# Chapter 3

# German light verb construction in the course of the development of machine translation

Shaimaa Marzouk
Johannes Gutenberg University Mainz

The German light verb construction (LVC) is commonly used despite its relative complexity. Different writing guidelines recommend avoiding LVCs and replacing them with the base verb constructions (BVCs). However, since not every LVC has an equivalent BVC, replacement is not always possible. The present study addresses two aspects: first, how the machine translation (MT) of LVC has evolved in light of recent progress in MT and the increasing dominance of neural machine translation (NMT), and second, whether the use of BVCs improves MT output compared to LVCs. The analysis of the MT output of both scenarios, LVC and BVC, is performed for different MT approaches in terms of number and types of MT errors, style and content quality ratings, and scores from two automatic evaluation metrics (AEMs). For this, a mixed-methods triangulation approach that includes error annotation, human evaluation, and automatic evaluation was applied and five MT systems were examined: a rule-based system (RBMT), a statistical system (SMT), two differently constructed hybrid systems (HMT), and a neural system (NMT). The study is conducted for the language pair German-to-English in the technical domain. The results show that systems that employ earlier MT approaches (RBMT, SMT, HMT) benefited from replacing the LVC with the corresponding BVC as their output was improved (i.e., MT errors were reduced; quality and AEMs scores were increased). On the contrary, the NMT system was able to produce MT with minimal number of errors both for LVCs and BVCs and recorded the highest quality levels in both scenarios among the analyzed MT systems.

Shaimaa Marzouk. 2021. German light verb construction in the course of the development of machine translation. In Tra&Co Group (ed.), *Translation, interpreting, cognition: The way out of the box*, 47–66. Berlin: Language Science Press. DOI: 10.5281/zenodo.4545033

## 1 Introduction

The German term *Funktionsverbgefüge* was coined by von Polenz (1963: 26); the English counterpart, *function verb constructions* or *light verb constructions*, hereafter LVCs, goes back to the linguist Jespersen (1942: 117). With this, Jespersen (ibid.) distinguishes between a light verb and a heavy verb (a.k.a. *full verb*, a verb that emphasizes the full meaning). Some examples of LVCs are: *eine Frage stellen* ('to pose a question'), where *stellen* is a light verb, as opposed to *fragen* ('to ask'), which acts as a full verb; *eine Handlung ausführen* ('to perform an action') to replace the full verb *handeln* ('to act'); *etw. zu Papier bringen* ('to put sth. on paper') instead of *schreiben* ('to write'), and *eine Entscheidung treffen* ('to make a decision') instead of *sich entscheiden* ('to decide').

As illustrated by these examples, an LVC is simply a combination of a verb and a noun that can only be correctly understood with both components. Strictly speaking, it is a complex predicate that consists of a semantically light verb and a deverbal noun (Jespersen 1964: 117). The verb in the LVC acts merely as a functional element, letting the noun represent the main predicate (Grimshaw & Mester 1988). The LVC is not just found in German, but in many other languages as well. In English, *make a decision* is sometimes used instead of *decide*. Similarly, in Arabic, one might say *yakhudh qraraan* ('make a decision') or *yuqrīr* ('decide').[1] Both variants also exist in Spanish *tomar una decisión* ('make a decision') and *decidir* ('decide').

The present study focuses on German LVCs that can take on one of the following forms: a verb plus a noun in the accusative case (e.g., *eine Handlung ausführen*) or a verb plus a prepositional phrase (e.g., *zu Papier bringen*). German LVCs are used predominantly in technical, scientific, legal, and official texts (Bruker 2013: 38f.), but despite their widespread use, they are criticized both in linguistics and translation. In linguistics, they are seen as a sign of "Umschreibungssucht" (addiction to reformulating) and "Verbaphobie" (verbaphobia) (Daniels 1963: 9f.) and have been described as "unnecessarily complicated" and "inelegant" (Storrer 2006). Because of the relative complexity of LVCs, several Controlled Language varieties and writing guidelines prompt writers to avoid them: (1) The rule "Avoid light verb constructions" is found in Leichte Sprache (Easy German Language), which is increasingly being applied to simplify legal, political, and administrative texts for people with low language skills or cognitive limitations (Hansen-Schirra & Gutermuth 2018). Here, the rule is included to reduce sentence complexity (Bredel & Maaß 2016). (2) The rule is also applied in Controlled

[1]The Arabic examples were transliterated by https://de.glosbe.com/transliteration/Arabic-Latin.

Languages used in technical documentation in order to keep sentences more concrete and direct (Gesellschaft für Technische Kommunikation, Tekom e.V. 2013: 107). (3) The same rule is present in the guidelines from the weekly German magazine *Die Zeit* entitled *Recommendations for prospective journalists*, which advise journalists to use full verbs instead of LVCs, as full verbs are usually clearer and more efficient (Die Zeit 2007).

However, despite their structural complexity, LVCs are widely used. This is partially due to the fact that some LVCs completely lack any equivalent, e.g., *in Ordnung halten* ('to keep in order'). Other LVCs have a more nuanced meaning that can be difficult to express using the base verb construction, hereafter BVC. One example of this kind of LVC is *eine Maschine in Betrieb setzen* ('to put a machine into operation'); this is a process that usually includes several different procedures depending on the complexity of the machine and is therefore much more than simply (p. 107) *eine Maschine einschalten* ('to turn on a machine') (Baumert & Verhein-Jarren 2012). Concretely, the LVCs can influence meaning in four ways, which are known as "action types" (Zifonun, Hoffmann & Strecker 1997: 704):

*Causative:* emphasize the initiator of an action, e.g., *der Starter setzt den Motor in Gang* ('the starter sets the engine in motion').

*Inchoative:* mark the beginning of an action, e.g., *endlich geht das Buch in Druck* ('finally, the book goes to press').

*Durative:* emphasize the duration of an action, e.g., *ein neues Modell ist bereits in Arbeit* ('a new model is already in production').

*Passive:* form a distinct passive meaning variation, e.g., *die neue Methode findet Anwendung bei dem Versiegelungsprozess* ('the new method is applied in the sealing process').

## 2 Machine translation of LVCs

As discussed, the usage of the LVCs can be indispensable in conveying a distinct nuance of meaning or because there is no BVC equivalent. Despite the existence of LVCs in several languages, there are a number of difficulties in MT of LVCs. Heine (2017) describes LVCs as "a typical example of phenomena that are neither explainable with (exclusively) grammatical rules nor lexical units" and how the sentence syntax as well as the lexical components of the MT system are decisive for an error-free MT output. Therefore, depending on the complexity of the sentence syntax and the MT system approach as well as the system capacity, the

primary challenge of translating an LVC is that the parser must first identify it as such. The system needs to be able to distinguish between *stellen* as a full verb, as in *etwas auf den Tisch stellen* ('to put something on the table') as opposed to *stellen* as a light verb in *etwas zur Verfügung stellen* ('to make something available'), *eine Frage stellen* ('to ask a question'), or *etwas in Rechnung stellen* ('to invoice something').

After identifying the LVC in the source language, a transfer problem between the source and target language may appear. Depending on the language pair, the LVC might be (best) translated using an equivalent LVC, a different function verb, a BVC, or a completely different construction (Bruker 2013: 96), e.g., translating *zur Verfügung stehen* as 'are available'. In addition, a syntactic translation problem may arise while translating LVCs with prepositional phrases that have no articles or with a preposition that can require different cases, such as *in*, *an*, *auf*, or *unter*, e.g., *in Betrieb nehmen* (accusative) vs. *in Betrieb bleiben* (dative) or *in Verhandlungen treten* (accusative) vs. *in Verhandlungen stehen* (dative). Such cases cannot be strictly morphologically differentiated. The correct case for the output of the syntax information for the nominal phrase in the LVC can probably only be determined by using appropriate lexicon entries for the function verb, as the verb selects the case of the prepositional phrase. (ibid.: 75) Another potential problem can be encountered on a morpho-syntactic level in processing LVCs that include compounds, e.g., *Verstellung vornehmen* in *Höhenverstellung vornehmen* or *Behandlung durchführen* in *Fleckenbehandlung durchführen*. In such cases, the LVC with the compound must first be morpho-syntactically analyzed and broken down into its component parts (Winhart 2005). For this, an exact semantic analysis of the compound is required for a correct processing of the LVC (Bruker 2013: 97).

The difficulties in the MT of LVCs as well as their frequent use in the German language make its relevance for Natural Language Processing evident. Nonetheless, the LVC has not yet received the attention it needs in computational linguistics, particularly in MT research. There is a number of linguistic studies that closely investigate the linguistic differences between LVCs and BVCs on the basis of corpora (Glatz 2006; Storrer 2007; 2006). Storrer (2006) shows how the influence of both constructions goes beyond their different pragmatic and stylistic impacts. Others investigated the properties of multiword predicates and developed automatic methods for distinguishing among literal, metaphorical, and idiomatic multiword predicates (Fazly & Stevenson 2005). North (2005) examined the productivity of LVCs that include predicative nouns and developed computational measures for quantifying the acceptability of LVCs. Kuhn (1994) analyzed how

the HPSG-based translation approach[2] handles LVCs. Marzouk & Hansen-Schirra (2019) analyzed the impact of avoiding LVCs among other German Controlled Language rules on machine translatability across different MT approaches and found out that the NMT system delivers in comparison to RBMT, SMT and hybrid MT systems mostly error-free output both before and after the application of the rules showing even a decrease in quality after applying the rules. Further studies on MT of German LVCs across different MT approaches including the NMT have not yet been conducted, to the best of my knowledge. In light of the proven linguistic differences between LVCs and BVCs (Glatz 2006; Storrer 2007; 2006) and the success achieved by NMT in improving MT output compared to earlier approaches (Bentivogli et al. 2016; Marzouk & Hansen-Schirra 2019; Popović 2018; Toral & Sánchez-Cartagena 2017), this study aims to track MT's progress in translating LVCs.

The remainder of this paper is organized as follows: Section 3 provides an overview of the empirical study including the dataset and the MT systems used. Section 4 outlines the methodology applied. Results are presented in Section 5 followed by a discussion in Section 6. Finally, Section 7 provides the conclusion, mentions the limitations of the study, and gives an overview of future work.

## 3 Description of study

The study analyzes two aspects of MT with regard to LVCs: (1) to what degree different MT approaches are able to translate LVCs and (2) whether the use of BVCs improves the MT compared to using LVCs. In the analysis, MT outputs of LVCs and BVCs are contrasted across four MT approaches, and the impact of each construction is measured in terms of number and types of MT errors, style and content quality ratings, and AEMs scores. The examined MT approaches are represented by five MT systems: Google Translate (an NMT system), Lucy LT KWIK Translator (an RBMT system), SDL Free Translation (an SMT system), and Bing by Microsoft and Systran (two differently constructed HMT systems).[3] The selection criteria of the systems were (1) to be an online freely available system, (2) to offer the language pair German-to-English, and (3) to cover different MT approaches.

For the analysis, a test suite was constructed that consists of 24 source sentences extracted from a corpus of German technical user manuals using the Con-

[2] HPSG: Head-driven Phrase Structure Grammar (Pollard & Sag 1994) was considered the best available grammar formalism at that time.

[3] The MT step was performed at the end of 2016. At that time, Bing became an HMT system by adding language-specific rule components to its original SMT system, and Systran was also developed from an RBMT system into a hybrid system.

trolled Language checker CLAT (Rösener 2010).[4] The 24 analyzed LVCs were as follows: twelve accusative LVCs and twelve prepositional LVCs. After reformulating the sentences using the BVCs, both versions (LVCs and BVCs) were machine translated into English using the aforementioned five MT systems, resulting in a dataset of 240 MT sentences (24 source sentences × 2 versions × 5 systems). In the source sentences, company-specific and specialist terms were replaced with common terms (e.g. *Gerät* instead of *Feinstzerkleinerer*; *Steckdose* instead of *Schutzkontaktsteckdose*). This modification was necessary for two reasons: (1) the MT systems used in the study were not trained in advance with specific relevant corpora; (2) to avoid human evaluators investing too much time investigating the translation of these types of uncommon terms during the human evaluation. Source sentences that included more than two specific terms were excluded entirely from the analysis to avoid application of multiple changes to the original source sentences.

## 4 Methodology

A mixed methods triangulation approach was applied that incorporates three evaluation methods: error annotation, human evaluation, and automatic evaluation. The analyses were conducted in a black box context, as the focus is on the comparison of the MT outputs of the LVC and BVC scenarios (and not on the internal processes of the systems). In the following, the analyses are demonstrated in detail.

### 4.1 Error annotation

The goal of the error annotation is to identify the MT errors in the use of LVCs (LVC scenario) and BVCs (BVC scenario) and compare them in terms of their number and type. The annotation was conducted by a qualified experienced German–English translator and checked by two professional German–English translators. Further, based on the existence or non-existence of MT errors, the data were divided into four groups, referred to as "annotation groups". These are: FF (for false-false): translation contains error(s) in both scenarios; FR (for false-right): translation contains error(s) only in the LVC scenario; RF (for right-false): translation contains error(s) only in the BVC scenario; RR (for right-right): no errors

[4] CLAT is one of the most well-known Controlled Language checkers in Germany developed by the society for the promotion of applied information sciences (IAI) at Saarland University; see: http://www.iai-sb.de/de/produkte/clat.

in either scenario. The error classification applied is mainly based on Vilar et al. (2006) and encompasses the error types shown in Table 3.1.[5]

Table 3.1: Error classification applied in the annotation

| Category | No. | Type |
|---|---|---|
| Orthography | OR.01 | Punctuation error |
| | OR.02 | Capitalization error |
| Lexis | LX.03 | Omission |
| | LX.04 | Addition |
| | LX.05 | Untranslated |
| | LX.06 | Consistency error (a word is repeated in the sentence and translated differently each time) |
| Grammar | GR.07 | Wrong word class |
| | GR.08 | Wrong verb tense / composition / person |
| | GR.09 | Wrong agreement gender / number / person |
| | GR.10 | Wrong word order |
| Semantics | SM.11 | Confusion of sense (output translation is possible, but not in the given context) |
| | SM.12 | Wrong choice (output translation is apparently wrong) |
| | SM.13 | Collocation error |

The error taxonomy of Vilar et al. (2006) was used as a basis for the error annotation due to its explicity, clarity and appropriate degree of granularity. However, further more extensive taxonomies, such as the multidimensional quality metrics (MQM) framework can be also used for the analysis. This would be particularly useful in case of examining fine-grained or more specific types of errors.

## 4.2 Human evaluation

The goal of the human evaluation is to compare the content and style quality of the MT in the LVC and BVC scenarios. Following the quality definition of Hutchins & Somers (1992), the *content quality* is the extent to which the translation reflects

[5] As the analysis of the LVCs and BVCs was part of a large-scale study that aimed to examine different German Controlled Language rules, it was necessary to add two further relevant error types to Vilar et al. (2006)'s taxonomy (capitalization and consistency) and to exclude two error types in Vilar et al. (ibid.) that were irrelevant for the study (idioms and style).

the information in the source text accurately; and the extent to which the translation is easy to understand (ibid.). The *style quality* is the extent to which the translation sounds natural and idiomatic in standard written English, is appropriate for the intention of its content (ibid.) as well as presented clearly in terms of orthography. The definition covers the orthography as an instrument for presenting the content in an adequate way that serves its intention.

Based on these definitions, the content quality (CQ) covers the criteria accuracy and clarity; the style quality (SQ) encompasses the criteria idiomaticity, appropriateness to the content intention as well as correctness and clarity of the orthographic presentation.

The human evaluation Figure 3.1 consisted of (1) evaluating the SQ and CQ of the MT (*) on two 5-point Likert scales; (2) selecting the relevant quality criteria that justify the assigned quality scores: accuracy and clarity under the CQ; idiomacy, appropriateness to the content intention as the content well as correctness and clarity of the orthographic presentation under the SQ; (3) providing the word or part of the translation relevant to each chosen criterion; (4) where many modifications were necessary, the participant had to enter an alternative translation for the whole sentence.

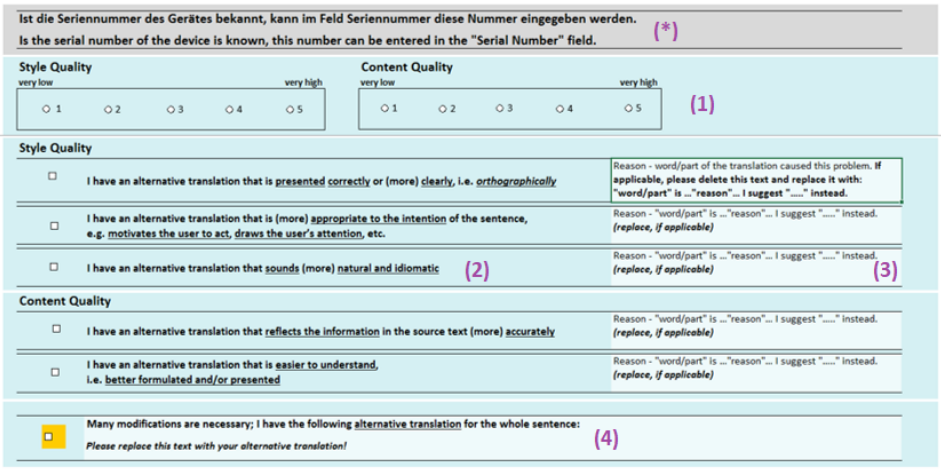


Figure 3.1: Interface of the human evaluation

Concerning the participants, different studies recommend recruiting more than 3–4 participants (Fiederer & O'Brien 2009). In this study, five participants initially conducted the tests and the number of participants was successively increased until the accumulated average of the quality values stabilized. After the eighth participant, the accumulated quality averages hardly changed. Accordingly, the number of participants was not increased anymore. The participants

are native English speakers and hold a bachelor's degree in translation. In addition, all participants were students in the last or penultimate semester of the master's degree program in translation. Participation was remunerated.

Regarding the test procedure, the analysis of the LVCs and BVCs was part of a large-scale study that aimed to examine different German Controlled Language rules (Marzouk & Hansen-Schirra 2019). Within the scope of the study, each participant evaluated in total 1,100 MT sentences that were randomized and split into 44 tests (the analysis of the LVC vs. BVC was a subset of this dataset). Each participant had the opportunity to choose whether to rate one, two or three tests per day, depending on his or her availability. The basic requirement was to evaluate at least one test daily, thus avoiding interruptions that could possibly have a negative effect on the intra-rater agreement. In addition, the participants were asked to take a break between the tests. The 44 tests were sent in a different randomized order to the participants (e.g. the 1st participant received test 40, test 8, test 5 consecutively). A decreasing motivation over a 3–4 week evaluation period is unavoidable. Therefore, this randomization ensured that no particular sentences were evaluated by all participants at the end of the evaluation. The tester received the answered tests every day and checked them for completeness (i.e. all sentences were rated and commented on if necessary). In case of any missing data, the participant was asked to complete them, then he or she received the new tests for the next day.

### 4.3 Automatic evaluation

The alternative translation obtained from the human evaluation acted as a reference translation for the automatic evaluation metrics (AEMs) in order to compare their scores in the LVC and BVC scenarios. Two reference translations per sentence were randomly selected for the comparison. The study applied the evaluation metrics TERbase and hLEPOR. The former is a basic edit distance metric that calculates the minimum number of edits needed to change the evaluated MT so that it exactly matches the reference translation and works without stemming, synonymy lookup and paraphrase support (Snover et al. 2006; Gonzàlez & Giménez 2014). It was necessary to consider the use of synonyms as an edit, as the participants quite often recommended the use of a certain synonym while evaluating the translation accuracy. At the same time, hLEPOR was applied as one of the advanced metrics that has proven to have a state-of-the-art correlation with human evaluation compared with metrics like BLEU, TER, and METEOR among others (Han et al. 2013). The calculation model of hLEPOR is based on three factors: an enhanced length penalty, an *n*-gram position difference penalty and the harmonic mean of precision and recall (ibid.).

# 5 Results

## 5.1 Analysis of the annotation groups (FF, FR, RF, RR) based on the error annotation

Comparing the LVC and BVC scenarios (Figure 3.2) showed that 42% of the sentences were translated correctly in both scenarios (group RR), while half of this percentage (21%) was translated incorrectly in both scenarios (group FF). At the same time, 29% of the sentences were translated incorrectly while using the LVCs and correctly after using the BVCs (group FR). On the other hand, 8% were only translated incorrectly while using the BVCs (group RF).

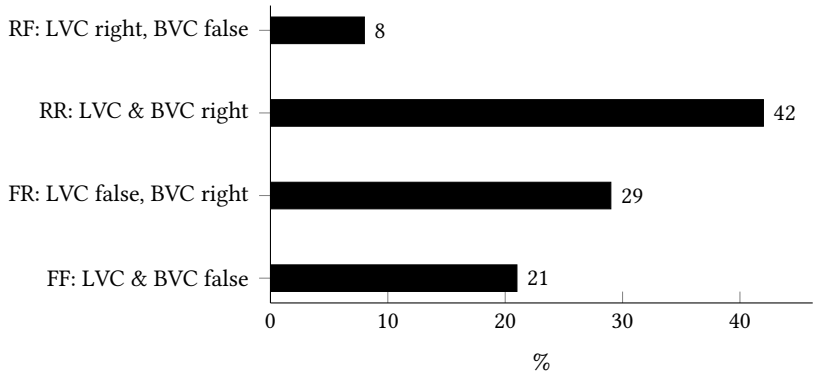


Figure 3.2: Distribution of annotation groups for all the MT systems

Based on the existence and non-existence of MT errors, the impact of using the BVC instead of the LVC on the MT output cannot be considered effectively positive. The only positive impact can be observed in the FR group (false in case of LVC – right in case of BVC). This group amounts to 29%. At the same time, the groups RF and FF together amount to 29%: In RF (right in case of LVC – false in case of BVC), there is a clear negative impact of using the BVC and in FF the usage of the BVC did not help produce an error-free MT.

Considering the groups RR and FF, since the translations were both in the LVC scenario and the BVC scenario correct (RR group) or incorrect (FF group), a positive impact of a certain scenario can only be justified if its quality values in these two groups were higher. In order to explore quality changes in each annotation group, the results of the error annotation and human evaluation were triangulated. The triangulated results showed no significant quality changes in the RR and FF groups. The only significant quality change was in a few cases of the

group FR, indicating that getting an incorrect MT of the LVC and a correct MT of the BVC led to significantly higher quality in case of the BVC.

## 5.2 Analysis of the error types

On the semantic level, the three semantic error types SM.11 confusion of sense, SM.12 wrong choice and SM.13 collocation error were affected in both scenarios. However, a significant change in the number of errors was only observed in error type SM.13 collocation error; this decreased significantly after replacing the LVC with a BVC. Furthermore, in few cases, the grammatical error types GR.08 wrong verb and GR.10 wrong word order and the lexical error type LX.04 addition were differently affected in both scenarios without showing a significant increase or decrease in a certain scenario. The remaining error types were not relevant.

## 5.3 Analysis of the quality changes based on the human and automatic evaluations

Although the analysis of the annotation groups did not reflect a substantial quality increase after using the BVC except for the aforementioned significant quality change in group FR, a significant increase in the MT quality in terms of style and content quality (SQ and CQ) as well as AEMs scores was detected based on the human and automatic evaluations where the BVC was used.

Furthermore, the Spearman test was conducted to investigate the correlation between the difference in the overall quality[6] and the differences in the AEMs scores in both scenarios. The test showed a significant positive strong correlation ($\rho > 0.5, p < 0.001$). Accordingly, the quality changes detected in both analyses (human and automatic evaluation) were in line with each other.

## 5.4 Comparison of the impact of replacing the LVC with the BVC at MT system level

So far, the results show that the MT of BVCs had a significant higher quality in terms of human scores of the SQ and CQ as well as AEMs scores. Subsequently, an analysis at MT system level was conducted in order to explore which MT systems exhibited these higher quality levels. The general positive impact of using BVCs instead of LVCs on the MT output at system level is shown in Figure 3.3 and Figure 3.4.

[6]The overall quality is the mean of SQ and CQ, as analyzing the correlation here requires no distinction between the quality parameters.

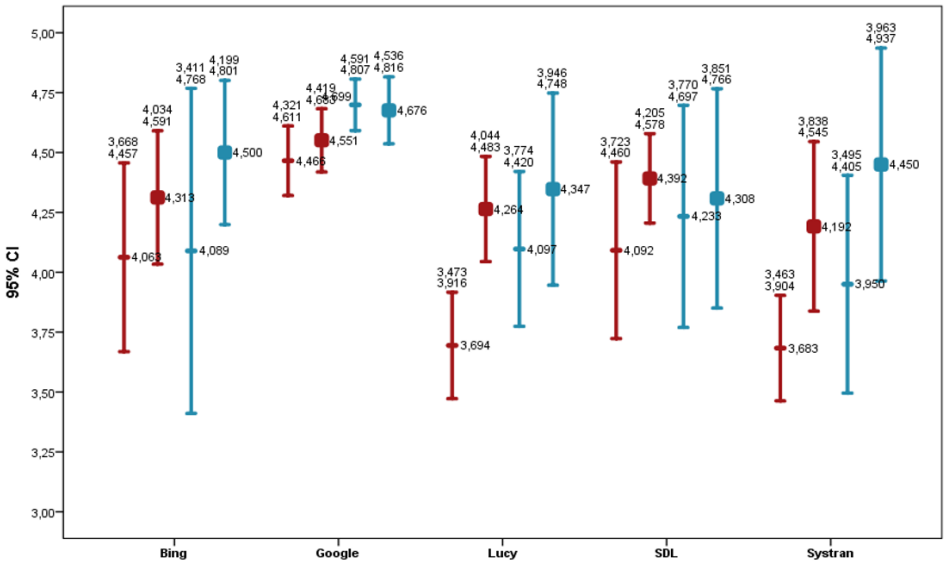


Figure 3.3: Style and content quality in case of using LVCs as opposed to LVCs

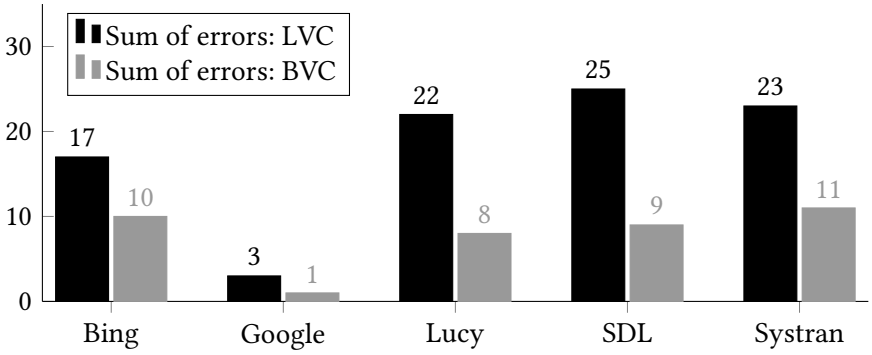


Figure 3.4: Number of MT errors in case of using LVCs as opposed to BVCs

For the RBMT system (Lucy) and one hybrid system (Systran), using the BVC was very advantageous in reducing the number of errors and increasing SQ significantly. In the other hybrid system (Bing) and the SMT system (SDL), the number of errors decreased and the SQ and CQ increased after using the BVC; however, the changes were not significant. The NMT system (Google Translate) showed distinct results: the number of errors was minimal (three errors in the LVC scenario; one error in the BVC scenario). GNMT was able to translate 88% of the sentences in both scenarios correctly, followed by Bing with 46%, and recorded the highest SQ and CQ among all systems in both scenarios as well.

### 5.5 Correlation between the error types and the quality values

The earlier MT approaches showed the following significant strong correlations between a decreased number of errors of the different error types and increased quality values when using a BVC: In Lucy, the decrease in the semantic errors SM.11 confusion of sense and SM.12 wrong choice correlated with the increase in SQ ($\rho = -0.521, p = 0.027$) and CQ ($\rho = -0.537, p = 0.021$) respectively. In Bing, there was a correlation between the error type LX.03 omission and CQ ($\rho = -0.565, p = 0.035$). In SDL, the correlation was observed between each of the error types LX.04 addition and GR.10 wrong word order and the SQ (for LX.04: $\rho = -0.594, p = 0.020$; for GR.10: $\rho = -0.641, p = 0.010$) as well as between each of the error types LX.04 addition and SM.12 wrong choice and the CQ (for LX.04: $\rho = -0.646, p = 0.009$; for SM.12: $\rho = -0.593, p = 0.020$). Finally, in Systran, the error type GR.07 wrong word class correlated with the CQ ($\rho = -0.511, p = 0.018$).

## 6 Discussion

The results show that using BVCs instead of LVCs enhanced the MT of the systems that apply earlier MT approaches (RBMT, SMT, and HMT). It was observed that BVCs simplified the sentence structure and provided an equivalent for German LVCs, which do not have an English counterpart. This section discusses some examples and contrasts the output of the earlier MT approaches with that of the NMT approach in order to gain a deeper insight into the quantitative results.

The first LVC *Höhenverstellung vornehmen* (example 1 in Table 3.2) poses two challenges for MT: including the compound *Höhenverstellung* and having no counterpart for *Verstellung vornehmen* in English. The usage of the BVC *Höhe verstellen* led to breaking down the compound *Höhenverstellung* and solved the collocation problem in English for the RBMT system Lucy. Concretely, it was associated with a correction of the collocation error (SM.13) and thus facilitated producing an error-free MT.

In example 2 in Table 3.3 and example 3 in Table 3.4, the LVCs include prepositional phrases. The LVC *zur Verfügung stehen* is a common German LVC. Although the SMT system SDL was able to parse it correctly, the MT included a wrong word order error (GR.10). The usage of the BVC simplified the sentence structure and was associated with a correction of the word order. The LVC in example 3 (Table 3.4) *zur Anwendung kommen*, on the contrary, is not as common as *zur Verfügung stehen* and was associated with a wrong verb error (GR.08) in the MT of the HMT system Systran. This error was corrected when the BVC was used.

In translating the LVC *zur Verfügung stellen* in example 4 (Table 3.5), the HMT system Bing exhibited semantic and lexical difficulties: a wrong choice error (SM.12) in 'represents' and an addition error (LX.04) in 'available'. Such semantic and lexical errors occur when the system translates the LVC literally (e.g., translating *zur Verfügung stellen* as 'represent available' instead of 'provide'). Using the BVC resolved these MT difficulties and was associated with a correction of both errors.

According to the human evaluation, correcting the MT errors in the abovementioned examples made the translation more appropriate for its intention, more attention-grabbing, and easier to understand, which led to the enhancement of the SQ and CQ.

While systems that apply earlier approaches were not able to identify the LVC in the source language as such in some cases and in other cases faced different transfer problems in the translation from German to English, GNMT was able to overcome these difficulties and handle all the aforementioned MT issues that the other systems encountered. As a result, GNMT produced translations with a minimal number of errors, if any, and recorded the highest SQ and CQ levels both in the LVC and the BVC scenarios.

## 7 Conclusion

The German LVC is a relatively complex construction on both a linguistic and translational level. In this study, I analyzed to which degree different MT approaches (RBMT, SMT, HMT, and NMT) are able to translate the LVC, and whether replacing LVCs with BVCs improves the MT output. The analysis was conducted based on a comparison of the number and types of MT errors, style and content quality ratings, and AEMs scores in the LVC vs. BVC scenario for five MT systems. The study focused on the target language English in the technical domain.

Table 3.2: Example 1. The LVC and BVC are presented in **bold**. *Italic* is used for correct tokens of the translation; underlining for the incorrect tokens.

| | |
|---|---|
| LVC | Die Höhen**verstellung** der Fronten können Sie mittels eines Schraubendrehers **vornehmen**. |
| Lucy | You can **carry out** the height ***adjustment*** of the fronts using a screwdriver. |
| GNMT | You can ***adjust*** the height of the fronts using a screwdriver. |
| BVC | Die **Höhe** der Fronten können Sie mittels eines Schraubendrehers **verstellen**. |
| Lucy | You can ***adjust*** the height of the fronts using a screwdriver. |
| GNMT | The height of the fronts can be ***adjusted*** by means of a screwdriver. |

Table 3.3: Example 2

| | |
|---|---|
| LVC | Auf der Startseite **stehen** die folgenden Funktionen zur Auswahl **zur Verfügung**. |
| SDL | On the Start page, **are** the following functions ***available*** to choose from.. |
| GNMT | The following functions ***are available*** for selection on the start page. |
| BVC | Auf der Startseite **sind** die folgenden Funktionen zur Auswahl **vorhanden**. |
| SDL | On the Start page, the following functions ***are available*** to choose from. |
| GNMT | The following functions ***are available*** for selection on the start page. |

Table 3.4: Example 3. The LVC and BVC are presented in **bold black**. *Italic* is used for correct tokens of the translation; underlining for the incorrect tokens.

| | |
|---|---|
| LVC | Somit kann die Fluggesellschaft nicht garantieren, dass die Gepäckregeln immer **zur Anwendung kommen**. |
| Systran | Thus, the airline cannot guarantee that the baggage rules always **apply**. |
| GNMT | Thus, the airline cannot guarantee that the baggage rules ***are*** always ***applied***. |
| BVC | Somit kann die Fluggesellschaft nicht garantieren, dass die Gepäckregeln immer **angewendet werden**. |
| Systran | Thus, the airline cannot guarantee that the baggage rules ***are*** always ***applied***. |
| GNMT | Thus, the airline cannot guarantee that the baggage rules ***are*** always ***applied***. |

Table 3.5: Example 4. The LVC and BVC are presented in **bold black**. *Italic* is used for correct tokens of the translation; underlining for the incorrect tokens.

| | |
|---|---|
| LVC | Der Navigationsbaum **stellt** alle vorhandenen Seiten der Konfigurierung **zur Verfügung**. |
| Bing | The navigation tree **represents** all existing pages of the configuration **available**. |
| GNMT | The navigation tree ***provides*** all existing pages of the configuration. |
| BVC | Der Navigationsbaum **stellt** alle vorhandenen Seiten der Konfigurierung **bereit**. |
| Bing | The navigation tree ***provides*** all the existing configuration pages. |
| GNMT | The navigation tree ***provides*** all existing pages of the configuration. |

The results of the earlier MT approaches (RBMT, SMT, and HMT) confirmed the complexity of LVCs on a translational level: the MT of LVCs was more error-prone, and the MT quality (SQ, CQ, and AEMs scores) increased with the usage of BVCs. For the RBMT, SMT, and HMT systems, if there were an equivalent BVC for each LVC, the MT problem would be eliminated. However, not all LVCs have an equivalent BVC. In addition, an LVC is, in some cases, needed to express a certain nuanced meaning that the BVC cannot convey as effectively. Since the LVC cannot always be avoided, there is a need to translate it properly. According to the results, the NMT approach provides a capable architecture that can handle the complexity of LVCs: GNMT system was able to translate 88% of the sentences correctly in both the LVC and the BVC scenarios. This was followed by Bing's mere 46%. GNMT system also recorded the highest SQ and CQ values of all systems (> 4.4 out of 5 points) in both scenarios. Therefore, using an NMT system, such as GNMT, allows for the flexibility to choose between LVC and BVC. This, in turn, gives room for the author to prioritize sentence semantics and focus more on the pragmatics.

This study has explored the MT of German LVC for different MT architectures, including NMT, which – to the best of my knowledge – has not yet been examined. However, the following limitations should be mentioned: The study was conducted only for one target language. Although the number of the source sentences was not high, the sentences were translated by five different MT systems, and the MT output was evaluated by eight subjects. In future work, I plan to explore how the NMT architecture tackles further common complex constructions in German based on a corpus analysis of different target languages.

## Abbreviations

| | | | |
|---|---|---|---|
| LVC | light verb construction | SQ | style quality |
| BVC | base verb construction | CQ | content quality |

## References

Baumert, Andreas & Annette Verhein-Jarren. 2012. *Texten für die Technik*. Berlin, Heidelberg: Springer Vieweg. DOI: 10.1007/978-3-662-47410-5.

Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo & Marcello Federico. 2016. Neural versus phrase-based machine translation quality: A case study. In *Proceedings of the 2016 conference on empirical methods in Natural Language Processing*, 257–267. Austin, Texas: Association for Computational Linguistics. DOI: 10.18653/v1/D16-1025.

Bredel, Ursula & Christiane Maaß. 2016. *Leichte Sprache: Theoretische Grundlagen. Orientierung für die Praxis* (Sprache im Blick). Berlin: Dudenverlag.

Bruker, Astrid. 2013. *Funktionsverbgefüge im Deutschen: Computerlexikographische Probleme und Lösungsansätze*. Hamburg: Bachelor Master Publ.

Daniels, Karlheinz. 1963. *Substantivierungstendenzen in der deutschen Gegenwartssprache: Nominaler Ausbau des verbalen Denkkreises* (Sprache und Gemeinschaft). Düsseldorf: Schwann.

Die Zeit. 2007. *Der Weg zum journalistischen Schreiben*. Tech. rep. 4. https://gymwil.files.wordpress.com/2007/10/zeit-journalistentipps5.pdf.

Fazly, Afsaneh & Suzanne Stevenson. 2005. Automatic acquisition of knowledge about multiword predicates. In vol. 1, 31–42. ACL.

Fiederer, Rebecca & Sharon O'Brien. 2009. Quality and machine translation: A realistic objective? *The Journal of Specialised Translation* 11(11). 52–74. http://www.jostrans.org/issue11/art_fiederer_obrien.pdf.

Gesellschaft für Technische Kommunikation, Tekom e.V. 2013. *Leitlinie "Regelbasiertes Schreiben" Deutsch für die Technische Kommunikation*. Tech. rep. Stuttgart.

Glatz, Daniel. 2006. Funktionsverbgefüge – semantische Doubletten von einfachen Verben oder mehr? In Kristel Proost, Gisela Harras & Daniel Glatz (eds.), *Domänen der Lexikalisierung kommunikativer Konzepte*, 129–178. Tübingen: Narr.

Gonzàlez, Meritxell & Jesús Giménez. 2014. *An open toolkit for automatic machine translation (meta-) evaluation*. Tech. rep. http://asiya.lsi.upc.edu/Asiya_technical_manual_v3.0.pdf.

Grimshaw, Jane & Armin Mester. 1988. Light verbs and $\theta$-marking. *Linguistic Inquiry* 19. 205–232.

Han, Aaron Li-feng, Derek F. Wong, Lidia S. Chao, Liangye He, Yi Lu, Junwen Xing & Xiaodong Zeng. 2013. Language-independent model for machine translation evaluation with reinforced factors. In Khalil Sima'an, Mikel L. Forcada, Daniel Grasmick, Heidi Depraetere & Andy Way (eds.), *Proceedings of the XIV machine translation summit*, 215–222.

Hansen-Schirra, Silvia & Silke Gutermuth. 2018. Modellierung und Messung Einfacher und Leichter Sprache. In Susanne Jekat, Martin Kappus & Klaus Schubert (eds.), *Barrieren abbauen, Sprache gestalten*, 7–23. Winterthur: ZHAW Zürcher Hochschule für Angewandte Wissenschaften.

Heine, Antje. 2017. Zwischen Grammatik und Lexikon: Ein forschungsgeschichtlicher Blick auf Funktionsverbgefüge. In *International conference – light verb constructions in Germanic languages*. Brussels: Université Saint-Louis.

Hutchins, William J. & Harold L. Somers. 1992. *An introduction to machine translation*. Cambridge: Academic Press.

Jespersen, Otto. 1942. *A modern English grammar on historical principles*. Part VI. Copenhagen: Ejnar Munksgaard.

Jespersen, Otto. 1964. *Essentials of English grammar*. Alabama: University of Alabama Press.

Kuhn, Jonas. 1994. *Die Behandlung von Funktionsverbgefügen in einem HPSG-basierten Übersetzungsansatz*. University of Stuttgart. (Doctoral dissertation).

Marzouk, Shaimaa & Silvia Hansen-Schirra. 2019. Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation* 33(1-2). 179–203. DOI: 10.1007/s10590-019-09233-w.

North, Ryan. 2005. *Computational measures of the acceptability of light verb constructions*. University of Toronto. (Doctoral dissertation).

Pollard, Carl J. & Ivan A. Sag. 1994. *Head-driven phrase structure grammar* (Studies in contemporary linguistics). Chiacgo: University of Chicago Press.

Popović, Maja. 2018. Language-related issues for NMT and PBMT for English–German and English–Serbian. *Machine Translation* 32(3). 237–253. DOI: 10.1007/s10590-018-9219-5.

Rösener, Christoph. 2010. Computational linguistics in the translator's workflow: Combining authoring tools and translation memory systems. In *Proceedings of the NAACL HLT 2010 workshop on computational linguistics and writing*, 1–6. Association for Computational Linguistics.

Snover, Matthew, Bonnie Dorr, Richard Schwartz, Linnea Micciulla & John Makhoul. 2006. A study of translation edit rate with targeted human annotation. *AMTA 2006: Proceedings of the 7th Conference of the Association for Machine Translation of the Americas: Visions for the Future of Machine Translation* 1. 223–231.

Storrer, Angelika. 2006. Funktionen von Nominalisierungsverbgefügen im Text: Eine corpusbasierte Fallstudie. In Kristel Prost & Edeltraud Winkler (eds.), *Von der Intentionalität zur Bedeutung konventionalisierter Zeichen*, 147–178. Tübingen: Narr.

Storrer, Angelika. 2007. Corpus-based investigations on German support verb constructions. In Christiane Fellbaum (ed.), *Collocations and idioms: Linguistic, lexicographic, and computational aspects*, 164–187. London: Continuum Press. http://www.studiger.tu-dortmund.de/images/Storrer_2007_Corpus-based_investigations_on_german_support-verb_constructions.pdf.

Toral, Antonio & Víctor M. Sánchez-Cartagena. 2017. A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. In *Proceedings of the 15th conference of the European chapter of the Association for Computational Linguistics*, vol. Volume 1: Long papers, 1063–1073. Valencia, Spain: Association for Computational Linguistics.

Vilar, David, Jia Xu, Luis Fernando D'Haro & Hermann Ney. 2006. Error analysis of statistical machine translation output. In 697–702. Genoa, Italy: European Language Resources Association (ELRA).

von Polenz, Peter. 1963. *Funktionsverben im heutigen Deutsch* (Wirkendes Wort). Düsseldorf: Schwann.

Winhart, Heike. 2005. *Funktionsverbgefüge im Deutschen: Zur Verbindung von Verben und Nominalisierungen*. Universität Tübingen. (Doctoral dissertation). https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/46248/pdf/Dissertation-Drive.pdf.

Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin: Walter de Gruyter.