MACHINE TRANSLATION AND SPECIALIZED TEXTS: A BRIEF OVERVIEW COMPARING GOOGLE TRANSLATE, DEEPL AND CHATGPT

Ivanka Sakareva

Abstract

Machine translation (MT) is essential in an era of globalization and fast information exchange, especially for translating specialized texts like technical documentation, scientific papers, and legal documents, which are crucial for international collaboration and knowledge sharing. These texts present unique challenges due to their specific terminology, complex structures, and nuanced contexts. Effective MT tools such as Google Translate, DeepL and ChatGPT need to deal not only with language equivalence but also with content-specific complexities. Translating specialized texts, therefore, requires more than just linguistic skills; it necessitates domain-specific knowledge and an appreciation of the complex relationship between language, structure, and context (based on a personal experience of translating legal, technical and scientific texts). As MT technology progresses, driven by advances in computational linguistics and AI, it aims to produce translations that are both technically accurate and contextually nuanced, addressing the challenges posed by the complexity and variability of human language, thus improving global communication and understanding.

Keywords: ***machine translation tools, specialized texts, terminology, globalization, Statistical Machine Translation, Neural Networks.***

Ivanka Sakareva –
South-West University
"Neofit Rilski", Blagoevgrad,
Bulgaria

doi.org/10.33919/flcy.24.4.11

MACHINE TRANSLATION AND SPECIALIZED TEXTS: A BRIEF OVERVIEW COMPARING GOOGLE TRANSLATE, DEEPL AND CHATGPT

In an age of globalization and rapid information exchange, machine translation (MT) is crucial for overcoming language barriers. Translating specialized texts, such as technical documentation, scientific papers, and legal documents, is essential for international collaboration and knowledge sharing. These genres pose challenges due to their specific terminology, complex structures, and nuanced contexts. Therefore, machine translation tools need to deal not only with linguistic equivalence but also with context-specific uses of language and structure. For instance, accurate translation in technical documentation is vital to correctly convey detailed technical information. In scientific publications, maintaining the precision and clarity of complex concepts is essential. Legal documents require translation tools to navigate legal jargon and preserve legal meaning across different jurisdictions. Hence, translating specialized texts demands more than just linguistic skills; it requires domain-specific knowledge and an appreciation of the complex relationship between language, structure, and context.

The concept of utilizing machines to assist with translation dates back to the 1940s. Warren Weaver, a pioneer in this area, proposed the use of computers for translation as early as 1949, inspired by the successful application of cryptography during World War II. This represented a major turning point towards investigating computational approaches to address the challenges of language translation. About twenty years later, early machine translation efforts centred on rule-based systems, which generated translations using extensive sets of linguistic rules and dictionaries for each language pair. Despite their initial promise, rule-based machine translation (RBMT) systems struggled with the complexity and variability of human language, resulting in rigid, literal translations that often lacked nuance. The technology underlying most of these early systems was rule-based machine translation, which relies on dictionary-based lexical substitutions and pre-programmed morphosyntactic rules for automated transfer operations. In the following decades, this transfer approach was replaced by more advanced, better-per-

forming, and more user-friendly technologies. (Jolley et al., p. 27)

At the beginning of the 21st century, the advent of powerful computers and the availability of large bilingual text corpora made Statistical Machine Translation (SMT) feasible. This approach analyses vast amounts of text to identify statistical correlations between words and phrases in different languages. SMT significantly improved translation quality by considering context and the probability of word and phrase matches.

There are several machine translation tools available today, including Google Translate, DeepL, and ChatGPT. Each tool uses different algorithms and methodologies, resulting in varying levels of effectiveness, particularly when translating specialized texts. It is crucial to evaluate and compare these tools to understand their respective strengths and limitations. Google Translate, which employs statistical methods and neural networks, is highly accessible and versatile for general use but may struggle with accurately translating domain-specific terminology in technical or legal documents. DeepL, with its strong neural network models, excels in maintaining accuracy, especially in European languages and specialized fields, making it a top choice for precise translations. ChatGPT, known for its versatility and contextual sensitivity, uses transformer-based models to deliver nuanced translations. However, its effectiveness with specialized texts can vary depending on the subject's complexity and specificity.

Evaluating machine translation tools for specialized texts requires a comprehensive assessment across several key dimensions: accuracy, fluency, adequacy, consistency, and domain appropriateness (Han et al., 2016, p. 1). These criteria are essential for determining the effectiveness of tools such as Google Translate, DeepL, and ChatGPT in providing reliable translations that satisfy the strict requirements of technical, scientific, and legal documents.

Mentioned by several scholars, the evolution of machine translation has been heavily influenced by the demand for accurate translations of specialized texts, driving progress in neural network models to handle the complex-

MACHINE TRANSLATION AND SPECIALIZED TEXTS: A BRIEF OVERVIEW COMPARING GOOGLE TRANSLATE, DEEPL AND CHATGPT

ities of technical languages. Such texts refer to technical, scientific, and legal documents that are critical carriers of information and present unique challenges for machine translation. These texts often require not only linguistic proficiency but also an understanding of the specific subject matter. As MT technology progresses, there is a growing focus on ethical AI practices, such as reducing biases in training data and ensuring the privacy and security of translated content. These ethical considerations are crucial in specialized fields, where the precision and fairness of translations can have significant consequences. The translation of specialized texts using machine translation is an area of ongoing research and development, situated at the crossroads of computational linguistics, domain-specific knowledge, and advanced AI technologies. As MT systems grow more advanced, the aim is to produce translations that are not only technically accurate but also contextually nuanced and ethically responsible, thereby enhancing the accessibility and understanding of specialized knowledge across different languages.

Machine translation of specialized texts encounters substantial difficulties primarily due to the complex nature of domain-specific language, which presents unique challenges that general-purpose MT systems often struggle to overcome. Fields encompassing legal, technical and scientific texts are characterized by their use of highly specific terminology and intricate sentence structures, which are not easily managed by standard MT systems based on a personal experience comparing notarial certification and crane description translated using GoogleTranslate, DeepL and ChatGPT. One significant challenge is the accurate handling of specialized jargon. MT systems that have not been specifically trained on domain-relevant terminology may generate translations that are not only incorrect but also potentially misleading, as they fail to capture the precise meaning of technical terms. In addition to terminology, understanding the nuanced context in which these terms are used is crucial for producing accurate translations, but this level of contextual comprehension is frequently beyond the capabilities of general MT models. These models often lack the depth of understand-

ing required to interpret specialized language correctly, leading to errors in the translated text.

The problem is further intensified by the scarcity of high-quality, domain-specific training data. MT systems rely on extensive amounts of relevant text to learn and improve, and the limited availability of comprehensive and carefully selected specialized corpora means that these systems often lack the necessary exposure to accurately translate domain-specific content. Complex sentence structures, which are common in specialized texts, add another layer of difficulty. Even when a translation is grammatically correct, it may still fail to convey the intended meaning due to the inherent complexity of the source material.

Moreover, specialized texts frequently contain cultural, legal, or contextual references that can be challenging for MT systems to interpret accurately (the legal text researched was full of terms and phrases that have no equivalent in the target language, in this case Bulgarian). These references may involve subtle distinctions or specific local knowledge that general models are not equipped to handle effectively (Wang and Sawyer, 2023). As a result, ensuring the quality of translations in these contexts usually requires significant post-editing by human experts who possess both linguistic skills and domain-specific knowledge. These experts are crucial for correcting inaccuracies and ensuring that the translation aligns with the original intent of the text. Machine translation is a challenging task due to the inherent complexity of natural languages. Words often have multiple meanings and potential translations, sentences can be interpreted in various ways, and the relationships between linguistic elements are frequently ambiguous (Okpor, 214, p.165).

Finally, the rapid evolution of specialized fields means that MT systems must be continuously updated to incorporate the latest terminology and concepts. This ongoing need for updates presents a significant challenge, as keeping an MT system current with new developments and emerging trends in specialized areas is both resource-intensive and technically demanding. Thus, main-

taining translation accuracy and relevance in specialized texts is a dynamic and complex task that requires continuous effort and adaptation. After all, translation tools each have distinct advantages and drawbacks. Google Translate is known for its broad language support and ease of use but can be inconsistent and less reliable with specialized texts. DeepL excels in providing accurate and fluent translations, particularly in European languages, making it ideal for professional and domain-specific use. ChatGPT offers contextual translation capabilities, though its performance can vary, suggesting it should be used alongside other tools for specialized content. For precise needs, such as some legal documents, based on personal experience, DeepL is the most reliable due to its ability to maintain accuracy and legal nuances. Google Translate and ChatGPT may require careful review and post-editing to ensure accuracy and correctness, highlighting the need to choose translation tools based on the specific requirements of the task.

References

Han, L. & Wong, D. (2016). Machine Translation Evaluation: A Survey. https://arxiv.org/abs/1605.04515.

Jolley, J. R, & Maimone, L. (2022). Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature. *L2 Journal: An Open Access Refereed Journal for World Language Educators*, 14(1). http://dx.doi.org/10.5070/L214151760 https://escholarship.org/uc/item/2df5d55c

Okpor, M. (2014). Machine Translation Approaches: Issues and Challenges. *IJCSI International Journal of Computer Science Issues*, Vol. 11, Issue 5, No 2, September 2014, pp.159-165. DOI: 10.13140/RG.2.2.12055.38561.

Wang, P., & Sawyer, D.B. (2023). *Machine Learning in Translation* (1st ed.). Routledge. https://doi.org/10.4324/9781003321538