universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Grasping at Straws: Comparing the Quality of Idiom Translation of Neural Machine Translation Systems

verfasst von | submitted by

Tamara Nedic BA BA MA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 070 331 342
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Translation Deutsch Englisch
Betreut von | Supervisor:	Assoz. Prof. Mag. Dr. Dagmar Gromann BSc

This page left blank intentionally

Acknowledgements

First and foremost, I would like to sincerely thank Professor Gromann for her guidance and support throughout my work. Her detailed feedback and dedication have been invaluable, and I am truly grateful for the time she put into supervising this master's thesis.

To my boyfriend, Dalibor, thank you for taking in all my rants with endless patience and understanding, and for standing by my side through it all.

To Katarina and Jelena, for always checking in on my writing progress and for firmly believing this day would come, even when the circumstances suggested otherwise.

A special debt of gratitude is owed to my dear friend, Ozan, for his unwavering support, guidance, and encouragement.

Finally, I dedicate this thesis to my late grandfather, Dragoš, without whom I wouldn't be here today.

To my mother, Slađana, and my brother, Nemanja –
thank you for everything.

Abstract (English)

Idioms are a fascinating linguistic phenomenon because they convey meanings that differ greatly from those of their individual components. Due to their idiomatic nature, translating this group of expressions is challenging, as it is difficult to accurately transfer the meaning of an idiom from one language to another that may not have an exact equivalent. In recent years, machine translation has made significant progress, largely due to the advent of Neural Machine Translation (NMT). Building on these developments, this thesis investigates the performance of state-of-the-art NMT systems when translating idioms. The focus is on Google Translate, DeepL, and SYSTRAN in particular. The analysis is based on a set of idioms that were tested both in and out of context. To evaluate the performance of each NMT system when translating idioms, the selected examples are annotated according to the leading theory on idiom translation strategies. Furthermore, the influence of context on machine translation output is examined, i.e., whether the presence and length of context impact the accuracy of the suggested translation. Using descriptive statistics and Cohen's Kappa to compare the obtained annotations, it is shown that DeepL performs best with regard to idiom translation. Google Translate also performed successfully, while SYSTRAN failed to meet expectations. Furthermore, this master's thesis demonstrates that the presence of context influences the three systems differently. While DeepL recognizes idioms in context and in isolation in a similar way, Google Translate delivers better results when context is given. SYSTRAN showed no difference in results with and without context, indicating an overall very low understanding of idiomatic expressions.

Abstract (Deutsch)

Idiome sind ein faszinierendes sprachliches Phänomen, da sie Bedeutungen vermitteln, die sich stark von denen ihrer einzelnen Bestandteile unterscheiden. Aufgrund ihrer idiomatischen Natur ist die Übersetzung dieser Ausdrücke eine Herausforderung, da es schwierig ist, die Bedeutung eines Idioms genau von einer Sprache in eine andere zu übertragen, in der es möglicherweise keine exakte Entsprechung gibt. In den letzten Jahren hat die maschinelle Übersetzung erhebliche Fortschritte gemacht, was vor allem auf die Einführung neuronaler maschineller Übersetzungssysteme (NMT) zurückzuführen ist. Aufbauend auf diesen Entwicklungen untersucht diese Masterarbeit die Leistungsfähigkeit führender NMT-Systeme bei der Übersetzung von Idiomen. Der Schwerpunkt liegt dabei auf Google Translate, DeepL und SYSTRAN. Die Analyse basiert auf einer Reihe von Idiomen, die sowohl im als auch außerhalb des Kontexts getestet werden. Um die Leistung jedes NMT-Systems bei der Übersetzung von Idiomen zu bewerten, werden die ausgewählten Beispiele anhand der führenden Theorie zu Übersetzungsstrategien für Idiome annotiert. Darüber hinaus wird der Einfluss des Kontexts auf das Ergebnis der maschinellen Übersetzung untersucht, d. h. ob das Vorhandensein und die Länge des Kontexts die Genauigkeit der vorgeschlagenen Übersetzung beeinflussen. Anhand deskriptiver Statistiken und des Cohen-Kappa-Wertes zum Vergleich der erhaltenen Annotationen wird gezeigt, dass DeepL bei der Übersetzung von Idiomen am besten abschneidet. Google Translate schnitt ebenfalls erfolgreich ab, während SYSTRAN die Erwartungen nicht erfüllen konnte. Darüber hinaus zeigt diese Masterarbeit, dass das Vorhandensein von Kontext die drei Systeme unterschiedlich beeinflusst. Während DeepL Idiome im Kontext und isoliert in ähnlicher Weise erkennt, schneidet Google Translate besser ab, wenn Kontext gegeben ist. SYSTRAN zeigte keinen Unterschied in den Ergebnissen mit und ohne Kontext und insgesamt lediglich ein sehr geringes Verständnis von Idiomen.

Table of Contents

List of Tables

1. Introduction

Machine translation (MT) has achieved remarkable progress in recent years, largely due to the advent of Neural Machine Translation (NMT) systems and the increasing availability of extensive multilingual data (Baziotis et al. 2023; Rambelli et al. 2023). A number of neural machine translation systems, such as Google Translate, DeepL or SYSTRAN, can now produce fluent translations in dozens of languages, often capturing both syntax and semantics with impressive accuracy. However, one area that remains particularly challenging is the translation of idiomatic expressions (Dankers et al. 2022; Tian et al. 2023). Idioms are fixed expressions whose meanings cannot be directly inferred from the literal meanings of their constituent words, which makes them especially difficult for computational models to process. When translated literally, idioms often lose their intended meaning, resulting in awkward or misleading translations. Addressing this challenge is crucial, as idiomatic language is pervasive in everyday communication and plays a significant role in conveying cultural nuances. Research into idiom translation thus not only enhances the accuracy of machine translation systems but also strengthens their ability to communicate cultural values across languages. This master's thesis examines how three leading neural machine translation systems, namely Google Translate, DeepL, and SYSTRAN, handle idiomatic expressions. In order to do so, this master's thesis deals with the broader field of machine translation, the concept of idioms, and finally, the specific issue of idiom translation. The first chapter provides an overview of the history of MT, its development, and the notion of (machine) translation quality, followed by an exploration of creative language use in MT. The second chapter examines idioms: their definitions, classifications, and various approaches to their analysis. The third theoretical chapter focuses on idiom translation, the specific challenges it poses, and strategies for handling these complex linguistic phenomena. Based on Keizer's (2016) classification of idioms, this master's thesis evaluates idioms as they occur in sentences from the Corpus of Contemporary American English (COCA) and in isolation, that is, without any context provided. Using descriptive statistics, this investigation seeks to answer the research questions as presented below and assess the extent to which leading MT systems succeed in translating idioms.

1.1 Research Questions and Assumptions

This investigation is guided by the following research questions:

1. How does the performance of different neural machine translation systems vary in translating idioms, as classified by a common framework of idiom translation strategies?

The working assumption is that the performance of the selected systems will differ, at least slightly, depending on the nature and the amount of the training data on which they rely. It is further assumed that specific types of translation strategies will be more frequent and prominent in the machine translation output than others. For instance, the use of another contextually and culturally fitting idiom in the place of the original idiom is expected to be less frequent across the systems than paraphrasing or literal translation.

2. To what extent does the length of context influence the ability of machine translation systems to accurately translate idiomatic expressions?

In this case, it is assumed that providing a longer context will generally lead to a more accurate translation of idiomatic expressions, since additional input provides crucial cues for disambiguating figurative meaning. NMT systems, designed to exploit contextual information, are therefore expected to benefit from sentences that situate idioms within clear semantic frames.

In the course of the analysis, this master's thesis will also examine whether there are differences in MT output depending on the idiom type as proposed by Keizer (2016), since this distinction is based on subtle cognitive nuances.

1.2 Motivation and Objectives

With the rapid advances in artificial intelligence, machine translation has developed at an unprecedented pace. Its accuracy across numerous languages has improved considerably, often in ways that were unthinkable only a decade ago. Nevertheless, idioms remain a particularly challenging aspect of language for both humans and machines. As a result, the problem of idiom translation in MT has not yet received sufficient academic attention, nor has a comprehensive solution been achieved. This thesis aims to contribute to investigating this research problem by systematically analyzing how idioms are translated by state-of-the-art MT systems and whether context and its length influence the translation output.

The expected results of this thesis are potentially of interest to developers of machine translation systems as they explore the ability of such systems to effectively handle highly creative and culture-specific language. The results could also be relevant to translators as they provide an insight into the prevalent translation strategies for idioms in machine translation

systems. From a theoretical perspective, this thesis contributes to the discussion on how idioms can be translated and how the type of idiom and specific context may affect the strategy.

2. Machine Translation

Machine translation was one of the first non-numerical applications of digital computers to emerge following World War II (Kenny 2022). Over the past decades, MT has undergone remarkable transformations, moving from early rule-based experiments to the sophisticated neural architectures used today. This evolution reflects broader developments in both linguistic theory and computational technology, as well as the growing demand for fast and reliable translation in an increasingly globalized world. Kenny (2022: 32) describes machine translation as a process that involves "the automatic production of a target text on the basis of a source-language text". As a form of automated text production, MT draws on insights from computer science, linguistics, and translation studies. To understand the success of MT today, it is important to look at its beginnings. This section therefore briefly touches upon the historical development of MT, outlines the main types of MT approaches, and examines different perspectives on machine translation quality.

2.1. Historical Development of Machine Translation

The origins of machine translation date back to the 1930s (Henisz-Dostert et al. 1979). George Artsrouni was issued a patent certificate in 1933 for a translation machine he called mechanical brain (Henisz-Dostert et al. 1979). Coincidentally, in the same year, Piotr Petrovich Trojanskij, unaware of Artsrouni's patent, envisioned and proposed a comprehensive machine-assisted translation process, focusing on translating words from one language to another (Molosnaja 1961). As Molosnaja (1961) explains, this detailed process consisted of three stages. The first stage involved analyzing the input text, intending to transform the given content into a translatable unit by representing all inflected words in their canonical, i.e., base forms. The second step was to replace the already transformed text with the text of the other language, transformed in the same way as the source text. The third and final stage concerned substituting the transformed text with the adequate form of the target language. According to Trojanskij's vision, the first and the third stages were to be performed by humans, while the machine's task was to carry out the second stage (Molosnaja 1961). Henisz-Dostert et al. (1979) summarize this process by stating that Trojanskij's concept basically consists of analysis, transfer, and synthesis. Although Tronajskij's project was limited to transferring the text from the source to the target language, he firmly believed in the automation of the entire translation process (Henisz-Dostert et al. 1979). However, according to Henisz-Dostert et al. (1979), Trojanski's process was lacking in linguistic proficiency, as he did not pay sufficient attention to the specificities of the

analysis and synthesis stages. Additionally, idiomatic expressions and lexical homonymy, i.e., words having the same spelling or pronunciation but different meanings, were dealt with only superficially. Nonetheless, Trojanskij's ideational contribution was undeniable (Henisz-Dostert et al. 1979). In practice, his idea was an apparatus that comprised a tape traversing a table, featuring equivalent words or units from different languages inscribed in vertical columns. The operational principle was entirely mechanical (Henisz-Dostert et al. 1979). In 1948, Trojanskij wanted to build his machine using an electromagnetic communication device. However, this plan failed due to a lack of technical support and knowledge at the time, which meant that the entire project could not be carried out as envisaged (Henisz-Dostert et al. 1979).

Another significant achievement occurred in 1949 following the publication of Weaver's well-known memorandum, which greatly advanced research in the field of machine translation (Henisz-Dostert et al. 1979). The first chapter of Weaver's (1949: 15) memorandum succinctly explains why it is important to seek solutions to translation problems worldwide and what motivates him to do so:

> There is no need to do more than mention the obvious fact that a multiplicity of language impedes cultural interchange between the peoples of the earth and is a serious deterrent to international understanding. The present memorandum, assuming the validity and importance of this fact, contains some comments and suggestions bearing on the possibility of contributing at least something to the solution of the world-wide translation problem through the use of electronic computers of great capacity, flexibility, and speed. (Weaver 1949: 15)

Weaver's memorandum instigated the emergence of a new field of research. Among the first scholars to engage with Weaver's proposals was Reifler, who, influenced by the memorandum, started the earliest machine translation research at the University of Washington (Henisz-Dostert et al. 1979). According to Henisz-Dostert et al. (1979), the groundbreaking events that impacted the development of machine translation include theoretical texts as well as practical experiments. Booth's (1958) article on the historical background of machine translation represents an example of the former. Booth was merely one of dozens of scholars who reacted to Weaver's memorandum and reflected on the core messages he delivered. Furthermore, Taube (1961) highlighted two aspects of the memorandum in particular. The first point is the dependency of meaning on context. He suggested examining this relationship by determining the extent to which context is necessary to reduce the ambiguity of meaning. In this case, the concept involves identifying the minimal context that imparts any meaningful significance (Taube 1961). The second aspect Taube (1961) addressed concerns Weaver's belief in the potential of computers for machine translation, given their effective implementation in cryptographic work

during World War II. Weaver (1949) believed that there are linguistic universals in all languages, which implies that all languages might have some common characteristics and a logical connecting structure. Based on these beliefs, Weaver (1949) perceived translation itself as a sort of cryptographic problem. According to this assumption, a book written in Chinese is nothing else but a book in English written in the Chinese code (Henisz-Dostert et al. 1979). Since computers had been able to help decrypt messages sent in foreign languages, they should also be able to decode one language's code and transfer it into another, that is, to translate words from one language to another. Such a simplified representation of the translation process as decrypting codes clearly does not hold, since cryptographic codes were unirelational, whereas languages display a wide range of variation on many levels. In essence, Weaver's memorandum put forward the following compelling arguments in favor of machine translation: the importance of facilitating global communication, the use of computers as effective tools to achieve this goal, the significance of context in distinguishing polysemous words, and the pursuit of linguistic universals applicable to all languages (Henisz-Dostert et al. 1979).

A number of experiments were carried out to put the theory into practice. The beginnings of machine translation were rather humble due to the technological constraints at the time (Kenny 2022). Therefore, it was not until 1955 that researchers could establish and demonstrate conclusively the feasibility of machine translation (Henisz-Dostert et al. 1979). According to Henisz-Dostert et al. (1979), this was made possible owing to the well-known Georgetown University IBM machine translation system experiment. During this experiment, more than sixty sentences were translated from Russian into English by a machine translation system developed by Dostert (1955). The experiment set out to illustrate translations based on a small set of 250 test sentences on different topics, ranging from organic chemistry to general subjects. The idea was to demonstrate grammatical and morphological problems and put forward ideas about possible future projects regarding machine translation (Hutchins 2004). Although this experiment was considered groundbreaking, it had several shortcomings. According to Garvin (1972: 46), who demonstrated automatic translations together with Dostert, the limitations of the translation algorithm were multifaceted:

> The search span of the recognition routine was restricted to the immediately adjacent item to the left or right; the command routine was restricted, for selection decisions, to a choice from among two equivalents, for arrangement decisions, to a rearrangement of the translations of two immediately adjacent items. (Garvin 1972: 46)

Further limitations were addressed by Henisz-Dostert et al. (1979), who noted that none of the Russian sentences had negative particles, interrogatives or compound sentences were not taken

into account, and nearly all verbs were formulated in third person (Henisz-Dostert et al. 1979). Actually, both Dostert and Garvin understood the experiment as a first effort and admitted that the demonstration might have been premature (Hutchins 2004). Nevertheless, the experiment aroused expectations that machine systems will be able to produce high-quality translations in the near future.

By the late 1960s and 1970s, machine translation had been primarily applied for defense purposes by organizations closely affiliated with the government and engaged in international affairs (Kenny 2022). It was only years later, in 1977, that machine translation became widely accessible to a large population of internet users with Babel Fish, the first freely available online translation service developed by SYSTRAN for AltaVista (SYSTRAN 2025b). As technology advanced and the internet expanded in the following years, machine translation also followed suit. Google Translate, originally a statistical machine translation service, was launched in 2006 (Och 2006). Ten years later, the next step in improving Google Translate was introduced when the Google Brain Team published a technical report on the implementation of neural machine translation (Wu et al. 2016). Finally, historical developments have led to the ever-growing presence of machine translation, and nowadays, machine translation has become an integral part of intercultural communication, accessible to everyone and easy to use. In order to understand this trend, the following chapter looks at different machine translation systems and elaborates on their development.

2.2. Types of Machine Translation Approaches

Kenny (2022) offers a brief overview of different machine translation approaches, namely Rule-Based Machine Translation (RBMT), Statistical Machine Translation (SMT), and the most recently developed Neural Machine Translation (NMT). RBMT works on the principle of linguistic rules that govern automated translation decisions. In order to make the machine process words, highly skilled linguists create a set of rules for a given language pair. This set of rules usually implies a list of all words in both the source and the target language, together with a set of rules that specify the way the words in question might be combined with other words or structures. The machine then matches the source and target words according to these rules. The stepwise procedure from the source to the target is framed as an algorithm. Okpor (2014) adds that this set of rules encompasses a morphological, syntactic, and semantic analysis of both the source and the target languages in a given translation task. Although this type of machine translation is outdated due to the great amount of manual effort in handcrafting rules, there is still a

small number of systems that are based on this type of machine translation (Kenny 2022). Furthermore, as Kenny (2022) states, the rule-based approach can also be found in hybrid forms with other types of machine translation. Okpor (2014) and Poibeau (2017) offer a narrower classification of RBMT approaches. These encompass direct, transfer-based, and interlingual machine translation approach. According to Okpor (2014), these approaches differ in the depth of the analysis of the source language and the extent to which they attempt to provide a language-independent representation of the meaning between the source and the target language. The direct machine translation approach represents the least prevalent subtype within RBMT. It is distinguished by the feature of executing the translation process at the granularity of individual words. Here, words from the source language are translated directly to the target language without an intermediate representation. In other words, these systems are dictionary-based and involve minimal syntactic and semantic analysis. Transfer-based machine translation approach emerged from the need to develop a system that integrates some kind of syntactic analysis. Going beyond the simplified word-to-word direct machine translation, the transfer-based translation can examine the source language sentence structure. As a result, the translation output is anticipated to exhibit a higher degree of idiomatic expression. Finally, Poibeau (2017: 21) characterizes the interlingual approach as "the most complex and ambitious" as well as "a more or less formal representation of the content to be translated". As the name itself implies, the interlingual approach requires an interlingua in addition to the source and the target language. These systems then transfer the content of the source language to interlingua before translating it from the interlingua to the target language. Okpor (2014) defines this interlingual language as a language that represents content in a neutral manner, i.e., a representation separate from any natural language. Yet, Okpor (2014) also adds that the definition of an interlingua poses a substantial obstacle because a fully neutral language represents a far-reaching goal linguists have not been able to attain.

Unlike RBMT that use hand-crafted rules and bilingual dictionaries, SMT works directly with a large amount of bilingual data and detects regularities regarding equivalences between the given languages by applying statistical analysis (Poibeau 2017). Kenny (2022) outlines two essential components in statistical machine translation: the translation model and the language model. The former represents a bilingual model with three sides: a source language side, a target language side, and the probability score of the given source-target language pair. The latter represents a monolingual model that concerns the target language only and measures the probability of correctly producing a certain sequence of words (Koehn 2009). In other words, the language model analyzes an English sentence and indicates how likely it is that it

represents the target language correctly. The language model not only enables fluent output sequences but also takes word order and word translation into account. For example, a probabilistic language model would suggest the word order 'the house is small' over the incorrect word order 'small the is house' (Koehn 2009: 181). The leading method for language models is *n*-gram language modelling. An *n*-gram is a sequence of *n*-words (Jurafsky & Martin 2025) that provides information about how likely it is that specific words are to follow each other (Koehn 2009). For example, a 2-gram (bigram) is a two-word sequence, e.g., 'the water', while a 3-gram (trigram) represents a three-word sequence of words like 'the water of', etc. (Jurafsky & Martin 2025). Therefore, a trigram language model estimates the probability of a three-word sequence occurring in a corpus by calculating the likelihood of a word given the two preceding words (Kenny 2022). Furthermore, Liu and Zhang (2023) identify three general approaches to SMT: word-based, phrase-based, and syntax-based machine translation. The word-based MT translates words in isolation (Koehn 2009). However, it performs different operations individually for each separate word. The second model, namely the phase-based approach, emerged out of necessity to address the problem of producing a 'word salad' – a translation in which many words are correct, but their order is "a confusing jumble" (Lopez 2008: 8). Thus, phrase-based SMT translates small word sequences, i.e., phrases at a time, as the name itself suggests (Koehn 2009; Lopez 2008). The third model, i.e., the syntax-based MT, addresses the challenge of translating between two syntactically different languages. To tackle this issue, a syntax-based MT approach applies syntactic knowledge within the framework of SMT, focusing on modeling hierarchical syntactic structures (Liu & Zhang 2023). The idea behind syntax-based models is to assign a parallel syntactic tree structure to a pair of sentences in different languages, with the aim of translating them by applying reordering operations to the trees (Liu & Zhang 2023).

While SMT models generate translations using probability and statistical models, the third approach to machine translation, namely NMT, uses artificial neural networks to mimic the human neural network and model the entire translation process (Tan et al. 2020). These neural networks represent an interconnected set of neurons and neuronal layers that recognize patterns in order to translate text from one language to another. The output or activation of these artificial units is dependent on the stimuli they receive from other neurons and the strength of the connection through which these stimuli are transmitted (Forcada 2017). Neural networks, being sets of connected neurons, are basically defined by their behavior (Forcada 2017). Unlike traditional machine translation systems that include various separate, modular components, neural models use an end-to-end model for producing translations (Lamb & Xie 2015). This means that NMT is able to learn the mapping directly from the input text to the associated

output text (Wu et al. 2016). However, in order to do so, NMT systems require very large training corpora and are thus known to be computationally expensive and demanding (Forcada 2017; Wu et al. 2016). NMT architectures were first introduced by Kalchbrenner and Blunsom (2013), Sutskever et al. (2014) and Cho et al. (2014b). These models follow a general paradigm that is based on the encoder-decoder architecture (Lamb & Xie 2015; Tang et al. 2018). An encoder processes the source language input text and performs the process of vectorization, i.e., transforming textual data into a numerical vector representation, enabling the computation of the data. The decoder then interprets the vectorized input and generates the target language text (Sen & Jamwal 2024). These early architectures adopt a fixed-length approach, meaning that the size of the source representation or the input to the decoder is fixed, regardless of the input sentence length (Tan et al. 2020). However, it was found that the translation performance of these early NMT models decreases as the length of the source sentence increases (Sountsov & Sarawagi 2016). Cho et al. (2014a) suggested that this shortcoming is due to the fixed-length source sentence encoding. To circumvent this problem, Bahdanau et al. (2014) introduced the concept of attention. In this context, attention refers to the decoder's ability to place attention on different parts of the source sentence that are meaningful for producing the next token (Stahlberg 2020).

Different architectures have been shown to be effective for building powerful encoders and decoders, which include recurrent, convolutional, and self-attention-based neural networks (Cheng et al. 2022; Stahlberg 2020; Tan et al. 2020; Tang et al. 2018). Recurrent Neural Networks (RNNs) are a type of neural network architecture that is mainly used to detect patterns in a sequence of data (Schmidt 2019). These neural networks pass information through a network with cycles, which allows them to transmit information back into itself. Furthermore, this functionality also enables them to take not only the current input into account, but also previous inputs (Schmidt 2019). Convolutional Neural Networks (CNNs) perform well in visual recognition tasks (Lamb & Xie 2015; Hujon et al. 2023). With CNNs, images go through multiple convolution and pooling layers, where different convolution filters learn to detect different patterns in images (Lamb & Xie 2015). In the context of machine translation, a sentence is modeled as a "matrix by concatenating its word embeddings as columns, and multiple convolution and pooling operations encode regional (*n*-gram) information and increasingly global information about the sentence" (Lamb & Xie 2015: 1). CNNs operate over a fixed-size window of the input sequence which makes simultaneous computation of all features for the source sequence possible (Gehring et al. 2017). According to Gehring et al. (2017), a succession of convolutional layers offers a shorter path for processing relationships between elements of sequences than

RNNs do. Despite these advantages, CNN is still employed less frequently in NMT than the other two dominant approaches (Hujon et al. 2023). The third popular NMT architecture is the self-attention model. Since its introduction in 2015, the concept of attention (Bahdanau et al. 2014) has been used in various NMT architectures (Stahlberg 2020). The application of the concept culminated in 2017, when a new network architecture called the Transformer was introduced (Vaswani et al. 2017). The Transformer architecture is based on an attention mechanism to draw global dependencies between input and output, thus completely dispensing with recurrence and convolution (Vaswani et al. 2017). This attention mechanism is called self-attention. It allows the model to relate different positions of a single sequence in order to compute its representation and capture relationships between elements (Stahlberg 2020). Furthermore, the Transformer uses multi-head attention, which enables the model to "jointly attend to information from different representation subspaces at different positions" (Vaswani et al. 2017: 4).

Until the advent of NMT, SMT was the undisputed state-of-the-art in machine translation (Forcada 2017). For example, in 2006, when Google Translate was first launched, it used phrase-based machine translation as the key approach enabling the translation service (Och 2006). Following the launch of NMT, a significant amount of research has been conducted to compare the translation output of NMT and SMT. The results show that NMT models are superior with regard to the overall translation quality in most settings, thus outperforming SMT (Kenny 2022; Stahlberg 2020). The first successful NMT model was published by Google in 2016, ten years after introducing the phrase-based Google Translate MT system (Wu et al. 2016). The researchers behind the implementation of Google's NMT, also known as GNMT, claim that their system's translation quality "approaches or surpasses all currently published results" (Wu et al. 2016: 19) as well as that GNMT "approaches the accuracy achieved by average bilingual human translators on some of [their] test sets" (Wu et al. 2016: 19). Google's NMT utilizes RNNs to learn the mapping directly from input text to the corresponding output text, addressing many issues associated with NMT architectures known until 2016. These challenges include, among others, massive computational resources and a significant amount of time to train an NMT system on a large-scale translation dataset (Wu et al. 2016). The GNMT model consisted of three components: an encoder, a decoder, and an attention. Four years later, in 2020, Google replaced the original GNMT with a hybrid architecture that combines a transformer encoder with an RNN decoder. This choice is supported by evidence that Transformers are generally more effective than RNNs in machine translation. However, research conducted at Google suggested that many of the identified benefits stem from the Transformer encoder, while the Transformer decoder did not perform significantly better than the RNN decoder, thus

resulting in a hybrid model (Caswell & Liang 2020).

Having provided an overview of the various approaches to machine translation and their underlying technologies, the next section will examine translation assessment from the perspectives of both translation studies and machine translation.

2.3. Translation Quality

To capture the many complexities intertwined with the concept of translation quality, it is beneficial to first define translation and quality separately. Castilho et al. (2018: 10) describe translation as a "complex cognitive, linguistic, social, cultural and technological process". Being a mixture of various components, defining translation seems to pose a true challenge. Freeman (2009) points to the simplified definition of translation as the transfer of meaning between languages, while Sontag (2001: 340) defines this process as "leading something across a gap, to make something go where it was not". The many definitions of translation are closely related to the plethora of different strands of translation, be it the Bible, technical translation, or literary translation (Freeman 2009). Attempts to craft a normative definition of translation, meaning what a good translation should be, add another layer to this variety of different perspectives on translation. Rather than counting and reviewing various definitions of translation, Koby et al. (2014) propose a landscape as a range on which multiple definitions of translation can be placed. This landscape is defined by two axes: the scope and the specifications, which offer two contrasting views on the definition of translation. The former ranges from narrow to broad, while the latter refers to absolute-to-relative specifications (Koby et al. 2014). With regard to the scope, Koby et al.'s (2014) broad definition treats translation as the creation of target content that corresponds to source content. The term content is intentionally used to represent the many different processes that go beyond the traditional text-centric understanding of translation. In this view, translation encompasses both machine and human translation, localization, transcreation, legal, scientific, literary translation, as well as overt and covert translation. Therefore, this definition addresses a broader range of non-textual adaptations that are perceived as translation processes.

By contrast, the narrow definition of translation differentiates between translation *per se* and other activities (Koby et al. 2014). While translation *per se* pertains to transferring a written source text into a written target text that fully conveys the meaning of the text, other activities do not necessarily carry over the full meaning of the source text. Rather, they include activities such as summarization and transcreation. With regard to specifications, Koby et al.

(2014) differentiate between absolute specifications that are to be applied to all translations and relative specifications that vary in accordance with the respective translation task. Within the broad definition of translation, the following parameters can vary according to the task in question: target language, audience, purpose, register, format, layout, style, and content. Conversely, the narrow definition assumes a core set of absolute specifications (Koby et al. 2014).

While some scholars tend to categorize diverse approaches to translation quality into two broader groups, which are equivalence-based approaches and functional approaches (Lauscher 2000; Wang 2017), others (House 2015) group and label them as psychosocial approaches, response-based approaches, and text and discourse-oriented approaches. Castilho et al. (2018) also address equivalence-based and functional approaches, considering descriptive translation quality as well, which House (2015) summarizes under text and discourse-oriented approaches. Additionally, Koby et al. (2014) offer an alternative view on translation quality that stems from a quality management perspective and Garvin's (1984) approaches to quality. The following subchapters incorporate Koby et al.'s (2014) approach to translation quality and follow Castilho et. al's (2018) classification for the remainder of this chapter.

2.3.1. Koby et al.'s Approach to Translation Quality

In order to provide a framework for defining translation quality, Koby et al. (2014) base their theory on quality management rather than relying on linguistic theories. To this end, Koby et al. outline Garvin's (1984) framework, which describes five approaches to determining quality: the transcendent approach, the product-based approach, the user-based approach, the production-based approach, and the value-based approach. Examining translation quality through the lens of quality management, Koby et al. (2014) suggest that Garvin's transcendent and product-based approach may be best represented by the concepts of accuracy and fluency. Accuracy relates to the extent to which the meaning of the original text is conveyed adequately in the target language, and it is typically defined as the extent to which translators succeed in this undertaking. Hence, Koby et al. (2014: 408) define it as "correspondence between the source and the target text". Fluency is commonly defined as the degree to which a translation adheres to the rules and norms of the target language, irrespective of the source or input text (Castilho et al. 2018; Koby et al. 2014). From the perspective of the transcendent approach, translators ought to deliver unmistakably recognizable products, meaning the utmost accuracy and fluency, while acknowledging the fact that these values cannot be measured. Advocates of the product-

based perspective, however, argue that many aspects of accuracy and fluency are to be quantified. Against this background, Castilho et al. (2018) identify the juxtaposition between accuracy and fluency as a commonality in the translation industry in terms of translation quality, which is often characterized by inherent incongruities. In other words, translation quality assessment is predominantly conducted by examining adequacy and fluency. The third, user-based approach, treats text quality according to the end-user's needs. If these needs are not met, the translation is considered to lack quality. Adhering to this perspective, it is the end-user that determines the quality of the translation output. The following production-based approach has, according to Koby et al. (2014), already substantially influenced the translation industry by emphasizing the importance of implementing functional processes in the translation process. This momentum is reflected in the international standard ISO 17100 (2015). Moreover, Castilho et al. (2018) draw attention to ISO 11669, the international standard that offers general guidance for all phases of a translation project (ISO 11669: 2012), which was later withdrawn and replaced by the 2024 edition (ISO 11669: 2024). According to the standard, translation quality is defined as "the degree to which a set of inherent characteristics of an object fulfills requirements", while quality check refers to an "examination of target language content to determine its conformance to translation project specifications" (ISO 11669: 2014). However, it is important to note that not all translations need to be the same in order to qualify as 'good' or 'according to standard'. The fifth and last approach, namely the value-based approach, appears to be a subject of much debate in the field (Koby et al. 2014). According to this approach, the value of products or services arises when they provide increased benefits relative to their costs. As such, products with higher value are considered to be of higher quality. This means that products or services that show the best performance might not be perceived as providing the highest value and thus also not the highest quality.

Expanding on their definition of translation and considering Garvin's (1984) approach, Koby et al. (2018) propose the following definition of a high-quality translation:

> A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture. (Koby et al. 2018: 417)

This quote illustrates the difficulty of providing a comprehensive definition of translation quality without taking into account its many complexities. These complexities, in turn, make the concept difficult to operationalize and measure. Due to the variety of approaches, perspectives,

and categories relating to translation quality, Drugan (2013) argues that there is widespread agreement in translation studies that objective quality measurement does not exist.

2.3.2. Equivalence-Based Models

Equivalence is one of the key concepts in translation studies, particularly present in early translation theory (Lauscher 2000). While agreeing that equivalence presents a central concern in translation studies, Palumbo (2009: 42) describes it as "the most problematic and divisive issue" in the field. The notion of equivalence is central to equivalence-based models that perceive translation as a process of reproducing the source text as closely as possible. In order to capture numerous aspects that need to be considered, translation studies scholars differentiate between various types of equivalence, such as dynamic and formal equivalence (Nida 1964), factual equivalence (Van den Broeck 1985), functional equivalence (House 1977, 1997, 2015) etc.

In 1971, Reiss (1971) introduced one of the first systematic approaches to translation quality management. She defined translation as a "bipolarer Vorgang" (Reiss 1971: 11) because it constantly needs to consult the source text while aiming to produce a target text. Equivalence is "zielsprachliche Entsprechung" (Reiss 1986: 12) which can be deemed optimal when it is chosen according to the linguistic and situational context, the linguistic and stylistic level, and the intention of the author and have the same value (ger. Wertigkeit) as the text in the source language (Reiss 1986: 12). In order to achieve optimum equivalence between the source and the target text, Reiss (1986) suggests analyzing both texts by employing the categories of text type, linguistic instructions, and extralinguistic determinants. Based on Bühler's (1934) expressive, appellative, and representative functions of language, Reiss (1986) distinguishes between expressive, operative, and informative texts. To account for text types that are not consumed exclusively in written form but rather serve as a basis for oral communication, e.g., songs, speeches, dramas, etc., Reiss (1986) introduced the fourth text type, namely the audio-medal category. Linguistic instructions refer to semantic, grammatical, and stylistic properties of the source language, while extralinguistic determinants can be generally captured by the notion of situational context, i.e., contextual and situational factors that ought to be taken into account (Reiss 1986: 71). According to Lauscher (2000), Reiss' model has been very influential, especially in Germany and Austria. The "three process-governing rules" (Lauscher 2000: 152) lead to a new understanding of translation evaluation that is different from the word-by-word comparison between the source and the target text. With respect to extralinguistic determinants in particular, Lauscher highlights that "the wording of the source text can no longer be deemed

the sole point of reference for translation and quality assessment" (Lauscher 2000: 152). However, with regard to practical application, Reiss' concept of optimum equivalence appears to be too vague of a yardstick for good translation (Lauscher 2000). Therefore, in order to be able to apply the model practically, the concept of optimum equivalence must be precisely defined, beyond stating that it means that both source and target text elements ought to be of equal value (Lauscher 2000).

Another model that had a major impact on the development of translation quality assessment is House's (1977) model. House's model of translation advocates keeping distance from subjective quality evaluation and moving towards an assessment of functional equivalence. For House (1977), the preservation of meaning constitutes the essence of translation between languages. House identifies three dimensions of meaning: semantic, pragmatic, and textual. Thus, she defines translation as "the replacement of a text in the source language by semantically and pragmatically equivalent text in the target language" (House 1977: 103). She finds the term equivalent to be the fundamental criterion of translation quality and, thus, the key term in her definition of translation. According to House's definition, an adequate translation delivers semantic and pragmatic equivalence with the source text. House's equivalence concept presupposes that the target text has a "function equivalent" (House 1977: 103) to that of its source text. She posits that there are elements in any text that disclose that text's function when applying appropriate analytical tools. The function of a text is "the application or use which the text has in the context of a situation" (House 1977: 103).

In order to establish functional equivalence between the source and the target text, it is first necessary to analyze the source text in detail and precisely identify what kind of equivalence needs to be achieved in the target text (House 1997). For this purpose, House (1977: 104) breaks down the notion of situation as embedded in her definition and proposes two dimensions that need to be taken into account when assessing translation quality: dimensions of language user (geographical origin, social class and time) and dimensions of language use (medium, participation, social role relationship, social attitude, and province). While the dimensions of language user are self-explanatory, the latter need to be illustrated in more detail. Medium can take up to two forms, simple (written) or complex (written to be spoken, e.g., speech, or written to be spoken as if not written, e.g. play). Similarly, participation can also be simple or complex. Simple participation refers to a monologue or a dialogue, while complex participation refers to various means of indirect participation elicitation and indirect addressee involvement (House 1977). The third aspect, namely the social role relationship between the addresser and the addressee, may be symmetrical (i.e., one of solidarity or equality) or asymmetrical (characterized

by some kind of authority relationship that differentiates them). The following dimension, i.e., social attitude, considers how social distance influences the degree of formality or informality. Province, the last dimension of language use, refers not only to the professional occupation of the text producer but also, broadly defined, to the field or the subject of the text and the details of the text production. The function is then the result of an analysis of the text according to the eight situational dimensions mentioned earlier.

To achieve optimal quality, it is desirable to have a match along the mentioned eight aspects between the source and the target text. By applying these eight dimensions, a certain textual profile can be construed that sets the norm against which the quality of the target text is to be assessed. Any deviation along the dimensions is regarded as an error. House terms such dimensional errors "covertly erroneous errors" (House 1977: 105). These are to be differentiated from "overtly erroneous errors" (House 1977: 105) that refer to deviations in the denotative meanings of elements or failure to conform to the linguistic system of the target language. The final quality assessment of the target text consists of a comparison of both covertly and overtly erroneous errors and a judgment regarding the relative match between the two functional components (House 2015). Building on the differentiation between overt and covert erroneous errors, House also distinguishes between overt and covert translation. While in the case of an overt translation the target text addresses are not being addressed directly, a covert translation "enjoys the status of the original source text in the target culture" (House 1977: 107).

In 1997, House revised her model. While the revised model employs the same theoretical basis, the definition of translation, and the textual function as the original model (1977), it introduces the classic Hallidayan register concepts of Field, Mode, and Tenor (House 2015). Fiel encapsulates the topic, the content of the text, or its subject matter. Tenor relates to the relation between the addresser and the addresses and captures their relationship in terms of social distance as well as the degree of emotional charge (House 2015: 64). Tenor also captures the text producer's temporal, geographical, and social provenance as well as their personal viewpoint towards the text being portrayed. Tenor also captures the degree of formality. Finally, mode reflects both on the medium, which, again, can be simple or complex, and the degree of potential or actual participation allowed for between writer and reader (House 2015).

A number of other authors (Nida 1964; Catford 1965; Koller 1995; 2011; Pym 2010) also perceive equivalence as an important concept. Contrary to this view, other scholars deem equivalence rather unnecessary (Reiss & Vermeer 1984; Hatim & Mason 1990) and advocate a functional approach to translation quality.

2.3.3. Functional Approach

As Lauscher (2000) explains, according to this approach, translation is considered a process of text production based on a source text, and the target text is seen as a text "in its own right" (Lauscher 2000: 155). For each translation, the function is then determined by the translation commissioner and the translator in consideration of the target text's use in the target culture. Reiss and Vermeer (1984) proposed a general theory of translational action termed the skopos theory, which is a subcategory of the general theory of action. The former begins with the source text, which is then translated/interpreted. Building on the view that a text is a piece of information offered to a recipient by its producer, the authors argue that a target text, being a text, is also an offer of information to a recipient. Thus, a translatum, i.e., the outcome of the translation or interpreting process, may be regarded as a text offering information in a particular way about another offer of information (Nord 2013: 18). The Greek word skopos stands for purpose or aim. Thus, the highest rule of this theory is the skopos rule, meaning that any action is determined by its purpose (Nord 2013). Hence, a transitional action is also governed by its purpose. In their skopos theory, Reiss and Vermeer (1984) downplay equivalence in favor of adequacy and argue that achieving adequacy sets the guideline for the translation process. The authors define adequacy as "[…] the relationship between a source text and a target text, where consistent attention is paid to the purpose (skopos) of the translation process" (Reiss & Vermeer 1984: 127). A translation is deemed adequate if the chosen target language signs align with the requirements of the translation purpose (Nord 2013). Hence, in Reiss and Vermeer's (1984) view, equivalence is rather a particular form of adequacy.

2.3.4. Descriptive Approach

Castilho et al. (2018) and House (2015) further identify descriptive approaches (Toury 1995) as one of the main strands in translation quality. These approaches consider equivalence to be "of little importance in itself" (Toury 1995: 85) and assume its existence per definitionem. In contrast to Reiss and Vermeer's (1983) view, equivalence is never a relationship between source and target text but a "functional-relational notion" (Toury 1995: 86). This concept refers to the "set of relationships which will have been found to distinguish appropriate from inappropriate modes of translation performance for the culture in question" (Toury 1995: 86). Toury (1995, 2012) emphasizes the importance of empirical investigation in translation studies and the analysis of the macro-context. This approach focuses on the actual translations within the system of the receiving culture and thus considers the original to be of subordinate importance

(Toury 1995, 2012). Translations are regarded as cultural facts, i.e, “facts of the culture which hosts them” (Toury 2012: 24).

2.4. Machine Translation Quality

While the previous chapter considers translation, quality, and translation quality in general, this chapter investigates approaches to Machine Translation Quality (MTQ) in particular. Perspectives on translation quality within translation studies are shaped by various theoretical approaches to both translation itself and the concept of quality. By contrast, research into the quality of machine translation focuses on practical applications and identifies three strands to evaluating the quality of MT output. These encompass machine translation metrics, Machine Translation Quality Assessment (MTQA), and Machine Translation Quality Estimation (MTQE).

2.4.1. Machine Translation Metrics

Generally, machine translation may be evaluated manually, i.e. by humans, and automatically (Castilho et al. 2018; Lee et al. 2023; Sepesy Maučec & Donaj 2020). However, human evaluation is considered costly, subjective, and time-consuming compared to automated quality measurement. Although the quality of the latter is not as good as human evaluation, it is cost-effective and statistically justified (Lee et al. 2023; Sepesy Maučec & Donaj 2020). Automatic metrics for evaluating translation quality are typically validated by measuring how well they correlate with human assessments, that is, how closely the machine translation output resembles a reference translation (Agrawal et al. 2024). Some of the most well-known and commonly used metrics are Bilingual Evaluation Understudy (BLEU) (Papineni et al. 2002), National Institute of Standards and Technology (NIST) metric (Doddington 2002), and Metric for Evaluation of Translation with Explicit Ordering (METEOR) (Banerjee & Lavie 2005).

BLEU, being a numerical metric that judges the quality of a machine translation by its closeness to one or more reference human translations, rests on two main components: a numerical translation closeness metric and a corpus of good-quality human reference translations (Papineni et al. 2002). The closeness metric is inspired by the word error rate but adapted to support multiple reference translations, allowing flexibility in word choice and order. The core idea is to use a weighted average of variable-length matches against the reference translations.

Moreover, BLEU is based on a precision measure. To compute precision, the number of candidate translation words (unigrams) that occur in any reference translation are counted up and divided by the total number of words in the candidate translation. However, in order to address the need to exhaust a reference word after identifying a matching candidate word, Papineni et al. (2002) propose modified unigram precision. Modified unigram precision is computed by counting the maximum number of times a word occurs in any single reference translation and limits how often a word in the candidate can count toward the score. If a word appears more times in the candidate than in any single reference, the extra occurrences are clipped (Papineni et al. 2002).

Moreover, a candidate translation should be neither too long nor too short. To a certain extent, the *n*-gram precision tends to this by penalizing spurious words in the candidate that do not appear in any of the reference translations. Additionally, as already introduced, modified precision penalizes using a word more frequently in a candidate translation than its maximum reference count. However, modified precision alone fails to ensure the proper translation length. In order to address this limitation, Papineni et al. (2002) introduced the multiplicative brevity penalty factor. This factor ensures that a high-scoring candidate translation must match the reference translation in length, word order, and word choice. The final BLEU score is then the geometric mean of the modified *n*-gram precision multiplied by the brevity penalty. This metric ranges from 0 to 1. Only translations that exactly match a reference will receive a BLEU score of 1. Having more reference translations increases the chance of matching valid phrases, which can lead to higher BLEU scores.

The Defense Advanced Research Projects Agency (DARPA) commissioned NIST to develop an MT metric based on Papineni et al.'s (2002) work. Thus, it can be said that the NIST metric was derived from BLEU. However, the main difference between the two metrics is in how they weigh *n*-gram matches and treat the brevity penalty. Unlike BLEU, NIST considers how informative a particular *n*-gram is and assigns more weight to informative *n*-grams, i.e., *n*-grams that occur less frequently (Doddington 2002). Moreover, NIST applies a different, more lenient brevity penalty, so small variations in translation length do not significantly impact the overall score. However, these metrics have certain weaknesses. They do not use recall, which refers to the proportion of the matched *n*-grams out of the total number of n-grams in the reference translation, but merely *n*-gram precision, leading to varying results depending on the number of references. Furthermore, BLEU merely considers exact word matching between the machine translation output under evaluation and one or more reference translations (Lee et al. 2023). Additionally, it cannot count stems or synonymous words, which leads to difficulties in

evaluating a translation of a language with various morphological variants (Banerjee & Lavie 2005; Lee et al. 2023). For example, BLEU would not match 'running' and 'runs' since they are not considered the same word in the *n*-gram matching process (Lee et al. 2023).

Metric for Evaluation of Translation with Explicit ORdering (METEOR) was explicitly designed to address the limitations of BLEU (Banerjee & Lavie 2005: 67). It is based on an explicit word-to-word matching between the MT output (also called system translation) and one or more reference translations. This metric also accounts for words that are simple morphological variants of each other, meaning words that share the same stem, as well as for synonymous words. Possible matches are based on unigram precision, unigram recall, and a direct measure of "how out-of-order the words of the MT output are with respect to the reference" (Banerjee & Lavie 2005: 66). METEOR creates an alignment between a pair of translations to be compared. Banerjee and Lavie (2005: 67) define alignment as "mapping between unigrams, such that every unigram in each string maps to zero or one unigram in the other string, and to no unigrams in the same string". In the first phase, the system lists all possible unigram mappings between the two strings. With METEOR, three modules map unigrams based on different criteria. The exact module maps two unigrams if they are exactly the same, e.g., 'computers' map to 'computers' but not to 'computer'. The Porter stem module maps two unigrams if they are the same after they are stemmed using the Porter stemmer, i.e., after removing the suffix to obtain the stem of a given word. This means that 'computers' would map to both 'computers' and 'computer'. The third, the WorldNet (WN) synonymy module, maps two unigrams if they are synonymous with each other according to WordNet (Banerjee & Lavie 2005). Then, the F mean is computed by combining precision and recall using a harmonic mean that assigns most of the weight to recall. This is due to the finding that recall correlates much better with human translation than precision, and that the combination of the two using the F mean metric allows for further improvements (Banerjee & Lavie 2005). In order to take longer matches into account, METEOR applies a penalty based on how well the system translation aligns with the reference translation in terms of longer phrases. In doing so, METEOR penalizes short contiguous matches, thus rewarding longer matches that are more likely to reflect a meaningful translation quality. The final METEOR score is calculated by applying this penalty to the F mean.

Marie et al. (2021) conducted an empirical investigation of several hundred research papers published between 2010 and 2020 that deal with machine translation metrics. Based on this review, the authors demonstrate that a great number of MT evaluations relied exclusively on differences between BLEU scores, despite the fact that this metric had been introduced about two decades ago and over a hundred other metrics have been developed since then, claiming to

be better than BLEU (Marie et al. 2021). In recent years, this landscape has been significantly altered due to the rise of metrics based on neural networks that allow for more experimentation (Juraska et al. 2024). According to Guerreiro et al. (2024), some of the most current and widely used evaluation metrics are Crosslingual Optimized Metric for Evaluation of Translation (COMET) (Rei et al. 2020) and Bilingual Evaluation Understudy with Representations from Transformers (BLEURT) (Sellam et al. 2020).

2.4.2. Machine Translation Quality Assessment

While acknowledging the scores the above-mentioned automatic metrics provide, Popović (2018) states that MT researchers often benefit from having additional information about their systems, such as what the most serious problems are in a translation system, what its strengths and weaknesses are, whether some types of errors are more difficult to post-edit than others, etc. Error classification and analysis techniques have emerged to address the relationship between such questions and the overall quality scores (Popović 2018). Actual errors in a translated text are identified and classified in order to deliver a better basis for decisions about a given task. Most commonly, error analysis pursues the aim of obtaining an error profile for a translation output, i.e., a distribution of errors over the defined error classes. Another common use is to compare different translation outputs by identifying the distribution of errors for each error class across different translations (Popović 2018). Error classification can be carried out manually, automatically, or semi-automatically. In terms of manual error classification, Popović (2018) argues that the most straightforward method for error analysis is to mark each erroneous word in the translation output and assign a corresponding error tag to it. One way of doing so is to assign a quality score to a given task, most commonly ranging from 1 (poor) to 5 (perfect) (Popović 2018). Additionally, annotators should at least obtain either the source text, a reference translation, or both.

Guzmán et al. (2015) conducted an investigation of diverse sources of information and the difference in assigning overall quality between bilingual and monolingual annotators. They found that monolinguals are slower but more consistent than bilinguals. At the same time, exposure to additional information in terms of the source language text resulted in all annotators becoming slower and less consistent, leading the authors to recommend the use of reference translation alone. In contrast, Fomicheva (2017) and Bentivogli et al. (2018) highlighted the issue that reference-based human evaluation might unjustly penalize accurate translations that differ significantly from the reference translation. Thus, the literature largely agrees a that

source-based evaluation is preferable to a reference-based one (Bentivogli et al. 2018; Fomicheva 2017; Läubli et al. 2020).

The first shift towards the use of explicit error classification and analysis was reported by Vilar et al. (2006). Their classification scheme has a hierarchical structure with three levels. The first level encompasses five large classes of errors: missing words, unknown words, punctuation errors, word order, and incorrect words. The first three classes of errors are rather self-explanatory. In terms of word order, the authors distinguish between word and phrase-based reordering. In the case of the former, a correct sentence can be generated by reordering individual words, whereas with the latter, blocks of consecutive words, i.e., phrases, should be moved together to produce a correct translation. Within the following category of incorrect words, Vilar et al. (2006) identify incorrect forms, extra words, styles, and idioms. Incorrect form occurs when the system is unable to produce the correct form of the word, despite the basic form being translated correctly. This is particularly present in inflected languages. The extra words class of errors is relevant when examining the translation of speech inputs. While style errors concern "a bad choice of words" (Vilar et al. 2006: 698) while preserving the meaning, the idioms pertain to idiomatic expressions that are not recognized as such by the system and are thus translated literally. In the following year, Vilar et al. (2007) proposed a ranking-based evaluation approach, i.e., a direct comparison of two sentences at a time performed by humans. This approach enables lower dependency on extensive evaluation guidelines, thus requiring less effort.

More recently, Popović (2020) proposed a novel method for manual evaluation of MT output to compensate for the lack of information about problematic parts of the translation about the given quality criteria. What is novel about this method is that evaluators do not assign any scores nor classify errors, but mark all problematic parts of the text, be it words, phrases, or sentences. The proposed method provides more detailed information while requiring less effort because it involves direct marking of the problematic segments rather than considering the overall rating based on the issues at hand. This evaluation process is based on two prominent quality criteria: comprehensibility and adequacy. While the former relates to the degree to which a translated review can be understood, the latter concerns the degree to which the translation conveys the meaning of the original text in the source language. Popović (2020) found that this method results in a high inter-annotator agreement because it does not include fine-grained error classification.

2.4.3. Machine Translation Quality Estimation

Next to ranking competing outputs from best to worst, annotations can also be conducted as direct estimates of the quality of the given translations (Graham et al. 2013). Unlike ranking-based evaluation, a direct estimation method of assessment considers not only which output is better, but also the degree to which the particular output outperformed the other(s). However, the commonly used interval-level scale for direct assessment, typically consisting of a small set of fixed categories, often forces human judges to choose between categories that do not faithfully reflect their judgment. Consequently, this may lead to inconsistencies in human judgments. These limitations motivated Graham et al. (2013) to address the continuous nature of translation quality assessment and propose a continuous rating scale for direct assessment. This method does not use discrete scales but assigns a continuous score between 0 and 100. Graham et al. (2013) compared human assessor consistency levels for judgments collected on a five-point interval-level scale to those collected on a 1–100 continuous scale. They found that the use of a continuous scale eliminates individual judge preferences, leading to a higher degree of inter-annotator consistency.

Lommel et al. (2013) present the Multidimensional Quality Metrics (MQM), a generic methodology for assessing quality that can be adapted to address numerous evaluation needs. This open and extensible system for declaring and describing translation quality metrics uses a shared vocabulary of issue types and can be applied not only to human and machine translation but also to any text type. The MQM error typology represents a system of seven high-level error type dimensions, which, then, include more specific error subtypes. The main idea is to have a large set of hierarchical error categories allowing the selection of any subset appropriate for the task at hand (Lommel et al. 2013; Popović 2018). The seven core dimensions include terminology, accuracy, linguistic conventions, style, locale conventions, audience appropriateness, design, and markup.

In 2018, Specia et al. (2018) argued that despite the widespread adoption of MT systems, the quality of automated translations was often still unsatisfactory. This created the need for automatic metrics that would predict the quality of a translated segment, known as Quality Estimation (QE). Unlike standard evaluation methods that compare system translations with reference translations, the main objective of QE is to provide an estimate of how good or reliable a translated text is without referencing human translations. In order to do so, QE uses more complex metrics that surpass matching source segments against previously translated content (Specia et al. 2018).

Zhao et al. (2024) provide a comprehensive overview of the state-of-the-art QE annotation methods, shared tasks, methodologies, and challenges. Since the Conference on Machine Translation (WMT) established QE as a separate task, the research has developed in three main directions: QE based on handcrafted features, e.g., QuEst, QuEst++, methods based on large language models, e.g., Generative Evaluation Metric with Bilingual Alignment, Fine-Tuned LLM for QE, and QE based on deep learning models. The latter further includes subcategories such as classic deep learning approaches, e.g., deepQUEST, DirectQE, and those incorporating pre-trained language models, e.g., Pre-Est Based Transformer, TransQuest. According to Zhao et al. (2024), methods based on large language models have been central in QE research. While Specia et al. (2018) focus predominantly on quality prediction for sentences, QE can also be conducted at other textual levels, such as word-level and document-level QE (Zhao et al. 2024). Furthermore, direct assessment and MQM, the previously introduced annotation methods, extend their usefulness also to the area of quality estimation. However, the authors also acknowledge several shortcomings to QE. These include data scarcity, insufficient interpretability, the rarity of word-level and document-level QE methods, and the lack of standardized evaluation metrics (Zhao et al. 2024).

Summarizing the outcome of the WMT24 shared task on automated MT evaluation, Freitag et al. (2024) argue that the results show the superiority of neural-based learned metrics over overlap-based metrics like BLEU. According to the official ranking of primary submissions to the WMT24 Metric Task, metrics such as MetaMetrics-MT (Anugraha et al. 2024; Winata et al. 2024), MetricX-24-Hybrid (Juraska et al. 2024), and XCOMET (Guerreiro et al. 2024) ranked best. The first-ranked model, namely MetaMetrics-MT, achieves state-of-the-art performance for reference-based metrics (Anugraha et al. 2024). This model also comes in a reference-free variant that shows a competitive correlation with human ratings. The second-ranked model, that is MetricX-24-Hybrid, is trained on a combination of source-only, reference-only, or both examples. This enables the model to operate in both modes, namely reference-based and QE, making it a hybrid model. Thus, both MetricX-24-Hybrid and MetricX-24-Hybrid-QE are in fact the same model, the only difference being that references are excluded in the QE version (Juraska et al. 2024). The last model, xCOMET, aims at addressing gaps identified in current, widely used metrics, such as previously mentioned COMET and BLEURT, and represents an open-source learned metric that exhibits state-of-the-art performance in all relevant vectors of evaluation, i.e., sentence-level, system-level, and error span detection (Guerreiro et al. 2024).

2.5. Machine Translation and Creative Language Use

In recent years, creative translation has become a relevant topic in translation studies (Bayer-Hohenwarter & Kussmaul 2020). Creativity in translation seems to be commonly understood as "what goes beyond any literal rendering" (Bayer-Hohenwarter & Kussmaul 2020: 311). On a similar note, Rojo (2017: 353) argues that "[c]reative solutions have to be novel and depart from conventional translation behavior but should also render meaning accurately and give solutions appropriate to a certain textual, situational, and cultural context". In light of recent developments, it appears that creative translators are those who understand how to achieve a high-quality human translation best and when it is needed, as well as how to best cater to clients' needs (Bayer-Hohenwarter & Kussmaul 2020). In their pursuit of a creative solution, translators will often resort to producing shifts, that is, using words that are not literal reproductions of the source text, but rather new expressions. Thus, in a similar vein, Bayer-Hohenwarter and Kussmaul (2020: 312) state that "a creative translation is a translation that often involves changes (as a result of shifts) when compared to the source text, thereby bringing something that is new and also appropriate to the task that was set, i.e., to the translation assignment".

With regard to technical communication and creative language use, Hardesty and Hollinger (2020: 45) argue that despite attempts to make technical communication more humanistic, the dichotomy between technical and scientific fields and art and creativity seems to persist. More recently, Slater and Rosselot-Merritt (2024) argued that many scholars have started to look at technical communication practice through an artistic lens. On a similar note, Imran et al. (2024) investigated the use of figurative language, such as metaphors and idioms, in the domain of software engineering. According to the authors, in software engineering, figurative language is often used to convey technical concepts. They provided evidence that demonstrates the relevance and impact of figurative language in software engineering communication and the potential benefits of fine-tuning large language models with figurative language (Imran et al. 2024).

Moreover, literary translation exemplifies a field that requires a high degree of creativity, wordplay, idiomatic expressions, and cultural nuances. Due to this, a number of scholars examine creativity in machine translation based on literary texts. Guerberof-Arenas and Toral (2022) juxtapose three translation outputs of the same short story, which include human translation, machine translation, and post-editing. Based on their study, the authors claim that MT might only be useful in literary translation for segments that are straightforward and mechanical. According to the authors, this is due to the fact that machine translation is still not intelligent

enough to adjust meaning to its co-text and therefore understand the underlying, non-literal message that the author of the text attempts to convey. As a result, MT output involves semantic errors and overly literal translations. Against this background, creativity might be perceived as a problem-solving process, i.e., tool human translators innately possess, which helps them come up with interesting, "unexpected, surprising or brilliant" solutions (Guerberof-Arenas & Toral 2022: 21). Another major argument the authors put forward with regard to MT is that it actually limits and confines creativity in translation by suggesting rather straightforward, plain translation solutions. As Guerberof-Arenas and Toral (2022) state, even if translators edit the MT proposal, it is still difficult for them to distance themselves mentally from the result suggested by the machine. Although the authors argue that MT might be useful only for particular text segments, they also suggest that post-editing might not be the best choice for achieving high-quality literary translation either, since these still heavily rely on human wit and creativity.

Furthermore, Wei (2024) addressed the implementation of both cultural and emotional intelligence in the translation process. Cultural intelligence refers to the ability of machine translation systems to identify and correctly translate culture-specific segments, such as idioms, customs, or norms. To care for this, the training of machine learning models is done not only on linguistic data but also on cultural annotations. In the realm of emotional intelligence, there are two key complex layers: emotional tone and subtlety. The ability to identify and convey emotional tones is particularly important with artistic and literary texts (Wei 2024). Zajdel (2022) argues that artificial intelligence is becoming "increasingly innovative" (Zajdel 2022: 116), demonstrating positive results in areas that have traditionally been considered human. While also acknowledging this, Wei (2024) argues that machine translation still needs considerable human supervision, especially with regard to cultural and emotional aspects of a translation.

3. Idioms

In order to 'come to grips' with idioms, it is essential to first understand their nature, meaning, and defining characteristics. Idioms represent a truly challenging group of expressions for a number of linguistic disciplines, ranging from theoretical linguistics and psycholinguistics to language acquisition (Keizer 2016). Fellbaum (1993: 280) acknowledges this issue by stating that "[a]n analysis of the lexical, morphological, and syntactic flexibility of idioms would represent a very considerable undertaking". Due to the complexity of idioms, there also seems to be disagreement among scholars regarding the definition of idioms that clearly differentiates between similar groups of expressions, such as idioms, idiomatic expressions, and fixed phrases (Barkema 1996; Moon 1998; Granger & Paquot 2008). For this reason, there are numerous definitions of idioms, depending on the approach to idioms advocated by the respective author. Initially, idioms were defined in terms of their non-compositionality because "[t]he essential feature of an idiom is that its full meaning, and more generally the meaning of any sentence containing an idiomatic stretch, is not a compositional function of the meanings of the idiom's elementary parts" (Katz & Postal 1963: 275). Similarly, Fraser (1970: 22) defines idioms as a "constituent or series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed". However, alternative views that drifted away from the strict, homogenous view of idioms to a more liberal and flexible perception have emerged over time. These views generally recognize that idioms may have different degrees of idiomaticity and therefore encompass more than one dimension (Cruse 1986). Rafatbakhsh and Ahmadi (2019: 3) term this a continuum of idiomaticity. This gradual shift away from non-compositionality eventually led to the compositional view, both of which will be examined in greater detail throughout this section. Prior to this, the upcoming section will briefly introduce key lexico-semantic relations essential for understanding the complexities of idioms.

3.1. Lexico-Semantic and Terminological Relations

According to Murphy (2010), semantic relations among words refer to the relation between their senses. Some semantic relations can also be lexical. In such cases, the relatedness goes beyond meaning and includes also other aspects, such as morphological form or collocational patterns (Murphy 2010). Regarding terminological relations, according to Murphy (2010), there are two broad types of relationships between words: paradigmatic and syntagmatic. The latter refers to relations between words that "go together in syntactic phrases" (Murphy 2010: 108), such as 'dogs' (noun) and 'bark' (verb), which often belong to different word classes, as shown

in the example. Unlike syntagmatic relations, paradigmatic relations belong to the same word class and share common characteristics. The words embedded in such relations can be said to form a paradigm, that is, "a set of examples that show a pattern" (Murphy 2010: 109). The most commonly known paradigmatic relations include synonymy, hyponymy, and antonymy (Murphy 2010). The term synonym is derived from Greek roots 'syn' (alike) and 'onym' (name) (Murphy 2010: 110), thus referring to words that have the same meaning. In order to check whether two words mean the same, a substitutability test is used. Words are considered synonymous if there is no change in the sentence's meaning when one word is substituted for the other. In relation to this, the classification of synonyms into absolute, propositional/cognitive, and near–synonyms has been adopted in linguistics (Gimaletdinova et al. 2021). The first category designates words that are substitutable in any possible context without changing any aspect of the meaning of the given context (Murphy 2010). Such synonyms are quite rare (Gimaletdinova et al. 2021; Murphy 2010). Cognitive synonyms have "certain common semantic properties" (Gimaletdinova et al. 2021: 668), but differ in the degree of expressivity, e.g., 'fiddle' and 'violin'. The third category, namely near synonyms, denotes the same concepts, but refers to words that are not are not completely interchangeable and thus do not allow for substitution in any possible context. An example would be the adjectives 'handsome' and 'pretty', as the former describes men, and the latter describes women (Gimaletdinova et al. 2021).

The second common type of paradigmatic relations, namely hyponymy, denotes inclusion relations, i.e., 'type-of' relations (Murphy 2010). In other words, these paradigmatic relations involve notions that contain or are contained in other meanings. An example would be 'house' and 'building', where 'house' is a hyponym of 'building' because a house is a type of building. Another example would be 'cheddar' being a subset of 'cheese'. While cheddar is also cheese, it cannot be said that all cheese is cheddar. This would mean that 'cheddar' is a hyponym of 'cheese', while 'cheese' is a hyperonym of 'cheddar'. The Greek prefixes hypo- (under) and hyper- (over) help picture this asymmetrical inclusion relation (Murphy 2010).

The third type of paradigmatic relations, namely antonymy, represents the only relation that holds between pairs of words, making it particularly binary (Murphy 2010). Antonymy represents a contrasting relationship in which two words are of opposite meanings, e.g., 'black' and 'white'. The literature identifies three main types of antonymy: contrary (scalar or gradable) antonymy, complementary (contradictory) antonymy, and converse antonymy (Gao & Zheng 2014; Hu 2001; Murphy 2010). An example of contrary idioms would be 'short/tall' or 'old/young'. The adjective 'gradable' in gradable antonyms indicates that the two opposing words may differ in terms of degree. This also means that if one feature is denied, the other

feature is not automatically asserted (Gao & Zheng 2014). For example, describing someone as 'tall' indicates that the person is 'not short', but stating that a person is 'not tall' does not entail the quality of being short. This is because 'tall' and 'short' represent two extremes, but the range between them includes measurements that are neither tall nor short. Unlike with gradable antonyms, there is no intermediate ground between complementary antonyms. Thus, complementary antonyms, such as 'awake/asleep', 'married/single', 'pass/fail' are those for which the negation of one entails the assertion of the other and vice versa (Gao & Zheng 2014; Murphy 2010). Converse antonyms refer to the same relation from different perspectives, e.g., 'husband/wife', 'doctor/patient', or 'teacher/student' (Gao & Zheng 2014). Murphy (2010) provides the following pattern to illustrate this relationship: if X is p to Y, then Y is q to X. For example, if Bill (X) is a parent (p) to Madeline (Y), Madeline (Y) is a child (q) of Bill's (X) (Murphy 2010: 120).

Another way to look at terminological relations is to classify them into hierarchical, equivalence, and associative types (Clarke 2001; Maia et al. 2021; Tudhope & Jones 2001). Hierarchical relationships can be further divided into generic and partitive relations. Generic relations, also named hyponym-hypernym relation (Dahlberg 1978), genus-species (Mazzocchi 2017), 'is-a' and 'type-of' relationship (Hjørland 2007; Khoo & Na 2006), illustrate the link between a class with its subclasses, as explained earlier. In the case of partitive relations (Dahlberg 1978), also termed whole-part (Zeng 2008), or meronym-holonym relations (Broughton et al. 2005; Peters & Weller 2008), complex concepts are broken down into smaller units. The subdivision represents a hierarchy, with the term denoting the whole representing the higher class and the term denoting the parts of the whole being the lower class (Beghtol 2001; Weller & Stock 2008). For example, 'wheel' would represent a meronym of 'car', while 'car' would be the holonym of 'wheel'. Equivalent relationships refer to concepts whose meanings are the same or are considered to be equivalent in a particular context. Therefore, these relations pertain to concepts that can be represented with synonyms (Peters & Weller 2008; Stock 2010). The last class, associative or non-hierarchical, is "a mental association of the concepts" (Maia et al. 2021). According to Zeng (2008), associative relationships are neither hierarchical nor equivalent, thus forming a separate category.

3.2. Non-Compositional View of Idioms

Langlotz (2006) argues that the generative treatment of idioms is strongly determined by their characterization as semantically non-compositional strings. This concept of non-composition-

ality arises from the observation that the semantic interpretation of certain linguistic constructions cannot be reconciled with Frege's (1884) principle of compositionality. This principle states that the meaning of a sentence is determined solely by the meanings of its constituent parts and their syntactic arrangement (Frege 1884). Consequently, in order to adhere to this principle, idiomatic expressions ought to be classified as linguistic anomalies and outliers that exist outside the boundaries of grammatical regularity within the language (Langlotz 2006). Katz (1973: 358) puts this in a nutshell by suggesting that "[i]dioms are the exceptions that prove the rule: they do not get their meaning from the meaning of their syntactic parts. If an idiom is treated as if it were compositional, false predictions are made about its semantic properties and relations".

Langlotz (2006) looks back at the early generativist classifications of idioms, which, resting on the criterion of non-compositionality, propose the distinction between lexical idioms and phrase idioms (Katz & Postal 1963). The essential difference between these two categories lies in their internal syntactic complexity. Lexical idioms are predominantly associated with one of the most basic word classes, e.g., nouns, verbs, and adjectives, thus including exocentric compounds such as 'telephone', 'photograph', or 'redneck' (Langlotz 2006). Bauer (2008) explains that endocentric compounds are characterized by having an internal centre or a 'head', whereas exocentric compounds have a 'head' that is external to the compound itself. Defining endocentric compounds through the lens of hyponymy, Bauer (2008) explains that an endocentric compound could be understood as a hyponym of its head constituent. An example would be 'doorknob'. Since 'doorknob' is a type of 'knob', a direct hierarchical relationship is established between the two words, with 'doorknob' being a hyponym of 'knob'. Conversely, in the case of 'turnkey', although the head element is the verb 'turn', the resulting compound functions as a noun, which precludes it from qualifying as a hyponym of the head element. Consequently, such compounds fall into the category of exocentric compounds (Bauer 2008). Accordingly, lexical idioms, encompassing exocentric compounds, exhibit syntactic behavior akin to conventional lexical units. Therefore, lexical idioms pose a minimal challenge to generative theories of idiom representation and variation, since they adhere to syntactic derivation similar to ordinary lexical items.

Conversely, a complex internal syntactic structure of phrase idioms like 'shoot the breeze', 'take advantage', or 'spill the beans' demonstrates a pattern on a phrase level. Unlike lexical idioms, listing phrase idioms individually in the lexicon would be inefficient, since these expressions follow regular patterns and therefore do not require repeated elaboration of their ele-

ments each time they occur in a phrase (Langlotz 2016). Although phrase idioms are semantically exceptional, they appear to be less intriguing from a syntactic perspective. They often adhere to regular syntactic patterns and are subject to inflectional agreement rules, as illustrated below (Langlotz 2006: 17):

(1) John and his friends are shooting the breeze.
(2) Advantage was never taken of the students.
(3) The beans were spilled, yesterday.

Moreover, Keizer (2016: 985) further demonstrates that some idioms also allow for syntactic manipulation, such as passivization, raising, and relativization:

(4) Edith spilled the beans/The beans were spilled by Edith.
(5) The cat is out of the bag/The cat seems to be out of the bag.
(6) The strings that Pat pulled got Chris the job.

Example (4) illustrates that the idiom 'spill the beans' allows for passivization. This means that this phrase idiom can be transformed from the active to the passive voice without altering the meaning of the expression as a whole. In other words, both example sentences stated in (4) convey the same meaning, although expressed differently. The following example (5) illustrates raising, a syntactic transformation in which a Noun Phrase (NP) appears to have been moved from its own clause into a higher one without changing the semantic role of the given NP (Chomsky 1981). In the given example, raising does not seem to influence the meaning of the idiom 'the cat is out of the bag'. The last example (6) shows that the idiom 'pull the strings' allows for the syntactic mechanism of relativization, i.e., forming a relative clause (de Vries 2002).

While some phrasal idioms, as shown above, allow for various syntactic operations, others tend to be more rigid and demonstrate a modest degree of morphosyntactic flexibility, as illustrated below (Keizer 2016: 986):

*(7) *The bucket was kicked by Matthew.*
*(8) *The bucket that Matthew kicked was empty.*
*(9) *Matthew kicked the empty bucket/several buckets.*

Unlike the previous examples, the idiom 'kick the bucket' does not allow for passivization, relativization, or modification, e.g., 'empty' or quantification, e.g., 'several', as exemplified in example (9). Based on the aforementioned examples, Keizer (2016) argues that idioms are regarded as semantically non-decomposable units within the non-compositionality view. In terms of syntax, however, they allow for flexibility in the same way as regular phrases. The lack of a viable framework that accounts for phrase idiom specificities led scholars to gradually drift away from the rigidity of the non-compositional view and look at idioms through the lens of the compositional view instead.

3.3. Compositional View of Idioms

Diverging from the non-compositional view, as previously introduced, a number of scholars (Nunberg et al. 1978; Wasow et al. 1983; Gazdar et al. 1985) proposed an alternative approach that also puts the feature of compositionality in the spotlight but examines it from a different perspective. Nunberg et al. (1994) can be considered advocates of the compositional view due to their strong opposition to the generative conception of idioms as non-compositional segments (Langlotz 2006). Rather than understanding idioms as "anomalies of interpretation" (Nunberg et al. 1994: 505), Nunberg et al. (1994) argue that the non-compositional view by far neglects the semantic structure of idioms and criticize it for being too simplistic. Thus, they contend that the relationship between the components of idiomatic expressions is essentially semantic and attribute many difficulties in idiom analysis to the lack of differentiation among three key semantic properties of idioms: their relative conventionality, their opacity, and their compositionality (Nunberg et al. 1994: 498). The relative conventionality of an idiom is determined by the discrepancy between its idiomatic meaning, and the meaning one would predict by consulting only the literal meanings of its constituents when they occur in isolation from one another. The opacity of idioms denotes the ease with which the motivation for the use can be recovered. Finally, compositionality refers to the degree to which the phrasal meaning can be analysed in terms of the idiom parts. Nunberg et al. (1994) find the traditional interpretation of idiomatic compositionality in terms of "whether a native speaker would recover the sense of the idiom on hearing it in an uninformative context" to be "quite unnatural and false to the way most idioms are presumably learned" (Nunberg et al. 1994: 495). According to the authors, such an interpretation presupposes that the listener asks and is told what an idiom means every time they encounter one, which is highly unlikely. In other words, they argue that the meaning of many idioms cannot be predicted when taken out of context but is fully understood when they are surrounded by relevant information that aids comprehension.

Furthermore, according to Nunberg et al. (1994), these features do not depend on one another; that is, they can appear in isolation and do not presuppose the presence of the other two. Therefore, under the assumption that all idioms are conventional, the presence of this feature does not imply that all idioms are both opaque and non-compositional. An example would be 'to kick the bucket'. It may be argued that this idiom exhibits a high degree of conventionality combined with a high degree of opacity and non-compositionality. It can be considered opaque since it might be difficult even for speakers familiar with the phrase to retrieve the motivation for its use, that is, to grasp why this idiom has the meaning it does. Furthermore, its high degree of non-compositionality can be explained by the difficulty in recognizing the connection of the segment's constituents to the overall meaning of the idiomatic phrase. Another example would be the idiom 'hit the ceiling'. In this case, the motivation for the use, illustrated by the category of opacity, may be easily recovered since the action described by the phrase, when understood literally, often correlates with the metaphorical association of becoming extremely angry (Keizer 2006). Yet, it remains unclear how and by what means the meaning of each constituent contributes to the overall idiomatic meaning of the phrase. Along the proposed categories, the idiom 'hit the ceiling' can be classified as conventional, highly motivated (or transparent), and non-compositional. The last example that clearly shows the practical application of the categories proposed above is 'spill the beans'. This idiom can be described as conventional, motivated, and compositional because the idiomatic meaning of the idiom's parts makes up the meaning of the overall idiom. Here, 'spill' relates to 'divulge' and 'the beans' relate to 'information'. Moreover, transparency is also provided, as the speaker can easily retrieve the motivation for the use based on the meaning of its constituents (Keizer 2006).

Moreover, Nunberg et al. (1994) distinguish between two classes of idioms based solely on the notion of non-compositionality: idiomatically combining expressions (ICEs) and idiomatic phrases (IPs). According to them, idiomatically combining expressions are regarded as compositional, while idiomatic phrases represent non-compositional idioms (Nunberg et al. 1994). On the one hand, according to Nunberg et al.'s (1994) definition, ICEs concern cases where "the conventional mapping from literal to idiomatic interpretation is homomorphic with respect to certain properties of the interpretation of the idiom's components" (Nunberg et al. 1994: 496). Homomorphism comes from the Greek word morphe, meaning form or structure, while the prefix homo- means similar. This means that the transition from the literal to the idiomatic meaning is not random, but exhibits structural correspondence between literal elements and elements in the idiomatic interpretation. An example of ICEs would be the phrase 'pull strings',

as it consists of individual word-level lexical entries ‘pull’ and ‘strings’, whereby each constituent contributes to the meaning of the idiom as a whole. Here, ‘pull’ as a literal element would mean ‘using one’s influence’ in the idiomatic interpretation, while the literal element ‘strings’ would represent ‘personal connections’ in the idiomatic interpretation (Bargmann & Sailer 2018). The idiom ‘kick the bucket’ exemplifies idiomatic phrases, expressions that are “syntactically and semantically monolithic” (Bargmann & Sailer 2018: 2). This means that the phrase as a whole is stored in the lexicon, and its meaning is dependent on the idiom in its entirety rather than on its constituents.

Moreover, Langlotz (2006) uses the notions of motivation and isomorphism as criteria for describing them. According to Langlotz (2006), an idiom is motivated if speakers, based on their symbolic and conceptual knowledge, are able to explain the connection between the literal scene and the idiomatic meaning. In other words, an idiom is motivated if it is possible to “make sense of these projections” (Langlotz 2006: 114). In this regard, Langlotz (2006) distinguishes between unmotivated idioms, e.g., ‘kick the bucket’, and motivated ones, such as ‘spill the beans’. As for the second dimension, Langlotz (2006) equates the quality of being isomorphic with being analyzable. The author follows Geeraerts’s definition that an expression is isomorphic if “a one-to-one correspondence between the parts of the semantic value of a compound meaning as a whole and the constituent parts can be detected” (Geeraerts 1995: 60). Furthermore, Langlotz (2006) argues that isomorphism cannot exist without some degree of motivation, whereas the opposite appears to be very frequent. On the one hand, ‘spill the beans’ would be an example of both a motivated and isomorphic idiom since, as mentioned earlier, ‘the beans’ relate to ‘information’, while ‘spill’ maps to ‘divulge’. Thus, its isomorphic structure is an effect of this motivation. On the other hand, the expression ‘spill one’s wheels’, meaning fail to achieve anything satisfactory, is motivated but not isomorphic. In this case, no isomorphism can be attributed because it is not possible to establish direct, analyzable correspondences between the elements of the idiom (spin, wheel) and conceptual counterparts in the idiomatic meaning. This idiom is, however, motivated, since it is possible to make sense of its meaning by juxtaposing a stuck person with a stuck car, i.e., a car that is spinning its wheels cannot move forward and thus fails to reach its destination. Accordingly, a stuck person cannot progress or reach their goals.

Since accounting for all perspectives and categories of idioms would be a highly challenging task that extends beyond the scope of this thesis, the following chapter will attempt to illustrate a brief overview of the most prominent categories of idioms.

3.4. Categories of Idioms

Rafatbakhsh and Ahmadi (2019: 4) argue that all definitions of idioms have three topics in common: compositionality, institutionalization, and degree of frozenness or fixedness. The notion of compositionality refers, as previously outlined, to the feature of idioms that their meaning does not represent the sum of the meanings of their parts. Institutionalization refers to the common usage of idioms by a large number of people, while frozenness or fixedness refers to the degree an idiom is subject to variation, which can differ depending on the subtype of the idiom. Fraser's (1970) classification of idioms is one of the first serious undertakings to categorize idioms (Katsarou 2011; Keizer 2016; Mäntylä 2004). This undertaking is illustrated by a scale Fraser (1970) calls the Frozenness Hierarchy. Influenced by the then-prevailing linguistic theory grounded in generative grammar, Fraser bases his classification on the following grammatical operations that may apply to a large number of idiomatic expressions (Fraser 1970: 39):

L6 - Unrestricted
L5 - Reconstitution
L4 - Extraction
L3 - Permutation
L2 - Insertion
L1 - Adjunction
LØ - Completely Frozen

Fraser (1970) introduced completely frozen idioms at the bottom of the hierarchy. This level (LØ) denotes idioms to which no operation whatsoever may be applied. Contrary to the Ø level, the top level (L6), refers to idioms that allow for any operation and are therefore considered fully unrestricted. The five levels in between encompass adjunction, insertion, permutation, extraction, and reconstitution. Adjunction refers to operations such as gerundive nominalization transformation. An example would be transforming the sentence 'John hit the ball' to 'John's hitting the ball' (Fraser 1970). Nominalization refers to transforming a verb or an adjective into a noun, and gerundive normalization, being a specific type of it, employs gerunds, that is, verb forms ending in -ing that function as a noun. Insertion represents placing a non-idiomatic constituent into the idiom sequence, e.g., changing 'John read the riot act to the class' to 'John read the class the riot act'. In this example, the indirect object 'the class' was inserted into the idiom sequence 'read the riot act'. Thus, in this example, indirect object movement transformation was applied. In cases of idioms such as 'can't teach new tricks to an old dog' that contain both

direct (‘new tricks’) and indirect object (‘an old dog’), the grammatical operation is called permuting. The difference between permuting and insertion is that the indirect object movement is applied to a sequence that already contains both a direct and an indirect object, rather than inserting a non-idiomatic segment into the idiom sequence, as is the case with insertion. Extraction encompasses various transformations, such as transforming ‘look up the information’ to ‘look the information up’, which illustrates particle movement by extracting the particle from the idiom. Another example would be altering the sentence ‘The father laid down the law to his daughter’ to ‘The law was laid down by her father’ by extracting the direct object (‘the law’) and placing it outside the idiom. The final grammatical operation, namely constitution, pertains to a transformation known as action nominalization. Illustrated by the example of transforming ‘He laid down the law to his daughter’ to ‘His laying down of the law to his daughter’, this operation alters the syntactic function of the entire idiom. Here, the nominalization of the verb phrase ‘laid down the law’ results in a noun phrase that expresses the action of the verb (‘laying down of the law’). As a result, the idiom no longer functions as a verb within a sentence but as a noun phrase instead. Although the hierarchy reflects different degrees of flexibility, one major disadvantage of Fraser’s exhaustive systematization is that certain grammatical operations cannot be applied to any expression, which speaks against the proposed idea of a general hierarchy. For example, permutation applies merely to expressions with phrasal verbs.

Similarly, McCarthy (1992) argues that fixedness may be expressed as “potential for subjection” (1992: 55) to the following syntactic processes: addition, deletion, transposition, and substitution. This potential is illustrated by applying these operations to the expression ‘hit below the belt’. The first process, i.e., addition, demonstrates the rigidity of a given expression in terms of adding elements to the phrase. This means that by adding the adverb ‘hard’, saying ‘he hit hard below the belt’, the expression would sound odd. The same applies to adding the adjective ‘leather’ to the belt, which makes the expression sound rather literal. The second operation, namely deletion, implies the obligatory presence of all elements of the idiomatic expression. An omission of any segment of the expression, here in the example ‘hit below the belt’, would erase the idiomatic meaning and be interpreted in its literal sense. Transposition, the third operation McCarthy (1992) suggests, stands for the change of category without any change in meaning (Ten Hacken 2015). McCarthy (1992) posits that transposition is normally excluded in the case of idioms or is strictly limited to certain types. For example, the phrase ‘hit the belt’ exhibits a minimal degree of nominalization, e.g., ‘that was really a case of hitting below the belt’, but not passivation, e.g., ‘the belt was hit below several times’. Finally, the last operation refers to substitution. It means that lexical and grammatical substitutes may normally

not be used: ‘hit below the sash’ cannot be regarded as a synonym for ‘hit below the belt’ (McCarthy 1992).

Looking at idioms from a semantic perspective (Grant 2003; Turaev 2025), Fernando (1996) outlines three different categories of idioms, which are pure idioms, semi-idioms, and literal idioms. Pure idioms represent fixed, non-literal multiword expressions. This means that their meaning cannot be derived from the meaning of their constituents. For instance, ‘spill the beans’ has, according to Fernando (1996), no connection to its literal components. The second type of idioms, namely semi-idioms, contains one or more literal parts and at least one segment that is not to be interpreted literally and which is usually exclusive to the given expression. An example would be ‘drop names’, where ‘names’ constitutes the literal element and ‘drop’ the non-literal one. However, ‘drop’ is only non-literal when it co-occurs with ‘names’, that is, when it is part of the idiom ‘drop names’ (Fernando 1996). Literal idioms, the last subclass of idioms as proposed by Fernando (1996), represent expressions that are rather straightforward in structure and meaning. As such, they merely represent a sum of constituents’ literal meaning in a phrase, e.g., ‘in any case’.

Furthermore, Grant (2003) offers a summary of the most prominent categories of idioms on the continuum of idiomaticity, semantically speaking. Next to the classification proposed by Fernando (1996), these include (Grant 2003: 25):

- semi-opaque idioms whose meaning might be inferred from the meaning of constituents, although this might not be straightforward, e.g., ‘sail too close to the wind’ (Fernando & Flavell 1981);
- pseudo idioms which include elements that have no independent meaning on their own, e.g. ‘spic and span’ (Makkai 1969);
- full idioms that constitute elements whose literal meaning is not related to the idiom’s semantic interpretation, e.g., ‘butter up’ (Newmeyer 1974);
- figurative idioms that might be understood twofold, both literally and idiomatically, e.g., ‘catch fire’ (Cowie et al. 1983).

Addressing idioms from a functional approach, Glucksberg and McGlone’s (2001) classification also takes into account the degree of compositionality and transparency. Along these two criteria, the authors propose the following subclasses of idioms: non-compositional, compositional-opaque, compositional-transparent, and quasi-metaphorical. With the first type of non-

compositional idioms, there is no relation between the meaning of the idiom and its constituents, as in the idiom 'cheesecake' referring to 'a pinup art'. Under the assumption that an idiom is compositional, there are two subclasses depending on the degree of transparency: compositional-opaque and compositional-transparent. With compositional-opaque idioms, the overall meaning of the idiom does not follow clearly from the meanings of its individual parts**;** however, the meaning of each particular constituent of the given idiom can still constrain both interpretation and use (Glucksberg & McGlone 2001). An example of this subclass of idioms would be the phrase 'kick the bucket' since the verb 'kick' constrains information about the duration or the swiftness of the action. Although 'kick' does not directly contribute to the idiom's figurative meaning, it nonetheless shapes its interpretation as involving a swift action, meaning 'dying swiftly'. For example, saying 'he kicked the bucket all week' would sound peculiar**,** as it clashes with the idiom's implied swiftness. The latter subclass concerns idioms that may be both compositional and transparent. In these idioms, the idiom's constituents semantically correspond directly to the idiomatic meaning. An example of this idiom subclass is 'break the ice'. This example exhibits a one-to-one semantic relation between the verb 'break' with the idiomatic notion of a 'prompt change of an unpleasant social situation', while 'ice' refers to the idiomatic sense of a 'social or interpersonal tension' (Glucksberg & McGlone 2001). The last subclass, namely quasi-metaphorical idioms, pertains to idiomatic expressions whose meaning is transmitted via their "allusional content" (Glucksberg & McGlone 2001: 75). This means that these types of idioms activate a prototypical or a stereotypical instance of a specific category, be it people, events, situations, or actions. The underlying mechanism is one of dual reference since they simultaneously refer to a prototype of a certain concept and characterize some event or a situation as an instance of that concept. An example would be 'my job is a jail', where the word 'jail' alludes to the prototypical image of a place that 'restricts freedom', thereby similarly characterizing the 'job' as an equally restrictive space (Glucksberg & McGlone 2001).

In a similar vein, Keizer (2016) distinguishes between three major cognitive-semantic types of idioms, namely unmotivated, semantically non-decomposable (Type 1), motivated, semantically non-decomposable (Type 2), and motivated, semantically decomposable idioms (Type 3). Keizer's (2016: 992) classification is as follows:

- Type 1: unmotivated, semantically non-decomposable idioms
  to kick the bucket, to shoot the breeze, to pop one's clogs
- Type 2: motivated, semantically non-decomposable idioms:

to hit the ceiling, to go around in circles, to grit one's teeth, to blow one's top, to flip one's lid, to foam at the mouth, to smoke the peace pipe

- Type 3: motivated, semantically decomposable:
  to spill the beans, to touch a nerve, to grasp the nettle, to swallow the bitter pill, to bury the hatchet

Type 1 idioms are unmotivated because there is no obvious link between the literal image and the idiomatic meaning, and semantically non-decomposable because the parts of the idiom do not contribute individually to the figurative meaning. Type 2, illustrated by idioms such as 'to hit the ceiling', includes motivated, semantically non-decomposable idioms. A link between the literal image and the idiomatic meaning can be established, i.e., one could imagine a person 'hitting the ceiling' because they are very angry. Thus, this group is considered to be motivated. With regard to their semantic non-decomposability, there is no direct mapping between the idiom's parts and the figurative meaning it conveys; that is, 'hit' and 'ceiling' do not relate neatly to 'get angry'. Finally, the third type encompasses motivated, semantically decomposable idioms. An example of this category is 'to spill the beans'. Being decomposable, these idioms' parts can be mapped to their figurative meaning. In this case, as introduced earlier, 'spill' and 'the beans' map to 'divulge' and 'a secret'. Similarly, one could link the literal meaning of 'spilling the beans' to 'divulging a secret'. This classification will be revisited in the methodology section, as the empirical investigation of this thesis builds on these examples.

3.5. Idiom Comprehension and Semantic Interpretation

Essentially, the literature distinguishes between two main classes of models for idiom comprehension: direct look-up and hybrid models (Beck & Weber 2020; Glucksberg 1993; Katsarou 2011). The former models suggest that idiomatic meaning is directly activated in the mental lexicon (Beck & Weber 2020; Glucksberg 1993), which is defined as "the cognitive system that constitutes the capacity for conscious and unconscious lexical activity" (Jarema & Libben 2007: 2). The traditional direct look-up models encompass the Idiom List Hypothesis (Bobrow & Bell 1973), the Lexicalization Hypothesis (Swinney & Cutler 1979), and the Direct Access Hypothesis (Gibbs 1984). The first model, namely the Idiom List Hypothesis, originates from psycholinguistics and suggests that idioms are represented in an idiom list stored in memory. According to this model, a linguistic string that might contain an idiom is initially interpreted in its literal sense. The comprehension process is considered complete if the literal interpretation of

the given sequence matches the context. However, if the literal interpretation contradicts the context, the idiomatic meaning is activated by direct retrieval from the mental idiom list. This model assumes that idioms are non-compositional and cannot be recognized or interpreted based on their first constituent solely. In other words, this model assumes that a segment's idiomaticity may only then be comprehended as such if its literal expression fails to be understood (Bobrow & Bell 1973). However, the underlying assumption of this model may be disproved by psycholinguistic time-measurement experiments. One such experiment shows that potentially ambiguous idioms tend to be processed more quickly when used figuratively, whereas their literal use tends to result in a slower processing speed (Ortony et al. 1978). This implies that the idiomatic meaning of an idiom may be accessed directly from memory before the literal interpretation of the idiomatic construction is fully processed.

With the Lexicalization Hypothesis, Swinney and Cutler (1979) argue that idioms are represented as long words and stored together with all the other individual words in the mental lexicon. According to this approach, linguistic processing proceeds normally when an idiom is encountered. As part of this linguistic processing, lexical access, i.e., retrieval from the lexicon, where idioms are stored, takes place. Here, phrases such as 'by and large' are naturally found in the mental lexicon next to their individual elements, 'by', 'and', 'large', etc. Whether the literal or the idiomatic meaning will be activated first depends on the relative speed with which the linguistic processing and idiom, i.e., lexical access can be completed. The authors assume that idiom access is completed more quickly as it does not require the lexical, syntactic, and semantic processing involved in full linguistic analysis. Therefore, familiar idioms would be generally understood more rapidly in their idiomatic than in their literal expression. Based on a series of experiments, Swinney and Cutler (1979) claim to have been able to confirm that idioms are stored and retrieved from the lexicon as any other word, thus rejecting the idea that there is a special idiom list or a special processing mode.

Finally, Gibbs' (1984) Direct Access Hypothesis represents "an extreme version of the Swinney and Cutler model" (Glucksberg 1993: 5). Rather than suggesting that linguistic processing and lexical access co-occur, Gibbs (1984) argues that the linguistic processing of an idiomatic expression may be completely bypassed if the expression is identified as an idiom immediately. In other words, Gibbs (1984) assumes that idiom access may be as fast as to preclude any linguistic analysis at all. Furthermore, based on several psycholinguistic experiments, Gibbs (1984) provided evidence that supports the idea that people have a tendency towards the nonliteral meaning of certain linguistic instances, such as idioms, when used in literal contexts.

Therefore, it can be said that idioms that are used idiomatically facilitate idiom processing rather than complicate it. As Beck and Weber (2020) put forward, merely in the case of the Direct Access Hypothesis, the activation of idiomatic meaning may commence immediately, without retrieving the literal meaning first. With regard to the other two models, literal computing constitutes an obligatory part of the activation process. The rigidity of these models led to the development of hybrid approaches to idiom comprehension.

The so-called hybrid approaches (Beck & Weber 2020) offer an alternative to the traditional direct look-up models. Some of these approaches include the Configuration Hypothesis (Cacciari & Tabossi 1988), and the Hybrid Model of Idiom Representation (Sprenger et al. 2006). The configuration model put forward by Cacciari and Tabossi (1988) assumes that idioms are not encoded in the mental lexicon as separate lexical items, but that they embody complex combinations of simple words. Their meaning is thus associated with certain configurations of simple words, and the idiomatic meaning becomes available whenever sufficient input of the phrase has been processed. Single words that are part of the configuration are the same lexical elements that are accessed during regular language processing, but in this case, they are configured together to form the overall idiomatic meaning of the phrase. In order to define what is meant by sufficient input, Cacciari and Tabossi (1988: 678) introduce the concept of the key. The key of an idiomatic phrase represents an element of that phrase that is semantically more relevant than other parts of the same string. These keys send signals to the listener and initiate the process of evoking the idiomatic configuration and consequently retrieving the full idiomatic meaning. Thus, the identification of the lexical key marks the switching point between idiomatic and literal interpretation. Before hearing the key, the listener attempts to interpret the idiomatic sequence in its literal meaning. However, as soon as the key has been identified, the idiomatic meaning gets activated (Cacciari & Tabossi 1988). Still, it is worth mentioning that this model does not account for the fact that the interpretation might depend on the listener and their cultural background. Moreover, this approach has been criticized for failing to specify how the inherent syntactic constraints of idioms are represented within the framework, as well as for not providing a satisfactory definition of the idiom key.

To address these issues, Sprenger et al. (2006) proposed the Hybrid Model of Idiom Representation, which more distinctly links the meaning of literal elements to the overall idiomatic meaning at a representational level. They introduce the notion of superlemma, which is defined as "a separate representation of the idiomatic phrase on the lexical-syntactic processing level" (Sprenger et al. 2006: 176). According to this model, the selection and processing of an idiom representation closely resembles the selection and processing of a single word. Because

idioms are here represented by their own superlemma, their production follows the same rules of lexical selection as individual words. Upon selection of the superlemma, the syntactic constraints associated with the idiom are also activated in the production system.

3.6. Recognition and Interpretation of Variant Idioms

Having looked at various models that attempt to account for idiom comprehension and processing, this chapter looks at the recognition of idioms when variation is present. Variation in this chapter refers to both idiom variation and context variation. McGlone et al. (1994) challenge the traditional direct look-up models mentioned before, which assume that idioms are non-compositional and stored in memory as fixed expressions. Instead, the authors propose that idioms exhibit lexical flexibility and semantic productivity, meaning that their individual components contribute to the overall figurative meaning and, as such, allow for the creation of novel, yet understandable, variations. For example, one could exchange 'break the ice' with 'crack the ice', exemplifying lexical flexibility that proves interchangeability possible without a substantial alteration of the original meaning. Moreover, the authors suggest that it is also possible to generate new idiomatic meanings that connect both the original meaning of a particular idiom with new words. This is illustrated by substituting 'break' with 'shatter' in the aforementioned idiom 'break the ice'. By using the verb 'shatter' instead of 'break', a new idiomatic meaning is created. This new meaning may be retrieved as "breaking down an uncomfortable and stiff social situation flamboyantly in one fell swoop" (McGlone et al. 1994: 170), illustrating not merely lexical flexibility, but also semantic productivity.

According to the traditional direct look-up models, idioms require merely one operation, which is accessing the idiomatic meaning in the mental lexicon. Under this assumption, the authors argue that comprehension of variant idioms would require are at least six operations (McGlone et al. 1994: 170):

1. Recognizing the utterance as an intentional variant of the original idiom, not simply a speaker error;
2. Retrieving the meaning of the original idiom;
3. Activating word meanings of both variant and original idioms;
4. Comparing the word meanings of the two idiom forms;
5. Identifying the relations between those word meanings and

6. Taking the relation(s) between the word meanings to infer, by analogy, the relation(s) between the meanings of the original and variant idioms.

Following this sequential processing of variant idioms, substituting a word with one that conveys the opposite meaning of an original component in a conventional idiom would lead to the variant idiom as a whole being interpreted with an opposite meaning, e.g. 'get up on the right side of the bed' vs. 'get up on the wrong side of the bed'. Accordingly, if the relation between the original idiom constituent and the new substitute is one of quantity, the meaning of the variant idiom will also be set in the same relation to the original overall meaning, e.g. 'spill a single bean' vs. 'spill the beans'. While accounting for these operations, direct look-up models do not account for compositional features of idioms or their semantic productivity. In order to understand how variant idioms are processed, McGlone et al. (1994) elaborate on experimental findings, which suggest that variants of familiar idioms are understood rather rapidly. The more familiar the idiom in its original form, the more comprehensible the variant. Moreover, it was found that context facilitates the understanding of both original and variant idioms to a similar extent; however, idioms in their canonical, i.e., original form, are processed more quickly than their variants. Additionally, the experimental findings suggest that idioms in their canonical form were processed more rapidly than their literal paraphrases. Interestingly, variant idioms were understood just as quickly as their literal paraphrases, highlighting the similarity in how both types are processed. A possible explanation for this might be that variant idioms require linguistic processing as their literal expressions do, while original idioms do not. Another explanation refers to the type and degree of variation applied to the original idiom. For example, in some cases, it might be more straightforward to access the meaning of the original idiom and then modify the aspect implied by the variant, e.g., accessing 'spill the beans' and then modifying the quantity 'spill a single bean', as mentioned earlier (McGlone et al. 1994).

3.7. Comprehension of Ambiguous Idioms in Context

Ambiguous idioms may be understood as segments that can be interpreted both literally, based on the compositional meanings of their elements, and figuratively, as holistic expressions with their own meaning (Colombo 1993). As mentioned previously, Ortony et al. (1978) showed that using potentially ambiguous idioms in their literal meaning tends to slow down processing compared to using them idiomatically. This result was overwhelmingly confirmed by other

psycholinguistic experiments (Swinney & Cutler 1979; Estill & Kemper 1982; Glass 1983; Gibbs & Gonzales 1985; McGlone et al. 1994), leading to the so-called idiom superiority effect. Against this background, Colombo (1993) looks into whether the activation of idiomatic meaning occurs automatically or if context is the key factor in preactivating idiomatic meaning. Hence, Colombo (1993: 170) refers to the most prominent models of lexical ambiguity, which are the exhaustive access model and the selective access model. According to the first model, all meanings are activated automatically regardless of context bias. Then, in the following step, the most fitting and suitable meaning is selected. In the view of the latter, it is the context that can activate the figurative meaning. Colombo (1993) further argues that when one of the meanings is more prominent or more frequent when compared to the others, it will be activated sooner, regardless of the context.

In order to empirically test the processing of ambiguous idioms and the role context plays in the activation of figurative meaning, Colombo (1993) conducted three experiments. In the first experiment, the idioms were presented in the form of sentences, two per idiomatic expression. One of the two sentences was biasing the literal meaning, whereas the second sentence was biasing the figurative meaning. This experiment employed a cross-modal priming paradigm, which refers to the influence a stimulus in one sensory modality has on the processing of a subsequent stimulus in a different modality (Scherer & Larsen 2011). It was found that priming can be obtained for targets related to both the literal and the figurative meaning of the idiom. However, the priming effect for figurative meanings was weaker, independent of the influence of context. This suggests that while both interpretations are accessible, the activation of the idiomatic meaning takes more time. In contrast, the second experiment employed a self-paced reading task to examine how neutral and literally biased contexts influence idiom processing. Here, Colombo (1993) found no significant differences in reading times between idiomatic expressions and their control counterparts, regardless of the context type. This indicates that in the absence of a clear figurative bias, the figurative interpretation is not strongly activated. Instead, readers tend to favor the literal interpretation of ambiguous idioms when no contextual cues support the figurative meaning. The third experiment built on the previous one by directly manipulating both literal and figurative contexts. It revealed that longer reading times occurred when the figurative meaning was biased, particularly in segments of the sentence following the idiom. The results of this experiment also indicate a variability in the extent to which the figurative meaning is activated. In contrast, no such delay was observed in literally biased contexts, implying that the literal interpretation is more readily processed unless the context strongly suggests otherwise.

More recent studies, such as Beck and Weber's (2020), show that context facilitates idiom comprehension by biasing readers toward either a figurative or literal interpretation. Based on a self-paced reading experiment, it has been shown that idioms were processed more quickly when the surrounding context supported the intended meaning, whether literal or figurative. However, contexts that encouraged a literal interpretation only facilitated the process when the idiom itself could plausibly be understood in a literal way.

Moreover, Tilmatine et al. (2021) focus on the readers' individuality, that is, the extent to which contextual facilitation varies from person to person with regard to idiom processing. Their study suggests that individuals vary significantly in their ability to recognize and comprehend idiomatic expressions. These differences are influenced by cognitive and linguistic factors like visual working memory, reading comprehension skills, and non-verbal processing speed.

In another self-paced reading experiment, Arnon and Lavidor (2022) presented the participants with pairs of almost identical sentences that included ambiguous idioms. The ambiguity was only resolved in the third part of the sentence, which was either literal or figurative. The researchers suggested that, if two meanings of an ambiguous idiom are activated until the more suitable one is selected and the other is inhibited, there should be a systematic difference in reaction times between the two types of sentences. Their experiment proved this hypothesis correct, as they found that reading times for figurative sentences were shorter than for literal ones.

4. Idiom Translation

As the previous chapter has shown, idioms are complex linguistic units that resist a unanimous definition and classification within academia. Questions regarding their compositionality and the viability of various classification schemes remain contested. In other words, the extensive literature on idioms reflects a wide range of perspectives on core features of idioms. Building on this discussion about the ongoing lack of consensus in defining and classifying idioms, the following chapter aims to contextualize these findings within the field of idiom translation.

4.1. Detecting Idioms in Translation

The translation of idioms presents a considerable challenge for translators, as it requires more than merely converting complex lexical items from one language to another. Indeed, idioms are a particularly tricky area for translators to navigate, as they often require a deep understanding of the cultural nuances and stylistic conventions of the source language to accurately convey the original meaning in the target language (Abdelaal & Alezzawie 2019). It is therefore imperative that translators take into consideration cultural and stylistic differences as well as diverse discourse strategies present in both the source and target languages (Abdelaal & Alezzawie 2019; Razmjou 2004). To this end, a detailed examination of the source text's underlying structure is of great importance and appears to be a prerequisite for a successful translation. Aldahesh (2013) adds that the semantic, syntactic, and pragmatic complexities inherent in idioms make idiom translation particularly challenging.

Although the existing literature on idiom translation also considers the various definitions of idioms presented in the previous chapter, a detailed examination of idioms through the lens of translation studies reveals slight differences in perspective on how idioms should be classified. For example, Adelnia and Dastjerdi (2011) offer a broader overview of linguistic segments that they classify under idioms. They argue that a translator must differentiate between idioms and non-idioms in order to be able to correctly identify, comprehend, and translate idioms. According to Adelnia and Dastjerdi (2011), colloquialisms, proverbs, slang, allusions, and phrasal verbs represent subtypes of idiomatic expressions. The authors define colloquialisms as being characteristic of and appropriate for casual, ordinary, familiar, or informal conversation, as opposed to formal speech or writing. Such expressions are employed in everyday discourse, e.g., ‘he died of laughter’ as an informal way of expressing the phrase ‘he laughed a lot’. Unlike colloquialisms, proverbs are employed to render speech more concrete and easily

understandable, thereby enhancing its reliability and validity. Some examples of proverbs include ‘in Rome do as the Romans do’ and ‘no pain, no gain’. Furthermore, Adelnia and Dastjerdi (2011) define slang as the use of highly informal words and expressions that are not considered to be the standard language. It is frequently employed as a means of circumventing words that are deemed inappropriate or taboo. The penultimate group of idioms, according to Adelnia and Dastjerdi (2011), refers to allusions, which are defined as a figure of speech that refers to a place, event, literary work, myth, or work of art, either directly or by implication. Finally, phrasal verbs are defined as a combination of a verb and one or more additional segments that frequently convey a meaning that differs from that of the original verb. They are typically employed in informal contexts, as opposed to more formal verbs, e.g., the informal use of ‘to go on’ instead of ‘to continue’, or the use of ‘to hand in’ instead of ‘to deliver’.

Baker (1992) offers a different classification and distinguishes between idioms, fixed expressions, proverbs, and collocations. She highlights the importance of distinguishing idioms from collocations due to differences in the transparency of their meaning and the flexibility of their linguistic patterns. Collocations, as Baker (1992) explains, are flexible and allow for variations in their form and use, exemplified by phrases such as 'deliver a letter', 'a letter has been delivered' and 'delivery of a letter' (Baker 1992: 62). This flexibility demonstrates the adaptable nature of collocations within the language system. By contrast, idioms are described as “frozen language patterns that allow for little or no variation and often carry meanings which cannot be deduced from their individual components” (Baker 1992: 63). Baker (1992) notes that under normal circumstances idioms resist change, such as changing word order, deleting or adding words, substituting words, or changing grammatical structure, except in special cases like jokes or puns. This rigidity highlights the fixed and unchangeable nature of idioms in language use. While fixed expressions such as ‘ladies and gentlemen’, ‘having said that’or ‘as a matter of fact’ also allow little to no variation in form, unlike idioms, they typically convey transparent meanings (Baker 1992).

Stressing the importance of differentiating between idioms and other expressions as noted above, Baker (1992) postulates two main areas to consider when engaging with idiom translation. These are the ability to recognize and interpret an idiom correctly, and the difficulties involved in conveying the various aspects of meaning that an idiom or a fixed expression carries in the target language. These difficulties are, however, much more pronounced in idioms than in fixed expressions. Therefore, when dealing with idioms, the first challenge for a translator is to recognize that there is an idiomatic expression in the first place, as this is not always

obvious. Idioms vary in their recognizability, with some being more easily detectable than others. Easily recognizable idioms often include those that violate truth conditions (Baker 1992: 65), such as ‘it's raining cats and dogs’, ‘throw caution to the wind’, ‘storm in a teacup’, ‘jump down someone's throat’, and ‘food for thought’. In addition, idioms that appear grammatically ill-formed, such as ‘trip the light fantastic’, ‘blow someone to kingdom come’, ‘the powers that be’, and ‘the world and his friend’ are also considered to be more conspicuous. This implies that simile-like structures, especially those beginning with ‘like’, such as ‘like a bat out of hell’, and ‘like water off a duck's back’, tend to signal non-literal interpretations. In general, the more challenging and nonsensical an expression appears in a given context, the more likely it is that a translator will identify it as an idiom. Therefore, opaque idioms that initially appear nonsensical can be advantageous if a translator has access to extensive reference materials and monolingual dictionaries of idioms or can consult a first language speaker.

Still, idioms are prone to misinterpretation by those who are not already familiar with them. Some idioms are misleading because they appear transparent, offering a plausible literal interpretation if contextual signals indicating their idiomatic meaning are missing. Moreover, many idioms in English, and probably in all languages, have both literal and idiomatic meanings, making them ambiguous, as introduced in the previous chapter. For example, ‘to go out with’, meaning ‘to have a romantic or sexual relationship with someone’, and ‘to take someone for a ride’, meaning ‘to deceive or cheat someone’, can be interpreted in either direction. Such idioms can be intentionally manipulated evoke both their literal and figurative meanings. In such cases, a translator unfamiliar with the idiom might accept the literal interpretation and miss the idiomatic nuance.

Furthermore, there may be one or more closely related expressions in the target language that are similar in form but completely or partially different in meaning. For example, the English idiom ‘Has the cat got your tongue?’ is used to encourage someone to answer or participate in a conversation, especially if their silence is becoming annoying. In French, a similar expression ‘donner sa langue au chat’, which literally translates to ‘give your tongue to the cat’, means to ‘give up’ (Baker 1992: 66). These instances of superficially and formally identical or similar idioms with different meanings across source and target languages pose significant challenges for translators unfamiliar with the source idiom, as they may mistakenly apply an incorrect target-language interpretation. Thus, translators need to recognize not only how speakers and writers use and manipulate certain idiomatic features, such as variant idioms discussed earlier, but also remain alert to possible confusion caused by seemingly similar expressions in the source and target languages.

Additionally, a translator must be aware of the collocational environment surrounding any expression with an ambiguous meaning. Idiomatic and fixed expressions have distinct collocational patterns, forming relationships with other items in the text as single units and belonging to lexical sets different from those of their individual elements. For example, the idiom 'to have cold feet' does not refer to the separate collocations of 'cold', associated with, for example, the words 'feel', 'weather' or 'winter', or 'feet', associated with 'socks' or 'smelly'. Instead, 'having cold feet' has an idiomatic meaning unrelated to its literal components and is therefore associated with different collocates. The ability to discern meanings through collocation is an invaluable asset for a translator working from a foreign language. This skill is often subsumed under the general practice of "relying on context to disambiguate meaning" (Baker 1992: 66), which involves using knowledge of collocational patterns to decipher the meaning of a word or phrase. While understanding collocational patterns may not always reveal the meaning of an idiom, it can often help identify the presence of an idiom, particularly when it has both literal and non-literal meanings. Given these difficulties, the next chapter will take a closer look at issues in idiom translation.

4.2. Issues in Idiom Translation

The previous chapter has already touched upon the challenges translators encounter when engaging with idiom translation. Next to these, it is typical for a source language idiom to lack an exact equivalent in the target language (Adelnia & Dastjerdi 2011). This is a significant factor translators should consider and tackle. Furthermore, as Adelnia and Dastjerdi (2011) explain, the expression of concepts and realities differs between languages, with each language employing its own distinctive linguistic conventions. Consequently, on occasion, the same concept or idea is realized differently in two distinct languages. The translators' objective is thus to convey the intended meaning in a way that does not result in any loss of comprehension (Adelnia & Dastjerdi 2011). Due to these reasons, among others, the translation of idioms has traditionally been regarded as a challenging decision-making process, necessitating a high degree of expertise and creativity. In order to reach the right decision and complete the translation process successfully, it is crucial to possess a profound understanding of both the source and the target language (Kovács 2016). Specifying the concept of profound understanding of a given language pair, several authors argue that it is essential for translators to work in their first language rather than their second language in order to produce a correct and successful translation (Baker 2018; Hervey & Higgings 2002; Kovács 2016; Reiss 2014). The underlying logic behind this recommendation is based on the widely acknowledged fact that individuals who are first language

speakers have a deeper understanding of the linguistic and cultural nuances of their first language than those who have studied a second language at school or university (Kovács 2016). Similarly, Hervey and Higgins (2002) highlight that translator training programs tend to prioritize translations into the first language, as this approach facilitates the production of higher-quality translations than those created through the process of translating into a second language. Within this background, Reiss (2014) emphasizes the importance of providing readers in the target language with translation that can be understood and perceived in a manner that is aligned with their cultural context. However, due to the growing market demand for translations, translators frequently find themselves obliged to translate into their second language. In such circumstances, it becomes even more crucial to prioritize the incorporation of culture-specific elements into translator training, as well as the examination of the diverse strategies that may be used in the translation process (Reiss 2014).

Like Baker (1992, 2018), Mollanazar (2004) also notes that the first step in idiom translation is to acknowledge and detect their presence in the text. Should this fail, translators might fall into the trap of translating idioms literally, which results in "word-for-word translation of idioms [that is] often nonsense or even sometimes amusing" (Mollanazar 2004: 52). Although this task may sound simple, it is often underestimated. While acknowledging this, Baker (1992) argues that the difficulties involved in translating an idiom from one language to another are more complex than those of simply interpreting its meaning. In a similar vein, Glucksberg and McGlone (2001) suggest that the process of recognizing, comprehending, acquiring, and translating idioms may be influenced by the various forms and contexts in which they may be encountered. They propose a notable distinction between the processes of acquiring and comprehending idioms, e.g., 'spill the beans', and their identification when presented in modified forms, e.g., 'he didn't spill a single bean'. The meaning of an idiom variant is not directly stored in memory, which means that complex cognitive processes must be initiated in order to recognize and comprehend an idiom as such (Kovács 2016). Therefore, Baker (2018) argues that focusing on a particular quality of idioms, such as transparency or opaqueness, appears to be misleading. Baker (1992) argues her case by stating that, from a translator's perspective, an opaque idiom might sometimes be an easier translatory undertaking than a transparent one.

Concretely, Baker (1992) addresses the following difficulties when translating idioms: (a) the lack of a direct equivalent in the target language, (b) an existing actual equivalent in the target language that cannot, however, be applied in the same context in the target language as the idiom from the source language, (c) an idiom having both literal and a figurative meaning used as wordplay, which creates difficulties when deciding on a translation strategy – be it

literal or figurative meaning, and finally, (d) the differences between stylistic conventions and the presence or absence of idioms in particular genres which are language-specific.

Firstly, because of the unpredictable ways in which languages convey meaning, idioms and fixed expressions in one language may not have direct equivalents in the target language. Meanings may be expressed by a single word in one language, by a transparent fixed expression in another, or by an idiom in a third. Therefore, dealing with the lack of equivalents for idioms and expressions across languages seems to be an integral part of a translator's job (Baker 2018). Moreover, similar to single words, idioms and expressions can be culturally specific. Examples include phrases associated with particular social or religious contexts, such as 'Merry Christmas' or 'say when', which may not have direct equivalents in other languages. Basnett-McGuire (1980: 21) notes that 'say when' is associated with the English social behavior and has no equivalent in French or German. Formal correspondence expressions such as 'yours faithfully' and 'yours sincerely' in English illustrate this problem further. However, this does not imply that idioms and fixed expressions that contain culturally specific elements are necessarily untranslatable. The challenge does not lie in the specific elements that make up an expression, but in its meaning and its cultural context. For example, the English expression 'to take coal to Newcastle', referring to the abundance of coal at Newcastle, has a German counterpart 'Eulen nach Athen tragen'. Both convey the idea of providing something superfluous (Grauberg 1989). Similarly, in Welsh, 'it rains old women and sticks' conveys the same meaning as the English idiom 'it rains cats and dogs' (Baker 2018).

Secondly, although an idiom or fixed expression may have an analogous counterpart in the target language, the context in which they are used may differ. For example, the two expressions may carry different connotations, or they may not be pragmatically interchangeable. For example, the English idiom 'to sing a different tune', which means to express a change of opinion that contradicts previous statements or actions, has a Chinese equivalent that literally means 'to sing different tunes/to sing a duet'. While both idioms essentially depict conflicting points of view, the Chinese version often has a strong political connotation and can sometimes imply that the perspectives are complementary rather than conflicting, e.g., 'to sing a duet' (Baker 2018). Similarly, the idiom 'to go to the dogs', which means to deteriorate in quality, has a German counterpart 'vor die Hunde gehen'. However, unlike the English idiom, which can refer either to a person or a place, the German expression is only used in relation to a person and often implies dying or perishing (Baker 1992).

Thirdly, another difficulty pertains to the dual meanings of some idioms. As Baker (1992) postulates, an idiom in the source text can sometimes be used to express both its literal

and figurative meaning at the same time. Achieving this in the target text is difficult unless the target idiom mirrors the source idiom in terms of both structure and meaning. To illustrate her point, Baker (1992) offers the following example:

> In creating Lord Peter Wimsey, Dorothy L. Sayers demonstrated all the advantages of the amateur detective. As a wealthy dilettante, he could follow clues without the mundane necessity of earning a living. His title as the younger son of a duke played to reader snobbery and the obsessive fascination some readers have with the aristocratic lifestyle, or at least their imagined version of it. He had enough influence to **poke his nose** [emphasis kept] into other people's private affairs where less aristocratic **noses** [emphasis kept] might have quickly become bloody. (Baker 1992: 70)

This idiomatic play can only be replicated in languages such as French or German, which have a similar idiom referring to meddling in other people's affairs and incorporating the notion of a nose. Abdelaal and Alezzawie (2019) support Baker's claim, providing an example based on the idiom 'birds of a feather flock together'. This is an example of an expression that can be taken literally and still be meaningful, despite having a different idiomatic meaning. This idiom illustrates that birds of the same species, such as sparrows with sparrows and robins with robins, tend to flock together, indicating a natural tendency towards homogeneity within species. However, the phrase is generally used figuratively to describe people, suggesting that individuals with similar characteristics, attitudes, or interests often 'flock' with each other. While this expression may have been historically used to refer to similarities in race or ethnicity, its modern usage often refers to like-mindedness or similarity of character. Conversely, idioms such as 'kill two birds with one stone', which refers to achieving two goals with one action, can sometimes be interpreted literally. This is, however, not a common occurrence (Abdelaal & Alezzawie 2019).

Finally, according to Baker (2018), there may be significant differences between source and target languages with regard to the idiom usage conventions in written discourse. In English, idioms are prevalent in different types of texts, although their presence is not universal. For example, idioms are rarely used in news reports in the quality press, but are quite common in advertisements, promotional material, and tabloid press. Furthermore, the use of idioms in English depends largely on stylistic preferences. In contrast, in languages such as Arabic and Chinese, where there is a clear separation between written and spoken discourse due to the high degree of formality associated with written texts, idioms are seldom employed in written discourse. Fernando and Flavell (1981: 85) emphasize the different rhetorical effects of idioms across source and target languages, concluding that "translation is an exacting art". Due to this, idioms, more than any other language element, require translators to be not only accurate but

also highly sensitive to the rhetorical nuances of a given language (Fernando & Flavell 1981). Having outlined the challenges of idiom translation, the following subchapter will explore strategies and techniques for addressing and overcoming these issues.

4.3. Strategies in Idiom Translation

Identifying the problems and difficulties in idiom translation represents merely the first step toward a successful and accurate translation. Recognizing the typical pitfalls is important; however, a successful application of strategies to overcome these issues is crucial for producing a quality translation. Which strategy is the most appropriate depends on several factors, the most significant of which is the context (Baker 1992). Before engaging with the different perspectives on idiom translation strategies, it is necessary to clarify the terminology in question. As Harmon (2021) rightly points out, across translation studies, scholars differentiate between a translation strategy (Hatim & Munday 2003; Jean Delisle et al. 2004), sometimes referred to as method (Newmark 2008; Vinay & Darbelnet 1995), and a translation technique (Molina & Hurtado Albir 2004), also known as procedure (Hatim & Mason 2005; Newmark 2008). Newmark (2008: 81) points out that "[w]hile translation methods relate to whole texts, translation procedures are used for sentences and the smaller units of language". Practically, this means that translation strategies refer to the general approach the translator resorts to when translating a text, whereas translation procedures can be understood as individual, concrete solutions to specific sections of the text. For the purpose of this thesis, the terms strategy and method will be used interchangeably.

To frame the discussion of translation strategies specific to idioms, this section first looks into translation strategies in general. Newmark (2008) proposes several translation solutions, which include word-for-word translation, literal translation, faithful translation, semantic translation, adaptation, free translation, communicative translation, and finally, idiomatic translation. Word-to-word translation concerns preserving the original word order and translating words individually by opting for their most common meanings regardless of the context. This method is useful to understand the "mechanics of the source language" (Newmark 2008: 46) or as a preliminary preparatory step before applying a different translation strategy. Similarly, in the case of literal translation, lexical items are also translated individually, out of context. This translation strategy pertains to a method where the translator looks for the closest equivalents of the source text's grammatical constructions in the target language, e.g., word to word, group to group, collocation to collocation etc. Newmark (2008) understands literal translation as the basic translation procedure, upon which both semantic and communicative translation are

based. Moreover, Hervey et al. (2013), who also suggest these two strategies, characterize them as source text-oriented methods.

Faithful translation seeks to precisely reproduce the contextual meaning of the original while adhering to the grammatical structures of the target language. Culturally specific terms are transferred, and possible grammatical or lexical irregularities, i.e., any deviation from source language norms given in the source text, are retained. By doing this, the translation attempts to be completely faithful to the intentions and the way the message is expressed in the source text by the original author. Furthermore, the mere difference between the faithful and semantic translation is their primary focus (Newmark (2008). While faithful translation focuses on accurately reproducing the contextual meaning of the source text within the grammatical system of the target language, semantic translation is more concerned with aesthetic values and, as such, focuses more on conveying the naturalness of the source language in the target language.

The strategy of adaptation represents the "freest form of translation" (Newmark 2008: 40). Due to the free spirit of this method, it is typically used for plays, where cultural nuances are translated, but the text allows for a certain degree of rewriting. Free translation, where free is used in a different sense than the freest translation, "reproduces the matter without the manner" (Newmark 2008: 40). This feature renders it a free method that does not adhere to the form of the source text, usually resulting in a paraphrase much longer than the original. Finally, while communicative translation seeks to convey the precise contextual meaning of the original so that both content and language are easily understandable to the target audience, idiomatic translation involves reproducing the message of the original text in the target language by using an idiom or collocation that was not present in the source text.

Narrowing the scope down to idiom translation, Veisbergs (1996: 241) noted that idioms represent "one of the stumbling blocks for translators". Nida and Tabes (2003) add that "idioms are some of the most obvious candidates for semantic adjustment, for the very fact that they are idioms means it is unlikely that the same type of distinctive form will have the same meaning in another language" (Nida & Tabes 2003: 106). Nida and Tabes (2003) classify these adjustments into three categories: (a) from idioms to non-idioms, (b) from idioms to idioms, and (c) from nonidioms to idioms. On a general note, Veisbergs (1996) argued that there is a broad consensus among linguists that idioms should be translated with idioms. This view is still strongly represented today in both academic and applied translation research (Chung 2024; Donthi et al. 2025; Nasiri Moghaddam & Fatehi Rad 2021; Zaitova et al. 2025). Furthermore,

there used to be agreement on the typical translation solutions for idioms. Earlier, these strategies encompassed translation by equivalent, analogue, substitute idiom, and translation loan, thereby agreeing to the possibility of idiom loss in the translation process (Burger 1973; Kade 1976; Barchudarow 1979; Fernando & Flavell 1981; Glaser 1986). Later on, Wang and Wang (2013) introduced a strategy for translating idioms that was partially aligned with the previous commonly accepted view. They proposed literal translation, free translation, abridged translation, and borrowing translation as possible solutions. The first strategy, namely literal translation, refers to keeping the original content and message unadjusted. The authors argue that this strategy, while keeping the original reference unaltered, might help "fully spread the primitive culture" (Wang & Wang 2013: 1695). This means that some idioms have traditionally been adopted into the target languages and are being used without speakers being aware of their actual origin. Clearly, literal translation cannot be applied to all idiom types, as equivalent idioms often do not exist in the target language, which can lead to misunderstandings of the original text. Thus, translators might opt for the second strategy, namely free translation, to accurately express and transfer the meaning of the original idiom. Abridged translation strategies, which concern the method of removing repeating sections, are relevant for languages like Chinese that have inherently repetitive grammatical structures. Although these repetitions sound natural in the source language, they would sound awkward or redundant when translated into the target language. Finally, the last strategy refers to borrowing translation. With this strategy, the original term from the source language is kept in the target translation, possibly without explanation. This is feasible in cases where two languages exhibit similarities in the way they express certain feelings and experiences.

Building on the challenges presented in the previous chapter, this section will provide a brief overview of the strategies Baker (1992, 2018) proposes for overcoming these obstacles. This choice is supported by the fact that Baker's (1992) seminal work on idiom translation remains the foundational reference in idiom translation studies. Her proposed strategies are consistently cited and applied across a vast majority of scholarly work on idiom translation (Chen 2009; Strakšienė 2010; Adelnia & Dastjerdi 2011; Motallebzadeh & Tousi 2011; Shojaei 2012; Akbari 2013; Kovács 2016; Abdelaal & Alezzawie 2019; Rad 2021), making her model one of the most prominent frameworks for analyzing idiomatic transfer. Overall, Baker (1992) offers several translation strategies to overcome the issues mentioned in the previous chapter.

The first strategy Baker (1992) proposes is using an idiom of similar meaning and form. As the name itself reveals, this strategy pertains to employing an idiom in the target language that carries approximately the same meaning and consists of equivalent lexical items as the

idiom in the given source language. However, although this strategy might sound logical and simplistic, it is rarely the case that a language pair exhibits idioms of similar form and meaning for the exact same occurrences. Baker (1992: 72) offers the following example to illustrate this strategy:

(1) Source text

The Sultan's magnificent income was distributed impulsively at his command. **The rain fell on the just and on the unjust** [emphasis kept].

Target text (French)

Le revenue fabuleux du Sultan était distribué sur un simple ordre de sa part. **La pluie tombait aussi bien sur les justes que sur les injustes** [emphasis kept]**.**

Back-translation:

The fantastic income of the Sultan was distributed on a simple order on his part. **The rain was falling on the just as well as on the unjust** [emphasis kept]. (Baker 1992: 72)

While resorting to this first strategy might be a natural and logical strategy for translators, it is important not to exaggerate in this endeavor by desperately seeking an equivalent idiom rather than focusing on the appropriateness of the selected translation solution. Fernando and Flavell (1981: 82) refer to this as "the strong unconscious urge in most translators to search hard for an idiom in the receptor-language, however inappropriate it may be". Additionally, Baker postulates "the significance of the specific lexical items which constitute the idiom" (Baker 1992: 72) that need to be taken into account when choosing a translation strategy. This significance is displayed in many ways, e.g., whether the idioms are used and referenced elsewhere in the text, or whether the choice of register carries an additional message that needs to be transported into the target language. As described in Chapter 3, in some languages, idioms are widely avoided in (formal) writing. Even if there is an equivalent idiom in the target language, depending on the register and the degree of formality required, it might be best to avoid the first, intuitive urge to look for an equivalent idiom and rather opt for another strategy.

The second strategy Baker proposes concerns using an idiom of similar meaning but "dissimilar form" (Baker 1992: 74). Unlike the previous strategy, this possibility pertains to idioms that convey a similar meaning in the target language, but consist of different lexical items; that is, their form does not resemble the form of the idiom in the source language. According to Baker (1992: 74), such equivalents are often possible to find. This can be illustrated by the following example:

(2) Source text:

Feel the force of my fist [emphasis kept], frozen friend!

Target text (German):

Dir werde ich einheizen [emphasis kept], du Scheusal!

Back-translation:

I will make this hot for you [emphasis kept], monster! (Baker 1992: 74)

As Baker explains, the referee is an ice monster and the German translation means literally "I will be putting heat on you" (Baker 1992: 74).

The third strategy concerns translation by paraphrase. As she argues, this is doubtlessly the most commonly employed translation method when the first two appear to fail; that is, when there are no equivalent idioms either by both meaning and form or just meaning. This technique also appears to be an appropriate choice in cases where the degree of formality, stylistic preferences, and the register do not allow for an idiom solution in the target language. Baker (2018) emphasizes, however, that paraphrases may not always appear accurate to an objective reader when comparing the original text with the source text. An example would be the following translation from English into French (Baker 1992: 76):

(3) Source text:

One frequent criticism of the Manitoba Government throughout the language controversy was that it never seemed to **get a handle on the issue** [emphasis kept].

Target text (French):

Tout au long de la controverse linguistique, on reprocha fréquemment au gouvernement du Manitoba de ne pas réussir, selon toute apparence, a **maitriser** [emphasis kept] la situation.

Back-translation:

For the whole length of the linguistic controversy, the government of Manitoba was reproached frequently for not succeeding, by all appearances, in **mastering** [emphasis kept] the situation. (Baker 1992: 76)

As this example illustrates, the translation solution does not contain an idiom, but rather a paraphrase in the form of a synonym of the given idiom expressed by one verb.

The fourth strategy proposed by Baker (1992) refers to translation by omission. As she explains, similarly to single words, idioms may also be omitted in the target text. This may be

due to a lack of an equivalent in the target language, or to encountering a sequence that does not lend itself to paraphrasing or cannot be stylistically transferred into the target language. Baker (1992: 77) displays her point by providing the following example of a translation from English to Arabic. The original Arabic translation has been omitted due to an inability to transfer the Arabic sentence correctly, as no digital version is available given the year of release:

(4) Source text:

It was bitter, but funny, to see that Professor Smith had doubled his own salary before recommending the offer from Fayed and added a pre-dated bonus **for good measure** [emphasis kept].

Back-translation from Arabic:

It was regrettable, even funny, that Professor Smith had been able to double his salary twice before offering his recommendation to accept Fayed's offer, and that he added to this a bonus, the date of which had been previously decided on. (Baker 1992: 77)

Baker (1992) offers another and fifth idiom translation strategy, namely translation by compensation. However, she highlights the complexity of this strategy, as well as the necessity to elaborate on it in greater detail. This strategy is the only one that is not exclusive to idiomatic expressions in Baker's (1992) representation. It involves either omitting or downplaying the idiomatic nature of a given segment, while ensuring that the idiomatic meaning is conveyed elsewhere in the text. Puns are, for example, a suitable candidate for this strategy. Manson (1982) states that due to the inability to translate puns accurately at the points where they occurred in the text, the translators of 'Asterix' have at times omitted them but have reinserted English puns of equivalent impact rather than equivalent meaning at other instances of the cartoon. Hatim and Mason (2005) reiterate Baker's argument, adding that scholars generally agree that puns and phonesthetic effects, that is, the effects related to the aesthetics of sound, are suitable elements for the compensation strategy. However, they disagree that compensation is merely limited to puns and phonesthetic effects and argue for a broader application of this strategy. While suggesting that all puns, alliterations, rhyme, slang, and metaphors can be compensated in translation, Newmark remarks that "compensation is the procedure which in the last resort ensures that translation is possible" (Newmark 1991: 143-144), making it a 'final tool' when other translation strategies fail. Generally, a number of authors reflect on compensation as a valuable translation method (Hervey & Higgins 1992; Hervey & Higgins 2002; Hatim & Mason 2005; Newmark 2008, Motallebzadeh & Tousi 2011). Hervey and Higgins (2002) explain that, when

a straightforward translation (whether literal or not) would result in a significant loss of meaning, this loss can be mitigated by intentionally introducing a less significant one. This way, the important effects of the source text are roughly preserved in the target text through different methods than those used in the source text. Put simply, one type of translation loss is balanced out by deliberately allowing another, less severe one. The authors stress that resorting to this strategy may be reduced to "a simple balancing loss against loss in order to do more justice" (Hervey & Higgins 2002: 44). They highlight the advantages of this strategy and encourage translators to make these compromises where fitting, taking into account the nature and the purpose of both source and the target language. Furthermore, compensation is more relevant in non-technical texts than in technical ones. In translating documents such as legal texts, financial statements, or scientific papers, the need for compensation is minimal. In these cases, the translator identifies and uses the appropriate target language expressions. However, in most other genres, a successful translation might require compensation, especially in the case of literary texts (Hervey & Higgins 2002). Building on this recognition, Motallebzadeh and Tousi (2011) apply and explore this method in the context of idiom translation in literary texts, arguing that it merits further investigation.

The outlined strategies for translating idioms assume that the translator understands the context, nuances and cultural elements of both the source and the target language. These strategies help translators weigh up multiple options and select the one that best fits the given translation task. However, these qualities are inherent to human translators. While human translators can usually interpret and render idioms effectively by drawing on contextual cues and cultural knowledge, machine translation systems often struggle in this area, frequently producing literal, inaccurate or nonsensical outputs (Baziotis et al. 2023; Dankers et al. 2022; Rambelli et al. 2023; Tian et al. 2023).

4.4. Issues in Machine Translation of Idioms

A number of scholars agree that idioms pose a highly challenging task for machine translation (Baziotis et al. 2023; Dankers et al. 2022; Rambelli et al. 2023; Tian et al. 2023; Volk 1998). While the previous two sections addressed issues in the human translation of idioms as well as strategies for overcoming them, this chapter aims to identify obstacles present in the machine translation of idioms. Generally, it can be said that the first step in machine idiom translation is the same as for human translation, which is to recognize and identify the presence of an idiom (Baziotis et al. 2023; Dankers et al. 2022; Volk 1998). Volk (1998) states that the recognition of idioms, both in machine and human translation, is difficult due to various modifications and

"discontinuous spreading" over a clause (Volk 1998: 1). As introduced earlier, idioms contradict the principle of compositionality in language since their meaning cannot be predicted from the meanings of their part (Liu et al. 2023). Once an idiom is identified, a translation system must translate it as a whole due to its complex nature (Volk 1998). Identifying non-compositional and partly compositional idioms should not be a difficult task, as their lexical features are relatively fixed. In contrast, compositional idioms pose a greater challenge since their variation must be taken into account (Volk 1998). Already in 1998, Volk noted that machine translation faces significant challenges due to the inherent ambiguities in natural language, which can only be resolved by applying semantic features, context, and world knowledge. However, as Volk (1998) argues, it is both difficult and labor-intensive to acquire this knowledge. As a result, typically, only the most salient semantic features are incorporated in commercial MT systems, which leads to limited translation quality in commercial MT systems. Interestingly, decades later, the research on machine translation of idioms still refers to Volk's work, highlighting the same underlying issues in both human and machine translation (Anastasiou 2010; Zaninello & Birch 2020; Meenal & Govindarajan 2023).

Traditional MT systems, such as rule-based machine translation systems, which Volk (1998) refers to in his work, rely heavily on word-by-word processing and fixed word order. To illustrate this, Volk (1998) addresses the proposed integration of the so-called local grammar rules. For example, in the case of the expression 'to take care of', the local grammar rules instruct the system to work from left to right and process the word 'take' as a regular verb. Then, the system encounters the noun 'care', which is marked in the lexicon as the head of an idiom and a local grammar rule. This rule, then, asks the system to look to the left for a form 'take'. In case 'take' is encountered, the system must check for the preposition 'of'. In this case, all conditions are stored with the noun 'care', together with the fact that the idiom as a whole functions as a verb. However, Volk (1998) argues that this approach is very much word-oriented and strict in its command to look left or right. This strictness is hardly applicable to languages like German due to the language's flexible syntax and variable verb positioning. Idioms such as 'jdn. mit Argusaugen beobachten', meaning 'to watch someone like a hawk', often appear in discontinuous forms, with their components separated across the clause. Indeed, German verbs may appear in different positions, depending on the clause type. These are illustrated below (Volk 1998: 16):

(5) Er **beobachtete** den Mann, der die Bank betrat, **mit Argusaugen** [emphasis kept].
He was watching the man, who entered the bank, like a hawk.

(6) Ich weiß, dass er ihn **mit Argusaugen beobachtet** [emphasis kept].
I know that he is watching him like a hawk.

In response, Volk proposes a more structurally aware approach, that is, MT systems should identify whether all elements of an idiom, both contiguous, e.g., 'mit Argusaugen', and discontinuous, e.g., 'beobachten', co-occur within the same clause and fulfil the idiom's syntactic requirements, such as needing an animate subject and a physical object. By shifting the focus from linear word order to clause-level syntactic structure, MT systems can more reliably detect and translate idioms in morphosyntactically complex languages like German. In a nutshell, Volk (1998) suggests that raising structural syntactic awareness would make idiom recognition and translation feasible.

In 1952, Bar-Hillel delivered his well-known statement on machine translation of idioms, arguing that "[t]he only way for a machine to treat idioms is – not to have idioms" (Anastasiou 2010: 142). Indeed, more than half a century later, it is argued that overly literal translations and lack of broader context are two reasons why machine translation is still not at parity with human translators, especially in the case of literary texts (Matusov 2019; Omar & Gomaa 2020; Poibeau 2022, Liu et al. 2023) Even for high-resource languages such as English, there is still little understanding on the way language models process idiomatic expressions (Boisson et al. 2022). In this regard, Dankers et al. (2022) discuss the state-of-the-art Transformer, on which all three of the NMT models investigated in this master's thesis are (partially) based. Dankers et al. (2022) conclude that NMT models often fail to translate idioms accurately, tending to over-produce compositional, literal translation. Although Dankers et al. (2022) and Volk (1998) address different machine translation technologies, namely Transformer-based NMT and rule-based MT, respectively, they share a common concern. This concern is based on the underlying argument that idioms are often translated word-by-word, resulting in a loss of their idiomatic meaning. This juxtaposition highlights the challenges of advancing machine translation of idioms, as the fundamental issues have remained unresolved for over two decades. Furthermore, Dankers et al. (2022) found that Transformer-based MT models fail to encode idioms as fixed expressions unless explicitly trained to do so, meaning that they tend to treat all phrases, including idioms, compositionally by default. Consequently, this tendency contributes to literal translations of idioms, rendering them a challenging task for NMT (Avramidis et al. 2019; Constant et al. 2017; Dankers et al. 2022; Isabelle et al. 2017). When a Transformer-based MT model produces a correct, non-literal translation, that is, when it successfully identifies an expression as idiomatic, the encoder tends to process the idiom more holistically, treating it as a single lexical unit. In other words, in cases when the model produces a correct translation, the

encoder treats the idiom non-compositionally. This behavior is reflected in the attention mechanism**,** where the components of the idiom are more tightly grouped, and there is reduced interaction with the surrounding context. In the decoder's cross-attention, idiomatic expressions receive less focused attention on specific source tokens**,** further indicating that the model does not rely on word-by-word alignment in these cases. However, Dankers et al. (2022) argue that this is an exception rather than the norm. Furthermore, the authors present a twofold argument: on the one hand, they argue that idiomatic expressions are omnipresent in natural language. On the other hand, they claim that idioms are used less frequently than their elements and, as such, require identification and comprehension before translation. This is especially true for potentially idiomatic expressions (PIEs), introduced earlier as ambiguous idioms. Whether PIEs should be translated literally or figuratively depends on the context.

In a similar vein, Baziotis et al. (2023: 1) comment that idiom translation is "a major open problem in neural machine translation" which still struggles with the translation of rare Multi-Word Expressions (MWE). In accordance with the previous findings, Baziotis et al. (2023) reiterate that NMT models tend to translate idioms literally, that is, word-by-word, which leads to inaccurate and misleading translations. Moreover, the authors also agree that idioms are a more challenging translation task than non-idiomatic occurrences and thus depend heavily on the context, which is also consistent with the previous assertions. Similar statements can be found with Zaninello and Birch (2020), who argue that MWE, including idioms, challenge current NMT systems and remain an open question. Having provided a brief introduction to machine translation, idioms, and the machine translation of idioms, the remainder of this thesis focuses on the empirical research.

5. Related Work

Over the past decade, there has been a growing consensus in academia regarding the increasingly accurate outputs of machine translation (Dankers et al. 2022; Liu et al. 2020; Santing et al. 2022; Wang et al. 2022). Recently, several Transformer models, such as Bidirectional Transformers for Language Understanding (BERT) (Devlin et al. 2019), Bidirectional and Auto-Regressive Transformer (BART) (Lewis et al. 2020), and Text-to-Text Transfer Transformer (T5) (Raffel et al. 2020) have reportedly achieved great success in tackling a variety of Natural Language Processing (NPL) tasks (Santing et al. 2022). These models allow for context representation, meaning they take not only the word itself into account, but also rely on the context, i.e., the surrounding words to the left and right. In almost every NPL task, these models have obtained groundbreaking results (Liu et al. 2020; Santing et al. 2022). State-of-the-art research on machine translation of idioms highlights the importance of context, following previously introduced theoretical approaches. Additionally, much research has been conducted in the field of paraphrasing idioms to replace them with their literal meaning (Liu & Hwa 2016; Tien-Ping & Jia Jun 2021; Zhou et al. 2022). Such approaches first identify the idiom in the source language, replace it with a paraphrased version, e.g., 'kick the bucket' to 'die', and translate the paraphrased version instead of the idiom itself. This helps MT systems to avoid delivering literal translations such as 'den Eimer treten', translating 'die' as 'sterben' instead.

Instead of paraphrasing, Salton et al. (2014) attempted to efficiently translate idioms by implementing a substitution method. Their work evaluates the quality of idiom translation in Statistical Machine Translation (SMT) using the language pair English-Brazilian-Portuguese. Their hybrid method is envisaged as a pre- and post-processing step. Relying on an idiom dictionary, they first identify idioms in the source sentences, which they then substitute in the given sentence with the idiom's literal meaning, noting down sentences that include a substitution. Following this, Salton et al. (2014) check if the original idiom present in the source sentence has an idiomatic equivalent in the target language by consulting a separate bilingual dictionary. This is done for all sentences that contain substitution. Should there be a corresponding idiom in the target language, the translation of the literal meaning of the source language idiom is replaced with the target language idiom. Should there be no equivalent idiom in the target language, this post-processing step is omitted, and the translation is considered done (Salton et al. 2014). However, this model relies heavily on available dictionaries. Furthermore, it does not offer a solution to the translation of ambiguous idioms, i.e., differentiating whether an idiom is used in its literal or figurative meaning.

Although the swiftly increasing accuracy of machine translation is generally indisputable, machine translation of idioms still appears to be a major challenge for NPL technologies, often resulting in inaccurate or unnatural translations (Santing et al. 2022; Škvorc et al. 2022). The reason for this may be attributed to the fact that datasets used for machine translation are not rich in idioms (Fadaee et al. 2018; Santing et al. 2022; Saxena & Paul 2020; Škvorc et al. 2022). For this purpose, a few authors have introduced their own datasets on machine translation of idioms. Agrawal et al. (2018) focus on idiom handling for various NPL tasks based on Indian languages. More precisely, they handle seven languages in addition to English, on which neural MT and statistical MT models were trained. They found that phrase-based SMT can handle idiomatic sentences better than NMT across their language set. Similarly, Fadaee et al. (2018) evaluate the quality of idiom translation of neural MT and phrase-based MT (PBMT) for the language pair English/German. Their primary contribution is the collection of a bilingual dataset, which they automatically built from an existing English-German parallel corpus. The results they obtained point to two conclusions. The first one implies that the performance of both PBMT and NMT in terms of idiom translation is lower than their performance with translations containing few idioms. The second conclusion indicates that the NMT experiments with special reference to the presence of idioms, i.e., special input token, perform better than PBMT, but still worse than the NMT baseline in terms of BLEU or idiom translation quality. Nevertheless, Fadaee et al. (2018) found that the NMT experiment using a special input token to indicate the presence of an idiom in a sentence performs much better than PBMT and NMT baseline in terms of unigram precision, meaning how many individual words were translated correctly, and word-level idiom accuracy score, that is, how well the idiom was translated at the word level. Based on the obtained results, Fadaee et al. (2018) conclude that an annotated dataset of idioms can be used to run targeted models to improve the neural translation of idioms. However, it is necessary to use evaluation metrics specific to idioms, e.g., an adapted form of measuring the accuracy of unigrams focused only on idioms.

6. Methodology

The first step of this empirical investigation involves selecting idioms and test sentences. The second step refers to explaining the choice of machine translation systems under investigation, which are Google Translate, DeepL, and SYSTRAN. Two language students, one of whom is the author of this thesis, annotated the MT translations of the chosen idioms and the respective test sentences according to translation strategies proposed by Baker (1992). The results obtained were evaluated using descriptive statistics, i.e., absolute and relative frequency. To ensure the validity of the chosen method, Cohen's kappa value was measured.

6.1. Data Selection

The empirical research of this master's thesis employs idioms as proposed by Keizer's (2016). She distinguishes between three major cognitive-semantic types of idioms, namely unmotivated, semantically non-decomposable (Type 1), motivated, semantically non-decomposable (Type 2), and motivated, semantically decomposable idioms (Type 3). The underlying differentiation between the three categories is cognitive in nature. With regard to the theoretical part concerning the compositionality of idioms, this master's thesis also aims at examining whether these features are reflected in Google Translate, DeepL and SYSTRAN's translation output, and if so, to what extent. Due to this, Keizer's (2016) typology is considered suitable for the intended investigation. Keizer's classification and the selected idioms this thesis employs are the following (Keizer 2016: 992):

- Type 1: unmotivated, semantically non-decomposable idioms
  to kick the bucket, to shoot the breeze, to pop one's clogs
- Type 2: motivated, semantically non-decomposable idioms:
  to hit the ceiling, to go around in circles, to grit one's teeth, to blow one's top, to flip one's lid, to foam at the mouth, to smoke the peace pipe
- Type 3: motivated, semantically decomposable:
  to spill the beans, to touch a nerve, to grasp the nettle, to swallow the bitter pill, to bury the hatchet

The chosen idioms were translated both in and outside of context. For translation in context, four test sentences per idiom that differ in register type and context length were extracted from the Corpus of Contemporary American English (COCA). COCA is the only large and balanced corpus of American English, allowing users to see the frequency of a feature across different

genres. In doing so, COCA offers a great insight into variation in the English language (Davies 2025). The corpus contains more than one billion words of text, having added 25+ million words each year between 1990-2019. Furthermore, it offers a wide range of genres, including spoken, fiction, popular magazines, newspapers, academic texts, TV and movie subtitles, blogs, and other web pages (Davies 2025). Given its size and the diversity of its genres, COCA seemed to be the right choice for extracting idiomatic sentence examples. Additionally, COCA enables users to specify their search design by allowing them to choose the most commonly identified language units researchers use, namely words, collocations, *n*-grams, and parts of speech (POS) tags for grammatical patterns (Davies 2025).

The choice of context is as follows: two test sentences are in a more formal register, such as news or journals, one example includes an informal context, such as spoken language or TV, while the last example is presented as a longer text. This is done to examine whether different registers and context lengths influence the translation output. These test sentences were extracted by allowing the corpus to suggest all sentences containing any form of the verb in question by capitalizing the verb, e.g., 'SPILL the beans'. In cases of reflexive verbs such as 'to pop one's clogs', the following tag 'POP * clogs' was employed, allowing for different tenses of the verb 'to pop' and any particle that might be encountered between the verb 'to pop' and 'clogs', which resulted, as expected, predominantly in pronouns.

6.2. Machine Translation System Selection

This master's thesis focuses on three leading commercial machine translation systems, namely Google Translate, DeepL, and SYSTRAN. Since all three machine translation systems currently use the same technology, namely neural machine translation, the criteria for choosing these three systems concern aspects other than their underlying technology. These include the number of languages supported, the familiarity of the system among nonprofessionals, and additional features.

Overall, Google Translate appears to be primarily used by non-professionals, whereas leading language services companies (LSC) tend to favour DeepL or SYSTRAN. A survey examining the use of various machine translation systems found that the vast majority of respondents (96.6%) use Google Translate as their primary tool (Kasperė et al. 2021). Among these respondents, 56.2% indicated using Google Translate very frequently, while 19.4% use this system frequently. Moreover, this survey indicated that 16% use DeepL, among whom 6.5% used it very often and 4.5% often (Kasperė et al. 2021). Another study confirms these

findings and reports that 95.2% of their respondents indicated using Google Translate for machine translations (Vieira et al. 2023). Based on these findings, it can be said that Google Translate is the most commonly used machine translation tool. Moreover, Google Translate has been the preferred machine translation service used by researchers to evaluate the accuracy of translating preexisting documents (Rao et al. 2024). In June 2024, Isaac Caswell, Senior Software Engineer for Google Translate, announced that another 110 will be incorporated into Google Translate, thus offering more than 200 language pairs for translation in both directions (Caswell 2024). In contrast, at the time of writing, SYSTRAN supports 55 languages (SYSTRAN 2025c), while DeepL works with more than 30 languages (DeepL 2025c).

However, with regard to translation accuracy, DeepL prides itself on being "the world's most accurate translator", offering "unbeatable translation quality for global business" (DeepL 2025b). DeepL claims to deliver unparalleled accuracy, outperforming Google Translate by 1.3x. The 2024 Industry Survey by the Association of Language Companies (ALC) revealed that 82% of language services companies now use DeepL (DeepL 2024a). This number is significantly higher than the number of those who use other language service providers, with Google Translate scoring 46%. Furthermore, DeepL has increased its share of the machine translation market by 44%, establishing itself as the market leader. This increase is particularly significant in the enterprise market (DeepL 2025b).

SYSTRAN was one of the earliest MT systems available online (Hutchins 2000). With more than 50 years of experience, SYSTRAN has achieved numerous breakthroughs in the field of MT (SYSTRAN 2025b) . It offers software solutions that cater to the needs of corporate and individual customers. Since 1968, SYSTRAN has also been a trusted partner of the US Government and numerous governments around the world (SYSTRAN 2025a). This MT system claims to be the preferred choice for its technological robustness, rapid translation speed, and consistently high translation quality (SYSTRAN 2025a). It prides itself on offering the fastest machine translation performance and providing the widest range of machine translation software solutions on the market, enabling quick and cost-effective customization (SYSTRAN 2025a). In sum, it can be said that Google Translate is the most dominant tool among non-professionals, DeepL the most precise one, while SYSTRAN is regarded as the fastest and the most cost-effective regarding customization.

6.3. Annotation Procedure

As mentioned earlier, the machine translation of the chosen idioms was assessed both in and outside of context. The four test sentences per idiom were classified according to their idiom

type (Type 1, 2, or 3), the place the given idiom occupies within the category, e.g., Type 1, idiom 1, and the sequence of the test sentence for the respective idiom (Type 1, idiom 1, test sentence 1, i.e., T1-1-1). Subsequently, the translations of the test sentences were annotated per MT system according to translation strategies proposed by Baker (1992). As introduced earlier, these include translations with an equivalent idiom, a similar idiom, translation by paraphrase, and omission. As the theory suggests, equivalent idioms are indistinguishable in both meaning and lexical items, while similar idioms have similar meanings but 'dissimilar form' (Baker 1992). Furthermore, paraphrasing is understood as a correct interpretation of the idiomatic meaning translated with a non-idiomatic expression. In addition to the strategies mentioned above, Baker (1992) also introduces compensation as the fifth strategy. However, this would require a rather long input text, which is beyond the scope of this master's thesis. For this reason, this strategy will not be included in the evaluation. Instead of compensation, the author added another category that pertains to incorrect or literal translations of the given idioms. It is important to note that the analysis focuses merely on machine translation of idioms and does not evaluate the accuracy of the whole sequence. Annotations were done on an MT system level, marking the relevant category with 1. With regard to translation outside of context, each idiom was translated in isolation by all MT systems. In this case, the annotation took place based on two categories: idiomatic and literal translation. The aim was to see whether Google Translate, DeepL, and SYSTRAN assume that the phrase should express idiomatic or literal meaning. Again, the relevant category was annotated with 1.

The annotations were made in an Excel document that contains three sheets, one for each MT system. Each sheet contains two tables with respective annotations. The first table includes four test sentences per idiom, totaling 60 examples in context, while the second table represents the translation of idioms outside of context. The evaluation was made by two language students. Annotator 1, the author of this master's thesis, is doing her master's in translation with a focus on Specialised Translation and Language Industry in English and German. Annotator 2 is currently working as a university assistant predoctoral at the University of Graz in the field of English linguistics. Annotator 2 received the Excel file containing three sheets with two tables each, i.e., original test sentences in English and their respective translations, as well as idiom translations without context. Annotator 2 was instructed to read the test sentence carefully and mark the relevant category with 1.

6.4. Analysis Method

The obtained annotations were analyzed using descriptive statistics, which include absolute and relative frequency. The absolute frequency refers to the raw count of occurrences, i.e., the number of times a particular value for a variable has been observed to occur. In this case, the absolute frequency represents the number of times Google Translate, DeepL, and SYSTRAN have delivered idiomatic translations, accurate paraphrases, omitted the idiomatic phrase, or delivered incorrect or literal translations. Unlike absolute frequency, relative frequency describes how often an event occurs in relation to the total number of events, i.e., it calculates the proportion or percentage of the dataset that a value or category represents.

6.5. Validation

As touched upon earlier, the annotations were done by a master's student and a university assistant predoctoral student on a set of 60 examples for each of the three MT systems. Inter-rater reliability and Cohen's kappa (κ) value were used to measure the validity of the method in terms of the agreement between the obtained annotations. Inter-rater reliability indicates the consistency of annotations provided by different annotators for the same item. It measures the extent of agreement between annotations and serves as an indicator of the quality and accuracy of the annotation procedure (Sun 2011). Cohen's κ is considered the most suitable and most widely accepted measurement of inter-rater reliability when the agreement is measured on a nominal scale (Sim & Wright 2005; Von Eye 2006). The formula for Cohen's κ is as follows (Cohen 1960):

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

P_o stands for observed agreements, that is, the proportion of times the raters agreed, while P_e indicates the expected agreement by chance. The two annotators worked independently with nominal, categorical data, which is why Cohen's κ was rendered a suitable measure to ensure the validity of the method. Cohen's κ is obtained as a number ranging from 0 to 1. To interpret the κ, Landis and Koch (1977) propose the interpretation scale depicted in Table 1:

Table 1: Interpretation scale of Cohen's κ (Landis & Koch 1977)

Kappa Value (κ)	**Strength of Agreement**
< 0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect (Excellent)

The obtained Cohen's κ was analyzed in Chapter 7.3 according to this interpretation scale.

7. Results

15 idioms, classified into three different types as suggested by Keizer (2016), were tested. The idioms were analyzed both in context and in isolation to determine whether context influences MT output. In the former case, each idiom was embedded in four test sentences of different lengths, totalling 60 test sentences per system. The following subchapters present the results in context, in isolation, and according to idiom type. To ensure the validity of the method employed, the results of Cohen's kappa are discussed.

7.1. Overview of the Results in Context

Table 2 represents an overview of the results obtained for all three MT systems according to Baker's (1992) translation categories. As the results show, when the translation was idiomatic, the systems predominantly suggested the correct equivalent, with a few instances of similar idiom suggestions. Therefore, the first two categories, i.e., equivalent and similar idioms, will be addressed collectively as idiomatic translations when describing the results, to provide a more comprehensive overview. Another reason for doing so is that the small number of occurrences of similar idioms does not provide relevant insights when considered as a single category. Paraphrasing, meaning correct interpretation of the original idiom, but non-figurative translation, can be interpreted as the next relevant category. Moreover, omission yielded no results. Finally, the number of incorrect translations provides a broader picture of how well the respective system detects and translates idiomatic expressions.

Table 2: Translation in context of Google Translate, DeepL and SYSTRAN in absolute counts and percentages

	Google Translate		**DeepL**		**SYSTRAN**	
	Count	%	Count	%	Count	%
Equivalent Idiom	26	43,33	31	51,66	18	30
Similar Idiom	4	6,67	4	6,67	0	0
Paraphrase	16	26,67	15	25	0	0
Omission	0	0	0	0	0	0
Incorrect/literal translation	14	23,33	10	16,67	42	70

With regard to relative frequency, as shown in Table 2, the results indicate that both Google Translate and DeepL have delivered accurate idiomatic translations in over 50% of the cases.

When also taking paraphrasing results into account, it can be said that both systems provided correct translations in approximately 75% of cases, whether idiomatic or non-figurative but accurate. SYSTRAN appears to perform the worst, with only 30% of the translations being equivalent, and zero results using a similar idiom or paraphrase. Thus, this MT system delivered inaccurate, literal translations of the given idioms in 70% of the cases.

7.2. Overview of the Results Without Context

With regard to idiom translation outside of context, the results indicate a rather scattered picture across the three systems, as depicted in Table 3:

Table 3: Translation in isolation of Google Translate, DeepL and SYSTRAN in absolute counts and percentages

	Google Translate		**DeepL**		**SYSTRAN**	
	Count	%	Count	%	Count	%
Idiomatic translation	6	40	11	73,33	3	0,20
Literal translation	9	60	4	26,67	12	0,80

From a relative frequency perspective, as depicted in Table 3, Google Translate appears to assume that phrases should be translated literally in 60% of cases. Conversely, DeepL indicates a much higher percentage of idiomatic translations, suggesting that it assumes the literal usage of the given idiomatic phrase in 27% of the cases. With SYSTRAN, the results are the other way around, assuming that 80% of the cases are meant to be understood literally.

To juxtapose the proportion of incorrect suggestions, Table 4 summarizes the results obtained in both categories across all three MT systems:

Table 4: Comparison between MT output of Google Translate, DeepL and SYTRAN in and outside of context in absolute counts and percentages

	Google Translate		**DeepL**		**SYSTRAN**	
	Count	%	Count	%	Count	%
Incorrect Translations (context)	14	23,33	10	16,67	42	70
Incorrect Translations (no context)	9	60	4	26,67	12	80

As mentioned earlier, this investigation considers 60 examples in context and 15 in isolation. The results indicate a relative similarity between the way DeepL and SYSTRAN treat idioms

in and outside of context, differing in both cases in ten percentage points. Google Translate, however, displays a larger gap between its output with and without any context.

7.3. Overview of Results According to Idiom Type

In her classification, Keizer (2016) differentiates between three types of idioms. As elaborated in Chapter 3.4, the first type refers to unmotivated, semantically non-decomposable idioms, exemplified by the idiom 'to kick the bucket'. Type 2 is illustrated by idioms such as 'to hit the ceiling' and 'to go around in circles', and includes motivated, semantically non-decomposable idioms. The third type encompasses motivated, semantically decomposable idioms. An example of this category is 'to spill the beans'.

Regarding Google Translate, the results depicted in Table 5 indicate that Type 1 idioms, such as 'to pop one's clogs', seem to confuse Google Translate, resulting in the highest proportion of incorrect translations (0,33%) across all types. In all example sentences within Type 1, equivalent idioms were not suggested. However, similar idioms and paraphrases still make up the majority of the output, meaning that opaqueness is not a definite obstacle for this MT system. Still, the fact that Type 1 is the category with the smallest number of idioms also needs to be taken into account. With regard to Type 2, e.g., 'to grit one's teeth', and Type 3, e.g., 'to swallow the bitter pill', which are both motivated, there seems to be almost no difference in the MT output, resulting in nearly identical portions of accurate and incorrect translations. Furthermore, Google Translate suggested the correct equivalent idiom in over 50% of the cases across Type 2 and Type 3. Correct interpretations with non-figurative translations follow with 18% and 25%, respectively, while incorrect translations account for 18% and 20% of cases.

Table 5: Results of Google Translate grouped by idiom type in absolute counts and percentages

	Type 1		**Type 2**		**Type 3**	
	Count	%	Count	%	Count	%
Equivalent Idiom	0	0	16	57,14	10	50
Similar Idiom	2	16,67	2	7,14	0	0
Paraphrase	6	50	5	17,86	5	25
Omission	0	0	0	0	0	0
Incorrect Translation	4	33,33	5	17,86	5	25

DeepL, as Table 6 shows, delivers slightly different results than Google Translate. Overall, Type 1 was translated correctly, resulting in 25% of translations being inaccurate. Unlike Google Translate, DeepL delivered almost all correct translations for Type 2, achieving an accuracy rate of 93%. However, Type 3 exhibits similar results to Google Translate, with 25% of translations being inaccurate and 75% being correctly transferred into German.

Table 6: Results of DeepL grouped by idiom type in absolute counts and percentages

	Type 1		**Type 2**		**Type 3**	
	Count	%	Count	%	Count	%
Equivalent Idiom	0	0	20	71,43	11	55
Similar Idiom	4	33,33	0	0	0	0
Paraphrase	5	41,67	6	21,43	4	20
Omission	0	0	0	0	0	0
Incorrect Translation	3	25	2	7,14	5	25

SYSTRAN delivered rather straightforward results, as illustrated in Table 7. With Type 1, there was not a single correct translation, resulting in 100% inaccurate, literal translations. As for Type 2, the system still predominantly delivered incorrect translations (64%), while in 36% of cases, the translation was idiomatically correct in the target language. Type 3 follows suit with 60% inaccurate suggestions and 40% correct idiomatic interpretations.

Table 7: Results of SYSTRAN grouped by idiom type in absolute counts and percentages

	Type 1		**Type 2**		**Type 3**	
	Count	%	Count	%	Count	%
Equivalent Idiom	0	0	10	35,71	8	40
Similar Idiom	0	0	0	0	0	0
Paraphrase	0	0	0	0	0	0
Omission	0	0	0	0	0	0
Incorrect Translation	12	100	18	64,29	12	60

Generally speaking, Type 2 idiom group contains several examples of the same idiom being translated differently depending on the context. With regard to Google Translate, an interesting example is ‘to hit the ceiling’. Across the four test sentences, this idiom was translated differently each time: once as an equivalent idiom, once as a similar idiom, once as a paraphrase, and once incorrectly. Another example is ‘to blow one's top’, where, surprisingly, longer contexts

resulted in incorrect translations, while shorter examples provided correct suggestions. The idiom ‘to flip one's lid’ follows suit. Furthermore, the four translations of the idiom ‘to flip one’s lid’ also demonstrated inconsistencies in the accuracy of the translation output.

As far as DeepL is concerned, within Type 1, it interpreted the idiom ‘to pop one's clogs’ incorrectly three times, while providing one correct paraphrase. In this case, the accurate suggestion was the example with the longest context. On the contrary, in the case of ‘to blow one's top’, only one sentence was translated incorrectly (‘sich verausgaben’). However, this sentence is similar in length to the one that was translated correctly, raising the question of whether it is the length of the context or rather the words immediately surrounding the idiom that point to the figurative meaning. The same goes for ‘to flip one’s lid’. In the case of ‘to touch a nerve’, three sentences were translated using an equivalent idiom, while the fourth example was an incorrect translation. Again, it is worth mentioning that the fourth example was the longest one. With SYSTRAN, there were no halfway solutions. It either recognized idioms and translated them with an equivalent expression or produced an incorrect translation.

As illustrated, several idioms were translated differently by the same system depending on the context. Interestingly, the length of the context did not play a crucial role. Examples could be observed where examples with the most text led to an incorrect translation, and vice versa. It was also noticed that one and the same idiom in two short test sentences was sometimes translated correctly and sometimes incorrectly. Therefore, it seems that idiomatic interpretations are produced correctly when the immediately surrounding words aid the figurative meaning unambiguously, regardless of context length.

7.4. Cohen’s κ Results

The Cohen’s Kappa value across all three systems amounts to 0.979 when idioms are translated in context, and 0.956 when translated without context, that is, in isolation. Looking at the individual systems, Google Translate shows a Kappa value of 0.958 in context and 1.000 out of context. DeepL achieves 0.979 in context, but the agreement drops to 0.867 out of context. SYSTRAN reaches a perfect agreement of 1.000 in both conditions. According to the interpretation framework of Landis and Koch (1977) as introduced in Chapter 6.5. and summarized in Table 9, these values indicate an almost perfect level of agreement across the systems.

Table 8: Cohen's k Results

	In context	In isolation
Google Translate	0.958	1.000
DeepL	0.979	0.867
SYSTRAN	1.000	1.000
All Systems	0.979	0.956

It should be noted, however, that an additional round of review and discussion was carried out after the initial annotations. This was necessary due to the dilemma of how narrowly the concept of idiom should be interpreted and in order to avoid inconsistencies in the final version. Although instructions had been provided in advance, the ambiguity of several examples led to uncertainty about how the categories should be applied. Because of this complexity, a further discussion based on concrete examples was required. This procedure may have contributed to slightly higher Kappa values than would otherwise have been obtained. In the case of idiom evaluation outside of context, the particularly high values can be explained by the fact that only two categories were applied.

8. Discussion

Overall, the results indicate that both Google Translate and DeepL are able to handle the chosen idiomatic expressions and translate them correctly from English into German. When comparing the accuracy of translations of equivalent idioms, similar idioms, and paraphrases, DeepL appears to perform slightly better than Google Translate, achieving an additional five percentage points of correct results. On the contrary, SYSTRAN failed to achieve the task and mainly delivered literal translations of the given idioms. Additionally, contrary to the assumption, the examined NMT systems predominantly provided idiomatic translations, i.e., equivalent or similar idioms rather than paraphrase or literal translation, when observed as separate categories.

With regard to context, it seems that DeepL was the only system that assumed that most expressions ought to be understood idiomatically when no context was given. When context was provided, DeepL increased its accuracy rate by ten percentage points. In a similar vein, SYSTRAN exhibited a slight increase in accuracy when given context but translated the majority of the examples literally. In the case of Google Translate, there seems to be no straightforward connection between the two categories. While it produced more than half of the literal interpretations of idioms in isolation, it did fairly well when translating the same expressions in context. Thus, it can be said that Google Translate's output appears to be most influenced by context, resulting in greater accuracy between translations in than out of context. Nevertheless, no general conclusion that applies to all three systems can be drawn. It can be said, however, that DeepL tends to produce accurate translations of idioms it recognizes as such out of context, providing slightly better results when context is present. This occurs regardless of the length of the context provided. Differences in the type of translation, be it equivalent, similar, or paraphrase, could be observed, but there were extremely few occurrences where context influenced the accuracy of the translation output for idioms previously recognized as such in isolation. The opposite example, exemplified by SYSTRAN, also confirms this underlying assumption. In this case, nearly all idioms that were not identified as idioms in isolation were also never translated idiomatically in context. With Google Translate, the translation output varied depending on the context, resulting in different suggestions across the equivalent idiom, similar idiom, and paraphrase categories. However, no direct link between the context length and the accuracy of the output could be established.

Concerning different idiom types, it seems that unmotivated, semantically non-decomposable idioms, such as 'to kick the bucket', confuse Google Translate and SYSTRAN, leading to the lowest score across the different idiom types. While Google Translate managed to suggest similar idiomatic and non-figurative correct translations, SYSTRAN failed to translate any of

the given idioms correctly. It is interesting to note that all idioms SYSTRAN translated correctly have an equivalent idiom in German according to Baker's (1992) definition, e.g., 'grit one's teeth' (die Zähne zusammenbeißen), 'smoke the peace pipe' (den Friedenspfeife rauchen), swallow the bitter pill' (die bittere Pille schlucken) and 'bury the hatchet' (das Kriegsbeil begraben). Taking this into consideration, it may be argued that SYSTRAN is far behind DeepL and Google Translate with regard to identification, comprehension, and accurate translation of idiomatic expressions from English into German. Moreover, the aforementioned examples are scattered across Type 2 and Type 3 but are all motivated. Additionally, Google Translate provided relatively similar results for the two types. This may imply that the motivation and opacity of idiomatic expressions play a role in MT rather than their semantic (de-)composition. Still, DeepL managed to translate almost all test sentences in Type 2 accurately. It scored highly with Type 3 as well, but it obtained nearly 20 percentage points less than with Type 2. These results suggest that DeepL is particularly good at detecting idiomatic nuances.

Moreover, not a single case of the omission strategy was observed. Given that recent research on NMT, as introduced earlier, argues that machine translation often produces overly literal and word-for-word translations, this result is not surprising. Machines focus on providing an output translation for each input word, whereas understanding when an omission would be appropriate requires human understanding. Similarly, this also indicates that not all strategies Baker (1992) proposed for human translators are applicable to machines. The reason behind this is that humans are able to differentiate between subtle cognitive and cultural nuances and adjust the target text accordingly. In fact, out of the categories Baker (1992) suggested, it might be said that only two are relevant for machine translation: accurate idiomatic translation and paraphrasing, while adding inaccurate translation as an additional criterion.

It is worth noting that the rapid expansion of these NMT systems poses a true methodological challenge, since identical sentences may be translated differently at different times. In connection with this, it is important to note that if the test sentences were to be translated again by the same systems at a later point, the results may be completely different.

9. Conclusion

Machine translation has developed rapidly over the last decade. This progress is largely driven by the advent of neural machine translation systems, which use neural networks to mimic certain aspects of human cognition during the translation process. This groundbreaking development has produced remarkable results. However, the translation of idiomatic expressions remains a significant challenge. Idioms, as highly complex linguistic units, are difficult even for human translators to render accurately. Consequently, machine translation systems often fail to capture their figurative meaning, instead producing awkward or overly literal translations.

This master's thesis set out to investigate the performance of three leading NMT systems, namely Google Translate, DeepL, and SYSTRAN, in translating idioms from English into German. The focus was twofold: first, to determine which system performs best when evaluated against a common idiom translation framework, and second, to examine whether the presence and length of context influence translation quality. The results revealed that DeepL proved to be more effective at recognizing and translating idiomatic expressions. Furthermore, the role of context varied across systems. While DeepL showed almost no discrepancy between translations of idioms in isolation and idioms in context, Google Translate's output improved when contextual information was available. Interestingly, SYSTRAN consistently struggled to handle idiomatic expressions, regardless of whether they were presented in or out of context. This indicates that only in the case of Google Translate did context have a notable impact on translation output, contrary to the assumption that all systems would benefit from additional information. Additionally, it was shown that context length did not influence the accuracy of the MT output.

A promising direction for future research would be a large-scale investigation of idiom translation in NMT, drawing on a broader set of idioms and a larger number of example sentences. The translation of idiomatic language remains an open challenge, and continued research in this field is essential for the development of more reliable and culturally sensitive machine translation systems.

10. Bibliography

Abdelaal, Noureldin & Alezzawie, Abdulkhaliq (2019). Translation strategies in the translation of idioms in Shakespeare's Romeo and Juliet. *Utopía y Praxis Latinoamericana* 24 (5), 275–90.

Adelnia, Amineh & Dastjerdi, Hossein Vahid (2011). Translation of idioms: A hard task for the translator. *Theory and Practice in Language Studies* 1 (7), 879–883. DOI: 10.4304/tpls.1.7.879-883

Agrawal, Ruchit; Kumar, Vighnesh Chenthil; Muralidharan, Vigneshwaran & Misra Sharma, Dipti (2018). No more beating about the bush: a step towards idiom handling for Indian language NLP. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), o.S.

Agrawal, Sweta; Farinhas, António; Rei, Ricardo & Martins, André F. T. (2024). Can Automatic Metrics Assess High-Quality Translations? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 14491–14502, Miami, Florida, USA. Association for Computational Linguistics.

Akbari, Monireh (2013). Strategies for translating idioms. *Journal of Academic and Applied Studies* 3 (8), 32-41.

Aldahesh, Ali Yunis (2013). On idiomaticity in English and Arabic: a cross-linguistic study. *Journal of Languages and Culture* 4 (2), 23–29. DOI: 10.5897/JLC2013.0220

Anastasiou, Dimitra (2010). *Idiom treatment experiments in machine translation*. Newcastle Upon Tyne: Cambridge Scholars.

Anugraha, David; Kuwanto, Garry; Susanto, Lucky; Wijaya, Derry Tanti & Winata, Genta Indra (2024). MetaMetrics-MT: Tuning Meta-Metrics for Machine Translation via Human Preference Calibration. In *Proceedings of the Ninth Conference on Machine Translation*, pages 459–469, Miami, Florida, USA. Association for Computational Linguistics.

Arnold, D. (ed.) (1994). *Machine translation: an introductory guide*. Manchester: NCC Blackwell.

Arnon, Tamar & Lavidor, Michal (2023). Cognitive control in processing ambiguous idioms: evidence from a self-paced reading study. *Journal of Psycholinguistic Research*, 52 (1), 261–281 DOI: 10.1007/s10936-022-09861-z

As-Safi, Abdul B. (2011). *Translation Theories, Strategies and Basic Theoretical Issues.* Al Manhal.

Avramidis, Eleftherios; Macketanz, Vivien; Strohriegel, Ursula & Uszkoreit, Hans (2019). Linguistic evaluation of German-English machine translation using a test suite. In: *Proceedings of the Fourth Conference on Machine Translation*, 445–454.

Bahdanau, Dzmitry; Cho, Kyunghyun and Bengio, Yoshua (2014). Neural Machine Translation by Jointly Learning to Align and Translate. (Version 7). arXiv:1409.0473v7 (Stand 18.06.2025).

Baker, Mona (2018). *In other words: a coursebook on translation*. 3rd edition. London: Routledge, Taylor & Francis Group.

Banerjee, Satanjeev & Lavie, Alon (2005). Meteor: an automatic metric for MT evaluation with improved correlation with human judgments. In: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. Ann Arbor, MI, 65–72.

Barchudarov, Leonid S. (1979). *Sprache und Übersetzung: Probleme der allgemeinen und speziellen Übersetzungstheorie*. Moskau: Verlag Progress.

Bargmann, Sascha & Sailer, Manfred (2018). The syntactic flexibility of semantically non-de

composable idioms. In: Sailer, Manfred & Markantonatou, Stella. *Multiword expressions: insights from a multi-lingual perspective.* Berlin: Language Science Press, 1-29.
Barkema, Henk (2008). Idiomaticity And terminology: a multi-dimensional descriptive model. *Studia Linguistica* 50 (2), 125–160. DOI: 10.1111/j.1467-9582.1996.tb00347.x
Bassnett, Susan (1980). *Translation studies*. London: Routledge.
Bauer, Laurie (2008). Exocentric compounds. *Morphology* 18 (1), 51–74. DOI: 10.1007/s11525-008-9122-5
Bayer-Hohenwarter, Gerrit and Kußmaul, Paul (2020). Translation, creativity and cognition. In: Alves, Fabio, & Jakobsen, Arnt Lykke (Eds.) *The Routledge Handbook of Translation and Cognition*. 1st ed. London: Routledge, 310–325.
Baziotis, Christos; Mathur, Prashant & Hasler, Eva (2023). Automatic evaluation and analysis of idioms in neural machine translation. In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. Dubrovnik, Croatia: Association for Computational Linguistics, 3682–3700. DOI: 10.18653/v1/2023.eacl-main.267
Beck, Sara D. & Weber, Andrea (2020). Context and literality in idiom processing: evidence from self-paced reading. *Journal of Psycholinguistic Research* 49 (5), 837–863. DOI: 10.1007/s10936-020-09719-2
Beghtol, Clare (2001). Relationships in classificatory structure and meaning. In: Bean, Carol A. & Green, Rebecca (Eds.) *Relationships in the Organization of Knowledge*. Dordrecht: Springer, 99–113.
Bentivogli, Luisa; Bisazza, Arianna; Cettolo, Mauro & Federico, Marcello (2016). Neural versus phrase-based machine translation quality: a case study (Version 2). In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas, 257–267.
Bentivogli, Luisa; Federico, Marcello; Cettolo, Mauro and Federmann, Christian (2018). Machine Translation Human Evaluation: An investigation of evaluation based on Post-Editing and its relation with Direct Assessment. In *International Workshop on Spoken Language Translation*, pages 62–69, Brussels, Belgium. International Conference on Spoken Language Translation.
Bloomfield, Leonard (1993). *Language*. New York: Holt, Rinehart & Winston.
Bobrow, Samuel A. & Bell, Susan M. (1973). On catching on to idiomatic expressions. *Memory & Cognition* 1 (3), 343–346. DOI: 10.3758/BF03198118
Boisson, Joanne; Camacho-Collados, Jose & Espinosa-Anke, Luis (2022). *CardiffNLP-Metaphor at SemEval-2022 Task 2: Targeted Fine-tuning of Transformer-based Language Models for Idiomaticity Detection*. In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022),* pages 169-177. Seattle, United States: Association for Computational Linguistics.
Booth, Andrew D; Brandwood, Leonard; Cleave, J. P. (1958). *Mechanical resolution of linguistic problems*. London: Butterworths Scientific Publications.
Broughton, Vanda et al. (2005). *Knowledge organization*. Copenhagen: Royal School of Library and Information Science.
Bühler, Karl (1934). *Sprachtheorie. Die Darstellungsfunktion der Sprache*. Jena: Verlag von Gustav Fischer.
Bühler, Karl & Eschbach, Achim (2011). *Theory of Language: The representational function of language*. (Goodwin, Donald Fraser, Trans.) Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/z.164
Burchardt, Aljoscha; Lommel, Arle & Macketanz, Vivien (2021). A new deal for translation quality. *Universal Access in the Information Society* 20 (4), 701–715.
Burger, Harald & Jaksche, Harald (1973). *Idiomatik des Deutschen*. Tübingen: Niemeyer.
Cacciari, C. & Tabossi, Patrizia (eds.) (1993). *Idioms: processing, structure, and interpretation*.

Hillsdale, N.J: L. Erlbaum Associates.
Cacciari, Cristina & Tabossi, Patrizia (1988). The comprehension of idioms. *Journal of Memory and Language* 27 (6), 668–683. DOI: 10.1016/0749-596X(88)90014-9
Callison-Burch, Chris; Koehn, Philipp; Monz, Christoph and Zaidan, Omar (2011). Findings of the 2011 workshop on statistical machine translation. In: *Proceedings of the 6th Workshop on Machine Translation.* Edinburgh, Scotland, 22–64.
Castilho, Sheila; Doherty, Stephen; Gaspari, Federico & Moorkens, Joss (eds.) (2018). *Trans lation quality assessment: from principles to practice*. 1st edition. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-91241-7
Caswell, Isaac (2024). *110 new languages are coming to Google Translate*. https://blog.google/products/translate/google-translate-new-languages-2024/(Stand: 27.6.2024)
Caswell, Isaac & Liang, Bowen (2020). Recent advances in Google Translate. Software Engineers, Google Research. https://research.googleblog.com/2020/06/recent-advances-in-google-translate.html (Stand: 26.02.2025).
Catford, John C. (1965). *A linguistic theory of translation*. Oxford: Oxford University Press.
Chen, Linli (2009). On literal translation of English idioms. *English Language Teaching* 2, 164-166.
Cheng, Lei; Khalitov, Ruslan; Yu, Tong & Yang, Zhirong (2022). Classification of long sequential data using circular dilated convolutional neural networks (Version 2). arXiv:2201.02143v2 (Stand: 22.02.2025).
Cheragui, Mohamed Amine (2012). Theoretical overview of machine translation. *CEUR Workshop Proceedings* 867, 160-169.
Cho, Kyunghyun (2018). Deep Learning. In: Cho, Kyunghyun, Mitkov, Ruslan (Ed.) *The Oxford Handbook of Computational Linguistics 2nd edition*. Oxford University Press. DOI: 10.1093/oxfordhb/9780199573691.013.55
Cho, Kyunghyun; van Merrienboer, Bart; Bahdanau, Dzmitry & Bengio, Yoshua (2014a). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In: *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar. Association for Computational Linguistics.
Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger & Bengio, Yoshua (2014b). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.
Chomsky, Noam (1981). *Lectures on government and binding*. Dordrecht: Foris.
Chung, Dang (2024). Challenges of translating idiomatic expressions: a cross-linguistic analysis at a university in Hanoi, Vietnam. *International Journal of Social Science and Human Research* 7 (10), 7668-7678. DOI: 10.47191/ijsshr/v7-i10-39
Clarke, Stella G. Dextre (2001). Thesaural Relationships. In: Bean, Carol A., & Green, Rebecca (Eds.) *Relationships in the Organization of Knowledge*. Dordrecht: Springer Netherlands. Vol. 2, 37–52 DOI: 10.1007/978-94-015-9696-1_3
Cohen, Jacob (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20* (1), 37–46 DOI: 10.1177/001316446002000104
Colombo, L. (1993). The comprehension of ambiguous idioms in context. In: Cacciari, Cristina & Tabossi, Patrizia (eds.) *Idioms, processing, structure, and interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates, 163-200.
Colombo, L. & Williams, J. (1990). Effects of word- and sentence-level contexts upon word recognition. *Memory & Cognition* 18, 153–163.

Constant, Mathieu; Eryiğit, Gülşen; Monti, Johanna; Van der Plas, Lonneke; Ramisch, Carlos; Rosner, Michael & Todirascu, Amalia (2017). Multiword expression processing: a survey. *Computational Linguistics* 43 (4), 837–892.
Coughlin, Deborah (2003). Correlating automated and human assessments of machine translation quality. In: *Proceedings of MT Summit IX.* New Orleans, USA, 1-8.
Cowie, A. P., Mackin, R., & McCaig, I. R. (1983). *Oxford Dictionary of Current Idiomatic English: Phrase, Clause, and Sentence Idioms*. Oxford: Oxford University Press.
Cruse, Alan D. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
Cruse, D. Alan (2002). Hyponymy and its varieties. In: Green, Rebecca & Bean, Carol A. (Eds.) *The Semantics of Relationships*. Dordrecht: Springer, 3–21.
Cutler, Anne (1982). Idioms: the colder the older. *Linguistic Inquiry* 13, 317–320.
Dahlberg, Ingetraut (1978). Teoria do conceito. *Ciência da Informação* 7 (2), 1-7.
Dankers, Verna; Lucas, Christopher & Titov, Ivan (2022). Can Transformer be Too Compositional? Analysing Idiom Processing in Neural Machine Translation. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, 3608–3626. DOI: 10.18653/v1/2022.acl-long.252
Davies, Mark (2025). *The Corpus of Contemporary American English*. https://www.english-corpora.org/coca/
DeepL (2024).a *ALC Survey Results*. https://www.deepl.com/en/blog/alc-survey-results (Stand: 15.6.2025)
DeepL (2025).b *Quality - Unbeatable Translation Quality for Global Business*. https://www.deepl.com/en/quality (Stand: 13.6.2025)
DeepL (2025).c *Why DeepL?* https://www.deepl.com/en/whydeepl (Stand: 15.6.2025)
de Vries, Mark (2002). *The syntax of relativization*. Utrecht: LOT.
Delisle, Jean; Hannelore, Lee-Jahnke & Cormier, Monique C. (eds.) (2003). *Terminologie de la traduction/Translation terminology/Terminología de la traducción/Terminologie der Übersetzung*. Amsterdam/Philadelphia: John Benjamins.
Depraetere, Ilse (ed.) (2011). *Perspectives on translation quality.* Berlin: De Gruyter. DOI: 10.1515/9783110259889
Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton & Toutanova, Kristina (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: *Proceedings of the 2019 Conference of the North*. Minneapolis, Minnesota: Association for Computational Linguistics. 4171–4186 DOI: 10.18653/v1/N19-1423
dict.cc (2023). *Online Dictionary English-German*. https://www.dict.cc/?l=e. (Stand: 24.2.2023).
Doddington, George (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In: *Proceedings of the second international conference on Human Language Technology Research.* San Diego, California: Association for Computational Linguistics, 138-145. DOI: 10.3115/1289189.1289273
Donthi, Sundesh; Spencer, Maximilian; Patel, Om B.; Doh, Joon Young; Rodan, Eid; Zhu, Kevin and O'Brien, Sean (2025). Improving LLM Abilities in Idiomatic Translation. In: *Proceedings of the First Workshop on Language Models for Low-Resource Languages*. Abu Dhabi, United Arab Emirates. Association for Computational Linguistics, 175–181.
Dostert, Leon E. (1955). The Georgetown-I.B.M. experiment. In: Locke, William N. & Booth, A. Donald (Hg.) *Machine translation of languages*. Cambridge, MA: MIT Press, 124–135.
Drugan, Joanna (2013). *Quality in professional translation: assessment and improvement*. London: Bloomsbury. DOI: 10.5040/9781472542014

Dumas, Bethany K. & Lighter, Jonathan (1978). Is slang a word for linguists? *American Speech*, 53 (1), 1-5. DOI: 10.2307/455336

Elliott, Debbie; Hartley, Anthony & Atwell, Eric (2003). Rationale for a multilingual aligned corpus for machine translation evaluation. In: Archer, D.; Rayson P., Wilson A. & McEnery, Tony (eds.) *Proceedings of CL2003: International Conference on Corpus Linguistics*, 191–200.

Estill, Robert B. & Kemper, Susan (1982). Interpreting idioms. *Journal of Psycholinguistic Research*, *11* (6), 559–568 DOI: 10.1007/BF01067612

Estrella, Paula; Olivier, Hamon & Popescu-Belis, Andrei (2007). How much data is needed for reliable MT evaluation? Using bootstrapping to study human and automatic met rics. In: Maegaard, Bente (ed.) *Proceedings MT Summit* 11, 167–174.

Fadaee, Marzieh; Bisazza, Arianna & Monz, Christof (2018). Examining the tip of the iceberg: A data set for idiom translation. In: Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, Takenobu Tokunaga (eds.). Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan, 925-929.

Fellbaum, Christiane (1993). The determiner in English idioms. In: Cacciari, Cristina & Ta bossi, Patrizia (eds.) *Idioms, processing, structure, and interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates, 271–295.

Fernando, Chitra (1996). *Idioms and idiomaticity*. Oxford: Oxford Univ. Press.

Fernando, Chitra & Flavell, Roger (1981). *On idiom: critical views and perspectives*. Exeter: University of Exeter.

Fields, Paul; Hague, Daryl R.; Koby, Geoffrey S.; Lommel, Arle & Melby, Alan (2014). What Is quality? A management discipline and the translation industry get acquainted. *Tradumàtica Tecnologies de La Traducció* 12, 404–412. DOI: 10.5565/rev/tradumatica.75

Flores D'Arcais, Giovanni B. (1993). The comprehension and semantic interpretation of idioms. In: Cacciari, Cristina & Tabossi, Patrizia (eds.) *Idioms, processing, structure, and interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates, 79-99.

Fomicheva, Marina (2017). *The Role of Human Reference Translation in Machine Translation Evaluation*. Dissertation, Universitat Pompeu Fabra.

Forcada, Mikel L. (2017). Making sense of neural machine translation. *Translation Spaces* 6 (2), 291–309.

Fraser, Bruce (1970). Idioms within a transformational grammar. *Foundations of Language*, 6 (1), 22-42.

Freeman, Richard (2009). What is 'translation'? *Evidence & Policy* 5 (4), 429–447. DOI: 10.1332/174426409X478770

Frege, Gottlob (1884). *The foundations of arithmetic (Die Grundlagen der Arithmetik)*. Trans. by J. L. Austin. Oxford: Blackwell; Nachdruck: Evanston, IL: Northwestern University Press.

Freitag, Markus; Mathur, Nitika; Deutsch, Daniel; Lo, Chi-Kiu; Avramidis, Eleftherios; Rei, Ricardo; Thompson, Brian; Blain, Frederic; Kocmi, Tom; Wang, Jiayi; Adelani, Ifeoluwa David; Buchicchio, Marianna; Zerva, Chrysoula & Lavie, Alon (2024). Are LLMs breaking MT metrics? Results of the WMT24 metrics shared task. In: *Proceedings of the Ninth Conference on Machine Translation*. Miami, FL: Association for Computational Linguistics, 47–81.

Gao, Chunming & Zheng, Qianzhen (2014). A Linguistic Study of Antonymy in English Texts. *Journal of Language Teaching and Research* 5 (1), 234–238 DOI: 10.4304/jltr.5.1.234-238

Garvin, David A. (1984). What does product quality really mean? *Sloan Management Review* 26 (1), 25-43.
Garvin, Paul (1972). The Georgetown-IBM experiment of 1954: An evaluation in retrospect. *Papers in linguistics in honor of Dostert*. The Hague: Mouton, 46–56.
Geeraerts, Dirk (1995). Specialization and reinterpretation in idioms. In: Everaert, Martin; van der Linden, Erik; Schenk, Andrej & Schreuder, Robert (Eds.) *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates, 57–73.
Gehring, Jonas; Auli, Michael; Grangier, David and Dauphin, Yann N. (2017). A Convolutional Encoder Model for Neural Machine Translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 123–135, Vancouver, Canada. Association for Computational Linguistics.
Gibbs, Raymond W. (1984). Literal meaning and psychological theory. *Cognitive Science* 8, 275–304.
Gibbs, Raymond W. & Gonzales, Gayle P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, *20* (3), 243–259.
Gimaletdinova, Gulnara; Khalitova, Liliia; Solovyev, Valery & Bochkarev, Vladimir (2021). Lexicographic study of synonymy: clarifying semantic similarity between words. *Computación y Sistemas* 25 (3), 1-9.
Glass, Arnold L. (1983). The comprehension of idioms. *Journal of Psycholinguistic Research*, *12* (4), 429–442 DOI: 10.1007/BF01067624
Gläser, Rosemarie (1986). *Phraseologie der englischen Sprache*. Tübingen: Niemeyer.
Gledhill, Christopher (2011). A lexicogrammar approach to checking quality: looking at one or two cases of comparative translation. In: Depraetere, Ilse (ed.) *Perspectives on translation quality*. Berlin: De Gruyter, 71–98. DOI: 10.1515/9783110259889.71
Glucksberg, Sam (1993). Idioms meanings and allusional content. In: Cacciari, Cristina & Tabossi, Patrizia (eds.) *Idioms, processing, structure, and interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates, 3-26.
Glucksberg, Sam & McGlone, Matthew S. (2001). *Understanding figurative language: from metaphors to idioms*. Oxford: Oxford University Press.
Graham, Yvette; Baldwin, Timothy; Moffat, Alistair and Zobel, Justin (2013). Continuous Measurement Scales in Human Evaluation of Machine Translation. In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 33–41, Sofia, Bulgaria. Association for Computational Linguistics.
Granger, Sylviane & Paquot, Magali (2008). Disentangling the phraseological web. In: Granger, Sylviane, & Meunier, Fanny (eds.) *Phraseology*. Amsterdam: John Benjamins Publishing Company, 27–49. DOI: 10.1075/z.139.07gra
Grant, Lynn E. (2003). *A corpus-based investigation of idiomatic multiword units*. Dissertation, Victoria University of Wellington.
Grauberg, Walter (1989). Proverbs and idioms: mirrors of national experience? In: Gregory James (ed.) *Lexicographers and Their Works*. Exeter: University of Exeter Press, 94-99.
Guerberof-Arenas, Ana and Toral, Antonio (2022). Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces* 11 (2), 184–212.
Guerreiro, Nuno M.; Rei, Ricardo; Stigt, Daan van; Coheur, Luisa; Colombo, Pierre & Mart ins, André F. T. (2024). xCOMET: Transparent Machine Translation Evaluation through Fine-grained Error Detection. *Transactions of the Association for Computat ional Linguistics* 12, pages 979–995.
Guerreiro, Nuno M.; Rei, Ricardo; Stigt, Daan Van; Coheur, Luisa; Colombo, Pierre & Martins, André F. T. (2024). x COMET: Transparent Machine Translation Evaluation through Fine-grained Error Detection. *Transactions of the Association for Computational Linguistics* 12, 979–995 DOI: 10.1162/tacl_a_00683

Guzmán, F.; Abdelali, A.; Temnikova, I.; Sajjad, H. and Vogel, S. (2015). How do humans evaluate machine translation? In *Proceedings of the 10th workshop on Statistical Machine Translation (WMT 2015)*, pages 457–466, Lisbon, Portugal.

Hardesty, Kathleen Sandell & Hollinger, Andrew (2020). Uncommonly Common: Reconsidering Creativity and Beauty in Technical Communication. *Technical Communication (Washington)* 67 (4), 28–48.

Harmon, Lucyna (2021). Idiom as a translation technique: a theorical postulate. *Cadernos de Tradução* 41 (1), 125–147. DOI: 10.5007/2175-7968.2021.e71778

Hassan, Hany; Aue, Anthony; Chen, Chang; Chowdhary, Vishal; Clark, Jonathan; Federmann, Christian; Huang, Xuedong; Junczys-Dowmunt, Marcin; Lewis, William; Li, Mu; Liu, Shujie; Liu, Tie-Yan; Luo, Renqian; Menezes, Arul; Qin, Tao; Seide, Frank; Tan, Xu; Tian, Fei; Wu, Lijun; Wu, Shuangzhi; Xia, Yingce; Zhang, Dongdong; Zhang, Zhirui & Zhou, Ming (2018). Achieving Human Parity on Automatic Chinese to English News Translation. arXiv:1803.05567

Hatim, Basil & Mason, Ian (1990). *Discourse and the translator*. London/New York: Longman.

Hatim, Basil & Mason, Ian (2005). *The translator as communicator*. 1st edition. Florence: Taylor and Francis.

Henisz-Dostert, Bożena; Macdonald, R. R. & Zarechnak, Michael (1979). *Machine translation.* Berlin/New York: De Gruyter Mouton.

Hervey, Sándor & Higgins, Ian (1992). *Thinking translation: a course in translation method French-English*. London/New York: Routledge.

Hervey, Sándor & Higgins, Ian (2002). *Thinking French translation: a course in translation method:French to English*. 2nd edition. New York/Florence: Routledge Taylor & Francis Group.

Hervey, Sandor; Loughridge, Michael & Higgins, Ian (2013). *Thinking German translation*. 2nd edition. Hoboken: Taylor and Francis.

Hjørland, Birger (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology* 41 (1), 367–405.

House, Juliane (1977). *A model for translation quality assessment*. Tübingen: TBL. Verlag Gunter Narr.

House, Juliane (1997). *Translation quality assessment: a model revisited*. Tübingen: Narr.

House, Juliane (2015). *Translation quality assessment: past and present*. London: Routledge.

Hujon, Aiusha Vellintihun; Amitab, Khwairakpam and Singh, Thoudam Doren (2023). Convolutional Sequence to Sequence Learning for English-Khasi Neural Machine Translation. In: *4th International Conference on Computing and Communication Systems (I3CS),* pages 1-4, Shillong, India.

Hutchins, John W. (2004). The Georgetown-IBM experiment demonstrated in January 1954. In: *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pages 102–114, Washington, USA. Springer.

Hutchins, John W. (ed.) (2000). *Early Years in Machine Translation: Memoirs and biographies of pioneers*. Amsterdam: John Benjamins Publishing Company. Vol. 97. DOI: 10.1075/sihols.97

Imran, Mia Mohammad; Chatterjee, Preetha and Damevski, Kostadin (2024). Shedding Light on Software Engineering-specific Metaphors and Idioms. In: *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*. Lisbon: Portugal, 1–13.

International Organization for Standardization (2012). Translation projects — General guidance (ISO Standard No. 11669: 2012). https://www.iso.org/standard/50687.html (Stand 20.03.2025).

International Organization for Standardization (2015). Translation services — Requirements for translation services (ISO Standard No. 17100: 2015). https://www.iso.org/standard/59149.html (Stand 20.03.2025).
International Organization for Standardization (2024). Translation projects — General guidance (ISO Standard No. 11669: 2024). https://www.iso.org/standard/79089.html (Stand 20.03.2025).
Isabelle, Pierre, Cherry, Colin & Foster, George (2017). A challenge set approach to evaluating machine translation. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, 2486–2496.
Jarema, Gonia & Libben, Gary (2007). *The Mental Lexicon: Core Perspectives*. BRILL. DOI: 10.1163/9780080548692
Jurafsky, Daniel & Martin, James H. (2025). *Speech and Language Processing.* 3rd edition. Stanford University.
Juraska, Juraj; Deutsch, Daniel; Finkelstein, Mara & Freitag, Markus (2024). MetricX-24: The Google Submission to the WMT 2024 Metrics Shared Task. In: *Proceedings of the Ninth Conference on Machine Translation*, pages 492–504, Miami, Florida, USA. Association for Computational Linguistics.
Kade, Otto (1976). Phraseologismen als Übersetzungsproblem. In: *Aktuelle Probleme der Phraseologie.* Leipzig: Wissenschaftliche Beiträge der KMU Leipzig, 49-62.
Kalchbrenner, Nal & Blunsom, Phil (2013). Recurrent continuous translation models. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Seattle, WA, 1700–1709.
Kasperė, Ramunė; Horbačauskienė, Jolita; Motiejūnienė, Jurgita; Liubinienė, Vilmantė; Patašienė, Irena & Patašius, Martynas (2021). Towards Sustainable Use of Machine Translation: Usability and Perceived Quality from the End-User Perspective. *Sustainability* 13 (23), 13430 DOI: 10.3390/su132313430
Katsarou, Eirene (2011). *Descriptive and theoretical aspects of English idioms*. Saarbrücken: LAMBERT Academic Publishing.
Katz, Jerrold J. (1973). Compositionality, idiomaticity, and lexical substitution. In Anderson, S. & Kiparsky, P. (eds.) (1973). *A Festschrift for Morris Halle*. New York: Holt, Rinehart, and Winston, 357–376.
Katz, Jerrold J. & Postal, Paul (1963). Semantic interpretation of idioms and sentences containing them. *Quarterly Progress Report of the MIT Research Laboratory of Electronics* 70, 275– 282.
Keizer, Evelien (2016). Idiomatic expressions in Functional Discourse Grammar. In: Inge, Genee & Keizer, M.E. *The Lexicon in Functional Discourse Grammar*. Linguistics 54 (5) (Special Issue), 981-1016. DOI: 10.1515/ling-2016-0022
Kenny, Dorothy (2022). Machine translation for everyone. Empowering users in the age of artificial intelligence. *Quaderns IEE* 2 (1), 148–152.
Khoo, Christopher S. G. & Na, Jin-Cheon (2006). Semantic relations in information science. *Annual Review of Information Science and Technology* 40 (1), 157–228.
Koby, Geoffrey S.; Fields, Paul; Hague, Daryl R.; Lommel, Arle & Melby, Alan (2014). Defining the Landscape of Translation. *Tradumàtica Tecnologies de La Traducció* 12, 392–403.
Koby, Geoffrey S.; Fields, Paul; Hague, Daryl R.; Lommel, Arle & Melby, Alan (2014). What Is Quality? A Management Discipline and the Translation Industry Get Acquainted. *Tradumàtica Tecnologies de La Traducció* 12, 404–412.
Koby, Geoffrey S.; Fields, Paul; Hague, Daryl R.; Lommel, Arle & Melby, Alan (2014). Defining translation quality. *Tradumàtica Tecnologies de La Traducció* 12, 413–420.
Koehn, Philipp (2009). *Statistical machine translation*. 1st ed. Cambridge: Cambridge University Press.

Koller, Werner (1995). The concept of equivalence and the object of translation studies. *Target* 7, 191–222.
Kovács, Gabriella (2016). About the definition, classification, and translation strategies of idioms. *Acta Universitatis Sapientiae Philologica* 8 (3), 85–101.
Lamb, Andrew and Xie, Michael (2015). *Convolutional Encoders for Neural Machine Translation*. Stanford University.
Langlotz, Andreas (2006). *Idiomatic creativity: a cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam, Philadelphia: J. Benjamins.
language models. 3rd ed. https://web.stanford.edu/~jurafsky/slp3 (Stand: 18.06.2025).
Läubli, Samuel; Castilho, Sheila; Neubig, Graham; Sennrich, Rico; Shen, Qinlan & Toral, Antonio (2020). A set of recommendations for assessing human-machine parity in lan guage translation. *Journal of Artificial Intelligence Research* 67*,* pages 653–672.
Lauscher, Susanne (2000). Translation quality assessment: Where can theory and practice meet? *The Translator* 6 (2), 149–168.
Lee, Seungjun; Lee, Jungseob; Moon, Hyeonseok; Park, Chanjun; Seo, Jaehyung; Eo, Sugyeong; Koo, Seonmin & Lim, Heuiseok (2023). A Survey on Evaluation Metrics for Machine Translation. *Mathematics*, 11 (4), 1006. DOI: 10.3390/math11041006
Lewis, Mike; Liu, Yinhan; Goyal, Naman; Ghazvininejad, Marjan; Mohamed, Abdelrahman; Levy, Omer; Stoyanov, Veselin & Zettlemoyer, Luke (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics. 7871–7880 DOI: 10.18653/v1/2020.acl-main.703
Liu, Changsheng & Hwa, Rebecca (2016). Phrasal Substitution of Idiomatic Expressions. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, 363–373. DOI: 10.18653/v1/N16-1040
Liu, Emmy; Chaudhary, Aditi & Neubig, Graham (2023). Crossing the Threshold: Idiomatic Machine Translation through Retrieval Augmentation and Loss Weighting. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing,* pages 15095–15111. Singapore: Association for Computational Linguistics. DOI:10.18653/v1/2023.emnlp-main.933
Liu, Qi; Kusner, Matt J. & Blunsom, Phil (2020). A Survey on Contextual Embeddings. (Version 2). arXiv:2003.07278v2
Liu, Yang & Zhang, Min (2023). Statistical machine translation. In: Chan, Sin-Wai (Eds.) *Routledge Encyclopedia of Translation Technology*. 2nd ed. London: Routledge, 208–218.
Lommel, Arle Richard; Burchardt, Aljoscha and Uszkoreit, Hans (2013). Multidimensional quality metrics: a flexible system for assessing translation quality. In *Proceedings of Translating and the Computer* 1-7, London, UK.
Lopez, Adam (2008). Statistical machine translation. *ACM Computing Surveys* 40 (3), 1–49.
Maia, Lucinéia Souza & Ângela De Lima, Gercina (2021). A Semantic-relations taxonomy for knowledge representation. *Brazilian Journal of Information Science: Research Trends* 15, DOI: 10.36311/1981-1640.2021.v15.e02123
Makkai, Adam (1969). The two idiomaticity areas in English and their membership: A stratificational view. *Linguistics*, 7 (50) DOI: 10.1515/ling.1969.7.50.44
Mäntylä, Katja (2004). *Idioms and language users: the effect of the characteristics of idioms on their recognition and interpretation by native and non-native speakers of English*. Dissertation, University of Jyväskylä.
Marie, Benjamin; Fujita, Atsushi & Rubino, Raphael (2021). Scientific credibility of machine

translation research: a meta-evaluation of 769 papers. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, 7297–7306.

Matusov, Evgeny (2019). The challenges of using neural machine translation for literature. In: *Proceedings of the Workshop on the Qualities of Literary Machine Translation*. Dublin, Ireland: European Association for Machine Translation, 10–19.

Mazzocchi, Fulvio (2017). Relations in KOS: Is it possible to couple a common nature with different roles? *Journal of Documentation* 73 (2), 368–383.

McCarthy, Michael (1992). English idioms in use. *Revista Canaria de Estudios Ingleses* 25, 55-65. Cambridge: Cambridge Univ. Press.

McGlone, Matthew S.; Glucksberg, Sam & Cacciari, Cristina (1994). Semantic productivity and idiom comprehension. *Discourse Processes* 17 (2), 167–190. DOI: 10.1080/01638539409544865

Meenal, T S, & Govindarajan P. (2023). The Challenges of Using Machine Translation While Translating Polysemous Words. *Studies in Media and Communication* 11 (2), 63. DOI: 10.11114/smc.v11i2.5944

Molina, Lucía & Hurtado Albir, Amparo (2004). Translation techniques revisited: a dynamic and functionalist approach. *Meta* 47 (4), 498–512. DOI: 10.7202/008033ar

Mollanazar, Hossein (2001). *Principles and methodology of translation*. 4th edition. Tehran: SAMT.

Molosnaja, Tat'jana N. (2000). My memoirs of MT in the Soviet Union. In: Hutchins, John W. (Ed.) *Studies in the History of the Language Sciences*. Amsterdam: John Benjamins Publishing Company. Vol. 97, 227–232. DOI: 10.1075/sihols.97.20mol

Moon, Rosamund (1998). *Fixed expressions and idioms in English: a corpus-based approach*. Oxford/New York: Clarendon Press/Oxford University Press.

Motallebzadeh, Khalil & Tousi, Seika (2011). Employing Compensation Strategy in Translation of Idioms: A Case Study of the Translation of Mark Twain's Adventures of Huckleberry Finn in Persian. *International Journal of Linguistics* 3 (1), 1-9. DOI: 10.5296/ijl.v3i1.1206

Murphy, M. Lynne (2010). Lexical and semantic relations. In: *Lexical Meaning*. Cambridge Textbooks in Linguistics. Cambridge University Press, 108-132.

Nasiri Moghaddam, Abdol-Majid & Fatehi Rad, Neda (2021). Translating Idioms: A Study on English Translation of The Date Palm. *Journal of Language Culture and Translation* 4 (1), 45-67. DOI: 10.30495/lct.2021.1938504.1042

Newmark, Peter (1991). *About translation*. Clevedon/Philadelphia: Multilingual Matters.

Newmark, Peter (2008). *A textbook of translation*. 12th impression. Harlow: Longman.

Newmeyer, F.J. (1974). The regularity of idiom behavior. *Lingua* 34, 327-342.

Nida, Eugene (1964). *Toward a science of translation*. Leiden: Brill.

Nida, Eugene A. & Taber, Charles R. (2003). *The theory and practice of translation*. 4th impression. Leiden Boston: Brill.

Nunberg, Geoffrey; Sag, Ivan A. & Wasow, Thomas (1994). Idioms. *Language* 70 (3), 491-538.

Och, Franz (2006). Statistical machine translation live. https://research.google/blog/statistical-machine-translation-live/ (Stand: 22.02.2025).

Okpor, M. D. (2014). Machine translation approaches: issues and challenges. *IJCSI International Journal of Computer Science Issues* 11 (5), 159–165.

Omar, Abdulfattah & Gomaa, Yasser (2020). The machine translation of literature: implications for translation pedagogy. *International Journal of Emerging Technologies in Learning (iJET)* 15, 228-235.

Ortony, Andrew; Schallert, Diane L.; Reynolds, Ralph E. & Antos, Stephen J. (1978). Interpreting metaphors and idioms: Some effects of the context on comprehension. *Journal of Verbal Learning and Verbal Behaviour* 17, 465–477.

Palumbo, Giuseppe (2009). *Key terms in translation studies*. London/New York: Continuum.

Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei Jing (2002). BLEU: a method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA, 311–318.

Peters, Isabella & Weller, Katrin (2008). Paradigmatic and syntagmatic relations in knowledge organization systems. *Information Wissenschaft und Praxis* 59 (2), 100–107.

Poibeau, Thierry (2017). *Machine translation*. Cambridge, Massachusetts: The MIT Press.

Poibeau, Thierry (2022). On "human parity" and "super human performance" in machine translation evaluation. In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*. Marseille: European Language Resources Association (ELRA), 6018–6023.

Popović, Maja (2018). Error classification and analysis for machine translation quality assessment. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (eds.) *Translation Quality Assessment*. Cham: Springer, 129–158. DOI: 10.1007/978-3-319- 91241-7_7

Popović, Maja (2020). Informative manual evaluation of machine translation output. In: *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*. Barcelona: International Committee on Computational Linguistics, 5059–5069.

Pym, Anthony (2010). *Exploring translation theories*. London: Routledge.

Rafatbakhsh, Elaheh & Ahmadi, Alireza (2019). A thematic corpus-based study of idioms in the Corpus of Contemporary American English. *Asian-Pacific Journal of Second and Foreign Language Education* 4 (1), 1-11. DOI: 10.1186/s40862-019-0076-4

Raffel, Colin; Shazeer, Noam; Roberts, Adam; Lee, Katherine; Narang, Sharan; Matena, Michael; Zhou, Yanqi; Li, Wei and Liu, Peter J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21 (140), 1–67.

Rambelli, Giulia; Chersoni, Emmanuele; Senaldi, Marco S. G.; Blache, Philippe & Lenci, Alessandro (2023). Are Frequent Phrases Directly Retrieved like Idioms? An Investigation with Self-Paced Reading and Language Models. In: *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023).* Dubrovnik, Croatia: Association for Computational Linguistics, 87–98. DOI: 10.18653/v1/2023.mwe-1.13

Rao, Pavithra; McGee, Lauren M. & Seideman, Casey A. (2024). A Comparative assessment of ChatGPT vs. Google Translate for the translation of patient instructions. *Journal of Medical Artificial Intelligence* 7, 11–11. DOI: 10.21037/jmai-24-24

Razmjou, Leila (2004). To be a good translator. *Translation Journal*, 8 (2), 1-5.

Rei, Ricardo; Stewart, Craig; Farinha, Ana C and Lavie, Alon (2020). COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2685–2702.

Reiss, Katharina (1971). *Möglichkeiten und Grenzen der Übersetzungskritik. Kategorien und Kriterien für eine sachgerechte Beurteilung von Übersetzungen*. München: Max Hueber Verlag.

Reiss, Katharina (1986). Texttypen, Übersetzungstypen und die Beurteilung von Übersetzungen. *Lebende Sprachen* 22 (1–4), 97–100.

Reiss, Katharina (2000). *Translation criticism, the potentials and limitations: categories and criteria for translation quality assessment*. Manchester/New York/St. Jerome Pub: American Bible Society.

Reiss, Katharina; Vermeer, Hans J.; Nord, Christiane & Dudenhöfer, Marina (2013). *Towards a general theory of translational action: Skopos theory explained*. Manchester: St. Jerome.

Rojo, Ana (2017). The Role of Creativity. In: Schwieter, John W., & Ferreira, Aline (Eds.) *The Handbook of Translation and Cognition*. 1st ed. Hoboken: Wiley Blackwell, 350–368.

Salton, Giancarlo; Ross, Robert & Kelleher, John (2014). An empirical study of the impact of idioms on phrase based statistical machine translation of English to Brazilian-Portuguese. In: *Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra)*. Gothenburg, Sweden: Association for Computational Linguistics, 36–41.

Santing, Lukas; Sijstermans, Ryan; Anerdi, Giacomo; Jeuris, Pedro; Ten Thij, Marijn & Batista-Navarro, Riza (2022). Food for Thought: How can we exploit contextual embeddings in the translation of idiomatic expressions? In: *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)*. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics. 100–110. DOI:10.18653/v1/2022.flp-1.14

Scherer, Laura D. & Larsen, Randy J. (2011). Cross-modal evaluative priming: Emotional sounds influence the processing of emotion words. *Emotion* 11 (1), 203–208 DOI: 10.1037/a0022588

Schmidt, Robin M. (2019). Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview. (Version 1). arXiv:1912.05911v1 (18.06.2025).

Sellam, Thibault; Das, Dipanjan & Parikh, Ankur P. (2020). BLEURT: Learning Robust Metrics for Text Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892.

Sen, Vijay Singh & Jamwal, Shubhnandan S. (2024). Transcending linguistic boundaries: A technical exploration of the evolution and future trajectory of corpus-based machine translation. *Journal of Electrical Systems* 20 (7), 2502–2509.

Sepesy Maučec, Mirjam & Donaj, Gregor (2020). Machine Translation and the Evaluation of Its Quality. In: Sadollah, Ali, & Shishir Sinha, Tilendra (Eds.) *Recent Trends in Comp utational Intelligence*. IntechOpen. DOI: 10.5772/intechopen.89063

Shojaei, Amir (2012). Translation of idioms and fixed expressions: Strategies and difficulties. *Theory and Practice in Language Studies* 2 (6), 1220–1229.

Sim, Julius & Wright, Chris C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, *85* (3), 257–268.

Škvorc, Tadej; Gantar, Polona & Robnik-Šikonja, Marko (2022). MICE: Mining Idioms with Contextual Embeddings. *Knowledge-Based Systems* 235 (107606), 1-11. DOI: 10.1016/j.knosys.2021.107606

Slater, Jarron and Rosselot-Merritt, Jeremy (2024). The Technical Communicator as Artist: Rhetoric, Aesthetics, and Form in the Workplace. *Technical Communication* 71 (2), 58–71.

Sontag, Susan (2001). *Where the stress falls: essays*. New York: Farrar, Straus, and Giroux.

Sountsov, Pavel & Sarawagi, Sunita (2016). Length bias in Encoder Decoder Models and a Case for Global Conditioning. In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1516–1525, Austin, Texas.

Specia, Lucia; Scarton, Carolina & Paetzold, Gustavo Henrique (2018). *Quality estimation for machine translation*. Synthesis Lectures on Human Language Technologies. Cham: Springer.

Sprenger, Simone; Levelt, Willem & Kempen, Gerard (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54 (2), 161–184.

Stahlberg, Felix (2020). Neural Machine Translation: A Review. *Journal of Artificial Intelligence Research* 69, 343–418.
Stock, Wolfgang G. (2010). Concepts and semantic relations in information science. *Journal of the American Society for Information Science and Technology* 61 (10), 1951–1969.
Strakšienė, Margarita (2010). Analysis of idiom translation strategies from English into Russian. *Kalbų Studijos* 17, 29–33.
Sun, Shuyan (2011). Meta-analysis of Cohen's kappa. *Health Services and Outcomes Research Methodology*, 11 (3–4), 145–163 DOI: 10.1007/s10742-011-0077-3
Sutskever, Ilya; Vinyals, Oriol & Le, Quoc V. (2014). Sequence to sequence learning with neural networks. In: *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 3104–3112.
Swinney, David A. & Cutler, Anne (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior 18* (5), 523–534. DOI: 10.1016/S0022-5371(79)90284-6
SYSTRAN (2025).a *Advatages of SYSTRAN Techonology*. https://www.systransoft.com/systran/translation-technology/advantages-of-systran-technology/ (Stand: 13.1.2025)
SYSTRAN (2025).b *Our history*. https://www.systransoft.com/systran/ (Stand: 13.6.2025)
SYSTRAN (2025).c *Your gateway to seamless multilingual communication*. https://www.systransoft.com/ (Stand: 14.6.2025)
Tan, Zhixing; Wang, Shuo; Yang, Zonghan; Chen, Gang; Huang, Xuancheng; Sun, Maosong & Liu, Yang (2020). Neural machine translation: a review of methods, resources, and tools. *AI Open* 1, 5–21.
Tang, Gongbo; Müller, Mathias; Rios, Annette & Sennrich, Rico (2018). Why Self-Attention? A Targeted Evaluation of Neural Machine Translation Architectures. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, 4263–4272. DOI: 10.18653/v1/D18-1458
Taube, Mortimer (1961). *Computers and common sense: the myth of thinking machines*. New York: Columbia University Press. DOI: 10.7312/taub90714
Ten Hacken, Pius (2015). Transposition and the Limits of Word Formation. In: Bauer, Laurie, Körtvélyessy, Lívia, & Štekauer, Pavol (Eds.) *Semantics of Complex Words*. Cham: Springer International Publishing. Vol. 3, 187–216. DOI: 10.1007/978-3-319-14102-2_10
Tian, Ye; James, Isobel & Son, Hye (2023). How Are Idioms Processed Inside Transformer Language Models? In: *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (SEM 2023)*. Toronto, Canada: Association for Computational Linguistics, 174–179. DOI: 10.18653/v1/2023.starsem-1.16
Tien-Ping, Tan and Jia Jun, Dong (2021). Translating Idioms using Paraphrasing, Machine Translation and Rescoring. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12 (3), 1942–1946. DOI: 10.17762/turcomat.v12i3.1027
Tilmatine, Mesian; Hubers, Ferdy & Hintz, Florian (2021). Exploring Individual Differences in Recognizing Idiomatic Expressions in Context. *Journal of Cognition*, 4 (1), 37 DOI: 10.5334/joc.183
Toury, Gideon (1995). *Descriptive translation studies and beyond*. Amsterdam Philadelphia: J. Benjamins.
Tudhope, D., Alani, H., & Jones, C. B. (2001). Augmenting thesaurus relationships: possibilities for retrieval. *Journal of Digital Information* 1(8), 1-20.
Turaev, Shakhzod (2025). Classification of English idioms. *Tadbirkorlik Va Pedagogika. Ilmiy-Uslubiy Jurnal* 4, 10-17.
Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Ł. ukasz & Polosukhin, Illia (2017). *Attention is All you Need*. In: Guyon,

I., Luxburg, U. Von, Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., & Garnett, R. (Eds.) *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Vol. 30, 1-11.

Veisbergs, Andrejs (1996). Idiom transformation, idiom translation and idiom dictionaries. In: Zettersten Arne & Pedersen, Viggo Hjornager (eds.) *Symposium on Lexicography VII*. Berlin/Munich/Boston: De Gruyter, 241–245. DOI: 10.1515/9783110937916-019

Vieira, Lucas Nunes; O'Sullivan, Carol; Zhang, Xiaochun & O'Hagan, Minako (2023). Machine translation in society: insights from UK users. *Language Resources and Evaluation* 57 (2), 893–914. DOI: 10.1007/s10579-022-09589-1

Vilar, David; Leusch, Gregor; Ney, Hermann and Banchs, Rafael E. (2007). Human evaluation of machine translation through binary system comparisons. In: *Proceedings of the Second Workshop on Statistical Machine Translation*, 96–103, Prague, Czech Republic.

Vilar, David; Xu, Jia; D'Haro, Luis Fernando and Ney, Hermann (2006). Error Analysis of Statistical Machine Translation Output. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy.

Vinay, Jean-Paul & Darbelnet, Jean (1995). *Comparative stylistics of French and English: a methodology for translation*. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/btl.11

Volk, Martin (1998). The automatic translation of idioms. Machine translation vs. translation memory systems. In: Weber, Nico. *Machine translation: theory, applications, and evaluation*. *An assessment of the state of the art*. St. Augustin: Gardez!-Verlag, 167-192. DOI: 10.5167/UZH-19070

Von Eye, Alexander (2006). An Alternative to Cohen's κ. *European Psychologist*, 11 (1), 12–24 DOI: 10.1027/1016-9040.11.1.12

Wang, Haifeng; Wu, Hua; He, Zhongjun; Huang, Liang & Church, Kenneth Ward (2022). Progress in Machine Translation. *Engineering* 18, 143–153. DOI: 10.1016/j.eng.2021.03.023

Wang, Jun-song (2017). Translation Quality Assessment: Bridging the gap between theory and practice. *Journal of Literature and Art Studies* 7 (10), 1701-1706.

Wang, Lanchun & Wang, Shuo (2013). A study of idiom translation strategies between English and Chinese. *Theory and Practice in Language Studies* 3 (9), 1691–1697. DOI: 10.4304/tpls. 3.9.1691-1697

Weaver, Warren (1999). Warren Weaver Memorandum, July 1949. *MT News International* 22, 5–15.

Wei, Bai (2024). Advancements and challenges in AI-driven creative translation: A comprehensive analysis. *Applied and Computational Engineering* 82 (1), 82–87.

Winata, Genta Indra; Anugraha, David; Susanto, Lucky; Kuwanto, Garry & Wijaya, Derry Tanti (2025). MetaMetrics: Calibrating Metrics For Generation Tasks Using Human Preferences. arXiv DOI: 10.48550/arXiv.2410.02381

Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le, Quoc V.; Norouzi, Mohammad; Macherey, Wolfgang; Krikun, Maxim; Cao, Yuan; Gao, Qin; Macherey, Klaus; Klingner, Jeff; Shah, Apurva; Johnson, Melvin; Liu, Xiaobing; Kaiser, Łukasz; Gouws, Stephan; Kato, Yoshikiyo; Kudo, Taku; Kazawa, Hideto; Stevens, Keith; Kurian, George; Patil, Nishant; Wang, Wei; Young, Cliff; Smith, Jason; Riesa, Jason; Rudnick, Alex; Vinyals, Oriol; Corrado, Greg; Hughes, Macduff & Dean, Jeffrey (2016). Google's neural mac hine translation system: Bridging the gap between human and machine translation (Vers ion 2). arXiv:1609.08144v2 (Stand: 18.06.2025).

Zaitova, Iuliia; Abdullah, Badr M.; Xue, Wei; Klakow, Dietrich; Möbius, Bernd & Avgustinova, Tania (2025). It's Not a Walk in the Park! Challenges of Idiom Translation in Speech-to-text Systems. arXiv:2506.02995v1

Zajdel, Alicja (2022). Catching the meaning of words. In: Luke Hadley, James, Taivalkoski-Shilov, Kristiina, S. C. Teixeira, Carlos, & Toral, Antonio. *Using Technologies for Creative-Text Translation*. 1st ed. New York: Routledge, 116–138.

Zaninello, Andrea & Birch, Alexandra (2020). Multi-word expression aware neural machine translation. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*. Marseille: European Language Resources Association (ELRA), 3816–3825.

Zarechnak, Michael (1979). The history of machine translation. In: Henisz-Dostert, Bożena; Macdonald, R. R. & Zarechnak, Michael. *Machine Translation*. De Gruyter Mouton, 1–88. DOI: 10.1515/9783110816679.1

Zeng, Marcia Lei (2008). Knowledge organization systems (KOS). *Knowledge Organization* 35 (2–3), 160–182.

Zhao, Haofei; Liu, Yilun; Tao, Shimin; Meng, Weibin; Chen, Yimeng; Geng, Xiang; Su, Chang; Zhang, Min & Yang, Hao (2024). From handcrafted features to LLMs: A brief survey for Machine Translation Quality Estimation. In *2024 International Joint Confere nce on Neural Networks (IJCNN)*, pages 1–10.

Zhou, Jianing; Zeng, Ziheng; Gong, Hongyu & Bhat, Suma (2022). Idiomatic Expression Paraphrasing without Strong Supervision. *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (10), 11774–11782. DOI: 10.1609/aaai.v36i10.21433

11. Appendix

11.1. Test Sentences in Context (overview)

ID	Test sentence	Google Translate	DeepL	SYSTRAN
T1-1-1	Ongoing treatment in psychotherapy, rehab centers or 12-step groups encourages temporary runs of sobriety, but it's easier to **kick the bucket** than to kick the habit (Science News).	Eine fortlaufende Behandlung in Psychotherapie, Reha-Zentren oder 12-Stufen-Gruppen fördert vorübergehende Phasen der Nüchternheit, aber es ist einfacher, **den Löffel aufzugeben**, als die Gewohnheit aufzugeben.	Eine kontinuierliche Behandlung in Psychotherapien, Reha-Zentren oder 12-Schritte-Gruppen fördert eine vorübergehende Nüchternheit, aber es ist leichter, **den Löffel abzugeben** als die Gewohnheit zu überwinden.	Die laufende Behandlung in Psychotherapie, Reha-Zentren oder 12-Schritt-Gruppen fördert vorübergehende Abläufe von Nüchternheit, aber es ist einfacher, **den Eimer zu kicken** als die Gewohnheit zu kicken.
T1-1-2	I can't say when I'm going to **kick the bucket** but any suggestions you may have that will allow me to get some immediate financial assistance will be greatly appreciated.	Ich kann nicht sagen, wann ich **den Löffel abgeben** werde, aber ich würde mich über jeden Vorschlag, der mir eine sofortige finanzielle Unterstützung ermöglichen könnte, sehr freuen.	Ich kann nicht sagen, wann ich **ins Gras beißen** werde, aber für jeden Vorschlag, der es mir ermöglicht, sofortige finanzielle Unterstützung zu erhalten, wäre ich Ihnen sehr dankbar.	Ich kann nicht sagen, wann ich **den Eimer kicken** werde, aber alle Vorschläge, die Sie haben, die es mir ermöglichen, eine sofortige finanzielle Unterstützung zu erhalten, werden sehr geschätzt.
T1-1-3	Cheryl had rejected her, her mother was long gone, and old Mr. Feldman **kicked the bucket** just when she was warming up to him.	Cheryl hatte sie zurückgewiesen, ihre Mutter war schon lange tot und der alte Mr. Feldman **starb**, gerade als sie sich für ihn erwärmte.	Cheryl hatte sie abgewiesen, ihre Mutter war schon lange tot, und der alte Mr. Feldman hatte gerade dann **ins Gras gebissen**, als sie sich für ihn erwärmte.	Cheryl hatte sie abgelehnt, ihre Mutter war längst weg, und der alte Mr. Feldman **trat den Eimer**, gerade als sie sich vor ihm erwärmte.
T1-1-4	But then he decided he wanted to manage the whole estate! Like he pictured himself in an office with a secretary and a desk, and the rest of us paying him to do the bookkeeping that the bank was doing anyway! And nobody	Aber dann beschloss er, den ganzen Nachlass zu verwalten! Er stellte sich vor, er säße in einem Büro mit Sekretärin und Schreibtisch, und wir anderen bezahlten ihn für die Buchhaltung, die die Bank sowieso	Aber dann beschloss er, dass er das ganze Anwesen verwalten wollte! Er stellte sich ein Büro mit einer Sekretärin und einem Schreibtisch vor, und wir anderen bezahlten ihn dafür, dass er die Buchhaltung erledigte, die oh-	Aber dann entschied er sich, das ganze Anwesen zu verwalten! Als ob er sich in einem Büro mit Sekretärin und Schreibtisch vorgestellt hätte und der Rest von uns ihn bezahlt hätte, um die Buchhaltung zu machen, die

	wanted that. So then they started to go after me. I wasted so much money on lawyers, you've no idea. Your parents were scared I'd run away with the money or something. And I'm really not the type! In fact I watched the money carefully. But they kept on suing. I've got papers around here somewhere. They're waiting for me to **kick the bucket** cause then they'll get more money.	machte! Und das wollte niemand. Also fingen sie an, mich zu verfolgen. Ich habe so viel Geld für Anwälte verschwendet, du hast keine Ahnung. Deine Eltern hatten Angst, ich würde mit dem Geld abhauen oder so. – Und ich bin wirklich nicht der Typ dafür! Ich habe sogar gut auf das Geld aufgepasst. Aber sie verklagten mich immer wieder. Ich habe hier irgendwo Unterlagen. Sie warten darauf, dass ich **abkratze**, denn dann kriegen sie mehr Geld.	nehin von der Bank erledigt wurde! Und das wollte niemand. Und dann waren sie hinter mir her. Ich habe so viel Geld für Anwälte verschwendet, das kannst du dir gar nicht vorstellen. Deine Eltern hatten Angst, dass ich mit dem Geld abhauen würde oder so. Und ich bin wirklich nicht der Typ dafür! Ich habe das Geld genau beobachtet. Aber sie haben immer weiter geklagt. Ich habe hier irgendwo Papiere. Sie warten darauf, dass ich **ins Gras beiße**, dann bekommen sie mehr Geld.	die Bank sowieso gemacht hat! Und niemand wollte das. Also fingen sie an, mich zu verfolgen. Ich habe so viel Geld für Anwälte verschwendet, du hast keine Ahnung. Deine Eltern hatten Angst, dass ich mit dem Geld davonlaufe. Und ich bin wirklich nicht der Typ! Tatsächlich habe ich das Geld genau beobachtet. Aber sie klagten weiter. Ich habe hier irgendwo Papiere. Sie warten darauf, dass ich **den Eimer kicke,** denn dann bekommen sie mehr Geld.
T1-2-1	Brady Hair, a solo practitioner and president of the Charleston County Bar Association, sometimes ducks into Brent's on his way to court or to visit a colleague. He inevitably finds someone he knows there and spends a few minutes **shooting the breeze**. "Most of the lawyers in Charleston know each other," he says. "We don't have big firms here." Although specialization is beginning to creep in, many of Charleston's lawyers are still general practitioners in the traditional	Brady Hair, Einzelanwalt und Präsident der Anwaltskammer von Charleston County, schaut auf dem Weg zum Gericht oder zu einem Kollegen manchmal bei Brent vorbei. Dort trifft er unweigerlich jemanden, den er kennt, und **plaudert** ein paar Minuten. „Die meisten Anwälte in Charleston kennen sich“, sagt er. „Wir haben hier keine großen Kanzleien.“ Obwohl sich die Spezialisierung langsam durchsetzt, sind viele Anwälte in Charleston	Brady Hair, Einzelanwalt und Präsident der Charleston County Bar Association, schaut auf dem Weg zum Gericht oder zum Besuch eines Kollegen manchmal bei Brent's vorbei. Er trifft dort unweigerlich jemanden, den er kennt, und **verbringt ein paar Minuten mit einem Plausch**. Die meisten Anwälte in Charleston kennen sich untereinander", sagt er." Wir haben hier keine großen Firmen. "Obwohl sich die Spezialisierung immer mehr durchsetzt, sind viele der Anwälte in Charleston	Brady Hair, ein Solo-Praktizierender und Präsident der Charleston County Bar Association, duckt sich manchmal bei Brent ein, auf dem Weg zum Gericht oder zum Besuch eines Kollegen. Er findet dort unweigerlich jemanden, den er kennt, und verbringt ein paar Minuten damit, **die Brise zu erschießen.**

	mold - because firms tend to be small, they tend to take just about any reasonable case that crosses their thresholds.	immer noch Allgemeinmediziner im traditionellen Sinne – da die Kanzleien meist klein sind, nehmen sie praktisch jeden vernünftigen Fall an, der ihnen über den Weg läuft.	immer noch Allgemeinmediziner im traditionellen Sinne - weil die Kanzleien eher klein sind, nehmen sie so ziemlich jeden vernünftigen Fall an, der ihre Schwelle überschreitet.	
T1-2-2	Whether you're on a date or **shooting the breeze** with a higher-up at work you'd like to impress, our cheat sheet will have you slinging common literary and political terms with the best of them.	Ganz gleich, ob Sie ein Date haben oder sich mit einem Vorgesetzten bei der Arbeit **unterhalten**, den Sie beeindrucken möchten, mit unserem Spickzettel können Sie sich mit den Besten über gängige literarische und politische Begriffe austauschen.	Egal, ob Sie ein Date haben oder mit einem Vorgesetzten bei der Arbeit **plaudern**, den Sie beeindrucken möchten - mit unserem Spickzettel können Sie die gängigen literarischen und politischen Begriffe mit den Besten austauschen.	Egal, ob Sie auf einem Date sind oder mit einer höheren Person auf der Arbeit **die Brise schießen**, die Sie beeindrucken möchten, unser Spickzettel wird Sie dazu bringen, gemeinsame literarische und politische Begriffe mit den besten von ihnen zu schleudern.
T1-2-3	But if you can walk into the subway and see a group of security personnel **shooting the breeze** together like before, you'll know we're not serious.	Aber wenn Sie in die U-Bahn gehen und eine Gruppe von Sicherheitsleuten sehen, die wie zuvor miteinander **plaudern**, wissen Sie, dass wir es nicht ernst meinen.	Aber wenn Sie in die U-Bahn einsteigen und eine Gruppe von Sicherheitskräften sehen, die wie früher miteinander **plaudern**, dann wissen Sie, dass wir das nicht ernst meinen.	Aber wenn Sie in die U-Bahn gehen und sehen, wie eine Gruppe von Sicherheitsleuten **die Brise** zusammen **schießt**, wie vorher, werden Sie wissen, dass wir es nicht ernst meinen.
T1-2-4	To pass time, some guys play cards. Others sit back and **shoot the breeze.**	Manche Leute vertreiben sich die Zeit mit Kartenspielen, andere lehnen sich zurück und **quatschen**.	Um sich die Zeit zu vertreiben, spielen einige Jungs Karten. Andere lehnen sich zurück und **unterhalten sich**.	Zum Zeitvertreib spielen einige Jungs Karten. Andere lehnen sich zurück und **schießen die Brise.**
T1-3-1	If it weren't for band, he'd **pop his clogs**.	Wenn es keine Band gäbe, **würden seine Clogs platzen**. Ich habe mich gefragt, ob du Lust auf etwas Leckeres hast?	Wenn es die Band nicht gäbe, würde er sich **die Hacken ablaufen**.	Wenn es keine Band gäbe, würde er seine **Clogs knallen lassen**.

T1-3-2	I want to have another 20 years before I **pop my clogs**.	Ich möchte noch 20 Jahre haben, bevor mir **die Puste ausgeht.**	Ich möchte noch 20 Jahre leben, bevor mir **die Klötze platzen.**	Ich will noch 20 Jahre, bevor ich meine **Clogs knalle.**
T1-3-3	Get regular exercise, maintain strong bonds with your family/friends/community, and don't overindulge in food or drink. And for heaven's sake, don't smoke. The rest is genetics and you can't stop that. If you do **pop your clogs** a year or two earlier, well, at least if you're successful you can retire younger and not have to eat cat food in your retirement years.	Treiben Sie regelmäßig Sport, pflegen Sie enge Bindungen zu Ihrer Familie, Ihren Freunden und Ihrer Gemeinde und übertreiben Sie es nicht mit Essen und Trinken. Und um Himmels willen, rauchen Sie nicht. Der Rest ist genetisch bedingt und lässt sich nicht ändern. Wenn Sie ein oder zwei Jahre früher **in Rente gehen**, können Sie, wenn Sie erfolgreich sind, zumindest früher in Rente gehen und müssen im Ruhestand kein Katzenfutter essen.	Treiben Sie regelmäßig Sport, pflegen Sie enge Beziehungen zu Ihrer Familie/Freunden/Gemeinschaft und nehmen Sie nicht zu viel zu essen oder zu trinken. Und, um Himmels willen, rauchen Sie nicht. Der Rest ist genetisch bedingt, und das können Sie nicht verhindern. Wenn Sie ein oder zwei Jahre früher **sterben**, können Sie sich zumindest jünger zur Ruhe setzen und müssen im Ruhestand kein Katzenfutter essen.	Machen Sie sich regelmäßig sportlich, pflegen Sie starke Bindungen zu Ihrer Familie/Freunden/Gemeinschaft und lassen Sie sich nicht zu sehr mit Essen oder Trinken verwöhnen. Und um Himmels willen, rauche nicht. Der Rest ist Genetik und das kann man nicht aufhalten. Wenn Sie Ihre **Clogs** ein oder zwei Jahre früher **platzen** lassen, nun, zumindest wenn Sie erfolgreich sind, können Sie jünger in Rente gehen und müssen in Ihren Ruhestandsjahren kein Katzenfutter essen.
T1-3-4	In fact we had seen better only last week when, in the hospital wing of The Maze Prison, IRA commander Bobby Sands had finally **popped his clogs.**	Tatsächlich hatten wir erst letzte Woche Besseres gesehen, als IRA-Kommandant Bobby Sands im Krankenflügel des Maze-Gefängnisses **endlich seinen Job machte.**	Tatsächlich hatten wir erst letzte Woche etwas Besseres gesehen, als der IRA-Kommandant Bobby Sands im Krankenflügel des Maze-Gefängnisses endlich seinen **Kopf aus der Schlinge zog.**	Tatsächlich hatten wir es erst letzte Woche besser gesehen, als IRA-Kommandant Bobby Sands im Krankenhausflügel des Maze-Gefängnisses endlich seine **Clogs platzen ließ**.
T2-1-1	Civil rights leaders **hit the ceiling** after seeing the plan and the White House quickly repudiated it.	Als Bürgerrechtler den Plan sahen, **stießen sie auf Ablehnung**, und das Weiße Haus lehnte ihn schnell ab.	Bürgerrechtler **gingen an die Decke**, als sie den Plan sahen, und das Weiße Haus wies ihn schnell zurück.	Bürgerrechtsführer **stießen an die Decke**, nachdem sie den Plan gesehen hatten, und das Weiße Haus lehnte ihn schnell ab.

T2-1-2	It is amazing to see politicos, who joy in the occasional "photo op" when children show up to learn the governmental process at "mayor for a day" and similar establishment charades, **hit the ceiling** when kids actually try to take part in the process.	Es ist erstaunlich zu sehen, wie Politiker, die sich über gelegentliche „Fototermine“ freuen, wenn Kinder auftauchen, um bei „Bürgermeister für einen Tag“ und ähnlichen Scharaden des Establishments den Regierungsablauf kennenzulernen, **an ihre Grenzen stoßen,** wenn Kinder tatsächlich versuchen, an den Prozessen teilzunehmen.	Es ist erstaunlich zu sehen, wie Politiker, die sich über den gelegentlichen „Fototermin“ freuen, wenn Kinder bei „Bürgermeister für einen Tag“ und ähnlichen Scharaden des Establishments den Regierungsprozess kennen lernen, **an die Decke gehen**, wenn Kinder tatsächlich versuchen, sich an dem Prozess zu beteiligen.	Es ist erstaunlich zu sehen, wie Politiker, die Freude an der gelegentlichen „Foto-OP“, wenn Kinder auftauchen, um den Regierungsprozess bei „Bürgermeister für einen Tag“ und ähnliche Establishment-Scharaden zu lernen, **an die Decke stoßen**, wenn Kinder tatsächlich versuchen, an dem Prozess teilzunehmen.
T2-1-3	When I told you what happened with my job, I expected you to **hit the ceiling**, Liz.	Als ich dir erzählte, was mit meinem Job passiert ist, erwartete ich, dass du **an die Decke gehen** würdest, Liz.	Als ich dir erzählt habe, was mit meinem Job passiert ist, habe ich erwartet, dass du **an die Decke gehst**, Liz.	Als ich dir erzählte, was mit meinem Job passiert ist, erwartete ich, dass du **die Decke triffst**, Liz.
T2-1-4	All the calls were screened through Yoko. He spoke to Yoko. He told her, apparently, that he was going to Japan and he was going to stay in the Lennon's suite that they usually stayed in when they went to Tokyo - the presidential suite at the Hotel Okura. And when Yoko heard this she **hit the ceiling** because, as far as she was concerned- or, at least, she told John and John told me that this would spoil	Alle Anrufe wurden von Yoko abgehört. Er sprach mit Yoko. Er erzählte ihr offenbar, dass er nach Japan reisen und in der Lennon-Suite übernachten würde, die sie normalerweise auf ihren Reisen nach Tokio nutzten – der Präsidentensuite im Hotel Okura. Als Yoko das hörte, **rastete sie aus**, denn ihrer Meinung nach – oder zumindest sagte sie es John und John sagte mir – würde dies ihr Hotel-	Alle Anrufe wurden über Yoko abgewickelt. Er sprach mit Yoko. Er erzählte ihr offenbar, dass er nach Japan fahren und in der Lennon-Suite wohnen würde, in der sie normalerweise wohnten, wenn sie nach Tokio fuhren - die Präsidentensuite im Hotel Okura. Und als Yoko das hörte, **ging** sie **an die Decke,** weil sie der Meinung war - oder zumindest sagte sie es John und John sagte mir, dass dies ihr Hotelkarma für immer verderben würde. Das war also sehr beunruhigend für sie.	Alle Anrufe wurden über Yoko abgehört. Er sprach mit Yoko. Er erzählte ihr, dass er nach Japan gehen würde und er in der Lennon-Suite bleiben würde, in der sie normalerweise wohnten, wenn sie nach Tokio gingen - der Präsidentensuite im Hotel Okura. Und als Yoko das hörte, **schlug** sie **an die Decke,** weil sie, was sie betraf -- oder zumindest, sie erzählte John und John mir, dass dies ihr Hotel-Karma für immer verderben würde. Das war für sie sehr alarmierend.

	their hotel karma forever. So this was very alarming to them.	karma für immer ruinieren. Das war für sie sehr beunruhigend.		
T2-2-1	Gabriel rolled his eyes in frustration, his leg started to fidget, remembering all the ones that Aziraphale has written that he had to read because it was his job. Saying nothing at all, **going around in circles** and reporting nothing. It gotten worse when the reports had to be a certain length or more other than writing one sentence reports.	Gabriel verdrehte frustriert die Augen, sein Bein begann zu zappeln, als er sich an all die Berichte erinnerte, die Aziraphale geschrieben hatte und die er lesen musste, weil es sein Job war. Er sagte nichts, **drehte sich im Kreis** und berichtete nichts. Es wurde schlimmer, als die Berichte eine bestimmte Länge oder mehr haben mussten, anstatt nur einzeilige Berichte zu schreiben.	Gabriel rollte frustriert mit den Augen, sein Bein begann zu zappeln und erinnerte sich an all die Berichte, die Aziraphale geschrieben hatte und die er lesen musste, weil es sein Job war. Sie sagten überhaupt nichts, **drehten sich im Kreis** und berichteten nichts. Noch schlimmer wurde es, wenn die Berichte eine bestimmte Länge oder mehr haben mussten, als nur einen Satz zu schreiben.	Gabriel rollte frustriert mit den Augen, sein Bein fing an zu zappeln und erinnerte sich an all die, die Aziraphale geschrieben hat, die er lesen musste, weil es sein Job war. Überhaupt nichts sagen, **im Kreis herumlaufen** und nichts berichten. Schlimmer wurde es, wenn die Berichte eine bestimmte Länge oder mehr haben mussten, als nur einen Satz Berichte zu schreiben.
T2-2-2	Come on, Eric. We're **going around in circles**. If we can get to the heart of why she's fighting you, we could have a chance of winning.	Komm schon, Eric. Wir **drehen uns im Kreis**. Wenn wir herausfinden, warum sie gegen dich kämpft, haben wir vielleicht eine Chance zu gewinnen.	Komm schon, Eric. **Wir drehen uns im Kreis.** Wenn wir herausfinden, warum sie gegen dich kämpft, haben wir eine Chance zu gewinnen.	Komm schon, Eric. Wir **drehen uns im Kreis.** Wenn wir zum Kern kommen, warum sie gegen dich kämpft, könnten wir eine Chance haben zu gewinnen.
T2-2-3	We have had three years going around in circles trying to work out how to make Brexit happen.	Wir **drehen uns** seit drei Jahren **im Kreis** und versuchen herauszufinden, wie wir den Brexit bewerkstelligen können.	Wir **haben uns** drei Jahre lang **im Kreis gedreht,** um herauszufinden, wie der Brexit zustande kommen soll.	Drei Jahre lang **haben** wir **im Kreis versucht herauszufinden**, wie wir den Brexit verwirklichen können.
T2-2-4	Dr. Friedman suggests professional help if your discussions constantly descend into name-calling or **keep going around in circles** without getting anywhere.	Dr. Friedman empfiehlt professionelle Hilfe, wenn Ihre Diskussionen ständig in Beschimpfungen versinken oder **sich** ständig **im Kreis drehen**,	Dr. Friedman empfiehlt professionelle Hilfe, wenn Ihre Diskussionen ständig in Beschimpfungen ausarten oder **sich im Kreis drehen**, ohne etwas zu bewirken.	Dr. Friedman schlägt professionelle Hilfe vor, wenn Ihre Diskussionen ständig in Beschimpfungen absinken oder **sich im Kreis bewegen**, ohne irgendwohin zu gelangen.

		ohne weiterzukommen.		
T2-3-1	When Sarkisian brought USC defensive coordinator with him, Carroll kind of **gritted his teeth**.	Als Sarkisian den USC-Verteidigungskoordinator mitbrachte, **biss** Carroll **die Zähne zusammen**.	Als Sarkisian den Defensivkoordinator der USC mitbrachte, **biss** Carroll **die Zähne zusammen**.	Als Sarkisian den USC Defensive Coordinator mitbrachte, **biss** Carroll irgendwie **die Zähne zusammen**.
T2-3-2	The amazing thing was that she'd just **gritted her teeth** and nursed through it for three weeks.	Das Erstaunliche war, dass **sie** einfach **die Zähne zusammengebissen** und drei Wochen lang durchgehalten **hatte**.	Das Erstaunliche daran war, dass sie **die Zähne zusammengebissen** und drei Wochen lang durchgehalten **hatte**.	Das Erstaunliche war, dass sie sich nur **die Zähne zusammenbeißt** und drei Wochen lang gepflegt hat.
T2-3-3	Hortensia **gritted her teeth** and held her breath before speaking, so it would come out calm.	Hortensia **biss die Zähne zusammen** und hielt den Atem an, bevor sie sprach, damit es ruhig klang.	Hortensia **biss die Zähne zusammen** und hielt den Atem an, bevor sie sprach, damit es ruhig klang.	Hortensia **biss die Zähne zusammen** und hielt den Atem an, bevor sie sprach, damit es ruhig herauskam.
T2-3-4	Despite my throbbing muscles, or even because of them, I continued to offer up the pain for him and others with this terrible blood cancer. I **gritted my teeth**, turned up my music player, and stumbled toward the finish line. In my near-delirium during those last miles, I could almost hear a sweet Irish accent encouraging my schoolmates and me to " offer it up " when we had a headache or skinned knee many years ago.	Trotz meiner pochenden Muskeln, oder gerade deswegen, ertrug ich weiterhin meinen Schmerz für ihn und andere, die an diesem schrecklichen Blutkrebs leiden. Ich **biss die Zähne zusammen**, drehte meinen Musikplayer auf und stolperte der Ziellinie entgegen. In meinem Delirium auf den letzten Kilometern konnte ich fast einen süßen irischen Akzent hören, der meine Schulkameraden und mich ermutigte, „es zu opfern“, als wir vor vielen Jahren Kopfschmerzen oder ein	Trotz oder gerade wegen meiner pochenden Muskeln opferte ich weiter den Schmerz für ihn und andere mit diesem schrecklichen Blutkrebs. Ich **biss die Zähne zusammen**, drehte meinen Musikplayer auf und stolperte auf die Ziellinie zu. In meinem Beinahe-Delirium auf den letzten Kilometern konnte ich fast hören, wie ein süßer irischer Akzent meine Schulkameraden und mich ermutigte, „aufzubegehren“, wenn wir vor vielen Jahren Kopfschmerzen oder ein aufgeschürftes Knie hatten.	Trotz meiner pochenden Muskeln, oder sogar wegen ihnen, habe ich ihm und anderen mit diesem schrecklichen Blutkrebs weiterhin den Schmerz geopfert. Ich **biss die Zähne zusammen**, drehte meinen Musikplayer auf und stolperte in Richtung Ziellinie. In meinem Beinahe-Delirium während dieser letzten Meilen konnte ich fast einen süßen irischen Akzent hören, der meine Schulkameraden und mich ermutigte, „es anzubieten“, wenn wir vor vielen Jahren Kopfschmerzen oder ein gehäutetes Knie hatten.

		aufgeschürftes Knie hatten.		
T2-4-1	There would be that time when he would **blow his top** and get that steely look in his eye, and you think to yourself, I don't want to be around this Navy fighter pilot for a moment or two.	Irgendwann würde es die Zeit geben, in der er **seinen Kopf auffliegen ließ** und diesen stählernen Blick in seinen Augen bekam, und Sie denken selbst: „Ich möchte nicht einen oder zwei Augenblicke in der Nähe dieses Marine-Kampfpiloten sein.	Es gab diese Momente, in denen er **sich aufregte** und diesen stählernen Blick hatte, und man dachte sich, dass man diesen Navy-Kampfpiloten nicht für einen Moment oder zwei in seiner Nähe haben wollte.	Es würde eine Zeit geben, in der er **sein Oberteil in die Luft sprengt** und diesen stählernen Blick in sein Auge bekommt, und du denkst, ich möchte nicht für ein oder zwei Augenblicke in der Nähe dieses Navy-Kampffliegers sein.
T2-4-2	A guy who **blows his top** in negotiations often comes out on top.	Wer bei Verhandlungen **die Fassung verliert**, gewinnt oft.	Ein Mann, der **sich** in Verhandlungen **verausgabt**, geht oft als Sieger hervor.	Ein Typ, der sich in Verhandlungen **die Spitze verpasst**, kommt oft an die Spitze.
T2-4-3	I don't know why you insist on making me **blow my top**.	Ich weiß nicht, warum Sie darauf bestehen, mich **ausrasten** zu lassen.	Ich weiß nicht, warum du darauf bestehst, dass ich mich **aufrege**.	Ich weiß nicht, warum du darauf bestehst, dass ich mein **Top vermassle.**
T2-4-4	It was just a regular day at home with my three sons. I **blew my top** over some trivial little thing. I can't even remember what. Maybe the kids had left toys lying all over the house, or they were practising new wrestling moves on each other, or maybe they had run in the house with muddyboots. Whatever it was, I lost my cool and hollered at all three kids till I was red in the face. Then I sent them up to their bedroom.	Es war ein ganz normaler Tag zu Hause mit meinen drei Söhnen. Ich bin wegen einer Kleinigkeit **völlig aus dem Häuschen geraten**. Ich weiß nicht mehr, was. Vielleicht hatten die Kinder überall im Haus Spielzeug liegen lassen, oder sie übten neue Wrestling-Techniken aneinander, oder vielleicht waren sie mit schmutzigen Stiefeln durchs Haus gerannt. Was auch immer es war, ich verlor die Fassung und	Es war ein ganz normaler Tag zu Hause mit meinen drei Söhnen. Ich **habe mich** über irgendeine banale Kleinigkeit **aufgeregt**. Ich kann mich nicht einmal daran erinnern, was es war. Vielleicht hatten die Kinder ihr Spielzeug im ganzen Haus liegen lassen, oder sie übten neue Wrestling-Moves aneinander, oder sie waren mit matschigen Stiefeln ins Haus gelaufen. Was auch immer es war, ich verlor die Fassung und brüllte alle drei Kinder an, bis ich rot im Gesicht war. Dann	Es war nur ein normaler Tag zu Hause mit meinen drei Söhnen. Ich habe mein **Oberteil** über irgendein triviales kleines Ding **geblasen**. Ich weiß nicht mal mehr, was. Vielleicht hatten die Kinder überall im Haus Spielzeug liegen lassen, oder sie übten neue Wrestling-Bewegungen aufeinander, oder vielleicht waren sie mit Schlammstiefeln im Haus gelaufen. Was auch immer es war, ich verlor die Ruhe und brüllte alle drei Kinder an, bis ich rot im Gesicht war. Dann

		schrie alle drei Kinder an, bis ich rot im Gesicht war. Dann schickte ich sie in ihre Schlafzimmer.	schickte ich sie auf ihr Zimmer.	schickte ich sie in ihr Schlafzimmer.
T2-5-1	He had positively **flipped his lid** the time Georgie had unlocked the door of his room and had taken a little stroll to stretch his legs in the cool of the evening.	Als Georgie die Tür seines Zimmers aufgeschlossen hatte, hatte er förmlich **den Deckel umgedreht** und einen kleinen Spaziergang gemacht, um sich in der Kühle des Abends die Beine zu vertreten.	Er **war regelrecht ausgeflippt**, als Georgie die Tür seines Zimmers aufgeschlossen hatte, und hatte einen kleinen Spaziergang gemacht, um sich in der Kühle des Abends die Beine zu vertreten.	Als Georgie die Tür seines Zimmers aufgeschlossen hatte, hatte er förmlich **den Deckel umgedreht** und einen kleinen Spaziergang gemacht, um sich in der Kühle des Abends die Beine zu vertreten.
T2-5-2	Aber die Augen in der Sphäre nicht. Und ich wusste nicht mal, wessen Augen sie waren! Ich wollte es unbedingt herausfinden. Ich erzählte meinem großen Brud+B31er von der Kugel und fragte, ob er so etwas jemals gesehen hätte. Seine Antwort war zuversichtlich, aber nicht allzu tröstlich: Er lachte und sagte mir, ich hätte meinen **Deckel umgedreht**. Ich versuchte es mit meiner Mutter. Auch sie lachte. “ Was für eine Fantasie! “, sagte sie.	Aber die Augen in der Kugel wollten nicht. Und ich wusste nicht einmal, wessen Augen das waren! Ich wollte es unbedingt herausfinden. Ich erzählte meinem großen Bruder von der Kugel und fragte ihn, ob er so etwas schon einmal gesehen habe. Seine Antwort war zuversichtlich, aber nicht gerade tröstlich: Er lachte und sagte, ich **sei ausgeflippt**. Ich versuchte es bei meiner Mutter. Auch sie lachte. „Was für eine Fantasie!“, sagte sie.	Aber die Augen in der Kugel wollten nicht. Und ich wusste nicht einmal, wessen Augen das waren! Das wollte ich unbedingt herausfinden. Ich erzählte meinem großen Bruder von der Kugel und fragte ihn, ob er so etwas schon einmal gesehen habe. Seine Antwort war zuversichtlich, aber nicht sehr tröstlich: Er lachte und sagte mir, ich **hätte den Kopf verloren**. Ich versuchte es bei meiner Mutter. Auch sie lachte. "Was für eine Fantasie! ", sagte sie.	Aber die Augen in der Sphäre nicht. Und ich wusste nicht mal, wessen Augen sie waren! Ich wollte es unbedingt herausfinden. Ich erzählte meinem großen Bruder von der Kugel und fragte, ob er so etwas jemals gesehen hätte. Seine Antwort war zuversichtlich, aber nicht allzu tröstlich: Er lachte und sagte mir, ich hätte meinen **Deckel umgedreht**. Ich versuchte es mit meiner Mutter. Auch sie lachte. “ Was für eine Fantasie! “, sagte sie.
T2-5-3	My husband would **flip his lid** to see this place.	Mein Mann würde **ausflippen**, wenn er diesen Ort sehen würde.	Mein Mann würde **ausflippen**, wenn er diesen Ort sehen könnte.	Mein Mann **hat den Deckel umgelegt**, um diesen Ort zu sehen.
T2-5-4	You have **flipped your lid!**	Du bist **ausgerastet**!	Sie **haben** den Mund **aufgemacht**!	Du **hast deinen Deckel umgelegt!**

T2-6-1	It seems to me that this is how it always is - a Christian says something in a loving manner but it's declared "hate speech" and then the people shouting out "hate monger" turn around and name-call and start **foaming at the mouth.**	Mir scheint, das ist immer so: Ein Christ sagt etwas auf liebevolle Art und Weise, aber es wird als „Hassrede“ bezeichnet, und dann drehen sich die Leute, die „Hassprediger“ schreien, um, beschimpfen ihn und fangen an, **Schaum vor dem Mund zu haben.**	Ich habe den Eindruck, dass es immer so abläuft - ein Christ sagt etwas in einer liebevollen Art und Weise, aber es wird als „Hassrede“ deklariert, und dann drehen sich die Leute, die „Hassprediger“ rufen, um, beschimpfen ihn und fangen an, **Schaum vor dem Mund zu haben**.	Es scheint mir, dass es so ist, wie es immer ist... ein Christ sagt etwas auf liebevolle Weise, aber es wird „Hassrede“ erklärt und dann drehen sich die Leute, die „Hassmonster“ rufen, um und rufen an und beginnen, **vor dem Mund zu schäumen**.
T2-6-2	Traditional publishing is **foaming at the mouth** to publish 50 Shades because they need the revenue. And it's not illegal right? Unethical yes, but unless the author gets sued by the author of the original work, more fan fiction is going to be published unless there is a massive law suit.	Traditionelle Verlage **schäumen vor Wut**, 50 Shades zu veröffentlichen, weil sie die Einnahmen brauchen. Und es ist doch nicht illegal, oder? Unethisch, ja, aber solange der Autor nicht vom Autor des Originalwerks verklagt wird, werden mehr Fanfictions veröffentlicht, es sei denn, es kommt zu einer massiven Klage.	Die traditionellen Verlage **haben Schaum vor dem Mund**, um 50 Shades zu veröffentlichen, weil sie die Einnahmen brauchen. Und es ist ja nicht illegal, oder? Unethisch ja, aber solange der Autor nicht vom Autor des Originalwerkes verklagt wird, wird mehr Fan-Fiction veröffentlicht werden, es sei denn, es kommt zu einem massiven Gerichtsverfahren.	Traditionelles Publizieren **schäumt vor dem Mund**, 50 Shades zu veröffentlichen, weil sie die Einnahmen brauchen. Und es ist nicht illegal, oder? Unethisch ja, aber wenn der Autor nicht vom Autor des Originalwerks verklagt wird, wird mehr Fan-Fiction veröffentlicht werden, es sei denn, es gibt eine massive Klage.
T2-6-3	The project left many critics **foaming at the mouth**, calling it empty and sold-out.	Das Projekt sorgte bei vielen Kritikern für **Schaum vor dem Mund**, sie bezeichneten es als leer und ausverkauft.	Das Projekt hat viele Kritiker mit **Schaum vor dem Mund** zurückgelassen, die es als leer und ausverkauft bezeichneten.	Das Projekt **ließ** viele **Kritiker vor dem Mund schäumen,** nannte es leer und ausverkauft.
T2-6-4	Dunc was brought in, kicking, screaming, **foaming at the mouth**.	Dunc wurde hereingebracht, strampelnd, schreiend und **mit Schaum vor dem Mund.**	Dunc wurde hereingebracht, schreiend und mit **Schaum vor dem Mund.**	Dunc wurde hergebracht, trat, schrie, **schäumte vor dem Mund**.

T2-7-1	Maybe it's time for the Anderson group to **smoke the peace pipe** with District Attorney Lykos and her supporters.	Vielleicht ist es für die Anderson-Gruppe an der Zeit, mit Bezirksstaatsanwältin Lykos und ihren Unterstützern **die Friedenspfeife zu rauchen**.	Vielleicht ist es für die Anderson-Gruppe an der Zeit, mit Staatsanwältin Lykos und ihren Anhängern **die Friedenspfeife zu rauchen**.	Vielleicht ist es an der Zeit, dass die Anderson-Gruppe mit Staatsanwalt Lykos und ihren Unterstützern **die Friedenspfeife raucht**.
T2-7-2	Churchill evidently **smoked the peace pipe** and got himself declared a Native American.	Churchill **rauchte** offenbar **die Friedenspfeife** und ließ sich zum amerikanischen Ureinwohner erklären.	Churchill **hat** offensichtlich **die Friedenspfeife geraucht** und sich zum amerikanischen Ureinwohner erklären lassen.	Churchill **rauchte** offenbar **die Friedenspfeife** und ließ sich zum Indianer erklären.
T2-7-3	I brought you a present, let's **smoke the peace pipe!**	Ich habe dir ein Geschenk mitgebracht, lass uns **die Friedenspfeife rauchen!**	Ich habe ein Geschenk für dich, lass uns **die Friedenspfeife rauchen!**	Ich habe dir ein Geschenk mitgebracht, **rauchen wir die Friedenspfeife!**
T2-7-4	The panel is read beginning from the right side, with the arrival of the White Rivers and their initial meeting with the Uintahs. Both bands agree to **smoke the peace pipe.** The progression through time is shown as the smaller characters on the right gradually increase in size as the sequence moves left.	Das Bild beginnt von rechts, mit der Ankunft der White Rivers und ihrer ersten Begegnung mit den Uintahs. Beide Gruppen vereinbaren, **die Friedenspfeife zu rauchen**. Der zeitliche Verlauf wird durch die nach links zunehmende Größe der kleineren Figuren rechts dargestellt.	Die Tafel wird von der rechten Seite beginnend gelesen, mit der Ankunft der White Rivers und ihrer ersten Begegnung mit den Uintahs. Beide Gruppen vereinbaren, **die Friedenspfeife zu rauchen**. Das Fortschreiten der Zeit wird dadurch dargestellt, dass die kleineren Figuren auf der rechten Seite allmählich größer werden, während sich die Sequenz nach links bewegt.	Die Tafel wird von der rechten Seite beginnend gelesen, mit der Ankunft der White Rivers und ihrem ersten Treffen mit den Uintah. Beide Bands stimmen zu, **die Friedenspfeife zu rauchen**. Die zeitliche Abfolge wird angezeigt, wenn die kleineren Zeichen auf der rechten Seite allmählich größer werden, wenn sich die Sequenz nach links bewegt.
T3-1-1	The man confessed to the computer theft and then **spilled the beans** on the luxury car theft scheme.	Der Mann habe den Computerdiebstahl gestanden und dann die Hintergründe des Diebstahls eines Luxusautos **verraten**.	Der Mann gestand den Computerdiebstahl und **verriet** dann den Plan für den Diebstahl von Luxusautos.	Der Mann gestand den Computerdiebstahl und **verschüttete** dann **die Bohnen** nach dem Luxusauto-Diebstahlschema.
T3-1-2	After the vendor **spilled the beans** about selling contraband potatoes, the	Nachdem der Verkäufer den Verkauf von Schmuggelkartoffeln	Nachdem der Verkäufer **den Verkauf** von	Nachdem der Verkäufer **die Bohnen** über den

	city issued a written warning.	**offengelegt hatte**, erließ die Stadt eine schriftliche Verwarnung.	Schmuggelkartoffeln **aufgedeckt hatte**, sprach die Stadt eine schriftliche Verwarnung aus.	Verkauf von Schmuggelkartoffeln **verschüttet hatte**, gab die Stadt eine schriftliche Warnung aus.
T3-1-3	Also I would never want to break their trust and **spill the beans o**n things they've told me.	Außerdem würde ich niemals ihr Vertrauen missbrauchen und **Dinge ausplaudern** wollen, die sie mir erzählt haben.	Außerdem würde ich niemals ihr Vertrauen brechen und **Dinge ausplaudern** wollen, die sie mir erzählt haben.	Außerdem würde ich nie ihr Vertrauen brechen wollen und **die Bohnen** auf Dinge **verschütten**, die sie mir erzählt haben.
T3-1-4	And you were lying, sir, when you said you met him there, because he was never there. And you walked to Mr Staplin's house because you didn't want your car seen outside his house. And you killed Mr Staplin because he threatened to **spill the beans** about the favour you did him in' 69 unless you did him another favour.	Und Sie haben gelogen, Sir, als Sie sagten, Sie hätten ihn dort getroffen, denn er war nie dort. Und Sie gingen zu Fuß zu Mr. Staplins Haus, weil Sie nicht wollten, dass Ihr Auto vor seinem Haus gesehen wurde. Und Sie haben Mr. Staplin getötet, weil er drohte, den Gefallen, den Sie ihm 1969 erwiesen hatten, **auszuplaudern**, wenn Sie ihm nicht noch einen Gefallen täten.	Und Sie haben gelogen, Sir, als Sie sagten, Sie hätten ihn dort getroffen, denn er war nie dort. Und Sie sind zu Fuß zu Mr. Staplins Haus gegangen, weil Sie nicht wollten, dass Ihr Auto vor seinem Haus gesehen wird. Und Sie haben Mr. Staplin getötet, weil er gedroht hat, den Gefallen, den Sie ihm 69 getan haben, **zu verraten**, wenn Sie ihm nicht einen weiteren Gefallen tun.	Und Sie haben gelogen, Sir, als Sie sagten, Sie hätten ihn dort getroffen, weil er nie da war. Und Sie gingen zu Mr. Staplins Haus, weil Sie nicht wollten, dass Ihr Auto vor seinem Haus gesehen wird. Und Sie haben Mr. Staplin getötet, weil er gedroht hat, **die Bohnen** über den Gefallen zu **vergießen**, den Sie ihm in '69 getan haben, es sei denn, Sie haben ihm einen anderen Gefallen getan.
T3-2-1	Obama **touched a nerve** with folks on the right Tuesday night because they know deep down that people are suffering from a lack of economic fairness.	Obama hat am Dienstagabend bei den Rechten **einen Nerv getroffen**, weil sie tief in ihrem Inneren wissen, dass die Menschen unter einem Mangel an wirtschaftlicher Gerechtigkeit leiden.	Obama hat am Dienstagabend **einen Nerv** bei den Rechten **getroffen**, weil sie wissen, dass die Menschen unter einem Mangel an wirtschaftlicher Fairness leiden.	Obama hat am rechten Dienstagabend bei den Leuten **einen Nerv berührt**, weil sie tief im Inneren wissen, dass die Menschen unter einem Mangel an wirtschaftlicher Fairness leiden.
T3-2-2	The debate **touches a nerve**. Polls show that very few people favor higher immigration,	Die Debatte **berührt einen Nerv**. Umfragen zeigen, dass nur	Die Debatte **berührt einen Nerv**. Umfragen zeigen, dass nur sehr wenige Menschen eine höhere	Die Debatte **berührt einen Nerv.** Umfragen zeigen, dass nur sehr wenige Menschen eine höhere

	even if many are content with the current level or divided about what to do about those already here.	sehr wenige Menschen eine höhere Einwanderung befürworten, auch wenn viele mit dem aktuellen Niveau zufrieden sind oder sich uneinig sind, was mit den bereits hier lebenden Menschen geschehen soll.	Einwanderung befürworten, auch wenn viele mit dem derzeitigen Niveau zufrieden sind oder uneins darüber sind, was mit den bereits hier lebenden Menschen geschehen soll.	Einwanderung befürworten, auch wenn viele mit dem aktuellen Niveau zufrieden sind oder sich uneins darüber sind, was mit den bereits hier Anwesenden geschehen soll.
T3-2-3	This really **touched a nerve** with me, as I have been battling depression myself.	Das **hat** mich wirklich **berührt**, da ich selbst mit Depressionen zu kämpfen hatte.	Das hat bei mir **einen Nerv getroffen**, da ich selbst mit Depressionen zu kämpfen habe.	Das **berührte** bei mir wirklich **einen Nerv**, da ich selbst Depressionen bekämpfe.
T3-2-4	There are many things that are hard to do, and probably the hardest is to pass the knowledge acquired by oneself to other people, the main reason I see for me is that wrong conclusions could bring a lot of confusion, spreading of misinformation and probably other consequences. I am trying to apply this knowledge in some practical ways, but I feel like not doing much, and am not sure if it is this the reason of me being here. I think you've **touched a nerve** here. I believe most everyone on this blog is trying to figure out what their calling is and how to do more. Its a frustrating situation.	Vieles ist schwer zu bewerkstelligen, und am schwierigsten ist es wahrscheinlich, das selbst erworbene Wissen an andere weiterzugeben. Der Hauptgrund dafür ist meiner Meinung nach, dass falsche Schlussfolgerungen zu großer Verwirrung, der Verbreitung von Fehlinformationen und möglicherweise weiteren Konsequenzen führen können. Ich versuche, dieses Wissen praktisch anzuwenden, habe aber das Gefühl, nicht viel zu tun, und bin mir nicht sicher, ob das der Grund für meine Anwesenheit ist. Ich glaube, du hast hier **einen Nerv getroffen**. Ich glaube, fast jeder in diesem	Es gibt viele Dinge, die schwer zu tun sind, und wahrscheinlich ist es am schwierigsten, das von einem selbst erworbene Wissen an andere Menschen weiterzugeben. Der Hauptgrund, den ich für mich sehe, ist, dass falsche Schlussfolgerungen eine Menge Verwirrung, die Verbreitung von Fehlinformationen und wahrscheinlich andere Konsequenzen mit sich bringen könnten. Ich versuche, dieses Wissen auf praktische Weise anzuwenden, aber ich habe das Gefühl, nicht viel zu tun, und bin mir nicht sicher, ob das der Grund ist, warum ich hier bin. Ich glaube, Sie haben hier **einen Nerv getroffen.** Ich glaube, dass die meisten Menschen in diesem Blog versuchen, herauszufinden, was ihre	Es gibt viele Dinge, die schwer zu tun sind, und wahrscheinlich ist das Schwierigste, das Wissen, das man selbst erworben hat, an andere Menschen weiterzugeben. Der Hauptgrund, den ich für mich sehe, ist, dass falsche Schlussfolgerungen viel Verwirrung, Verbreitung von Falschinformationen und wahrscheinlich andere Folgen bringen könnten. Ich versuche, dieses Wissen auf einige praktische Weise anzuwenden, aber ich habe Lust, nicht viel zu tun, und ich bin mir nicht sicher, ob dies der Grund dafür ist, dass ich hier bin. Ich glaube, Sie **haben** hier **einen Nerv berührt**. Ich glaube, dass die meisten auf diesem Blog versuchen herauszufinden, was ihre Berufung ist und

		Blog versucht herauszufinden, was seine Berufung ist und wie er mehr erreichen kann. Es ist frustrierend.	Berufung ist und wie sie mehr tun können. Das ist eine frustrierende Situation.	wie man mehr tun kann. Es ist eine frustrierende Situation.
T3-3-1	It might be supposed that it would be in Japan's interest to clarify its position by **grasping the nettle** of constitutional revision (or reinterpretation).	Man könnte annehmen, dass es im Interesse Japans wäre, seine Position **durch den Griff** zur Verfassungsrevision (oder Neuinterpretation) zu klären.	Man könnte meinen, dass es im Interesse Japans wäre, seine Position zu klären, indem es **sich in die Nesseln** einer Verfassungsrevision (oder Neuinterpretation) **setzt**.	Man könnte annehmen, dass es im Interesse Japans wäre, seine Position zu klären, indem man **die Nessel** der Verfassungsänderung (oder Neuinterpretation) **ergreift**.
T3-3-2	A rhetorical method would **grasp the nettle** and acknowledge the ethical and political interests with which arguments are constructed.	Eine rhetorische Methode würde **das Problem angehen** und die ethischen und politischen Interessen berücksichtigen, mit denen Argumente konstruiert werden.	Eine rhetorische Methode würde **in die Trickkiste greifen** und die ethischen und politischen Interessen anerkennen, mit denen Argumente konstruiert werden.	Eine rhetorische Methode würde **die Nessel greifen** und die ethischen und politischen Interessen anerkennen, mit denen Argumente konstruiert werden.
T3-3-3	Tomorrow, perhaps -- You'd do better to **grasp the nettle.**	Morgen vielleicht – Sie sollten lieber **die Initiative ergreifen.**	Morgen vielleicht - Es wäre besser, wenn Sie **die Nessel packen** würden.	Morgen vielleicht -- Sie sollten besser **die Brennnessel greifen.**
T3-3-4	I have found the sting a highly variable quality, which is why there is disagreement about the meaning of the old adage: **Grasp the nettle.** An entirely rational British friend of mine insists that if you really grab hold of young nettles, they won't sting. Perhaps she has been lucky. In any case, I believe that whoever first advised **grasping the nettle** meant something like what we mean	Ich habe festgestellt, dass der Stich sehr unterschiedlich ist, weshalb es Uneinigkeit über die Bedeutung des alten Sprichworts gibt: „**Greif die Brennnessel an**." Eine völlig vernünftige britische Freundin von mir behauptet, junge Brennnesseln würden nicht brennen, wenn man sie richtig anfasst. Vielleicht hatte sie	Ich habe festgestellt, dass die Qualität des Stachels sehr unterschiedlich ist, weshalb man sich über die Bedeutung des alten Sprichworts „**Greif die Brennnessel**" nicht einig ist. Eine ganz vernünftige britische Freundin von mir besteht darauf, dass junge Brennnesseln nicht stechen, wenn man sie richtig festhält. Vielleicht hat sie Glück gehabt. Auf jeden Fall glaube ich, dass derjenige, der zuerst dazu riet, **die Brennnessel zu**	Ich habe den Stich als sehr variabel empfunden, weshalb es Uneinigkeit über die Bedeutung des alten Sprichworts gibt: **Greif die Brennnessel.** Ein völlig rationaler britischer Freund von mir besteht darauf, dass, wenn man sich wirklich junge Brennnesseln schnappt, sie nicht stechen. Vielleicht hatte sie Glück. Auf jeden Fall glaube ich, dass derjenige, der zuerst geraten hat, die **Brennnessel zu greifen,** so etwas wie

	when we say bite the bullet. Once you've touched the nettle and felt the sting, you might as well grasp the plant securely and pull it out of the ground.	Glück. Ich glaube jedenfalls, dass derjenige, der zuerst riet, **die Brennnessel anzufassen**, so etwas meinte wie das, was wir meinen, wenn wir sagen: „In den sauren Apfel beißen“. Sobald man die Brennnessel berührt und den Stich gespürt hat, kann man die Pflanze genauso gut festhalten und aus dem Boden ziehen.	**packen**, etwas Ähnliches meinte wie wir, wenn wir sagen: „Beiß in den sauren Apfel“. Wenn man die Brennnessel berührt und den Stich gespürt hat, kann man die Pflanze genauso gut festhalten und aus dem Boden ziehen.	das meinte, was wir meinen, wenn wir sagen, beiß in die Kugel. Sobald man die Brennnessel berührt und den Stich gespürt hat, kann man die Pflanze auch sicher fassen und aus dem Boden ziehen.
T3-4-1	However, he returns to his senses and his religion, composes himself and **swallows the bitter pill** of having to turn one's back on them.	Doch er kehrt zur Besinnung und zu seiner Religion zurück, fasst sich und **schluckt die bittere Pille** herunter, ihnen den Rücken kehren zu müssen.	Aber er besinnt sich auf seinen Verstand und seine Religion, beruhigt sich und **schluckt die bittere Pille**, ihnen den Rücken kehren zu müssen.	Er kehrt jedoch zu seinen Sinnen und seiner Religion zurück, komponiert sich und **schluckt die bittere Pille**, sich von ihnen abwenden zu müssen.
T3-4-2	He who, just ten minutes earlier, had to **swallow the bitter pill** of bourgeois humiliation, looms like a tyrant and makes these unhappy creatures swallow the bitter pill of their supposed inferiority.	Er, der nur zehn Minuten zuvor **die bittere Pille** der bürgerlichen Demütigung **schlucken** musste, steht wie ein Tyrann da und lässt diese unglücklichen Kreaturen die bittere Pille ihrer vermeintlichen Minderwertigkeit schlucken.	Derjenige, der noch zehn Minuten zuvor die bittere Pille der bürgerlichen Demütigung schlucken musste, tritt wie ein Tyrann auf und lässt diese unglücklichen Geschöpfe **die bittere Pille** ihrer vermeintlichen Minderwertigkeit **schlucken.**	Wer nur zehn Minuten zuvor **die bittere Pille** der bürgerlichen Demütigung **schlucken** musste, wirkt wie ein Tyrann und lässt diese unglücklichen Geschöpfe die bittere Pille ihrer vermeintlichen Unterlegenheit schlucken.
T3-4-3	Thanks again for that entry, Ben. It really helps with **swallowing the bitter pill** of rejection.	Nochmals vielen Dank für diesen Eintrag, Ben. Er hilft wirklich, **die bittere Pille** der Ablehnung zu **schlucken**.	Nochmals vielen Dank für diesen Beitrag, Ben. Er hilft wirklich dabei, **die bittere Pille** der Ablehnung zu **schlucken**.	Nochmals vielen Dank für diesen Eintrag, Ben. Es hilft wirklich beim **Schlucken der bitteren Pille** der Ablehnung.
T3-4-4	Every day, no matter how bright the sun or	Jeden Tag, egal wie hell die Sonne oder	Jeden Tag, egal wie hell die Sonne scheint oder	Jeden Tag **schluckst** du, egal wie hell die Sonne

	how cool the trades, you **swallow the bitter pill** that with eyes so opened it turned out you were blind. Yes, it happened here on paradise where nothing more than God's sweet sun and gentle trades are supposed to happen. Since the fall of paradise, things you thought you believed, you find no longer to be true. You've come up with new guidelines: Not all people can be trusted.	wie kühl der Wind ist, **schluckst du die bittere Pille**, dass du mit offenen Augen blind warst. Ja, es geschah hier im Paradies, wo eigentlich nichts anderes als Gottes süße Sonne und sanfter Wind geschehen sollte. Seit dem Untergang des Paradieses entpuppt sich, was du zu glauben glaubtest, als falsch. Du hast dir neue Regeln ausgedacht: # Nicht allen Menschen kann man trauen.	wie kühl der Handel ist, **schluckst** du **die bittere Pille**, dass du mit so offenen Augen dastehst, dass du blind bist. Ja, das ist hier im Paradies passiert, wo nichts weiter als Gottes süße Sonne und sanfte Geschäfte passieren sollten. Seit dem Untergang des Paradieses sind die Dinge, die du zu glauben glaubtest, nicht mehr wahr. Du hast dir neue Richtlinien ausgedacht: # Man kann nicht allen Menschen trauen.	oder wie kühl die Geschäfte sind, **die bittere Pille**, die sich mit so offenen Augen herausstellte, dass du blind bist. Ja, es geschah hier im Paradies, wo nichts mehr als Gottes süße Sonne und sanfte Geschäfte passieren sollen. Seit dem Fall des Paradieses findet man Dinge, von denen man glaubte, sie zu glauben, nicht mehr wahr. Sie haben sich neue Richtlinien einfallen lassen: Nicht allen Menschen kann man vertrauen.
T3-5-1	They agreed that the task at hand was greater than any ill will or bad press and decided **to bury the hatchet**, at least professionally.	Sie waren sich einig, dass die bevorstehende Aufgabe größer war als jeder böse Wille oder die schlechte Presse, und beschlossen, **das Kriegsbeil zu begraben**, zumindest beruflich.	Sie waren sich einig, dass die anstehende Aufgabe wichtiger war als jeglicher böser Wille oder schlechte Presse, und beschlossen, **das Kriegsbeil zu begraben**, zumindest auf professioneller Ebene.	Sie waren sich einig, dass die anstehende Aufgabe größer war als jeder böse Wille oder jede schlechte Presse und beschlossen, **das Kriegsbeil zu begraben**, zumindest professionell.
T3-5-2	Even more worryingly, Lugo quickly indicated his desire to **bury the hatchet** with Bolivia and flew to La Paz to meet with Morales personally.	Noch beunruhigender ist, dass Lugo rasch seinen Wunsch signalisierte, **das Kriegsbeil** mit Bolivien zu **begraben**, und nach La Paz flog, um sich persönlich mit Morales zu treffen.	Noch besorgniserregender war, dass Lugo schnell seinen Wunsch äußerte, **das Kriegsbeil** mit Bolivien zu **begraben**, und nach La Paz flog, um sich persönlich mit Morales zu treffen.	Noch beunruhigender war, dass Lugo schnell seinen Wunsch äußerte, **das Kriegsbeil** mit Bolivien **zu begraben**, und nach La Paz flog, um sich persönlich mit Morales zu treffen.
T3-5-3	But I hope we can make a fresh start, **bury the hatchet**.	Aber ich hoffe, dass wir einen Neuanfang machen und **das**	Aber ich hoffe, wir können einen Neuanfang machen und **das Kriegsbeil begraben.**	Aber ich hoffe, wir können einen Neuanfang machen, **das Kriegsbeil begraben.**

		<u>Kriegsbeil begraben können.</u>		

11.2. Machine Translation Output without Context (overview)

Idiom	Google Translate	DeepL	SYSTRAN
to kick the bucket	*den Löffel abgeben*	*ins Gras beißen*	*den Eimer treten*
To shoot the breeze	*schießen Sie die Brise*	*ein bisschen plaudern*	*die Brise abschießen*
To pop one's clogs	*die Schuhe platzen lassen*	*sich die Klinke in die Hand geben*	*die Holzschuhe knallen lassen*
To hit the ceiling	*an die Decke stoßen*	*die Decke treffen*	*an die Decke gehen*
go around in circles	*sich im Kreis drehen*	*sich im Kreis drehen*	*im Kreis laufen*
grit one's teeth	*die Zähne zusammenbeißen*	*die Zähne zusammenbeißen*	*Zähne beißen*
blow one's top	*sich in die Luft jagen*	*sich die Blöße geben*	*sich die Oberseite sprengen*
flip one's lid	*den Deckel umdrehen*	*die Klappe halten*	*den Deckel umlegen*
foam at the mouth	*vor dem Mund schäumen*	*Schaum vor dem Mund zu haben*	*am Mund schäumen*
smoke the peace pipe	*die Friedenspfeife rauchen*	*um die Friedenspfeife zu rauchen*	*die Friedenspfeife rauchen*
spill the beans	*die Bohnen verschütten*	*die Wahrheit sagen*	*die Bohnen verschütten*
touch a nerve	*einen Nerv berühren*	*einen Nerv treffen*	*einen Nerv berühren*
grasp the nettle	*die Brennnessel fassen*	*in den sauren Apfel beißen*	*die Brennnessel fassen*
swallow the bitter pill	*die bittere Pille schlucken*	*die bittere Pille schlucken*	*die bittere Pille schlucken*
bury the hatchet	*das Kriegsbeil begraben*	*um das Kriegsbeil zu begraben*	*das Beil begraben*