs to be considered. For the interested reader, let us indicate that the use of LFs and derivatemes in DSyntSs serves to maximally reduce the size and complexity of lexical transfer bilingual indexes; a fully compositional derived lexeme is not supposed to be included into the transfer index, cf. Sect. 5.

- N is the set of nodes and A the set of directed arcs (or *branches*) such that these nodes and arcs form a dependency tree, with the initial node n^{init} and the end node n^{end} defined for each arc;
- $\lambda_{l_s \to n}$ is a function that assigns to each $n_i \in N$ an $l_i \in L_d$;
- $\rho_{r_s \to a}$ is a function that assigns to each $a_i \in A$ an $r_i \in R_{dsynt}$;
- $\gamma_{n \to g}$ is a function that assigns to the name of each LU associated with a node $n_i \in N, l_i \in \lambda_{l_s \to n}(N)$, a set of corresponding grammemes $G_{sem_i}, G_{sem_i} \subset G_{sem}$;
- $\delta_{n \to d}$ is a function that assigns to the name of a given LU associated with a node $n_i \in N, l_i \in \lambda_{l_s \to n}(N)$, a derivateme $d \in D_{sem}$.

Definition 6 (*Isomorphism of semantically equivalent DSyntSs*)

Let there be two DSyntSs:

$$S_1 = \langle N_1, G_{sem_1}, D_{sem_1}, A_1, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g}, \delta_{n \to d} \rangle$$

and

$$S_2 = \langle N_2, G_{sem_2}, D_{sem_2}, A_2, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g}, \delta_{n \to d} \rangle.$$

S_1 and S_2 are isomorphic iff

(a) $\forall n_i \in N_1$: A node correspondence function ν maps n_i onto the node $n'_j \in N_2$, i.e. $n'_j = \nu(n_i)$, and thus ensures a one-to-one correspondence between the nodes of S_1 and S_2;

and

(b) the following four types of one-to-one correspondence functions hold: (b.1) lexical correspondence function τ_{l_s}, (b.2) syntactic correspondence function τ_{a_s}, (b.3) semantic (quasi-)grammeme correspondence function τ_{g_s} and (b.4) semantic derivateme correspondence function τ_{d_s}.

(b.1) $\forall n_i \in N_1$: The lexical correspondence function τ_{l_s} maps the lexical label l_i of n_i onto the lexical label l'_i of the corresponding node $n'_i \in N_2$, i.e. $l'_i = \tau_l(l_i)$, and thus ensures the correspondence equivalence of the LUs labeling the corresponding nodes n_i and n'_i.

(b.2) $\forall (n_i^{init}, n_i^{end}, a_i) \in S_1$: The (syntactic) arc correspondence function τ_{a_s} establishes a correspondence between $(n_i^{init}, n_i^{end}, a_i)$ and $(n_i'^{init}, n_i'^{end}, a'_i) \in S_2$.

(b.3) $\forall n_i \in N_1$: The (quasi-)grammeme correspondence function τ_{g_s} establishes the correspondence between a (quasi-)grammeme g_i of the lexical label l_{n_i} and the (quasi-)grammeme g'_i of the lexical label l'_{n_j} of the corresponding node n'_j.

(b.4) $\forall n_i \in N_1$ for which $\delta_{n \to d}$ is defined: The derivateme correspondence function τ_d establishes the correspondence between the lexical label $l_{n_i} \oplus d_i$ (i.e. the union of a lexical label l_{n_i} and a derivateme d_i) associated with n_i and the lexical label of the corresponding node n'_j, this label being either l'_{n_j} (a lexeme) or $l'_{n_j} \oplus d'_j$ (i.e. the union of a lexeme and a derivateme).

In the transfer context, we are interested in the isomorphism and the linguistic mismatches between DSyntSs of $\mathcal{L}_S$ and $\mathcal{L}_T$. We speak of a "linguistic mismatch" between $DSyntS_S$ and $DSyntS_T$ if the isomorphism between them is violated:

Definition 7 (*Linguistic mismatch*)

Let there be two semantically (near-)equivalent DSyntSs S_S and S_T.

There exists a linguistic mismatch between S_S and S_T iff they are not isomorphic.

More precisely, a linguistic mismatch obtains when a linguistic item $\xi_S \in S_S$ corresponds either to a configuration of linguistic items $\{\xi'_T\}$ or to a single linguistic item ξ'_T of a different linguistic type than ξ_S, and vice versa.

The cases where ξ_S is a deep LU or a DSynt-relation configuration represent syntactic mismatches, which result from the lack of the isomorphism between the DSyntS$_S$ and its semantically equivalent DSyntS$_T$ (see above). If ξ_S is an LU, we observe the correspondence between an LU$_S$ and a phrase, cf. Serb. UJAK ≡ MATERNAL UNCLE, Rus. ÈLEKTRIČKA 'suburban train' ≡ Fr. TRAIN DE BANLIEUE, Ger. SONNEN-BRILLEN 'sunglasses' ≡ Rus. SOLNEČNYE OČKI, etc. If ξ_S is a DSynt-relation configuration, the correspondence holds between this configuration and a different configuration—such as, for instance, USED–**II**→READ ≡ Fr. LIRE 'read'–**ATTR**→ HABITUELLEMENT 'usually'. The resolution of any syntactic mismatch consists of the transduction of DSyntS$_S$ into DSyntS$_T$, necessarily involving some adjustments in accordance with contextual conditions (Mel'čuk and Wanner 2006).

Let us now come to the case where ξ_S is a grammatical signification.

Definition 8 (*Morphological mismatch*)

A linguistic mismatch is a *morphological* mismatch if ξ_S is a (quasi-)grammeme or the union of a lexical label and a derivateme.

According to Definition 8, the above typology of GS-correspondences shows the following three types of morphological mismatches:

(i) $g_S \Leftrightarrow LU_T$ (see 3.2.1.1)
(ii) $d_S \Leftrightarrow LU_T$ (see 3.2.1.2) and
(iii) $g_S \Leftrightarrow$ — (see 3.3).

In contrast, the cases $g_S \Leftrightarrow g_{T_1} \mid g_{T_2} \mid g_{T_3} \mid \ldots, d_S \Leftrightarrow d_{T_1} \mid d_{T_2} \mid d_{T_3} \mid \ldots$ and $d_S \Leftrightarrow LUd_{T_1} \mid LUd_{T_2} \mid LU_{T_3} \mid \ldots$, which also show a violation of the one-to-one correspondence, are not mismatches (in our technical sense).

4.2 Reduction of morphological mismatches in DSyntS

Someone knowledgeable in natural language morphology might come up with examples of morphological divergences not foreseen in the above typology. However, in Mel'čuk and Wanner (2006), it was shown that several divergences between syntactic structures of $\mathcal{L}_S$ and $\mathcal{L}_T$ often considered as mismatches disappear if the transfer is carried out at the DSynt-level. This concerns phenomena of surface syntax that are taken care of by analysis and synthesis modules of $\mathcal{L}_S$ and $\mathcal{L}_T$, respectively. In a similar vein, the recourse to DSyntS helps eliminate several morphological divergences, which are, thus, morphological pseudo-mismatches.

Below we present three sample cases of what can be easily mistaken for a morphological mismatch.

4.2.1 Pluralia/singularia tantum

The correspondence of the type Ger. *die Hose*$_{SG}$ 'pants' versus Rus. *brjuki*$_{PL}$, Ger. *die Schere*$_{SG}$ 'scissors' versus Rus. *nožnicy*$_{PL}$, Ger. *Waage*$_{SG}$ 'scales' versus Rus. *vesy*$_{PL}$ (these nouns are *pluralia tantum* in English and Russian) may strike the reader as an obvious case of morphological mismatch. However, in the DSyntS, a noun receives its grammatical number from the semantic structure. Therefore, a *plurale tantum* appears in the DSyntS with the 'SG' or 'PL' grammeme depending on the actual number of corresponding objects: thus, in the DSyntS, we can have *scissors*$_{SG}$ if this instance of the LU refers to one single object (a pair of scissors), and *scissors*$_{PL}$ if the reference is to more than one such object. The English (or Russian) monolingual dictionary contains the indication that this LU must always appear in the SSyntS in the plural, independently of its "deep" grammatical number.

The case of *singularia tantum* (*the summers*$_{PL}$ versus Rus. *leto*$_{SG}$, which has no plural) is analogous and does not require further elaboration here.

4.2.2 Plural versus singular in numeral phrases

In some languages, the number of a noun quantified by a numeral is governed by the numeral, which means that the grammatical number may contradict the "semantic" number. For instance, in Russian, within an NP in the nominative or the accusative, a noun quantified by DVA 'two', TRI 'three', and ČETYRE 'four' must be in the singular: *dva stola* lit. 'two of.table', *tri karandaša* 'three of.pencil', *četyre mal'čika* 'four of.boy' (plural: *stoly* 'tables', *karandaši* 'pencils', *mal'čiki* 'boys').

In German, certain masculine measure nouns remain in the singular when quantified by a numeral: *zehn Kilo Mehl* lit. '10 kilo flour', *zwanzig Pfund Äpfel* lit. '20 pound apples', *drei Sack Kartoffeln* lit. 'three bag potatoes' (plural: *Kilos* 'kilos', *Pfunde* 'pounds', *Säcke* 'bags').

In Turkish (as in all Turkic languages), a noun quantified by a numeral is always in the singular: *üç ev* lit. 'three house' or *beş kitap* lit. 'five book' (plural: *evler* 'houses', *kitaplar* 'books').

As already pointed out for the *pluralia tantum*, a noun in the DSyntS receives the semantic number grammeme, and thus after a numeral, except for '1', a noun is always 'PL'. In the SSyntS, this grammeme is replaced, when needed, by 'SG' in accordance with the requirements of the numeral in the specific $\mathcal{L}_T$.

4.2.3 Thematic/rhematic markers

In Japanese and Korean, a noun that expresses the semantic Theme of the sentence must be supplied with the grammeme 'THEME', while in a host of other languages no corresponding grammatical signification is available. Consider the following example:[27]

---

[27] This example illustrates another problem, in that Japanese does not necessarily indicate the exact semantic relation that links the semantic Theme to the rest of the sentence. However, in English, this relation must be expressed explicitly; that is why the English translation contains the preposition IN (Japan is the place

(33) a. Jap. *Nippon+wa*$_\text{THEME}$ *natsuga atsui*
      lit. 'Japan, summer hot.is'.
      ≡
      In Japan, summers are hot.

One is tempted to interpret this phenomenon as a correspondence between a gram-meme and nothing, i.e. as a morphological mismatch. However, this is not the case. In the Japanese DSyntS, the 'THEME' grammeme is not present, because the thematicity of a noun is specified not in the DSyntS but in the DSynt-Communicative Structure (DSynt-CommS) of the sentence; the corresponding grammeme appears only at the morphological level as a suffixal "materialization" of this specification. In this way, no morphological mismatch occurs.

The situation is identical in Korean and analogous, for example, in Somali, where the rhematicity of LUs is obligatorily marked by special particles.

5 A Model of morphological transfer

Before defining the morphological transfer model at the DSynt-level, let us consider the transfer of DSyntSs in general.

5.1 The transfer of DSyntSs

The transfer of DSyntSs consists of mapping a given DSyntS$_S$ onto a semantically equivalent DSyntS$_T$.[28] This mapping is implemented via four *translation* (= *corre-spondence*) *functions* τ:[29]

1. The lexical translation function τ_{l_s} (which formalizes lexical translation; see Sect. 1) maps an LU of DSyntS$_S$ onto a corresponding element of DSyntS$_T$; this element can be an LU, a syntactic subtree, or a grammatical signification, i.e. a derivateme or a (quasi-)grammeme; cf. Sect. 3.
2. The syntactic translation function τ_{a_s} (which formalizes syntactic translation) maps DSynt-relations of DSyntS$_S$ onto corresponding DSynt-relations of DSyntS$_T$.
3. The grammeme translation function τ_{g_s} (which formalizes the first part of mor-phological translation) maps semantic (quasi-)grammemes of DSyntS$_S$ onto cor-responding elements of DSyntS$_T$; such an element can be a (quasi-)grammeme, an LU, or an element of the rhetorical structure.
4. The derivateme translation function τ_{d_s} (which formalizes the second part of mor-phological translation) maps semantic derivatemes of DSyntS$_S$ onto correspond-ing elements of DSyntS$_T$; such an element can be a derivateme or an LU.

Footnote 27 continued

where the summers are hot), for which there is no equivalent in Japanese. It must be inferred from the context of the given structure. This is a particular case of a syntactic mismatch.

[28] In many cases, because of lexical and grammatical divergences, this equivalence is necessarily approx-imate.

[29] The order of the functions in the list by no means implies the order of their application. Their execution is controlled by the linguistic constraints built into the underlying rules.

τ_{l_s} essentially uses a lexical bilingual index, τ_{g_s} and τ_{d_s}, a grammatical signification index (see below); τ_{a_s} is based on a set of graph-rewriting rules.

In what follows, we focus on the grammatical signification translation functions and their linguistic realizations.

Definition 9 (*DSynt-Transfer*)

Let there be two DSyntSs defined over the alphabets of $\mathcal{L}_S$ and $\mathcal{L}_T$, respectively:
$S_S := \langle N_1, G_{sem_1}, D_{sem_1}, A_1, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g}, \delta_{n \to d} \rangle$
and
$S_T := \langle N_2, G_{sem_2}, D_{sem_2}, A_2, \lambda_{l_s \to n}, \rho_{r_s \to a}, \gamma_{n \to g}, \delta_{n \to d} \rangle$
Furthermore, let s_S be a subtree of S_S and s_T be a subtree of S_T.

We say that S_T is the result of the transfer of S_S iff for any s_S there exists an s_T and for any s_T there exists an s_S such that $\tau(s_S) = s_T$.

Definition 9 implies that (i) s_S is minimal because it corresponds to the argument of one of the four τ-functions, and (ii) $\bigcup_{s_{i_S}} = S_S$, i.e. S_S is fully processed by the translation functions.

According to Mel'čuk and Wanner (2006), the transfer at the DSyntS-level in a Meaning-Text model presupposes the existence of three types of resources:

1. Formalized monolingual *Explanatory Combinatorial Dictionaries* (ECDs) for the languages involved. These lexica are independent of the pair of languages considered and are direction-neutral, in the sense that each can be used either as a source- or as a target-language lexicon.
2. Bilingual lexical and grammatical signification correspondence indexes for the language pairs involved. A bilingual index (BI) is specific to each pair of languages; a lexical BI is a direction-neutral list of pairs of translationally equivalent LUs of $\mathcal{L}_S$ and $\mathcal{L}_T$; a grammatical signification BI is a direction-neutral list of pairs of translationally equivalent (quasi-)grammemes and derivatemes of $\mathcal{L}_S$ and $\mathcal{L}_T$. Lexical and grammatical signification BIs for a given pair of languages might show an overlap because some grammatical significations in $\mathcal{L}_S$ may correspond to LUs in $\mathcal{L}_T$ (and vice versa).[30]
3. Transfer rules that carry out the mapping between semantically equivalent syntactic representations of any $\mathcal{L}_S$ and $\mathcal{L}_T$. The core of this set of rules ensures the correspondence between DSyntS$_S$ and DSyntS$_T$; the other rules take care of the correspondences between DSynt-CommSs and DSynt-Rhetorical Structures of both languages.

5.2 The morphological transfer

The model of morphological transfer specifies how the grammatical signification translation function, or *morphological transfer*, can be implemented.

---

[30] This means that an equivalence between a GS and an LU can be considered equally as belonging to the lexical BI or the grammatical signification BI.

Morphological transfer is essentially based on the grammatical signification BI, which we consider in Sect. 5.2.1. ECDs are needed for morphological transfer to the extent that they supply information on the morphological properties of particular LUs. ECD-encoded morphological information is used as conditions on grammatical signification equivalences in the corresponding BI and in transfer rules (cf. Sect. 5.2.2).

5.2.1 Bilingual grammatical signification index

Similarly to lexical correspondences, GS-correspondences are captured in our approach by means of bilingual GS-indexes, specific to each pair $\mathcal{L}_S$ and $\mathcal{L}_T$. Given that the correspondences between GSs of $\mathcal{L}_S$ and $\mathcal{L}_T$ are strict equivalences, the GS-BI is not oriented, i.e. it can be used for translation between these two languages in both directions.

An entry of the GS-BI has one of the two following forms:

– for (quasi-)grammemes:

$g_S \Leftrightarrow \Psi$ | COND(itions)

where Ψ is one of the four following elements: (1) a g_T, (2) an LU_T with the corresponding syntactic substructure, (3) a rhetorical parameter, or (4) — (empty).

– for derivatemes:

$d_S \Leftrightarrow \Upsilon$ | COND

where Υ is either (1) a d_T or (2) an LU_T with the corresponding syntactic substructure.

It is important to emphasize that the order of the elements in an entry of a BI is irrelevant; either the first or the second element can serve as input. Thus, a BI is direction-neutral even if we use a "directed" source-target notation ($_S$ and $_T$ can be freely exchanged).

Let us illustrate each of these six entry types of the GS-BI.

(i) $g_S \Leftrightarrow g_T$ | COND

By default, in German–Russian, English–Arabic, etc. translation, for instance, the nominal 'SG_S' grammeme is mapped onto 'SG_T', the nominal 'PL_S' grammeme onto 'PL_T', verbal '$PRES_S$' grammeme onto '$PRES_T$', and so on:

'SG'[N_{Ger}] $\Leftrightarrow$ 'SG'[N_{Rus}] | no other rule applies

'PL'[N_{Ger}] $\Leftrightarrow$ 'PL'[N_{Rus}] | no other rule applies

'PRES'[V_{Ger}] $\Leftrightarrow$ 'PRES'[V_{Rus}] | no other rule applies

No specific conditions are associated with these default correspondences; they come to bear if none of the special-case rules apply. However, the default correspondences are often violated; under certain semantic or even pragmatic conditions, a grammeme may be mapped on a different grammeme of the same category. The verification of these conditions can be guaranteed only by using a complete Semantic Representation, including the referential structure. That is, within the restricted framework of our present model, human intervention

must be foreseen, cf. the following two samples for German–Russian transla-
tion:

'SG'[N_{Ger}] ⇔ 'PL'[N_{Rus}] │ N denotes an individual and is used in the
│ sentence S to refer to the class of such indi-
│ viduals, i.e. N has a generic reference

'PRES'[V_{Ger}] ⇔ 'FUT'[V_{Rus}] │ V refers to a future event and …

A more complex case is found in translating 'PL' from English (and, in general,
any SAE language) into Arabic (cf. (18), Sect. 3.2.1.1).

'PL'[N_{Eng}] ⇔ 'PL'[N_{Ar}] │ no other rule applies

'PL'[N_{Eng}] ⇔ 'DU'[N_{Ar}] │ N refers to a paired object (e.g., a pair of legs
│ or eyes)

'PL'[N_{Eng}] − **ATTR** → TWO ⇔ 'DU'[N_{Ar}] │ N does not refer to a paired
│ object

The last correspondence is, as a matter of fact, a combination of two elementary
cases; the Arabic 'DU' corresponds not simply to an English grammeme, but to
a combination of a grammeme ('PL') and a lexeme (TWO).

(ii) g_S ⇔ LU_T | COND

Hungarian–English translation of a semantic locative grammeme ('SUPERES-
SIVE' is a nominal case with the meaning 'being on the surface of N', cf. also
(18)):

'SUPERESSIVE'[N_{Hung}] ⇔ ON − **II** → [N_{Eng}] │ no other rules apply

Similar meaningful local cases are widespread in Finno-Ugric and Daghe-
stanian languages. Under translation to an SAE language they correspond to
meaningful prepositions.

(iii) g_S ⇔ < rhet.parameter >| COND[31]

Korean–English translation, a politeness grammeme (cf. (31) and Appendix
A1.2 for some details on politeness in Korean):

'HON'[N_{Kor}] ⇔ **formal**$_{Eng}$ | —

(iv) g_S ⇔ − | COND

This case covers the translation of grammemes of inflectional categories absent,
e.g., from such languages as Vietnamese, as in (34):

(34)   *I want to wash my face*$_{SG}$/*hands*$_{PL}$.
        ≡

---

[31] <rhet. parameter> stands for stylistic labels **formal**, **informal**, **colloquial**, etc.

Tôi muốn rúa mặ/tay
lit. 'I want wash face/hand'.

The 'SG' of 'face' and the 'PL' of 'hands' are not distinguished in Vietnamese. (If the speaker absolutely needs to distinguish whether one or both hands are to be washed, he has to add an explicit numeral: 'one' or 'two'.)

'SG'$[N_{Eng}] \Leftrightarrow -[N_{Viet}]$ | —
'PL'$[N_{Eng}] \Leftrightarrow -[N_{Viet}]$ | —

(v) $d_S \Leftrightarrow d_T$ | COND

Again, both straightforward (= non-conditioned) and more complex translation derivateme equivalences must be considered, cf. the translation of diminutives from Spanish into German:

'DIM'$[N_{Sp}] \Leftrightarrow$ 'DIM'$[N_{Ger}]$ | —

and from Spanish into Italian:

'DIM'$[N_{Sp}] \Leftrightarrow$ 'DIM**1**'$[N_{It}]$ | 'DIM'$[N_{Sp}]$ is positive
'DIM'$[N_{Sp}] \Leftrightarrow$ 'DIM**2**'$[N_{It}]$ | 'DIM'$[N_{Sp}]$ is negative

A positive or negative interpretation of Spanish diminutives is to be inferred from the context.

(vi) $d_S \Leftrightarrow LU_T$ | COND

Translation of the diminutive from German into English:
'DIM'$[N_{Ger}] \Leftrightarrow [N_{Eng}] -$ **ATTR** $\rightarrow$ NICE $-$ **QUASI-COORD** $\rightarrow$ LITTLE
Cf. *Hotelchen* ≡ *nice little hotel*, *Büchlein* ≡ *nice little book*.

As will be seen in Sect. 6, the bilingual grammatical signification index entries can also be specified in terms of rules.

5.2.2 Morphological transfer engine

The Morphological Transfer Engine (MTE) includes two major components:

1. The *transfer component proper*, which has recourse to the grammatical significa- tion BI in order to select the appropriate target grammatical signification (while observing the available conditions), lexeme, or rhetorical parameter. The selected grammatical signification or lexeme is inserted into the DSyntS$_T$; the rhetorical parameter triggers necessary transformations during synthesis (e.g. adding such lexical politeness markers as PLEASE or COULD YOU …?, etc. to the DSyntS$_T$).
2. The *missing grammeme provider*, which computes the necessary target semantic grammemes not obtained during the morphological transfer proper. The comput- ing is done in two steps: (i) an inflectional well-formedness checker identifies the missing inflectional categories in the DSyntS$_T$; (ii) an inflectional informa- tion retriever searches through the source structures or situational characteristics (if available) for clues on the missing grammemes.

The first component of the MTE is pressed into service whenever $\mathcal{L}_S$ shows semantic grammatical significations, which have to be transferred (e.g. English nominal numbers into French or a German agent suffix into Russian); the second one is used whenever $\mathcal{L}_T$ possesses a semantic inflectional category that is absent from $\mathcal{L}_S$ (as is the case of nominal numbers in English under Vietnamese–English translation, definiteness under Russian–German translation, etc.).

Let us emphasize that if an inflectional category is absent from either $\mathcal{L}_S$ or $\mathcal{L}_T$, the morphological transfer inevitably leads to a DSyntS$_T$ that is not fully equivalent to the DSyntS$_S$. Thus, when translating from Chinese into English, the grammemes of nominal number must be computed and introduced into the English DSyntS. Since they are absent from the Chinese DSyntS, the DSyntS$_S$ and DSyntS$_T$ are, strictly speaking, semantically not fully equivalent.

6 Two case studies: definiteness and aspect

As already mentioned, in what follows we examine in more detail one major type of morphological mismatch, namely, $g_S \Leftrightarrow -$, which occurs if $\mathcal{L}_S$ or $\mathcal{L}_T$ lacks the corresponding inflectional category. We consider two specific examples of this type of mismatch: the German category of definiteness, which is absent from Russian, and the Russian category of aspect, which is absent from German. Note that our choice of the language pair, i.e. German–Russian, and of the mismatch examples is of a purely illustrative nature. No parts of our proposal, except the transfer rules themselves, are language-specific. Any other language pair between which morphological mismatches occur could thus have been taken. We chose German and Russian because as native speakers we master both languages well enough to judge the correctness of the outcome of our experiment.

We deal with both definiteness and aspect in the "difficult" direction: computing the definiteness grammemes in the process of Russian–German translation, and aspect grammemes in the process of German–Russian translation.[32]

Both categories are notoriously difficult for description from the semantic and syntactic viewpoints. In spite of an extremely rich literature on both topics, a clear-cut and exhaustive picture of the use of the corresponding grammemes is not available. Neither is it our goal to shed new light on the problem nor even offer an overview of the state-of-the-art. We merely attempt to use some existing results in order to present examples of rules needed for morphological transfer that deal with determiners and aspects. These rules are no more than an illustration; on several occasions they must use conditions that cannot be formally verified, given the limited capabilities of today's implementations to access situational and encyclopedic information.

Our mini-study is based on the systematic comparison of short parallel texts in Russian and German from the news domain, with an eye to which conditions control

---

[32] Definiteness and aspect are, as a matter of fact, semantically linked. Thus, a definite single direct object tends to combine with a perfective verb; see, for instance, Mehlig (1988), Leiss (2000) for an overview and Kabakčiev (1984) for a special study based on Bulgarian. A previous study, practically parallel to ours, was conducted by Gawrońska-Werngren (1990), Gawrońska (1992), who considered the introduction of articles under Russian–Swedish translation and of aspects under Swedish–Russian translation.

the use of definiteness grammemes in German and aspect grammemes in Russian. The result of this study is a set of morphological transfer rules, which have been checked against a small parallel corpus; see Appendix 2 for the application of these rules to the texts.

6.1 Definiteness

In order to compute definiteness grammemes in German, two unequal classes of cases have to be considered: (i) the definiteness of a noun N is specified in the German dictionary without recourse to N's context; (ii) the definiteness of N is established based on N's context.[33]

We assume that the category of definiteness in German has three grammemes: 'DEF' ∼ 'INDEF' ∼ 'NON-DEF' (i.e. no article).

6.1.1 Dictionary-specified definiteness

Two subcases can be distinguished: the definiteness of a noun inside a phraseme (idiom or collocation) and the definiteness of a noun with a particular type of meaning.

6.1.1.1 *Definiteness within a phraseme* The divergence of definiteness even in structurally close equivalent idioms of two closely related languages (such as English and German) is a well-known fact; cf. some examples in which the articles are boldfaced:

(35)

English	German
facts of life	*Gegebenheiten des Lebens*
***an** eye for **an** eye*	*Auge um Auge*
cast pearls before swine	*Perlen vor die Säue werfen*
live from hand to mouth	*von der Hand in den Mund leben*

However, such divergences do not lead to morphological mismatches. Since an idiom is represented in the DSyntS as a single node, its internal nouns do not appear as separate nodes. Therefore, the question of their definiteness does not even arise at this level. Being stored in the dictionary with all its g_{DET} explicitly indicated in the lexical entry, an idiom is expanded into a full subtree at the SSynt-level, with all definiteness markers (i.e. articles) in place.

The use of definiteness grammemes within collocations is also often phraseologized; cf. a few English–German examples: *be in **an** awkward position ≡ in **der** Klemme stecken*, *make allusion to N ≡ **eine** Anspielung auf N machen*, *in connection with N ≡ **im** Zusammenhang mit N*, *make significant headway ≡ **einen** bedeutenden Fortschritt erzielen*, etc. (see also Anscombre 1986). However, contrary to idioms, in

---

[33] For contrastive studies of definiteness in article and article-less languages, cf. Birkenmaier (1979), Gladrow (1979), among others.

Table 2 Definiteness of the noun in German verb–noun collocations

• Noun is always 'DEF'	*zum Ausdruck bringen* lit. 'bring into.the expression', *zur Abstimmung bringen* (as opposed to *bring to **a** vote*), etc.
• Noun is always 'NON-DEF'	*Ärger auslösen* 'provoke anger', *in (regem) Gebrauch sein* 'be in (intense) use', *ohne (jegliche) Zweifel* 'without (any) doubt', *mit (großer) Freude* 'with (great) joy', *unter Feuer* 'under fire', *ohne Hilfe* 'without help', *voller Angst* lit. 'full.of fright', etc.
• Noun is 'NON-DEF' with no modifier and 'INDEF' with a modifier	*Abschied nehmen* 'take leave', *Anspruch erheben* 'lay claim', *Anwendung finden* 'find application' vs. *einen raschen Abschied nehmen* 'take a quick leave', *einen absoluten Anspruch erheben* 'lay an absolute claim', *eine neue Anwendung finden* 'find a new application', etc.
• Noun is 'DEF' with no modifier and 'DEF' or 'INDEF' with a modifier—depending on the modifier	*im Zusammenhang* lit. 'in.the connection', *in dem besagten Zusammenhang* lit. 'in the said connection', *in einem engen Zusammenhang* lit. 'in a close connection', etc.

collocations, definiteness can be controlled by the context. Thus, in verb–noun collocations, the presence/absence of an adjectival modifier with the noun may condition its definiteness. Among many theoretical possibilities, the cases listed in Table 2 are most common.

The corresponding information must be specified in the monolingual dictionary in the subentry of the collocation. Let us consider, for illustration, the collocations *Ärger* [*von* N$_{DAT}$] *auslösen* '[to] provoke N's anger' and *im Zusammenhang* [*mit* N$_{DAT}$] lit. 'in.the connection [with N]'.

In *Ärger* [von N$_{DAT}$] *auslösen*, the noun *Ärger* does not take an article even if modified by an adjective: *heftigen Ärger auslösen* lit. 'provoke intense anger' rather than [?]***einen** heftigen Ärger auslösen* lit. 'provoke an intense anger'. In *im Zusammenhang* [*mit* N$_{DAT}$], the grammeme of definiteness is also lexicographically specified, although conditionally:

  in ZUSAMMENHANG, 'DEF' | Z. has no adjectival modifier
  in ZUSAMMENHANG, 'INDEF' | Z. has an adjectival modifier

Compare *im Zusammenhang mit dem Geschehen* lit. 'in.the connection with the events' vs. *in einem unmittelbaren Zusammenhang mit dem Geschehen* lit. 'in an immediate connection with the events'.

Although rather complex, the specification of definiteness of nouns in collocations is relegated to the SSynt-level; it takes place along with the selection of a specific collocate. Therefore, definiteness in collocations does not lead to morphological mismatches, either. In fact, definiteness divergences related to phrasemes constitute typical pseudo-mismatches mentioned in Subsect. 4.2.

6.1.1.2 *Nouns with a semantic component of definiteness* The meaning of a noun can contain a semantic component of definiteness, which constraints the combination of the noun with grammemes of definiteness. Without seeking exhaustiveness, we will indicate three types of such nouns.

Nouns with 'definiteness' component in its meaning: For instance, the German lexeme *Zeit* '[the] time scale in geographical location Y' in the compound *Ortszeit* 'local time' is 'NON-DEF' (because it includes the meaning of 'definiteness': in a given location, the local time scale is unique and thus definite).

Nouns with unique referents: Nouns with unique referents—but not proper names—include such nouns as Ger. *die Welt* 'the world', *die Menschheit* 'the humanity', *der Mond* 'the moon', *das Universum* 'the universe', *der Himmel* 'the sky', etc., as well as common nouns that designate individuals occupying a unique position: Ger. *der Ministerpräsident von Frankreich* 'the prime minister of France', *die Amtsführung* 'the administration of an office', *die Kanzlerin* 'the chancellor', etc. This property must be explicitly reflected in their definition.

Obviously, this statement is an oversimplification: numerous lexical expressions have a unique referent only in a given context, which must be established by specific rules, cf. Rule 1 below.

Proper nouns: Several types of proper nouns have a special relationship with definiteness grammemes. For instance, names of beings (person or animal) and settlements (city, town, village, …) are always non-definite, which is indicated in their lexical entries. Names of rivers and mountains and names of geopolitical formations behave differently as we illustrate below, 6.1.2.2.

6.1.2 Context-specified definiteness: transfer rules

In Russian–German translation, context-specified definiteness grammemes must be computed by the missing-grammeme provider; this is necessary because the target language has an inflectional category that does not exist in the source language.

In German, the choice of a definiteness grammeme is, roughly speaking, controlled by the "local" syntactic-lexical context and/or by the global semantic and communicative structures of the sentence.

6.1.2.1 *Semantic-communicative rules for computing definiteness*   As is well known, definiteness is intimately related to the semantic and communicative properties of the sentence under translation. As far as semantics is concerned, the grammeme 'DEF' on a noun marks the uniqueness of its referent in the given speech situation. In other words, the Addressee can uniquely identify the referent of a noun marked as 'DEF'; 'INDEF', on the contrary, marks the non-uniqueness of the referent or even non-referentiality of the corresponding noun.

Rule 1 IF N has a unique referent in the given sentence
THEN N_{DEF}
ELSE N_{INDEF}

Cf. *Wir suchten den Vater Marias/das Buch, von dem Du gesprochen hast* vs. *Wir suchten Feuerholz*$_{INDEF}$/*ein Buch*$_{INDEF}$ $\sim$ *Bücher*$_{INDEF}$ 'We looked for fire wood/for a book/books'.

We assume that in German the indefiniteness of a mass noun or plural count noun is expressed by the zero article.

As far as the communicative structure is concerned, a noun which is Given, i.e. whose referent is present in the active consciousness of the Addressee, is most often marked 'DEF', and 'INDEF' otherwise.

Rule 2 IF N is Given
　　　　THEN N_{DEF}
　　　　ELSE N_{INDEF}

Since we are dealing with translation at the DSynt-level, the corresponding information has to be found in the DSynt-CommS of the target sentence. However, the construction of the DSynt-CommS is a very complex linguistic task, which is beyond our goals in this paper. Therefore, let us simply indicate two very particular phenomena, where the Russian sentence supplies some information concerning definiteness.

Rule 3 IF a Russian N_1 has a dependent $-II \rightarrow N_2$
　　　　AND
　　　　N_1 denotes a relation, function, property, etc. of N_2
　　　　THEN German $\tau_l(N_1)_{DEF}$

If the referent of N_1 is not unique in the given situation, Russian often uses the construction *odin iz N_1-ov N_2-a* 'one of the N_1s of N_2'. For instance, *prišel brat Ivana* lit. 'came brother of.Ivan', the correct German translation is *Ivans Bruder kam* (where *Bruder* is definite). Otherwise, the Russian expression would be *prišel odin iz brat′ev Ivana* lit. 'came one of Ivan's brothers', and a possible German translation is ***einer von Ivans Brüdern***.

Rule 4 IF a Russian N_{human} has a dependent $-ATTR \rightarrow NUM_{collective}$
　　　　THEN German $\tau_l(N)_{DEF}$

Cf. *troje druzej* ≡ ***die drei Freunde***. TROJE is a collective numeral.[34]

6.1.2.2 *Syntactic-lexical rules for computing definiteness* To illustrate the computation of definiteness grammemes, twelve rules are cited (RULE 5–RULE 16).

Name of a body of water or a mountain:

Rule 5 IF $N_{(prop)}$ is the name of a body of water (river, lake, sea, …)
　　　　or of a mountain
　　　　THEN IF $N_{(prop)}$ is in apposition to an N′
　　　　THEN $N_{(prop)NON-DEF}$
　　　　ELSE $N_{(prop)DEF}$

Cf. Ger. ***die Wolga*** versus ***der Fluß Wolga***, ***der Baikal*** versus *der See Baikal*; ***das Matterhorn*** versus *der Berg Matterhorn*, ***der Everest*** versus ***der Berg Everest***, etc.

Name of a geopolitical formation:

---

[34] For numbers between 2 and 10, Russian has special numerals called "collective," which express definiteness when applied to humans.

Rule 6 IF N$_{(prop)}$ is the name of a geopolitical formation (country, state,
  province, …)
  THEN IF N$_{(prop)}$ is in apposition to an N′
  THEN N$_{(prop)\text{NON-DEF}}$
  ELSE IF N$_{(prop)}$ is modified by an adjective
    THEN determine its definiteness as in the general case
      for common nouns
    ELSE N$_{(prop)\text{DEF}}$ or N$_{(prop)\text{NON-DEF}}$: consult the dictionary

Cf. *die Region Rheinland* 'the region of Rheinland', *das Departement Auvergne* 'the department of Auvergne'; *in Frankreich* 'in France', *in Russland* 'in Russia' versus *in der Schweiz* lit. 'in.the Switzerland', *im Irak* lit. 'in.the Iraq'; *in Bayern* 'in Bavaria', *in Ontario* versus *im Rheinland* 'in.the Rhineland', *in der Auvergne* 'in the Auvergne', but *in einem modernen Ontario* 'in a modern Ontario', *im poststalinistischen Russland* lit. 'in.the post-Stalinist Russia', etc.

Presence of a superlative or an ordinal numeral:

Rule 7 IF N has a modifying adjective in the superlative or an ordinal numeral
  AND
  N does not have a modifying, possessive or negative determiner
  THEN N$_{\text{DEF}}$

Cf. *das interessanteste Buch* 'the most.interesting book' (**interessantestes Buch*/**ein interessantestes Buch*)

Presence of a non-article determiner:

Rule 8 IF N has a modifying demonstrative, possessive or negative determiner
  THEN N$_{\text{NON-DEF}}$

Cf. Ger. *dieses/mein Buch* 'this/my book' vs. ***das** diese/meine Buch, ***ein** dieses/mein Buch. In contrast, cf. Rus. *odna moja kniga* lit. 'a my book'.

  In German and English, the anteposed genitive form is incompatible with determiners, forcing them out; cf. **das/ein Margas Buch* *'the/a Marga's book' versus *Margas Buch* 'Marga's book', *das/ein Buch Margas* lit. 'the/a book Marga's'. This phenomenon is treated at the SSynt-level where the anteposed genitive is assigned a special surface-syntactic relation.

Presence of a numeral:

Rule 9 IF N has a modifying numeral
  AND
  N does not have a modifying adjective in the superlative
  THEN N$_{\text{NON-DEF}}$

Cf. Ger. *Johanns zehn Bücher*$_{\text{NON-DEF}}$ 'John's ten books', *meine/diese zehn Bücher*$_{\text{NON-DEF}}$ 'my/these 10 books', *zehn Bücher*$_{\text{NON-DEF}}$ '10 books' versus *die (zehn) billigsten Bücher*$_{\text{DEF}}$ 'the (ten) cheapest books'.

Presence of a proper noun in apposition:

Rule 10 IF N has as apposition a proper name of the entity denoted by N
                THEN N$_{\text{DEF}}$

Cf. Rus. *gorodok Xibxib* 'little.town Hibhib' ≡ Ger. ***die Kleinstadt Hibhib*** 'the little.town Hibhib'.

Title:

Rule 11 IF N is a title depending on a human proper name
                THEN N$_{\text{NON-DEF}}$

Cf. Rus. *Fel′dmaršal Rommel′* ≡ *Feldmarschall Rommel*[35]

Noun in an elective construction:

Rule 12 IF N is a DSyntA **II** of an elective lexeme L
                THEN N$_{\text{DEF}}$

L is an elective lexeme if its DSyntA **II** denotes the set out of which the elements characterized by L are taken: Ger. *einer **der** Brüder* 'one of the brothers', *wenige von **den** Überlebenden* 'few of the survivors', *die vernünftigsten **der** Ärzte* lit. 'the most.reasonable of.the doctors', etc.

Government pattern restrictions: Definiteness of a noun N can also be affected by the government pattern restrictions in the lexical entry of the LU that syntactically governs N, cf. the following four rules.

1.   For N as DSyntA **II** of the conjunction ALS 'as':

Rule 13 IF N is a DSyntA **II** of the conjunction ALS 'as'
                AND
                N has no characterizing dependent
                THEN N$_{\text{NON-DEF}}$
                ELSE N$_{\text{DEF}}$

A characterizing dependent of N is opposed to a classifying dependent of N: the former specifies a particular characteristic of N (e.g. in ***expensive*** computer), while the latter indicates a subclass of Ns (e.g. in ***desktop*** computer).

(36)   a. *Als Minister war Hochberg sehr umstritten*
           lit. 'As minister, Hochberg was very controversial'.
           versus
           *Als **der** von der CDU ernannte Minister war Hochberg sehr umstritten*
           lit. 'As the minister nominated by the CDU, Hochberg was very controversial'.

---

[35] At the surface, this type of apposition is anteposed.

b. *Die Polizei hat ihn als Mörder gesucht*
lit. 'The police wanted him as murderer'
'The police wanted him for murder'.
versus
*Die Polizei hat ihn als **den** Mörder des Mädchens gesucht*
'The police wanted him as the murderer of the girl'.

In both examples, the dependent of the N under consideration denotes a characteristic of N's referent. If, however, the dependent classifies N, N is 'NON-DEF':

(37) a. *Als Minister für Internationale Angelegenheiten war Hochberg sehr umstritten*
'As Minister for Foreign Affairs, Hochberg was very controversial'.
b. *Die Polizei hat ihn als besonders gefährlichen Mörder gesucht*
'The police wanted him as an especially dangerous murderer'.

2. For N as DSyntA **II** of a verb of the type VORWERFEN '[to] charge / accuse' (i.e. a verb that has a corresponding government pattern):

Rule 14 IF N is DSyntA **II** of a verb of the type VORWERFEN
AND
N has no characterizing dependent
THEN N$_{\text{NON-DEF}}$
ELSE N$_{\text{DEF}}$

(38) a. *Ihm wurde Brandstiftung vorgeworfen*
'He was charged with arson'.
b. *Ihm wurde vorsätzliche Brandstiftung vorgeworfen*[36]
'He was charged with premeditated arson'.
c. *Ihm wurde **die** Brandstiftung im Theater vorgeworfen*
'He was charged with the arson in the theater'.

3. For N as DSyntA **II** of a verb of the type of BESCHULDIGEN '[to] accuse':

Rule 15 IF N is DSyntA **II** of a verb of the type of BESCHULDIGEN
AND
N has no dependent
THEN N$_{\text{DEF}}$
ELSE N$_{\text{DEF}}$ or N$_{\text{INDEF}}$

(in the 'ELSE'-branch, the choice between 'DEF' and 'INDEF' is made according to general rules for the use of articles).

(39) a. *Man hat ihm Verrat vorgeworfen*
'He was charged with treason'.
versus

---

[36] The dependent *vorsätzlich* is classifying rather than characterizing.

  b. *Man hat ihn **des** Verrats beschuldigt*
    lit. 'He was accused of.the treason'.

   The above two cases illustrate the impact of the government pattern on the use of articles: synonymous verbs VORWERFEN and BESCHULDIGEN (both roughly 'accuse') impose different definiteness grammemes on their DSyntA **II**.

4.   For N as DSyntA **II** of the copula SEIN 'be' or WERDEN 'become':

Rule 16 IF N is DSyntA **II** of the copula SEIN or WERDEN
      THEN IF N denotes a social class of human beings (ethnicity,
        religion, profession, …)
            THEN IF N has no modifier
                THEN N$_{\mathrm{NON\text{-}DEF}}$
                ELSE N$_{\mathrm{INDEF}}$
          ELSE IF N denotes a class different from a social class of human
            beings
      THEN N$_{\mathrm{INDEF}}$

(40)   *Er ist Franzose/Christ/Maler*
     lit. 'He is Frenchman/Christian/painter',
     but
     *Er ist **e**in typischer Franzose/Christ/Maler*
     lit. 'He is a typical Frenchman/Christian/painter'.

(41)   a. *Der Elephant ist **ein** Säugetier*
        'The elephant is a mammal'.
     b. *Eine Pistole ist **eine** Feuerwaffe*
        'A pistol is a firearm'.

6.2 Aspect

6.2.1 Some remarks on russian aspect

Our sketch of the aspect-establishing rules for Russian is mainly based on the results of Wierzbicka (1967); Glovinskaja (1982, 2001); Kabakčiev (1984); Apresjan (1988); Padučeva (1996)). Of special relevance are also works concerning the Russian aspect from the perspective of German (e.g. Scheljakin and Schlegel 1970; Mehlig 1978, 1989).
   In order to make our proposal realistic, we do not consider two important features of the aspect category.

–   The difference between Russian aspects is essentially semantic. Thus, for instance, the difference between a teleological prolonged action (expressed by the imperfective aspect of V) and its achieved result (expressed by the perfective aspect of V) should be encoded in terms of corresponding semantemes in the semantic structures. However, since we assume the transfer to happen at the DSynt-level and a

source DSyntS of a language without aspect (such as German) does not encode this information, we have to use syntactic "hints" for a basically semantic distinction. In other words, we attempt to reduce a semantic problem to a syntactic one.[37]

– Different inflectional forms of a verb do not behave in the same way as far as aspect is concerned. Thus, deverbal adverbs (*deepričastija*), passive participles, and the imperatives have special properties in this respect. There is also a link between the aspect of the passive verb and its formal pattern, etc. However, it is far beyond this paper to account for such subtle differences.

The proposed rules for computing Russian aspect grammemes are based on a classification of verbs in three major semantic families:[38]

– *punctual* verbs, whose meanings prototypically refer to momentary events: 'explode', 'hit', 'die', 'recognize', … (cf. Padučeva 1998)
– *processual* (or *atelic*) verbs, whose meanings prototypically refer to lasting events without an explicit result: states such as 'sleep', 'love', 'possess', …; processes such as 'burn', 'rotten', 'grow', …; and activities such as 'run', 'drink' [= 'be a drinker'], 'work', …
– *resultative* (or *telic*) verbs, whose meanings prototypically refer to lasting events necessarily leading to a result: 'build' [a bridge], 'dress', 'drink' [a glass of water], 'run' [two miles], …

From the viewpoint of aspects, these verb families have the following properties.

A punctual verb tends to be used in the perfective aspect if it does not refer to a (non-quantified) repetition or temporal extension of an essentially momentary event: *On srazu zaplatil*$_{\text{PERF}}$ lit. 'He paid immediately' versus *On vsegda platil*$_{\text{IMPERF}}$ *srazu* lit. 'He always paid immediately'.

A processual verb tends to be used in the imperfective aspect if it does not refer to a limited or completed 'portion' of a lasting event: *On zavtrakal*$_{\text{IMPERF}}$ *tol'ko fruktami* lit. 'He ate.for.breakfast only fruit-usually or always' versus *On pozavtrakal*$_{\text{PERF}}$ *tol'ko fruktami* lit. 'He ate.for.breakfast fruit-on a particular occasion'.

Therefore, if the above conditions are fulfilled, a punctual verb is given the perfective aspect and a processual verb the imperfective aspect; otherwise, the aspects are inverted. The verification of the conditions is based on context elements: perfective and imperfective triggers (see below).

As for a resultative verb, it can refer either to a lasting event (in this case, it is imperfective) or to the fact of achieving the result of this event (in this case, it is perfective). Following Wierzbicka (1967) and Kabakčiev (1984), we suppose that the number and definiteness of the direct object of a German verb translated by a Russian resultative verb correlate with the result being achieved/not achieved. A singular definite direct object is a strong indicator of the resultativity and, therefore, of the perfective aspect on

---

[37] The only two papers on the topic of computing the Russian aspect grammemes known to us (Nikolaeva 1959; Kobozeva 1980) concern general semantic characteristics of aspects and contain no data relevant to the DSynt-level.

[38] This classification is close to Vendler's four-pronged semantic classification of verbs. Thus, the family of resultative verbs corresponds roughly to Vendler's Accomplishments and Achievements. However, there is no direct correspondence between the two classifications.

the Russian translation equivalent; a plural indefinite/non-definite object is as strong an indicator of the verb referring to the lasting action.[39]

The preceding considerations are approximate and must be controlled by contextual clues. These clues are of two major types: (i) information found in the government pattern of the lexeme that governs the verb in question, (ii) information supplied by elements of the sentence carrying particular semantic components.

Similarly to definiteness, aspect grammemes on the verb V can be computed in two basic ways: (i) on the basis of a dictionary entry, or (ii) on the basis of the context.

6.2.2 Dictionary-specified aspect

When a Russian verb is part of an idiom, its aspect may be fixed; cf. *Jabločko ot jabloni ne daleko padaet*$_{\text{IMPERF}}$ lit. 'a little.apple from the apple.tree does not fall far.away', *Ego kak vetrom sdulo*$_{\text{PERF}}$ lit. 'Him as with.wind [it] blew.away' (= 'He disappeared as if blown away by the wind'), etc. In this case, the verb does not appear as a separate node in the DSyntS, and, thus, we do not deal with it at this level.

Many Russian verbs exist only in one aspect. Sometimes this is due to the verb's semantics; thus, *ljubit'*$_{\text{IMPERF}}$ 'love' or *znat'*$_{\text{IMPERF}}$ 'know' do not have the perfective because they denote states. Sometimes this is an idiosyncratic property: *očutit' sja*$_{\text{PERF}}$ 'find oneself somewhere' (no imperfective) versus *okazat' sja*$_{\text{PERF}}$/*okazyvat' sja*$_{\text{IMPERF}}$ 'find oneself somewhere'; *obraščat' sja*$_{\text{IMPERF}}$ 'treat someone, handle something' (no perfective) versus *obxodit' sja*$_{\text{IMPERF}}$/*obojtis'*$_{\text{PERF}}$ 'treat someone, handle something'. In the case of a *perfectivum/imperfectivum tantum* (which is, of course, specified in the dictionary), the problem of aspect determination obviously does not arise. However, the fact that a verb lacks an aspect may play a role in the lexical choice.

6.2.3 Context-specified aspect: Transfer rules

6.2.3.1 *Contextual syntactic triggers* Contextual determination of the aspect grammeme is possible because in combination with adverbials characterizing the manner in which the event, action, etc. is taking place, the aspect of the verb sometimes duplicates the relevant meaning elements:

– For the imperfective aspect, contextual clues are expressions that mean 'repetition' ('every time', 'several times', 'again and again', 'used to …', 'constantly', etc.) or 'temporal extension' ('when' [something is happening], …, 'gradually', 'for a long time', 'during N', etc.).
– For the perfective aspect, the indicators are expressions that mean 'moment' ('immediately', 'at once', 'right away', etc.) or 'limited portion' ('in NUM+ N$_{\text{time-measure}}$' (e.g. 'in three days'), etc.

The former are referred to as *imperfective triggers*, the latter as *perfective triggers*. A more complete list of such triggers is obviously needed.

---

[39] These heuristics are overridden by explicit contextual clues (imperfective and perfective triggers; see below), as in *On dolgo/mnogo raz/vsë ešče perestraival*$_{\text{IMPERF}}$ *svoj dom* lit. 'He for a long time/many times/still rebuilt his house'.

These contextual clues are sought in the DSyntS of the Russian target sentence under production. The specific cases where they have to be sought in the original German sentence are explicitly indicated in the rules.

To these two general cases, a list of particular cases should be added. For instance:

- A Russian imperative verb with negation must be in the imperfective.[40]
- The verb in a Russian interrogative sentence must be in the imperfective if the question bears on a life experience (which can be indicated by the presence of an adverbial such as *kogda-nibud'* $\approx$ 'once'):

  (42) Ger. *Hast du schon mal so etwas gesehen?*
        lit. 'Have you already seen once something like that?'
        $\equiv$
        Rus. *Ty kogda-nibud' videl*$_{\text{IMPERF}}$ *čto-libo takoe?*

Given that Russian aspects are not distinguished in the present tense, a general condition in all rules is: 'Verb V is not in the present indicative'.

Punctual verbs:

Rule 17 IF V is a punctual verb
   THEN IF the Russian DSyntS does not contain imperfective triggers
     THEN V_{PERF}
     ELSE V_{IMPERF}

(43) *Zwei Explosionen erschütterten die Stadt*
      'Two explosions rocked the town'.
      $\equiv$
      Dva vzryva sotrjasli$_{\text{PERF}}$ *gorod.*

The verb SOTRJASAT' 'rock' is punctual. The sentence in (43) has no imperfective triggers, so that SOTRJASAT' must appear in the perfective. In contrast, (44) contains a clear marker of repetition (*v tečenii celogo dnja* 'over the whole day'); therefore, SOTRJASAT' appears in the imperfective:

(44) *Explosionen erschütterten die Stadt im Laufe des ganzen Tages*
      'Explosions rocked the town over the whole day'.
      $\equiv$
      Vzryvy sotrjasali$_{\text{IMPERF}}$ *gorod v tečenie celogo dnja.*

Processual verbs:

Rule 18 IF V is a processual verb
   THEN IF the Russian DSyntS does not contain perfective triggers
     THEN V_{IMPERF}
     ELSE V_{PERF}

---

[40] Formally, the perfective is also possible in the imperative, but then the speech act is a warning rather than an order or request: *Ne upadite*$_{\text{PERF}}$ 'Be careful not to fall' versus *Ne strojte*$_{\text{IMPERF}}$/**Ne postrojte*$_{\text{PERF}}$! 'Do not build!'

(45) *Er frühstückte immer sehr früh*
      lit. 'He ate.breakfast always very early'.
      ≡
      On vsegda zavtrakal$_{\text{IMPERF}}$ *očen' rano.*

The verb ZAVTRAKAT' is processual; since in (45) there is no perfective trigger, it is in the imperfective. In (46), there is a perfective trigger (*za desjat' minut* 'in 10 minutes') and the verb is in the perfective:

(46) *Er frühstückte in zehn Minuten*
      lit. 'He ate.breakfast in 10 minutes'.
      ≡
      On pozavtrakal$_{\text{PERF}}$ *za desjat' minut.*

Resultative verbs:

Rule 19 IF V is a resultative verb
          THEN IF its German source has a direct object with the
          grammeme 'DEF' or with a non-article determiner
               THEN IF in the Russian DSyntS there are no
                        imperfective triggers
                    THEN V$_{\text{PERF}}$
       ELSE V$_{\text{IMPERF}}$
    ELSE V$_{\text{PERF}}$

(47) a. *Der Mechaniker reparierte das/mein Auto*
         'The mechanic repaired the/my car'.
         ≡
         Mexanik počinil$_{\text{PERF}}$ *(moju) mašinu.*
      b. *Während ich Einkäufe machte, reparierte der Mechaniker mein Auto*
         'While I was shopping, the mechanic was repairing my car'.
         ≡
         V to vremja, kak ja delal pokupki, mexanik činil$_{\text{IMPERF}}$ *moju mašinu.*
      c. *Der Mechaniker reparierte bereits zwei Mal mein Auto*
         'The mechanic already repaired my car twice'.
         ≡
         Mexanik uže dvaždy činil$_{\text{IMPERF}}$ *moju mašinu.*
      d. *Der Mechaniker reparierte Autos*
         'The mechanic repaired cars'.
         ≡
         Mexanik činil$_{\text{IMPERF}}$ *mašiny.*

The form of the source (German) verb is also a factor for determining the aspect of its Russian equivalent (see Sacker 1983). Thus, the German perfect is a strong indication at least for certain verbs that the Russian equivalent must be perfective. However, since this correlation is complex and requires further details, we cannot use it as a rule.

6.2.3.2 *Government pattern determination of the aspect*

– A Russian phasal verb takes the governed infinitive in the imperfective:

   (48)   *načat′ <perestat′/prodolžat′/...> stroit′*$_{\text{IMPERF}}$ ~ **postroit′*$_{\text{PERF}}$
          lit. 'begin <cease/continue/...> to build'

– The conjunction *kak by ne* lit. 'that [something not might happen]' requires that the governed verb be in the perfective:

   (49)   *Kak by on ne ušel*$_{\text{PERF}}$ ~ **uxodil*$_{\text{IMPERF}}$*!*
          lit. 'That he does not leave!' = 'If only he does not leave!'

Rule 20 IF V has aspect X specified in the government pattern of V's
            syntactic governor
            THEN V$_X$

7 Summary

Morphological translation has often been misjudged in MT because of the "easy" transfer of purely syntactic grammemes (such as person and number, controlled by agreement with the subject), which are readily covered in any MT system. However, the transfer of meaningful grammatical significations is incomparably more difficult. In its general form, it requires access to complex and abstract semantics, which still defies a rigorous description. This is why the interlingual correspondences between semantic grammemes are hard to formulate.

In this paper, we propose a typology of morphological divergences between languages and identify the divergences that give rise to morphological mismatches and thus to potential problems for the transfer. To handle morphological translation in general and morphological mismatches in particular, we suggest a morphological translation module that consists of a bilingual grammatical signification index and a rule-based morphological transfer engine. Both the index and the engine are intended to form an integral part of a translation model that furthermore comprises both a lexical and a syntactic transfer engine (as described in Mel'čuk and Wanner 2001, 2006), which operate on the DSyntS of an MTT-model. In the framework of this model, mismatches between a DSyntS$_S$ and its equivalent DSyntS$_T$ are interpreted as violations of the isomorphism between them.

To buttress our proposal, we have examined in more detail the transfer rules for two examples of the '$g_S \Leftrightarrow$ —' mismatch in German–Russian translation: the German category of definiteness, which is absent from Russian, and the Russian category of aspect, which is absent from German. For illustration, we apply these rules to two sample texts retrieved from the internet. Their apparent sufficiency demonstrates that high-quality morphological translation is possible, at least in the case of human-aided MT.

A possible criticism of our approach is that it has little chance to be scaled up. Scaling up is a problem if the number of phenomena to be addressed by a model is very big or if the (lexical) resources to handle the phenomena are unlikely to be available. As far as the first possible obstacle is concerned, we should be aware that the number of morphological mismatches is limited such that they can be controlled by our model. With respect to the second possible obstacle, it can be argued that the community

has already recognized the vital importance of lexical resources for a whole range
of NLP applications, and the number of large-scale lexical resources is constantly
increasing; cf., for instance, WordNet (Fellbaum 1998), EuroWordnet (Vossen (ed.)
1999), MEANING (Rigau et al. 2002), KYOTO (Vossen et al. 2008) for linguistically
motivated large-scale resources and CYC[41] for an AI-motivated resource. Therefore,
it is rather plausible to assume the availability of resources as required by our model
in the near future.

Appendix 1: Linguistic excursuses

A1.1 Proposed description of German conjugation

As a result of the organization of German verbal forms along two independent axes,
one for tense and the other for perfectivity, as suggested in Subsect. 3.1.2.1, each
verb form in the active of the indicative is represented in the DSyntS by two
grammemes:

Trad. name	Proposed encoding
Präsent	'PRES', 'NON-PERF'
Imperfekt	'PAST', 'NON-PERF'
Futurum I	'FUT', 'NON-PERF'
Perfekt	'PRES', 'PERF'
Plusquamperfekt	'PAST', 'PERF'
Futurum II	'FUT', 'PERF'

Cf. the forms of the verb FRAGEN 'ask' in 1SG:

FRAGEN$_{PRES,NON-PERF}$	⇔	*frage*
FRAGEN$_{PAST,NON-PERF}$	⇔	*fragte*
FRAGEN$_{FUT,NON-PERF}$	⇔	*werde fragen*
FRAGEN$_{PRES,PERF}$	⇔	*habe gefragt*
FRAGEN$_{PAST,PERF}$	⇔	*hatte gefragt*
FRAGEN$_{FUT,PERF}$	⇔	*werde gefragt haben*

This is a purely formal grouping since there is no semantic parallelism between
parallel forms. Thus, for instance, 'PRES','NON-PERF' : 'PRES','PERF' ≠ 'PAST','NON-PERF' : 'PAST','PERF'. However, assigning meaning to grammeme combinations is in
any case a problem apart; for our purposes—for establishing correspondences between
grammemes of different languages—such a "decomposed" description is more conve-

[41] http://www.opencyc.org/

nient than the "global" description of the traditional grammar. In particular, in this case it allows us to avoid one possible type of mismatch: the syntagmatic correspondence of one grammeme to a string of grammemes. Thus, under the traditional description we see Ger. 'IMPERFEKT' ⇔ Rus. 'PAST', 'IMPERFECTIVE'; in the decomposed description, we have Ger. 'PAST' ⇔ Rus. 'PAST', Ger. 'NON-PERF' ⇔ Rus. 'IMPERFECTIVE'.

A1.2 Korean politeness marking

Politeness level	DECLARATIVE	INTERROGATIVE	IMPERATIVE
Plain	*po + n +* **ta**	*po +* **ni***?*	*po + a +* **la***!*
Intimate	*po + a + Ø*	*po + a + Ø?*	*po + a + Ø!*
Familiar	*po + n +* **ey**	*po +* **na***?*	*po +* **key***!*
Blunt	*po + o*	*po + o?*	*po + o!*
Polite	*po + a +* **yo**	*po + a +* **yo***?*	*po + a +* **yo***!*
Deferential	*po + p + ni +* **ta**	*po +* **na***?*	*po + si + p + si +* **o***!*

There are six levels of politeness in Korean; in order to show the complexity of its morphological expression, we cite in the table above the verbal suffixes that distinguish politeness forms of the verb PO- 'see' in the three types of sentences: DECLARATIVE, INTERROGATIVE and IMPERATIVE.

Some forms include additional suffixes: *-n/-ni*: indicative mood, *-a/-e*: infinitive. Suffixes in the deferential imperative form an ending all elements of which must be present.

Appendix 2: Texts

Text 1

Russian

Večerom v sredu, točno v 18 časov 16 minut po mestnomu vremeni dva moščnyx vzryva sotrjasli gorodok Xibxib v vos′mi kilometrax k severu ot Bakuby. Dva otdel′no stojaščix zdanija, okružennye pal′movymi roščami, byli prevraščeny v grudu razvalin: ot nix ne ostalos′ kamnja na kamne. Soldaty sil bezopasnosti, kotorye ždali ètoj minuty so včerašnego dnja, ustremilis′ k mestu padenija bomb— i udostoverilis′: na ètot raz on popalsja. Abu Mussab al-Zarkavi, odin iz samyx razyskivaemyx terroristov mira, zamestitel′ Bin Ladena v Irake, mertv.

German

Genau um 18 Uhr und 16 Minuten Ortszeit erschütterten am Mittwochabend zwei heftige Explosionen die Kleinstadt Hibhib, acht Kilometer nördlich von Bakuba. Zwei freistehende Gebäude, umgeben von Palmenhainen, wurden zu Schutt und Asche, kein Stein blieb auf dem anderen. Die Sicherheitskräfte, die auf diese Minute seit dem gestrigen Tag gewartet hatten, rasten zum Einschlagsort der Bomben—und waren sich sicher: Diesmal haben sie ihn erwischt. Abu

Mussab al-Sarkawi, einer der meistgesuchten Terroristen der Welt, Osama Bin Ladens Stellvertreter im Irak, ist tot.

English translation

Wednesday evening, exactly at 18:16 local time, two powerful explosions rocked the little town of Hibhib, 8 km north of Bakuba. Two free-standing buildings, surrounded by palm groves, were turned to rubble and ashes; no stone remained on the other. The security forces, who had been waited for this moment since yesterday rushed to the spot were the bombs hit—and knew for sure: this time they got him. Abu Mussab al-Zarkawi, one of the most wanted terrorists in the world, Osama Bin Laden's representative in Iraq, was dead.

The translation simulated in our study is from Russian into German, that is, in the difficult direction as far as the category of definiteness is concerned.

Here is the list of all German nouns, in order of their appearance in the text, supplied with the number of the rule that was applied to obtain the definiteness grammeme.

German noun	Rule number
Uhr$_{\text{NON-DEF}}$, Minuten$_{\text{NON-DEF}}$	9
Ortszeit$_{\text{NON-DEF}}$	(Definiteness in its meaning; dictionary)
Mittwochabend$_{\text{DEF}}$	1
Explosion$_{\text{NON-DEF}}$	9
Kleinstadt$_{\text{DEF}}$	10
Hibhib$_{\text{NON-DEF}}$	(Proper noun; dictionary)
Kilometer$_{\text{NON-DEF}}$	9
Bakuba$_{\text{NON-DEF}}$	(Proper noun; dictionary)
Gebäude$_{\text{NON-DEF}}$	9
Palmenhain$_{\text{NON-DEF}}$	2
Schutt$_{\text{NON-DEF}}$, Asche$_{\text{NON-DEF}}$, Stein	(Within idiom; dictionary)
Sicherheitskräfte$_{\text{DEF}}$	1
Minute$_{\text{NON-DEF}}$	8
Tag$_{\text{DEF}}$	1
Einschlagsort$_{\text{DEF}}$	1 and/or 2
Bombe$_{\text{DEF}}$	1 and/or 2
Abu Mussab al-Sarkawi$_{\text{NON-DEF}}$	(Proper noun; dictionary)
Terrorist$_{\text{DEF}}$	12
Welt$_{\text{DEF}}$	1
Osama Bin Laden$_{\text{NON-DEF}}$	(Proper noun; dictionary)
Stellvertreter$_{\text{NON-DEF}}$	8
Irak$_{\text{DEF}}$	6

Text 2

German

Was genau Kanzlerin Merkel Peter Struck gesagt hat, ist unklar. Regierungssprecher Ulrich Wilhelm erwähnte heute lediglich, sie habe mit dem Sozialdemokraten über dessen Äußerungen zur Amtsführung der Regierungschefin gesprochen. Struck hat heftigen Ärger in der Koalition ausgelöst, weil er in der FAZ erklärt hatte, er vermisse Gerhard Schröder als Bundeskanzler: Es wäre besser für Deutschland, wenn Schröder noch regieren würde. Struck hatte Schröders Entscheidungsfreude hervorgehoben. Merkel dagegen sei eine Regierungschefin, "die viel mehr auslotet".

Führende CDU-Politiker griffen Struck öffentlich heftig an. Hessens Ministerpräsident warf Struck mangelhafte Mitarbeit vor. Er warnte, die Koalitionspartner sollten "respektvoll" miteinander umgehen.

Russian

Čto imenno skazala kanzler Merkelʹ Peteru Štruku, nejasno. Predstavitelʹ pravitelʹsta Ulʹrix Vilʹgelʹm segodnja otmetil tolʹko, čto ona govorila so Štrukom o ego vyskazyvanijax po povodu eë metodov upravlenija.

Štruk vyzval serʹëznoe nedovolʹstvo v pravitelʹstvennoj koalicii, zajaviv v gazete "FAZ", čto emu ne xvataet Gerxarda Šredera v kačestve kanclera: "Dlja Germanii bylo by lučše, esli by Šreder ostalsja u vlasti". Štruk podčerknul rešitelʹnostʹ Šredera. Merkelʹ že javljaetsja glavoj pravitelʹstva, kotoraja dejstvuet gorazdo bolee ostorožno.

Veduščie politiki partii CDU publično napali na Štruka. Premʹer-ministr Gessena upreknul Štruka v nedostatočnom sotrudničestve. On predupredil, čto partnery po koalicii dolžny obraščatʹsja drug s drugom "uvažitelʹno".

English translation

It is not clear what exactly Chancellor Merkel said to Peter Struck. The government's spokesman Ulrich Wilhelm mentioned yesterday only that she had talked to Struck about his statements concerning her management style.

Struck provoked serious anger in the government coalition by declaring in the FAZ newspaper that he was missing Gerhard Schröder as Chancellor: "It would be better for Germany if Schröder were still in power." Struck underlined Schröder's determination in decision making. Merkel, on the contrary, is a head of government "who mulls over her decisions too much."

Leading CDU politicians publicly attacked Struck in force. Hessen's Premier reproached Struck for lack of collaboration. He warned that coalition partners should treat each other "respectfully."

Here the direction of the simulated translation is German-to-Russian.

Consider the list of Russian verbs from Text 2, where each one is supplied with the number of the rule that has been applied to obtain its aspect grammeme.

Russian verb	Rule number
govorit'$_{\text{PERF}}$ 'say'	17
otmečat'$_{\text{PERF}}$	17
govorit''$_{\text{IMPERF}}$ 'speak'	18
serdit'$_{\text{PERF}}$	19
zajavljat'$_{\text{PERF}}$	17
ostavat'sja$_{\text{PERF}}$	17
podčerkivat'$_{\text{PERF}}$	17
atakovat'$_{\text{PERF}}$	17
uprekat'$_{\text{PERF}}$	17
preduprеždat'$_{\text{PERF}}$	17
obraščat'sja'$_{\text{IMPERF}}$	(*Imperfectivum tantum*; dictionary)

References

Anscombre J-C (1986) Article zéro, termes de masse et représentation d'événements en français contemporain. In: David J, Kleiber G (eds) Déterminants: syntaxe et sémantique. Université de Metz [= Recherches linguistiques, XI], Metz, pp 5–34

Apresjan JD (1988) Glagoly momental'nogo dejstvija i performativy v russkom jazyke. In: Karaulov JN (ed) Rusistika segodnja. Jazyk: sistema i ee funkcionirovanie. Nauka, Moskva, pp 57–78

Apresjan JD, Boguslavskij IM, Iomdin LL, Lazurskij AV, Pertsov NV, Sannikov VZ, Cinman LL (1989) Lingvističeskoje obespečenie sistemy ÈTAP-2. Nauka, Moscow

Apresjan JD, Boguslavsky IM, Iomdin LL, Tsinman LL (2007) Lexical functions in actual NLP-applications. In: Wanner L (ed) Selected lexical and grammatical topics in the Meaning-Text theory. In honour of Igor Mel'čuk. Benjamins, Amsterdam, pp 203–233

Birkenmaier W (1979) Artikelfunktionen in einer artikellosen Sprache. Wilhelm Fink Verlag, München

Bisle-Müller H (1991) Artikelwörter im Deutschen. Semantische und pragmatische Aspekte ihrer Verwendung. Tübingen, Niemeyer

Dressler W, Merlini L (1994) Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages. de Gruyter, Berlin, New York

Fellbaum Ch (ed) (1998) WordNet: an electronic lexical database. The MIT Press, Cambridge, MA

Gawrońska-Werngren B (1990) Translation great problem—on the problem of inserting articles when translating from Russian into Swedish. COLING-90, Papers Presented to the 13th International Conference on Computational Linguistics, Helsinki, Finland, pp 133–138

Gawrońska B (1992) Aspect—a problem for MT. Proceedings of the Fifteenth International Conference on Computational Linguistics, COLING-92, Nantes, France, pp 652–657

Gladrow W (1979) Die Determination des Substantivs im Russischen und im Deutschen. VEB Verlag Enzylopädie, Leipzig

Glovinskaja MJa (1982) Semantičeskie tipy vidovyx protivopostavlenij russkogo glagola. Nauka, Moskva

Glovinskaja MJa (2001) Mnogoznačnost' i sinonimija v vido-vremennoj sisteme russkogo glagola. Azbukovnik/Russkie slovari, Moskva

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. Envisioning Machine Translation in the Information Future, 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000. LNAI 1934, Springer, Berlin, pp 40–53

Hansen B (1995) Die deutschen Artikel und ihre Wiedergabe im Türkischen. Arbeiten zur Mehrsprachigkeit Bd. 53. Germanistisches Seminar, Hamburg

Heim I (2001) Artikel und Definitheit. In: von Stechow A, Wunderlich D (eds) Semantik. (HSK 6). de Gruyter, Berlin, New York, pp 487–534

Himmelmann NP (2001) Articles. In: Haspelmath M, König E, Oesterreicher W, Raible W (eds) Sprachtypologie und sprachliche Universalien (HSK 20.1). de Gruyter, Berlin, pp 831–842

Kabakčiev K (1984) The article and the aorist/imperfect distinction in Bulgarian: an analysis based on cross-language 'aspect' parallelisms. Linguistics 22:643–672

Kobozeva IM (1980) Nekotorye pravila vybora vida glagola pri sinteze prostogo predloženija, vyražajuščego zadannyj smysl. Aktualʹnye voprosy strukturnoj i prikladnoj lingvistiki. Izdatelʹstvo Gos. Moskovskogo Universiteta, Moskva, pp 91–103

Kovtun O (2003) Zur unbestimmten Determination im Deutschen im Vergleich zum Ukrainischen und Russischen. In: Hoffmann L (ed) Funktionale Syntax: Prinzipien und Prozeduren. de Gruyter, Berlin pp 335–347

Leiss E (2000) Artikel und Aspekt: die grammatischen Muster von Definitheit. de Gruyter, Berlin

Marcu D, Wang W, Echihabi A, Knight K (2006) SPMT: statistical machine translation with syntactified target language phrases. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, pp 44–52

Mehlig HR (1978) Überlegungen zur Funktion und Determinierung der Aspekte im Russischen. Slavistische Linguistik 1978 (= Slavistische Beiträge 133). Sagner, München

Mehlig HR (1988) Verbalaspekt und Determination. Slavistische Linguistik 1987 (= Slavistische Beiträge 230). Sagner, München

Mehlig HR (1989) Thema-Rhema Gliederung und Verbalaspekt. Zum Aspektgebrauch in russischen Fragesätzen. Slavistische Linguistik 1988 (= Slavistische Beiträge 242). Sagner, München

Mel'čuk IA (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. Benjamins Academic Publishers, Amsterdam, Philadelphia, pp 37–102

Mel'čuk IA (2006) Aspects of the theory of morphology. Mouton de Gruyter, Berlin, New York

Mel'čuk I, Wanner L (2001) Towards a lexicographic approach to lexical transfer in machine translation (illustrated by the German–Russian language pair). Mach Transl 16(1):21–87

Mel'čuk I, Wanner L (2006) Syntactic mismatches in machine translation. Mach Transl 20(2):81–138

Nagao M (1984) A framework of a mechanical translation between Japanese and English by analogy principle. In: Elithorn A, Banerji R (eds) Artificial and human intelligence. Elsevier Science Publishers, Amsterdam, pp 173–180

Ney H, Matusov E, Mauser A, Vilar D, Zens R (2007) Statistical MT from TMI-1988 to TMI-2007: what has happened? Invited talk at TMI-2007. Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation. Skövde, Sweden

Nikolaeva TM (1959) Opredelenie vida russkogo glagola s pomoščʹju konteksta. Mašinnyi perevod i prikladnaja lingvistika. 2:38–44

Olsen M, Traum D, Van Ess-Dykema C, Weinberg A (2001) Implicit cues for explicit generation: using telicity as a cue for tense structure in a Chinese to English MT system. MT summit VIII: Machine translation in the information age. Santiago de Compostela, Spain, pp 259–264

Padučeva EV (1996) Semantičeskie issledovanija. Semantika vremeni i vida. Semantika narrativa. Škola "Jazyki russkoj kulʹtury", Moscow

Padučeva EV (1998) Semantičeskie istočniki momentalʹnosti russkogo glagola v tipologičeskom rakurse. In: Čertkova MJu (ed) Tipologija vida: problemy, poiski, rešenija. Škola "Jazyki russkoj kulʹtury", Moscow pp 332–342

Rigau G, Magnini B, Aggire E, Vossen P, Carroll JA (2002) MEANING: a roadmap to knowledge technologies. Proceedings of the Workshop A Roadmap for Computational Linguistics, International Conference On Computational Linguistics, COLING-02, Taipei, Taiwan, pp 1–7

Sacker U (1983) Aspekteuller und resultativer Verbalausdruck. 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. Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural 27:183–189

Scheljakin MA, Schlegel H (1970) Der Gebrauch des russischen Verbalaspekts I. Potsdamer Pädagogische Hochschule, Potsdam

Somers H (1999) Example-based machine translation. Mach Transl 14(2):113–157

Vater H (1997) Determination im Deutschen und Polnischen. In: Koniuszaniec G, Prokop I (eds) Festschrift für Andrzej Z. Bzdęga zum 70. Geburtstag (= Studia Germanica Posnaniensia XXII). Uniwersytet Poznańskij, Poznań, pp 215–229

Vater H (2002) The word class 'article'. In: Cruse A (ed) Lexikologie (HSK 15.1). de Gruyter, Berlin, pp 621–628

Vossen P (ed.) (1999) EuroWordNet. General document. Technical report. Amsterdam, University of Amsterdam

Vossen P, Aggire E, Calzolari N, Fellbaum C, Hsieh S, Huang C, Isahara H, Kanzaki K, Marchetti A, Monachini M, Neri F, Raffaelli R, Rigau G, Tescon M (2008) KYOTO: a system for mining, structuring, and distributing knowledge across languages and cultures. Proceedings of LREC 2008. Marrakech, Morocco

Wierzbicka A (1967) On the semantics of the verbal aspect in Polish. In: To honor Roman Jakobson. The Hague, Mouton, pp 2231–2249