Chapter 1

Predicting cognate translation

Silvia Hansen-Schirra

Jean Nitzke

Katharina Oster

FTSK Germersheim, Johannes-Gutenberg-Universtität Mainz

Empirically-based translation research has so far been developed within two major self-standing approaches: corpus-based work on properties of translated texts or translation universals (product) and experimental studies of translators' expert performance (process). Recently, advances in corpus architecture and multi-level corpus querying are combined with methods from psycholinguistics and cognitive science in order to determine predictors for translation candidate probabilities, which in turn may range from free to literal translation solutions. In the corpus-based realm, free translations lead to normalization effects, whereas literal ones trigger shining-through. Speaking from a cognitive point of view, shining-through can be related to the literal translation hypothesis, while normalization may occur due to monitoring processes.

This paper investigates the conditions under which cognates are translated into more literal or free translation candidates. Some of the influential factors are text internal (e.g. context) or external (e.g. language status); others are translation inherent, such as the expertise of the translator and the translation mode. The former are discussed from a product-based perspective, the latter are analyzed in a more process-oriented manner. Multi-method approaches including translation corpora and experimental data are used for predicting the probability of cognate variation in translation. As a consequence, the predictors are discussed against the background of the monitor model.

Silvia Hansen-Schirra, Jean Nitzke & Katharina Oster. Predicting cognate translation. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting*, 3–22. Berlin: Language Science Press. DOI:10.5281/zenodo.1090944

1 Cognition meets translation constraints

Toury (1995) identifies two laws of translational behavior: he explains that there is a law of growing standardization, i.e., that "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of (more) habitual options offered by a target repertoire" (Toury 1995: 268). However, Toury also suggests that translators tend to produce a translated utterance not by retrieving the target language via their own linguistic knowledge, but directly from the source utterance itself. The universality of discourse transfer is expressed through another translational law, the law of interference: "in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text" (Toury 1995: 275).

From a corpus-based perspective, the first law is also reflected in Baker (1996) universal feature of normalization: Normalization (or conservatism) means that translators tend to conform to the typical patterns of the target language or even to exaggerate their use. This universal feature also includes the tendency to normalize marked and ungrammatical structures. But if the status of the source language is significantly higher than the status of the target language (for example, English compared with other languages in the field of software), normalization in translations is weakened or the opposite tendency might even be observed. If this is the case, the typical patterns of the source language are still visible in the translations, which Teich (2003) calls shining-through.

The continuum between foreignization and domestication is also reflected in the choice of literal vs. more or less free translation strategies and procedures as well as formal vs. dynamic equivalence (Vinay & Darbelnet 1995; Newmark 1988). However, Tirkkonen-Condit (2005b) argues that literal translation is a default translation procedure, which is cognitively preferred to others. Chesterman (2011) and Halverson (2015) reintroduce the concept of literal translation, assuming that entrenchment effects strengthen the co-activation of linguistic patterns and thus reduce the cognitive load during translation for literal renderings (see Schaeffer & Carl (2014) for an empirical operationalization).

From a cognitive perspective, literal translation can be explained by the priming effect. When a translator reads a source text element, a specific element in the target language is primed due to close memory links. It can then be more easily produced than other translation solutions. These close memory links might exist on different linguistic levels. Elements of similar form, similar word class and similar meaning have strong links across language borders.

The monitor model was proposed by Tirkkonen-Condit (2005a). She assumes

that translators follow a predefined translation root, which is the easiest way to translate a text. But they constantly monitor production and as soon as a problem is encountered in this default translation root, they stop the literal translation process and try to find a better solution. This model has been tested by Carl & Dragsted (2012).

The continuum between monitoring and priming/literal translation could be another way to perceive Toury's laws of standardization and interference. The monitor model, however, still exhibits some shortcomings. It is, for example, not precise enough to determine which factors influence priming. As priming might exist on several linguistic levels, what determines its strength? Finding answers to these questions and thus creating a more elaborate monitor model could help to predict translational behavior.

For this purpose, we will investigate cognates (translation equivalents which share a similar form). Several studies have shown that the number of cognates in translations varies significantly depending on other factors such as language status of the respective languages (Vintar & Hansen-Schirra 2005) and translation mode (Oster 2017 [this volume]). Cognates are relatively easy to manage in experimental settings and can be investigated in many language pairs. We thus believe that they are a good basis for the investigation of the different priming roots.

In the following, we will examine different factors that might influence the production of cognates. Some are text internal, such as context or external such as language status of the respective languages, as well as historical developments. These constraints will be investigated from a product-based perspective. However, other factors are translation inherent, such as the expertise of the translator and the translation mode, which will be analyzed from a more process-oriented perspective. We will show how the translation of cognates can be predicted within the context of the different constraints and finally discuss how the predictors can be implemented into the monitor model.

2 Cultural-political predictors

Our hypothesis is that cultural-political predictors influence translation choices. In the following, we introduce two external factors that predict translation behavior: language status and socio-historical influences.

2.1 Language status

The first study deals with two language pairs for which we assume that the relation between the source and target languages and cultures differ: English-German and English-Slovene. Since 1945, German has seemed to be susceptible to influences from the English language (Carstensen 1965). In contrast, Slovene is less influenced and exhibits language protectionism on a political level (Vintar & Hansen-Schirra 2005).

The results discussed here were published in Vintar & Hansen-Schirra (2005), which includes English-German and English-Slovene translations as well as German and Slovene original comparable texts. The authors fully automatically extracted the cognate pairs from the parallel corpora compiled for the study from popular scientific texts using an implementation of the Levenshtein's edit distance algorithm in the Perl String::Approx module (for details see ibid.). The original comparable texts were used as a tertium comparationis for the cognate frequencies.

For the comparison of the cognate frequencies, a parallel English-German and English-Slovene subcorpus and a comparable German and Slovene subcorpus were created. These had to be as comparable as possible in terms of corpus size and register. For this reason, all subcorpora comprised 10,000 tokens of popular scientific texts. Following Biber (1995), each subcorpus was composed of ten text samples consisting of roughly 1000 tokens. This guarantees that the sub-corpora is as well-balanced as possible. The COSMAS corpus was used as a monolingual reference corpora for German, and the FIDA was used for Slovene (Vintar & Hansen-Schirra 2005).

The comparison of the cognate frequencies in Slovene and German translations and Slovene and German originals shows that, in general, German has more cognates than Slovene, and more specifically German translations exhibit the highest cognate frequency (see 1; $\chi^2 = 60.33, df = 1, p > .001$).

Table 1: Cognate frequencies normalized to a corpus size of 10,000 words

	Slovene	German
Original	254	356
Translation	189	652

These results illustrate that German is more susceptible to cognate use than Slovene, and this is even more prominent in translations. However, a contrary tendency can be observed for Slovene translations which have fewer cognates than Slovene original texts. This might be interpreted as a slight aversion towards the use of cognates in Slovene translations.

On the one hand, it can be said that the context of a word is very important for the choice between cognate and native word. For instance, the English word *action* was not only translated with its Slovene cognate *akcija*, but a series of non-cognate translations (*delovanje, tehnika, ukrepanje, aktivnost, izvedba, operacija, udejstvovanje*) could also be found in the corpus depending on the context of the word. On the other hand, repetitions in translations are avoided by using the cognate as well as the native words for stylistic purposes (e.g. English *volcanic activity*, German *vulkanische Aktivität*, *vulkanische Tätigkeit*, *vulkanische Ausbrüche*, *vulkanische Bewegung*).

Nevertheless, it seems that German is more receptive to the use of cognates than Slovene. The preference of cognates in German might be explained by two different tendencies: first, it might mirror the use of Anglicisms in German, which in turn reflects the strong influence English nowadays has on the German language (especially as lingua franca of science, Ammon 2001). On the other hand, the cognate use might be an indicator of the susceptibility of the German language towards internationalisms rooted in a common etymological history (Braun et al. 2003). In contrast, it might be the case that Slovene as a 'minor language' tries to avoid foreign language material by using only native words to protect itself from language change. The tendency for or against cognates might therefore be related to the overall language – and translation – policy in the target society. Thus, avoiding cognates might be a strategy of linguistic purism and protectionism.

2.1.1 Socio-historical influences

Social-historical factors might influence the use of cognates, as well. In the following, we will compare the development of cognates in different languages over the course of time with a bottom-up methodology using the Google Books *Ngram Viewer*.[1] This tool shows the frequency of words and phrases used in the selected book corpora and over the course of the selected years (between 1500 and 2008).

[1] https://books.google.com/ngrams, last accessed 13th August 2016

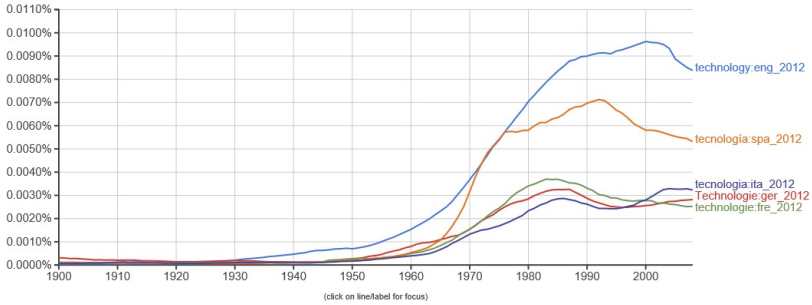

Figure 1: Diachronic development of *technology* and its cognate versions in German, Spanish, Italian, and French from 1900 to 2008.

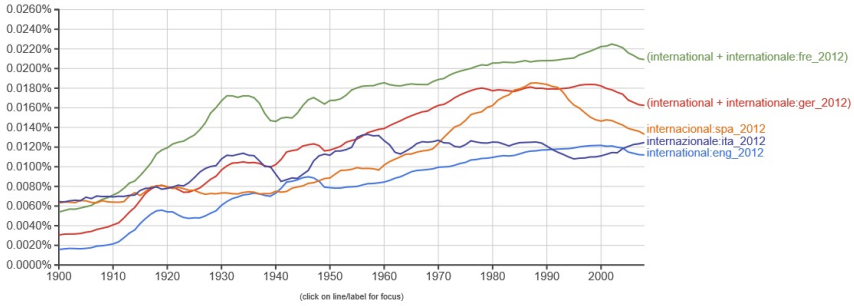

Figure 2: Diachronic development of *international* and its cognate versions in German, Spanish, Italian, and French from 1900 to 2008.

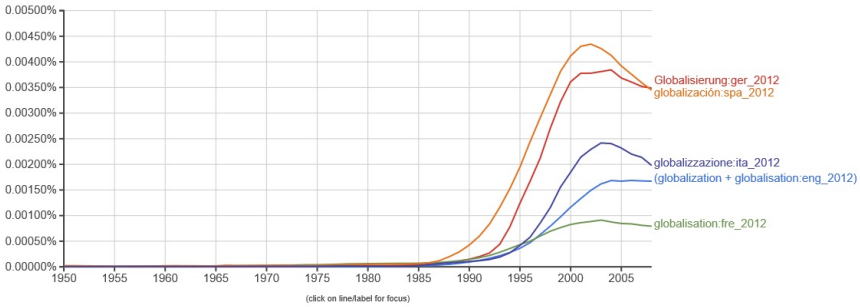

Figure 3: Diachronic development of *globalization (globalisation)* and its cognate versions in German, Spanish, Italian, and French from 1950 to 2008.

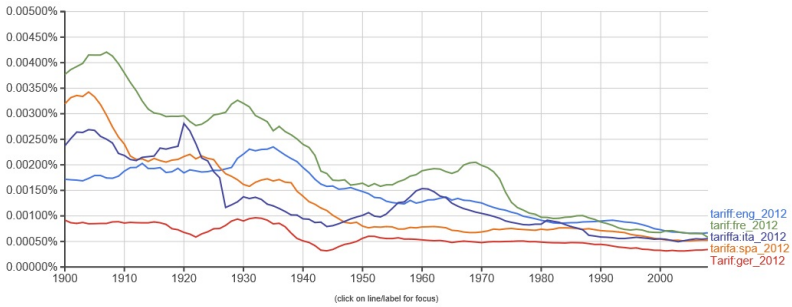

Figure 4: Diachronic development of *tariff* and its cognate versions in German, Spanish, Italian, and French from 1900 to 2008.

Figures 1-4 show the diachronic development of four cognate words in five different languages (English, French, German, Italian, Spanish) from 1900 to 2008 (apart from *globalization* – Figure 3 – because the word did not occur in the first half of the 20th century). All figures show similar developments of the presented words in the different languages over the course of time.

Technology and its multilingual cognate representations (Figure 1) hardly occurred in the corpora before the mid-60s, when the frequency of the words started to increase rapidly for the next decades. Although the term *technology* has existed since 1910 in the English language and originates from the Greek *tekhnologia*, the term *high technology* was coined only in 1964, which might also characterize the beginning of this linguistic development.[2]

The use of *international* and its multilingual cognate representations (Figure 2) increases steadily, but is not bound to a specific date or event. This indicates that international relations and economics – well known social developments – have become more important in our societies in the last century and hence affected the languages as well. In contrast to *technology*, *international* has English roots and was coined by the English social philosopher and solicitor J. Bentham.[3] However, the components of *international* (*inter*[4] and *national*[5]) have Latin roots, a language that influenced all examined languages. Hence, this might have promoted the inclusion and acceptance of the English word in the other languages.

[2] http://www.etymonline.com/index.php?allowed_in_frame=0&search=technology, last accessed 13th August 2016

[3] http://www.duden.de/rechtschreibung/international and http://www.etymonline.com/index.php?allowed_in_frame=0&search=international, last accessed 13th August 2016

[4] http://www.duden.de/rechtschreibung/inter_, last accessed 13th August 2016

[5] http://dwds.de/?view=1&qu=national, last accessed 13th August 2016

Globalization and its equivalents (Figure 3) show a similar development to *technology*, but the increase is more rapid and much later. The word globalization only emerged in 1961, although the verb *globalize* was first recorded in 1953, but not in the sense that refers to global economic systems.[6] Here, we can observe an interesting finding since the German and Spanish cognates appeared more frequently and earlier in time. This development cannot be attributed to the influence of English as lingua franca but rather to the fact that this internationalism derived from the Latin word "globus". This clearly shows that common etymological roots might trigger cognate usage as well.

In contrast to the other example, the use of *tariff* and its cognates decreases in the last decade in all five languages, albeit to different degrees. This might be caused by a restriction of meaning because the word *tariff* used to have an extended meaning, namely "prices" in general, whereas today it is mainly used within the context of taxes and wages.[7]

The examples discussed here indicate that the usage of cognates varies according to societal and technological development. The word might have popped up in one language, but due to common language roots it might be more easily accepted in other languages as well. Furthermore, language change, like extending or narrowing down the meaning of a word may also have an influence (Koselleck 1979).

3 Linguistic predictors

3.1 Linguistic context

The context, in which the words are embedded, is a very important factor for translation and translation choices – a phenomenon also known as intra-lingual communication. A *table* can, for example, be either furniture or a chart and the context in which the word is used usually clearly specifies which table is meant. We hypothesize that cognates are more frequently translated with a cognate when the translators are asked to translate a single word than when the cognate is integrated in a complete text.

To test this hypothesis, we ran a study with 67 participants, who had to translate single words in a list (with information on the word class) and a complete text

[6] http://www.etymonline.com/index.php?term=globalize&allowed_in_frame=0, last accessed 13th August 2016

[7] http://dwds.de/?qu=Tarif, last accessed 13th August 2016

that contained numerous cognates.[8] Both settings contained the same cognates. For the study, two political texts were chosen (190 and 186 words, respectively). A total of 20 cognates were isolated in each text and used to compose the cognate list. The participants were German native speakers who studied English and translation and were asked to translate one word list and one text. In addition, we set a time limit of three minutes for the list and 14 minutes for the texts, because we wanted the participants to first prepare a translation draft to ensure that they used the words first activated in their mental lexicon. The results are presented in Table 2.[9]

Table 2: Percentage of translations with cognates, with non-cognates, or no translation at all depending on an existing context

	Cognate	non-cognate	no translation
without context	57,39	32,24	10,37
with context	37,27	54,91	7,82

While cognates in the list are translated as cognates in over 57% of the cases, they were only translated with cognates in around 37% when they were presented in context. The picture is reversed for non-cognates translations (32% without context, 55% with context). In some instances, the translators were not able to produce a translation or chose to omit the word in the target text.

If we compare the translations of the same word with and without context, different patterns can be observed: Some words were translated by most participants with a cognate in the list condition, but were translated less often with a cognate in the text condition. For example, *priorities* was translated with a cognate in 93.6% of cases when it was only presented as a single word, or it was not translated at all (no participant translated the word with a non-cognate). In the text condition, however, *priorities* was translated as a cognate in only 52.9 % of cases, and 41.2 % of the participants chose a non-cognate translation. As another example, *shield* was mainly translated as a cognate (80.6 %) in the list condition

[8] The experiments in Section 3.1 and 4.1 were carried out at the FTSK. Translation students participated during a lecture in the different experiments. Since the experiments were part of their course, they did not receive any further credit for participation. The participants were informed that the results were treated anonymously and that they were only used for scientific purposes. The students were further informed that their participation had no influence on their grades and that they could withdraw from the experiment at any time.

[9] Thanks to Jan Skawski and Kai Schuhmacher who conducted the experiment and came up with first results in the context of a seminar paper.

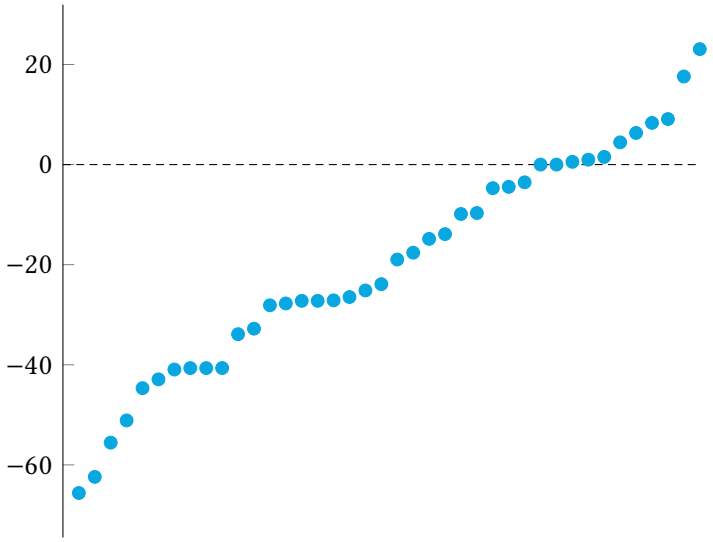

Figure 5: Distribution of decline (below zero) and increase (above zero) of cognate translation with context.

and never as a non-cognate, but it was only translated as a cognate in a quarter of the cases in the condition with context and as a non-cognate in 60 %. A point in Figure 5 represents one word of our texts/lists and the ratio of its decrease or increase (in percent) when translated in context compared to the single word translation. There were also instances for which it was the other way around (see Figure 5 and 6). For example, *diversity* was hardly translated with its cognate in the list task (3.6%), but the frequency increased considerably in the text task (26.3%). However, this is rather the exception than the rule, as can be seen in Figure 6, which shows how often the cognate use radically increased (> 10%), only slightly changed (±10%), or radically decreased (> 10%).

The analysis shows that the use of cognates in translations is dependent on the context of the translation. In general, the participants chose a cognate less frequently, when they were translating a whole text than when they only had to find German equivalents in a word list. This might indicate that the cognate translation is the "safest" without context, because the cognate is not only similar in meaning, but also in form. When a cognate is embedded in context, however, the translators are more secure about which translation choice to select.

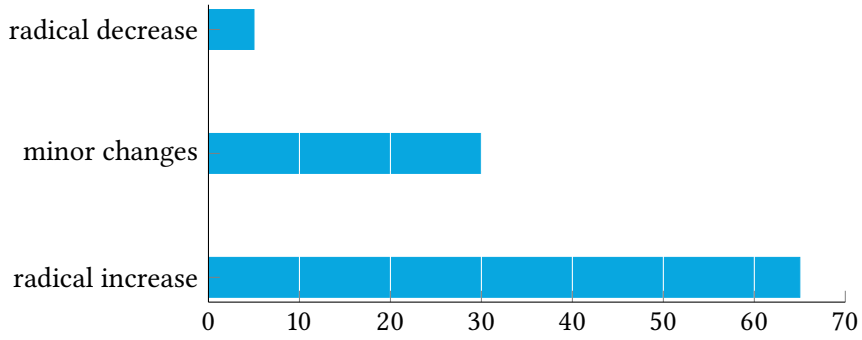

Figure 6: Change in translation strategy with context in percentage

3.2 Text type

As shown in the preceding section, context has an influence on cognate use. But why would e.g. *diversity* be translated more often as a cognate in a political text than in a list of single words? We assume that this behavior was triggered by the text type. Maybe the participants thought that the use of the cognate translation is more natural in the political context, although they are aware of a non-cognate alternative. Hence, we hypothesize that text types influence the use of cognates.

In the following, we used the statistics component of the online tool *DWDS*[10] to observe the intralingual influence of different text types on the use of cognates. We used the following pairs of cognates and non-cognates, and compared them for two different text types, namely newspapers (NP) vs. academic texts (AT). We chose the following example because we assumed that they might be used differently in the two text types. Further, we wanted to cover different word classes[11]:

- *komplex* (cognate), *kompliziert* (cognate) vs. *schwierig* (non-cognate)
- *original* (cognate) vs. *echt* (non-cognate)
- *publizieren/ Publikation* (cognate) vs. *veröffentlichen/Veröffentlichung* (non-cognate)
- *Maschine* (cognate), *Apparat* (cognate) vs. *Gerät* (non-cognate)

[10] „Digitales Wörterbuch der deutschen Sprache“ (Digital Dictionary of the German language), www.dwds.de

[11] We chose the most frequent non-cognates of the translation test in Section 3.1 to come up with these pairs. We neglected translations which only occurred once or twice.

- *spezifisch* (cognate), *charakteristisch* (cognate), *typisch* (cognate) vs. *besonders* (non-cognate), *deutlich* (non-cognate)

The results in Figure 7 show that, in general, there is no clear preference for cognates or non-cognates. However, when comparing different text types, we can see that cognates are preferred in academic texts compared to newspapers for the same cognate/non-cognate pair. This holds true for all our examples displayed in Figure 7, although the difference for the pair *komplex*, *kompliziert* (cognates) vs. *schwierig* (non-cognate) is only very small.

The interpretation of these results may be twofold:

First, it is possible to assume that German academic writing might be influenced by the lingua franca of science, which is English (Ammon 2001). Language contact might result in a higher frequency of Anglicisms, internationalisms and cognates in German academic writing. In addition, academic texts convey a high frequency of technical terms such as Latinisms, Grecisms and Anglicisms (Braun et al. 2003). At same time, these are the roots of cognates because they have typically been introduced into and established in different languages and language families.

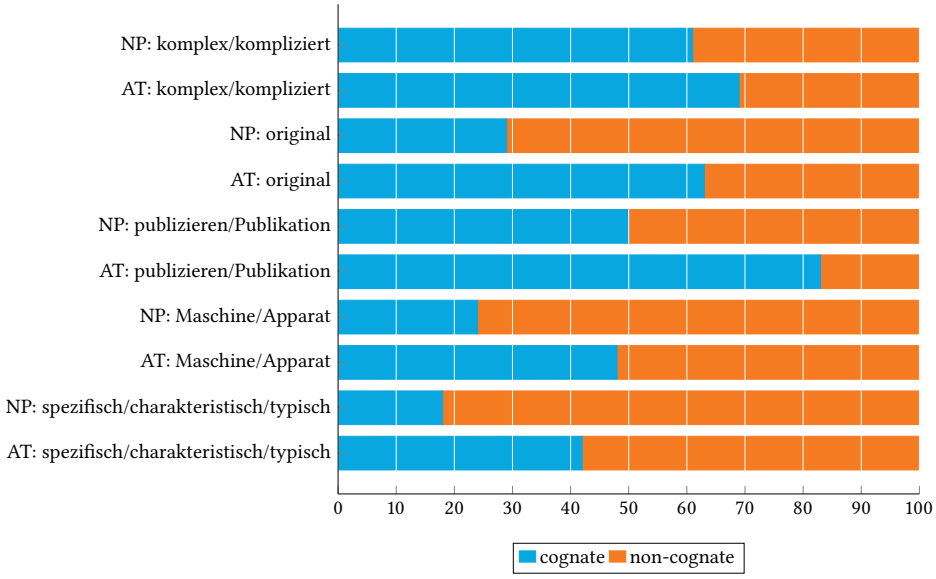

Figure 7: Examples for cognates and non-cognates in academic texts (AT) vs. newspapers (NP)

Secondly, the preference for non-cognates in newspaper texts might reflect a protectionary strategy of journalists towards their own language. They try to avoid cognates, which commonly have their routes in foreign languages, in favor of German synonyms (Liesem 2014). At the same time, shining-through effects of English constructions or internationalisms can also be found in popular-scientific texts translated from English to German (Hansen-Schirra et al. 2012) conveying a certain degree of technicality, which might be comparable to the academic text type under investigation.

In summary, typical preferences in terms of cognate usage can be identified for different text types. Further, we assume that a more in depth study might complete the picture. It seems, for example, reasonable that legal or technical texts – or in general very domain-specific texts – contain more cognates than newspaper texts or other general language texts.

4 Translation-inherent predictors

In the last part of the paper, we investigate characteristics of translators and the translation environments that might influence cognate use. These predictors can again be characterized as external.

4.1 Expertise

In the following study, we investigated whether cognate production changes during the translators' training. As Vandepitte et al. (2015) showed with respect to metonymic language, translation competence influences processing time and translation strategies. It can therefore be assumed that translation competence might also have an impact on cognate translation: with increasing translation experience, cognates might be used more consciously, because the translator is more aware of the potential meaning. If training and experience influence the number of cognates in translations, we take the factor experience as a variable for the processing of cognates in the translator's mind.

In total, 43 students of the FTSK in Germersheim participated in the experiment. They were all German native speakers and students of English. The text was taken from a news platform.[12] It dealt with home affairs in the United States[13] and was shortened in order to obtain a higher cognate density. The final text was

[12] http://www.foxnews.com/, last accessed 13th August 2016

[13] http://www.foxnews.com/politics/2013/03/03/obama-to-nominate-walmart-sylvia-matthews-burwell-for-budget-chief.html, last accessed 13th August 2016

187 words long and contained 49 English-German cognates which were analyzed in the target texts. The students translated the text in a lecture at the FTSK (see footnote 8).

We counted the number of cases in which participants decided to translate a source language cognate with a target language cognate. The number of cognates in the translations correlated significantly with the number of semesters (see also Figure 8): $r(41) = -0.42, p = 0.005$.

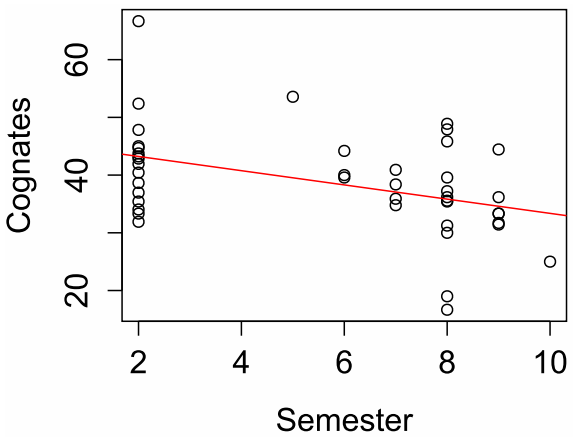

Figure 8: Usage of cognate correlates with expertise

These results suggest that a mechanism in the translator's mind develops during the translator training. This could be the mental lexicon, since it was shown that new words can also be easily learned in adulthood, and connections can be strengthened or weakened in its network-like structure (Aitchison 2012). But the reason could also be due to increased monitoring (see Oster 2017 [this volume] for the impact of monitoring and mental lexicon on the lexis of the target text). However, several studies concluded that monitoring does not develop anymore after childhood (Wiersema et al. 2007). It depends, however, on the mental resources available: motivation (Ganushchak & Schiller 2008) and time pressure (Ganushchak & Schiller 2006).

Our hypothesis is thus that the mental lexicon changes. It is reorganized; the connections between non-cognates become stronger since cognates are constantly filtered out by the monitoring process. Monitoring itself does not change. But as the translator needs less mental resources to activate non-cognates (their threshold is lowered over time), more mental resources are available for monitoring. This means that monitoring becomes stronger in translation tasks but not in

general settings. We have to keep in mind, however, that the results might not only be due to the translator training but also to increased expertise in the respective languages. This expertise goes hand in hand with the expertise in translation. But it could be worth investigating this factor in future studies.

4.2 Computer-aided translation

In the last decades, translation technologies have become more and more important as they make translations more consistent and the process more efficient. Translation memory systems and software for terminology management have been developed and established in most translation environments. A recent trend is the post-editing of a machine translated source text "by a human translator according to specific guidelines and quality criteria". (O'Brien 2011: 197) In this study, we hypothesize that the processing mode in which the translation is produced influences cognate use. We therefore compare human translation output and post-edited output. We hypothesize that machine translation generates more cognate translations and that the translator tends to adhere to the machine translation.

The experiments are part of the CRITT-TPR database[14] that collects translation process data for different tasks and in different languages. A total of 24 participants took part in the study used for this analysis: twelve professional and twelve semi-professional translators (students of the university with only little professional work experience). The texts were newspaper articles and sociology-related texts with different complexity levels. The length of the texts varies between 100 and 148 words. The participants were asked to translate two texts from scratch, post-edit two machine translated texts and monolingually edit two machine translated texts – from English to German respectively. For this study, we only looked at the post-edited and human translated target texts.

The tasks were conducted in Translog II,[15], a program used for recording mouse activity, key strokes and gaze data with the help of the Tobii eye-tracker, which also records the sessions, mouse activity, key-strokes and gaze data in Tobii Studio. There were no time restrictions and the participants could use the Internet freely as a research tool.

We determined the cognates from the source texts (58 cognates in all six source texts – some occurred more than once in one text or in a few texts) and extracted

[14] https://sites.google.com/site/centretranslationinnovation/tpr-db, last accessed 13th August 2016

[15] https://sites.google.com/site/centretranslationinnovation/translog-ii last accessed 13th August 2016

the realizations of these cognates in the MT output and in the target texts (human translation and post-editing). We differentiated between non-cognate and cognate translations. Further, we analyzed the varieties in the cognate realizations in the different translation modes.

Table 3 and 4 present the results of the cognate analysis. While Table 3 presents total numbers (e. g. 321 cognates were realized with a cognate translation in the translation from scratch mode), Table 4 shows the amount of variation in the different translations modes, independent of how often they occurred. Let us specify the counting procedure for Table 4 with some examples:

- The English cognate *motive* was realized as *Motiv* both in the translation from scratch and in the post-editing tasks. Hence, it was counted as TfS – Cognate: 1; TfS – Non-Cognate: 0; PE – Cognate: 1; PE – Non-Cognate: 0.
- The cognate *minimized* was realized as *minimieren*, *reduzieren*, *gering halten*, and *verringern* in the translation from scratch tasks and as *Minimierung*, *minimeren*, *Reduzierung*, and *Reduktion* in the post-editing tasks. It was counted as TfS – Cognate: 1; TfS – Non-Cognate: 3; PE – Cognate: 2; PE – Non-Cognate: 2.
- The cognate *analysts* was realized as *Analysten*, *Analytiker*, *Analysen*, and *Finanzexperten* in the translation from scratch tasks and as *Analysten* and *Experten* in the post-editing tasks. It was counted as TfS – Cognate: 3; TfS – Non-Cognate: 1; PE – Cognate: 1; PE – Non-Cognate: 1.

Table 3: Translation of Cognates in translation from scratch (TfS) and post-editing (PE) task.

	Cognate	Non-Cognate
TfS	321	127
PE	325	118

Table 3 shows that the distribution of English cognates realized as the German cognate-equivalent is quite similar in both translation modes: 71.7% in the translation from scratch task and 73.4% in the post-editing task. The chi-square test did not show significant differences between the two translation modes and the cognate realization: $\chi^2 = 0.2471, df = 1, p = 0.62$.

Table 4: Variations in translation from scratch (TfS) and post-editing (PE) task

	Cognate	Non-Cognate
TfS	59	91
PE	50	49

In the next step, we examined the variety in which the cognates were translated. While cognate variety is quite similar, the difference is remarkable in non-cognate variety. For the whole set-up, the chi-square test did not prove significance between the two translation modes and cognate realization: $\chi^2 = 2.59, df = 1, p = 0.11$. Next, we conducted Wilcoxon rank sum tests (the data was not distributed normally) for the differences in the variation in the cognate group and in the non-cognate group. The test did not prove significant for the cognate group ($W = 1883, p = 0.19$), but significant for the non-cognate group ($W = 2157.5, p = 0.005$).

Translations from scratch and post-edited target texts show a similar cognate and non-cognate usage, which is not in line with our hypothesis. By implication, this indicates that post-editing and human translation are very similar in this aspect. The machine translated cognate was not changed in 88.3% of instances (391 of 443) in the post-editing task. Interestingly, 67.9% (301 of 443) of the human translated cognates were congruent with the machine translation output. Hence, we assume that cognate/non-cognate translations are chosen in statistical MT system quite similar to human translation. The variety within non-cognate choices, however, is statistically higher in translations from scratch than in post-edited texts. When we take a closer look at the data, it turns out that the participants choose the MT in 87%of cases, and only 11% changed the MT. This explains why there is much more variety in human translations than in post-editing.

5 Enhancing the monitor model with translation predictors

The predictors presented in this study are not exclusive. Other translation-inherent constraints that influence the usage of cognates in translation can be skopos, time constraints, translation mode (Oster 2017; Gieshoff 2017 [this volume]), etc.

The results suggest that different mechanisms are responsible for the translation of cognates. When considering, for example, Levelt's speech production model (1989) as a basis for the processing of language during translation, the translation of words in general can be influenced by different steps. During the conceptualization phase, speakers adapt messages according to cultural and pragmatic norms. During formulation, the lexical selection in the mental lexicon can be primed by context and can depend on expertise.

When considering the translation of cognates, we can assume that according to the literal translation hypothesis (Halverson 2015), the translator always chooses the easiest path (the cognate translation). However, when considering cultural predictors for cognate translation, specific cultural norms are present at a translator's conceptual level causing monitoring (Tirkkonen-Condit 2005a). The same holds true for pragmatics. On a lexical level, the context pre-activates certain words (cognates or non-cognates). It causes thus less processing effort for the translator to choose the co-activated words than to look for alternatives. The mechanisms of controlling lexical choices might change with expertise according to Halverson (2015) gravitational pull hypothesis and thus lead to more pre-activation of non-cognates in experienced translators.

The findings related to the translation of cognates suggest that different priming roots exist and that the monitor model proposed by Tirkkonen-Condit should be adapted to these findings. The studies we presented are, however, pilot studies which were conducted in very natural settings. If we want to further explore the predictors of translations, we will need to conduct more controlled experiments in order to isolate different factors. However, the studies we presented in this paper can provide an overview of the different processes that might be involved. Future research might also consider other linguistic aspects such as syntax or pragmatics, and investigate how these features can be influenced by different conditions. This might help us to predict how a certain translator will translate a text in a certain situation.

References

Aitchison, Jean. 2012. *Words in the mind: An introduction to the mental lexicon.* Oxford: Wiley-Blackwell.

Ammon, Ulrich. 2001. *The dominance of english as a language of science: Effects on other languages and language communities.* Berlin: Walter De Gruyter.

Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager*, 175–186. Amsterdam: Benjamins.

Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Braun, Peter, Burkhard Schaeder & Johannes Volmert. 2003. *Internationalismen: Studien zur interlingualen Lexikologie und Lexikographie, band 2*. Tübingen: Niemeyer.

Carl, Michael & Barbara Dragsted. 2012. Inside the monitor model: Processes of default and challenged translation production. *Translation: Corpora, Computation, Cognition* 2(1). 127–145.

Carstensen, Broder. 1965. Englische Einflüsse auf die deutsche Sprache nach 1945. *Jahrbuch für Amerikastudien* (Heft 13).

Chesterman, Andrew. 2011. Reflections on the literal translation hypothesis. In Cecilia Alvstad, Adelina Hild & Elisabet Tiselius (eds.), *Methods and strategies of process research: Integrative approaches in translation studies*, 23–35. Amsterdam: Benjamins.

Ganushchak, Lesya Y. & Niels O. Schiller. 2006. Effects of time pressure on verbal self-monitoring: An ERP study. *Brain Research* 1125. 104–115.

Ganushchak, Lesya Y. & Niels O. Schiller. 2008. Motivation and semantic context affect brain error-monitoring activity: An event-related potential study. *NeuroImage* 39. 395–405.

Gieshoff, Anne Catherine. 2017. Audiovisual speech decreases the number of cognate translations in simultaneous interpreting. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting* (Translation and Multilingual Natural Language Processing 7), 313–330. Berlin: Language Science Press. DOI:10.5281/zenodo.1090978

Halverson, Sandra L. 2015. Cognitive translation studies and the merging of empirical paradigms: The case of 'literal translation'. *Translation Spaces* 4(2). 310–340.

Hansen-Schirra, Silvia, Stella Neumann & Erich Steiner. 2012. *Cross-linguistic corpora for the study of translations: Insights from the language pair english-german*. Berlin, New York: de Gruyter.

Koselleck, Reinhart. 1979. Historische Semantik und Begriffsgeschichte. In *Sprache und Geschichte*, vol. 1. Stuttgart: Klett-Cotta.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge: MIT Press.

Liesem, Kerstin. 2014. *Professionelles Schreiben für den Journalismus*. Wiesbaden: Springer.

Newmark, Peter. 1988. *A textbook of translation*. New York: Prentice Hall.

O'Brien, Sharon. 2011. Towards predicting post-editing productivity. *Machine Translation* 25(3). 197–215.

Oster, Katharina. 2017. The influence of self-monitoring on the translation of cognates. In Silvia Hansen-Schirra, Oliver Czulo & Sascha Hofmann (eds.), *Empirical modelling of translation and interpreting* (Translation and Multilingual Natural Language Processing 7), 23–39. Berlin: Language Science Press. DOI:10.5281/zenodo.1090948

Schaeffer, Moritz & Michael Carl. 2014. Measuring the cognitive effort of literal translation processes. In Ulrich Germann (ed.), *Workshop on humans and computer-assisted translation*, 29–37. Association for Computational Linguistics.

Teich, Elke. 2003. *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin: de Gruyter.

Tirkkonen-Condit, Sonja. 2005a. The monitor model revisited: Evidence from process research. *Translators' Journal* 50(2). 405–414.

Tirkkonen-Condit, Sonja. 2005b. The monitor model revisited: Evidence from process research. *Translators' Journal* 50(2). 405–414.

Toury, Guideon. 1995. *Descriptive translation studies and beyond*. Amsterdam: Benjamins.

Vandepitte, Sonia, Robert J. Hartsuiker & Eva Van Assche. 2015. Process and text studies of a translation problem. In Aline Ferreira & John W. Schwieter (eds.), *Psycholinguistic and cognitive inquiries into translation and interpreting*, 127–144. Amsterdam, Philadelphia: Benjamins.

Vinay, Jean-Paul & Jean Darbelnet. 1995. *Comparative stylistics of french and english: A methodology for translation*. Amsterdam, Philadelphia: Benjamins.

Vintar, Špela & Silvia Hansen-Schirra. 2005. Cognates: Free rides, false friends or stylistic devices? A corpus-based comparative study. In Geoff Barnbrook, Pernilla Danielsson & Michaela Mahlberg (eds.), *Meaningful texts: The extraction of semantic information from monolingual and multilingual corpora*, 208–221. London/New York: Continuum.

Wiersema, Jan R., Jaap J. van der Meere & Herbert Roeyers. 2007. Developmental changes in error monitoring: An event-related potential study. *Neuropsychologia* 45. 1649–1657.