Published in: Hansen-Schirra, Silvia/Neumann, Stella/Steiner, Erich (eds.): Cross-linguistic Corpora for the Study of Translations – Insights from the language pair English-German. - Berlin/Boston: de Gruyter, 2012. Pp. 133-146.
(Text, Translation, Computational Processing 11)

Sandra Hansen & Silvia Hansen-Schirra

7 Grammatical shifts in English-German noun phrases

1 Motivation: Translation shifts and translation-specific properties

When comparing source texts and their translations, translation shifts become apparent. Translation shifts have been discussed in translation studies since the 1950s (Vinay and Darbelnet 1958: 30ff. [English translation by Sager and Hamel 1995]; Catford 1965; Newmark 1988: 81ff.; van Leuven-Zwart 1989). The accounts are similar in that they categorize lexical, grammatical, and semantic shifts on various levels/ranks. Grammatical shifts are often called *transpositions* and refer to changing tense, number, person, part-of-speech or phrasal category. In semantic shifts, or *modulations*, a change of perspective occurs between source and target text. This may involve concretion, explication, negation of the opposite, (de-) passivization, etc. (cf. Vinay and Darbelnet 1958).

In computational linguistics, translation shifts of all types are a crucial issue for the development of MT systems. Identification, classification and formalization of translation shifts have received considerable attention in the MT community (e.g. in the Eurotra project, Copeland et al. 1991 or Barnett et al. 1991; Lindop and Tsujii 1991; Kinoshita, Phillips, and Tsujii 1992). Within this context, Dorr (1994) proposes a more fine-grained categorization of MT divergences. She distinguishes between thematic, promotional, demotional, structural, conflational, categorical and lexical divergences, thus using linguistic categories. Cyrus (2006) and Padó (2007) combine the two perspectives, focusing on semantic shifts.

Translation shifts may be due to cognitive factors, such as the translator's understanding, idiosyncratic preferences or constraints during the translation process, to contrastive differences between the languages involved or to different register characteristics. Thus, the resulting product, the translation, may differ on several linguistic levels compared to their source language texts, but also compared to original texts in the target language since they are based on a text in another language (interference). Recently, these translation properties have been empirically investigated using parallel and comparable corpora. For the language pair English-German, Teich (2003) finds a special kind of source language interference (the typical language use of the source language "shines

through" in the German translations). Further corpus studies involve the investigation of information structure in English and German texts (Doherty 1999), thematic structure (Hasselgård 1998), information packaging (Steiner 2002, 2004a; Fabricius-Hansen 1999), explicitation (Hansen-Schirra, Neumann, and Steiner 2007) and normalization (Hansen 2003). The research presented in section 3 tries to link up insights on translation shifts to the resulting translation properties using the concept of "grammatical metaphor", which is introduced in the following section.

2 Methodology: Grammatical metaphor in translated text

The categories mentioned above (transposition, modulation, etc.) are not fine-grained enough for a detailed description of translation shifts and their effect on the translation product. In most cases, they label the translation procedure, however, they are not sufficient to comprehensively describe the differences between source and target text. For instance, according to Vinay and Darbelnet (1958), a verbalization as well as a nominalization is labeled as "transposition". However, in order to understand what is going on during the translation process, it is important to know whether a translation is more nominal or rather verbal, that is to say the directionality of the change matters. Another disadvantage of the early notion of "transposition" is the fact that phenomena that have to do with different grammatical phenomena such as change of lexical vs. phrasal category, word order, dependency, lexical or phrasal features (such as number, person, gender, definiteness, voice), change of level of projection/rank and others all tend to be lumped together under the term "transposition", whereas they are, in fact, very different types of phenomena, both in terms of grammar and in terms of the processing issues that arise. Therefore, the concept of "grammatical metaphor" is introduced to account for translation shifts in greater detail, in finer granularity and with regard to the directionality of the relationships and processes involved.

Grammatical metaphor[1] (cf. Halliday 1985: 319ff. or Halliday and Matthiessen 1999: 227ff.) can be described as the encoding of the same ideational meaning by means of different phrasal categories (nominal vs. verbal vs. prepositional, etc.), or at different ranks, such as clause complex, clause, phrase, group, word or

1 We are leaving "interpersonal metaphors" aside at this point.

morpheme (cf. Steiner 2001b: 7ff.; 2004a: 139–166; 2004b: 141ff.). Within models of translation, the notion of grammatical metaphor has been used to describe differences between source and target language texts (cf. Steiner 2001b and Hansen 2003). Types of grammatical shifts, which can take place in grammatical metaphors across languages, are listed in Table 25, which is taken from Steiner (2001b: 14) and based on Halliday and Matthiessen (1999: 246ff.). Once again, it is important to mention that grammatical metaphor across languages is not restricted to the translation procedure termed "transposition" (cf. Vinay and Darbelnet 1995 and Newmark 1988), covering a much wider and also more theoretically motivated range of phenomena. It also implies a semantic tension between a congruent expression in a language A and its metaphorical realization in a language B (cf. Steiner 2001b). "Congruent" refers to a direct and transparent mapping from semantic onto grammatical categories, for example processes onto verbs, entities onto nouns, qualities of an entity to an adjective, etc., whereas other "indirect" mappings as in Table 25 below are classified as "metaphorical". The notions of "(in-)directness" of semantics-to-grammar mapping are widely used in the comparative and typological literature (cf. Hawkins 1986: 53ff. or König and Gast 2009: 98ff.). In example (1) below, both the English and the German version have the same degree of "metaphoricity", whereas in examples (2) and (3) below, the German versions are denser, less direct and therefore "more metaphorical".

Concerning the understanding of the SL text during the translation process, the following assumptions can be formulated (cf. Steiner 2001b and chapter 8 of this book): understanding within the translation process involves the unpacking of grammatical metaphor in the source text, i.e. moving from more metaphorical to more congruent variants, and re-metaphorizing it again in the target text. However, the process of re-metaphorization may not be fully completed during the production of the TL text. On the other hand, it may also result in an even more densely packed version resulting in metaphorization. Thus, the following effects may be found in the translation product (cf. also Hansen 2003):

- The term "metaphorization" is used in cases in which the TL expression is more metaphorical than the SL expression.
- The term "de-metaphorization" describes cases in which the SL is more metaphorical than the TL expression.
- The term "re-metaphorization" labels translations which show the same degree of grammatical metaphoricity as the SL expression.

In terms of translation properties, re-metaphorization goes hand in hand with shining through since the structures of the SL text are preserved in the

shift: metaphorical ⇒ congruent	English-German example
noun ⇒ adjective	*instability ⇒ instabil*
noun ⇒ verb	*transformation ⇒ transformieren*
noun ⇒ auxiliary	*the possibility of ⇒ können*
noun ⇒ semi-auxiliary	*the desire to ⇒ wollen*
noun ⇒ preposition	*accompaniment ⇒ mit*
noun ⇒ PP	*floor dust ⇒ Staub auf dem Boden*
noun ⇒ conjunction	*condition ⇒ wenn*
adjective ⇒ verb	*rising poverty ⇒ Armut steigt*
adjective ⇒ auxiliary	*the previous, past ... ⇒ war*
adjective ⇒ semi-auxiliary	*the initial ... ⇒ beginnen*
adjective ⇒ preposition	*the accompanying ⇒ mit*
adjective ⇒ PP	*superficial ... ⇒ auf der Oberfläche*
adjective ⇒ conjunction	*previous ⇒ vor*
verb ⇒ preposition	*replace ⇒ anstatt*
verb ⇒ PP	*to box ⇒ in eine Schachtel*
verb ⇒ conjunction	*follow ⇒ dann*
preposition ⇒ conjunction	*because of ⇒ weil*
PP ⇒ conjuction	*as a result ⇒ deshalb*
noun ⇒ 0	*the fact that ... ⇒ 0*
verb ⇒ 0	*have an influence ⇒ beeinflussen*
causative ⇒ 0	*X imposes work on Y ⇒ X lässt Y arbeiten*
phase auxiliary ⇒ 0	*to begin an examination ⇒ beginnen zu untersuchen*
post-modification ⇒ NP head	*a decision by the government ⇒ die Regierung entscheidet*
deictic ⇒ NP head	*the government's decision ⇒ die Regierung entscheidet*
classifier ⇒ NP head	*governmental decision ⇒ die Regierung entscheidet*
adjective ⇒ adverb	*a hasty decision ⇒ entscheidet hastig*
adjective ⇒ PP	*a lengthy argument ⇒ stritten für eine längere Zeit*
various ⇒ adverb	*yesterday's quarrel ⇒ stritten gestern*
various ⇒ PP	*departure for the station ⇒ fahren zum Bahnhof*

Table 25: Types of metaphorical shifts in translation

translation. In contrast, de-/metaphorization may result in normalization. If it is true that "the wording that is lower in rank will contain less information" (cf. Halliday and Matthiessen 1999: 231), it can be argued that grammatical metaphor is also related to the concept of simplification and explicitation (cf. Blum-Kulka 1986 and Baker 1996). These correlations are empirically investigated in the following section.

3 Empirical analysis of English-German noun phrases

The empirical analysis presented here is based on the ESSAY part of the CroCo Corpus. Furthermore, the CroCo reference corpora help to differentiate between register-specific and typologically driven language use. The CroCo annotation (see chapter 3) has been extended in such a way that all nominal heads have been marked and aligned (where applicable). This is illustrated in Figures 31 and 32: the head of a nominal phrase (e.g. *vision* or *markets* in Figure 31) is marked with the XML tag <head> and receives an "id". The "id" of the source text head is refered to in the target text (through "idref") and vice verca. This can be seen as the alignment of nominal heads, which can also result in an empty link (as explained in chapter 6).

Shifts in the structure from a source text head to a target text head are marked through the attribute "transfeat" (for translation feature), which indicates a "changed" or "unchanged" structure. In Figure 32, the embedded head *Märkte* (*markets*) is labeled with "unchanged" since it is premodified by an adjective in both languages. In contrast, the head *Vision* (*vision*) is analyzed as "changed", which is due to the shift from the English postmodifying PP to the German genitive postmodification.

```
<pre-modification type="determiner">Our</pre-modification>
<head id="en-10" idref="ge-9">vision</head>
<post-modification type="prepositional-phrase">for
  <pre-modification type="adjective">open</pre-modification>
  <head id="en-11" idref="ge-10">markets</head>
</post-modification>
```

Figure 31: Extended CroCo annotation of English originals

```
<pre-modification type="determiner">Unsere</pre-modification>
<head id="ge-9" idref="en-10" transfeat="changed">Vision</head>
<post-modification type="genitive">
  <pre-modification type="adjective">offener</pre-modification>
  <head id="ge-10" idref="en-11" transfeat="unchanged">Märkte</head>
</post-modification>
```

Figure 32: Extended CroCo annotation of German translations

The pre- and postmodifiers of the NPs are analyzed and labeled with the XML tag <pre-modification> or <post-modification> (see Figures 31 and 32). The following categories are included as premodifiers: adjective, adverb, determiner, genitive, noun (phrase), participle (phrase) and prepositional phrase. The following postmodifiers are annotated: adverb, apposition, genitive, infinitive (clause), noun (phrase), participle (phrase), proposition, prepositional phrase, relative clause and adjective phrase. On the basis of this extension to the CroCo annotation, shifts in the structure of source and target NPs can automatically be classified and quantified. Moreover, the detailed annotation of the NP presented here can be seen as an additional layer of the CroCo annotation, which can be combined with the other layers (e.g. part-of-speech or grammatical functions).

Comparing source texts in English and their translations into German, a chi-square test shows that the translations result in a significantly higher number of shifts compared to translations with identical realizations in the source and target texts ($\chi^2_{(1)} = 38{,}01$; $p < 0.001$).[2] This means that the structure of the English nominal group is only rarely preserved in the German translations, indicating normalization or hybridization rather than shining through.

Table 26 shows the typical (indicated by "+") and untypical (indicated by "–") pre- and postmodifiers of English (ER) and German (GR) while comparing the English and German reference corpora. Moreover, it shows which structures are typical (indicated by "+") and untypical (indicated by "–") for the English (EO_ESSAY) and German (GTrans_ESSAY) essays contrasting the register corpora to the reference corpora for English and German respectively. The last column indicates our findings when comparing the German translation corpus to the English originals (GTrans vs. EO) and to the German comparable corpus (GTrans vs. GO). Statistical nearness in terms of similarity between GTrans and EO is interpreted as shining through since the source language structures are preserved in the translations. Normalization can be found in cases of nearness between GTrans and GO in which the translations adhere to the norms of the comparable texts in the target language.

Note that Table 26 contains the significant differences only. For significance testing, a logistic regression analysis has been carried out (cf. Baayen 2008). This method is more precise when dealing with categorical data (cf. Jaeger 2008). It compares probabilities of feature frequencies under different conditions (sub-corpora in our case), thus paving the way for interpreting the dependent

2 "p" indicates the significance level, i.e. the error probability with which the observed event would occur, with $p < 0.05$ being significant, $p < 0.01$ being highly significant, etc.

		ER	GR	EO_ ESSAY	GTrans_ ESSAY	GTrans vs. GO
PREMODIFICATION	adjective			+	+	over-normalization
	adverb		+			normalization
	determiner		+	+	+	over-normalization
	genitive	+		–		normalization
	noun	+				normalization
	NP	+				normalization
	participle					pseudo-normalization
	participle phrase		+			normalization
POSTMODIFICATION	adverb	+				normalization
	apposition			–	–	over-normalization
	genitive		+	+	+	over-normalization
	infinitive clause	+				normalization
	NP	+				normalization
	participle phrase	+				normalization
	proposition	+				normalization
	PP				+	normalization
	relative clause	+				shining through

Table 26: Pre- and postmodifiers in English-German translations

variables (i.e. pre- and postmodifiers) against the independent ones (language, register, translation vs. original).[3]

Table 26 indicates significant differences for either the English and German sub-corpora or the German comparable corpus. No significant differences have been found for the following modifiers: premodifying prepositional phrase as well as postmodifying adjective phrase, clausal apposition, infinitive, noun and participle. Therefore, they are not included in Table 26. The significant results show that the following modifiers are typical of the English reference corpus (indicated by "+"): premodifying genitives, nouns and NPs as well as postmodifying adverbs, infinitive clauses, NPs, participle phrases, propositions and relative clauses. There are fewer modifiers typical of the German reference corpus: premodifying adverbs and determiners as well as postmodifying genitives. For both languages, adjectives, determiners and genitives as postmodifiers can be classified as register features for political essays. Prepositional phrases as post-

3 An interpretation of these results in terms of LSP (language for specific purposes) features can be found in Hansen-Schirra et al. (2009).

modifiers are characteristic of German political essays only. The analysis also reveals negative, i.e. rarely used, register features (indicated by "–"): appositions for English and German political essays and premodifying genitives for English political essays.

The contrastive differences are especially interesting in regard to the translation task since the translators have to decide whether they adhere to the structures of the source language or whether the norms of the target language are to be applied. The comparison of German translations to German originals of the same register reveals that the translators used the norms of the target language (which results in normalization) with respect to the following features (see Table 26): adverb, genitive, noun, NP, participle phrase (as premodifiers) and adverb, infinitive clause, NP, participle phrase, proposition, PP (as postmodifiers). This means that for the majority of modifiers, no significant differences have been found in the comparable corpus.[4]

Additionally, there are some special cases with respect to normalization: Determiners, adjectives and postmodifying genitives – all typical features of the German language or register – are more frequently used in the German translations compared to the German comparable corpus. This over-use of typical features results in an exaggerated normalization, which we call "over-normalization". The same holds for appositions: in this case, this negative register feature is less frequently used in the translations, exaggerating the norms of the target language. Again, this under-use results in an over-normalization.

Another case can be found when looking at the results for premodifying participles. Here, no significant differences could be found between the English and German reference and register corpora. There is, however, a significant over-use of this premodifier in the German translations compared to the German comparable corpus. Since the norms of English and German do not differ with respect to this feature, this effect cannot be interpreted as normalization. Still, it could be the case that the translator consciously or unconsciously copied the usage of premodifying participle phrases, which are characteristic of German. A high frequency of this premodifier in translated text results in a normalization effect (as described above). By imitating this behavior with participles only, the result is a pseudo-normalization of the translations.

4 The interpretation of non-effects is not unproblematic since the absence of a significant difference might be due to an insufficient data constellation. This does, however, not seem to apply to our corpus analysis because all relevant error probabilities are far below the 10% mark. Furthermore, the present sample size of $N > 10000$ indicates that existing differences must have been found.

Finally, only one case of shining through could be found in the comparable corpus: the frequency of relative clauses is not in accordance with the figures for the comparable corpus – the high number of relative clauses, which is typical of the English originals, is preserved in the translations.

In the following, some example cases will be discussed for the quantitative findings. It should be noted that the interpretation of examples with respect to translation properties is problematic since we are not yet able to classify the translation shifts on the basis of alignment figures. Therefore, the discussion presents individual translation shifts, which have to be quantified in future work. Let us start with the last of the phenomena introduced above, the shining through of relative clauses. Figures 33 and 34 display the results of the logistic regression analysis for relative clauses in the different sub-corpora.

Figure 33 shows that relative clauses are a contrastive feature of English (ER vs. GR: $p < 0.05$). They do, however, not have an effect on the register, as can be seen from contrasting the translations to the German reference corpus in Figure 34 (GTrans_ESSAY vs. GR: $p > 0.05$). There is, nevertheless, a significant difference between the translations and the comparable corpus (GTrans_ESSAY vs. GO_ESSAY: $p < 0.01$), showing that the source language structures (i.e. the high frequency of relative clauses) shine through in the translations. Example (1) illustrates such a shining through effect:

(1) a. *the growth that is essential to achieve that goal.* (EO_ESSAY)

b. *das Wachstum, das zur Erlangung dieses Ziels erforderlich ist* (GTrans_ESSAY)

A more idiomatic translation of the English relative clause could have been achieved through a premodifying participle phrase (e.g. *das zur Erlangung dieses Ziels erforderliche Wachstum*) resulting in a normalization. The higher frequency of relative clauses makes the German translations more explicit compared to German originals where grammatical metaphor can typically be found.

When looking at the most common phenomenon occurring in Table 26, i.e. that typical modifiers of English are not preserved in the translation but replaced by typical German modifiers, this normalization frequently co-occurs with grammatical metaphorization. In examples (2) and (3), postmodifying infinitive clauses, which are characteristic of English, are translated by nominal constructions (in both cases postmodifying PPs), resulting in grammatically more metaphorical structures. As the semantics of the postmodifiers in (2) and (3) below is that of a process, the nominal encodings of the German versions here are less direct and "more metaphorical" than the English ones. The process of making postmodification less explicit and thus more difficult to process is in

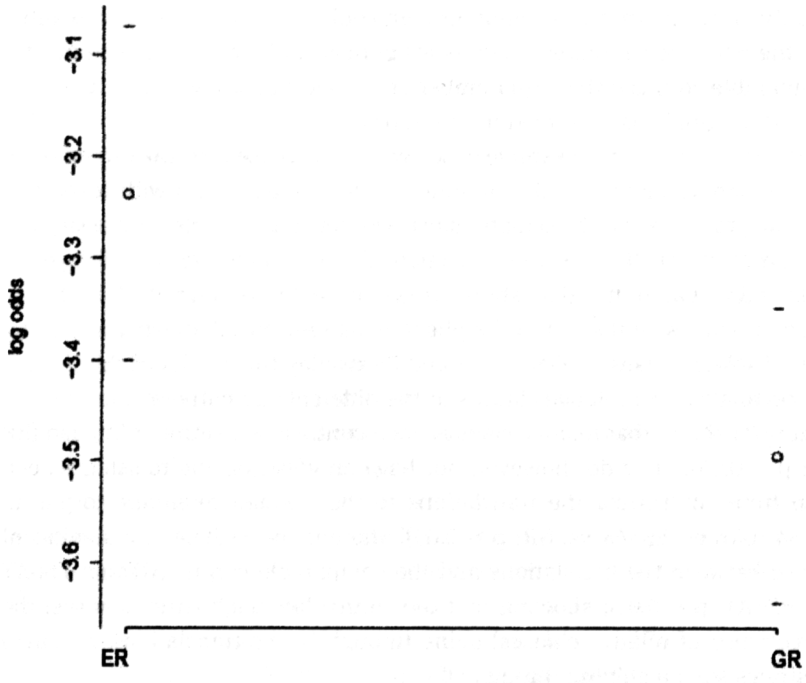

Figure 33: Logistic regression analysis for relative clauses in the reference corpora

both cases motivated by register constraints: a realization through an infinitive clause, which might have been possible in German as well (e.g. (2)a *Eine regionale, multilaterale Bestrebung, um die Volkswirtschaften der westlichen Hemisphäre zu vereinen*) is not characteristic of German, whereas the PPs, which are chosen in the translations (in the *b* variants in (2) and (3) below), are typical register features of German political essays.

(2) a. *a regional, multilateral effort to unite the economies of the Western Hemisphere* (EO_ESSAY)

b. *Eine regionale, multilaterale Bestrebung für die Vereinigung der Volkswirtschaften der westlichen Hemisphäre* (GTrans_ESSAY)

(3) a. *The candidate's effort to communicate with voters* (EO_ESSAY)

b. *den Bemühungen des Kandidaten um eine gute Kommunikation mit den Wählern* (GTrans_ESSAY)

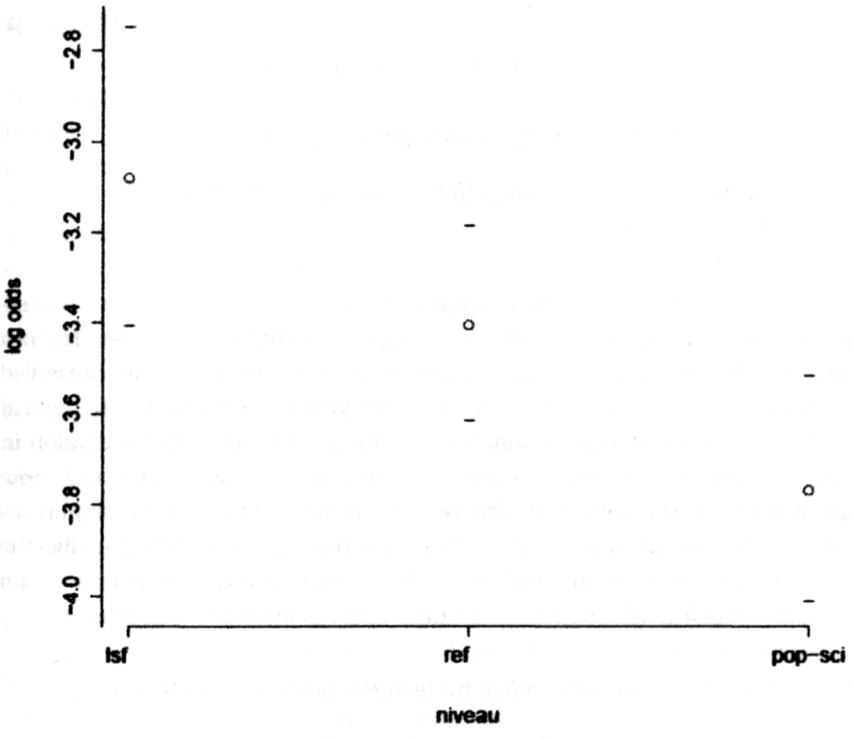

Figure 34: Logistic regression analysis for relative clauses in the German sub-corpora

Example (4) illustrates a similar strategy: the typical English postmodifier participle phrase is translated by a German premodifying participle construction. Again, the result is a more metaphorical structure in German. This shift is typologically motivated and thus obligatory because the preservation of the English structure is not possible in German. In consequence, the source and target structures of example (4) adhere to the norms of the respective languages.

(4) a. *the lines indicating support for two candidates* (EO_ESSAY)

b. *die Unterstützung für die beiden Kandidaten anzeigenden Linien* (GTrans_ESSAY)

Similarly, in example (5) and (6), adjectival English structures are again translated by premodifying participle phrases, which are typical of German. This normalization goes hand in hand with de-metaphorization since the German premodifications are more explicit, albeit more difficult to process due to their left-branching constituent structure.

(5) a. *rules-based trade* (EO_ESSAY)

b. *ein auf Regeln basierender Handel* (GTrans_ESSAY)

(6) a. *the majority of anti-trade arguments* (EO_ESSAY)

b. *die Mehrheit der gegen den Handel vorgebrachten Argumente* (GTrans_ESSAY)

Referring back to Table 26, examples (7) and (8) are illustrations of over-normalization, i.e. an exaggerated use of typical modifiers in German. In both cases English adjectival premodifiers, which could as well have been preserved in the target texts, are translated into postmodifying genitives, which are typical of the German register. Once again, these shifts illustrate de-metaphorization in translation, this time by the less metaphorical and thus more direct and more explicit encoding of entities in the German post-modifying nominal phrases compared to the English pre-modifying adjectival phrases. Additionally, the German partitive articles in genitive case here explicitate the semantic relation they have to their heads, where the English versions leave that implicit.

(7) a. *the U.S. commitment to free trade in this new era* (EO_ESSAY)

b. *das Engagement der Vereinigten Staaten für freien Handel in diesem Zeitalter* (GTrans_ESSAY)

(8) a. *votes among hard-core nonvoters* (EO_ESSAY)

b. *die Stimmen des harten Kerns der Nichtwähler* (GTrans_ESSAY)

Finally, it is important to keep in mind that some of the shifts reported in here are obligatory (e.g. example (4)) and some of them are optional (e.g. example (2)). Obligatory shifts are caused by systemically-based contrastive and in that sense typological differences between the source and the target language. Optional shifts are most commonly due to different register preferences of the languages involved, although they may just as frequently be due to requirements of language-specific information distribution, and in that sense typologically contrastive as well. And they may finally be due to other processing factors. Distinctive motivations of the translation shifts identified through the independent variables language typology and register constraints are only possible by using reference corpora as a tertium comparationis as has been done in this study.

4 Conclusion and outlook

This chapter attempted to quantify contrastive differences in the lexico-grammatical realization of the nominal group in English and German. The concept of grammatical metaphor was chosen as a means of describing this phenomenon. The statistical analysis of the parallel, comparable as well as reference corpora allows the explanation of the dependent variables against the independent ones. More specifically, contrastive features for English and German as well as typical register features for both languages can be detected. Furthermore, the analysis of the comparable corpus showed that in most cases the translations adhere to the typical norms of the target language. This results in (de-/re-) metaphorization of the translations, and accordingly in explicitation/implicitation, and in making the texts grammatically and informationally more or less dense, all of this depending on the direction of the processes just identified.

Future work involves the classification and annotation of translation shifts according to different degrees of grammatical metaphoricity. Such a multiply aligned and annotated parallel corpus can, on the one hand, serve as an empirical basis in translation studies – e.g. to investigate translation shifts, their sources within the translation process and their effects on the translation product. However, this kind of research has to be complemented by process-oriented, online research, i.e. psycholinguistic experiments (cf. for instance Hansen 2003 or Alves et al. 2010). On the other hand, such a database of grammatical translation shifts can also be used in translation practice and education – e.g. as a linguistically enriched translation memory that provides translation solutions for typological and register-specific translation problems – or for the development of transfer rules in machine translation.