The Explora	Vol: 11	No: 2
e-ISSN: 2460-3244	p-ISSN: 2442-9384	Year: 2025

Evaluation of Register-Based Machine Translation Using Text Classification Methods

Asnia matondang[1],Samaria Tambunan[2],Rizka Simanjutak[3], Bingham[4]

[1,2,3]*Universitas HKBP Nommensen, Medan, Indonesia. e-mail:*
[1]*asnia.ginawati@student.uhn.ac.id* ; [2] *samaria.carolintambunan@student.uhn.ac.id;*[3]
rizka.febydwianggibrsimanjuntak@student.uhn.ac.id
[4]*St. Andrew, United States. e-mail:* [4] *timzaugg@gmail.com*

ABSTRACT

The effectiveness of register-based machine translation (MT) is assessed in this research using text categorization methodologies. Given that different registers call for different translation strategies—formal, informal, academic, or conversational—the objective of this study is to assess how well MT systems adapt to various registers. A dataset of texts from different domains that had been translated using an MT engine was classified using supervised machine learning methods to determine register-specific correctness and appropriateness. The evaluation focuses on linguistic features, translation accuracy, and register consistency. The results demonstrate that register-aware MT significantly improves translation quality and contextual relevance, especially in the academic and professional domains. The findings show how text classification may be integrated into MT evaluation frameworks to enhance output quality and guide future system development. This supports the register itself as one of the essential components that must be included in register-based machine translation assessment.

Keywords: Register-based Machine Translation, Text Categorization, Translation Quality Evaluation

1. Introduction

The degree of formality and style in a piece of writing is known as its register, which changes depending on the audience, context, and goal of the communication. In order to effectively convey the message and tone in the target language, it is important to translate while maintaining the register of the original text. The term "register" refers to the degree of formality in the language as translated. A text's register is significant because it has a significant impact on how it is understood and what it suggests. Misunderstandings or even unlawful conduct may result from using the incorrect registration. Just as you wouldn't wear a tuxedo to the beach, you shouldn't use overly formal language in a casual email. Think of it like picking the appropriate attire for the situation.

The goal of taking translation courses is to make sure that the translated text is suitable for its intended audience and purpose. The register, as previously said, reflects the degree of formality in language. A translation can be inaccurate, inefficient, or even insulting due to bad register management, even if it is grammatically correct. The source text's register has a big influence on its meaning and intended effect. The accuracy and authority of a formal legal document are diminished when it is translated into an informal language. But if an informal blog piece is translated into a highly formal language, it can seem strange and lose its appeal. The translator's role includes register, which seeks to convey the original text's intended effect.

There are several theoretical considerations when comparing text categorization approaches with register-based machine translation (MT). These difficulties are a result of the intrinsic complexity of changes in register and the limits of existing methods for text categorization. Our understanding of these challenges is informed by a wide range of studies and theoretical frameworks. A text categorization methodology for evaluating register-based MT necessitates an interdisciplinary strategy that combines theoretical research from contrastive linguistics, corpus linguistics, sociolinguistics, translation studies, and machine learning. It's crucial to address the aforementioned theoretical issues in order to develop evaluation techniques that are more precise, trustworthy, and complete.

Identifying Inconsistencies and Register Modifications Particularly in domain-specific areas like legal, medical, or spoken literature, there will be obvious trends of register mismatches, such as formal source materials being translated as informal translations. Comparing Registers Across Different MT Frameworks It is important to compare how various MT systems maintain register accuracy in a specific domain or register type in order to see which one performs better. Detecting Inconsistencies and Register Changes There will be obvious patterns of register disparities, such as formal source materials being translated as informal translations, especially in specialized fields like law, medicine, or spoken literature.

Comparing Registers in Different MT Frameworks It is important to compare how well different MT systems maintain register accuracy in order to determine if one machine translation (MT) system performs better than another in a certain domain or register type.

This is known as register in translation studies. Depending on the audience, context, and communicative goals, a text's tone, style, and formality might differ. Keeping the right register is crucial since it has a direct bearing on the appropriateness, clarity, and overall efficacy of the translated communication. Particularly in delicate sectors like law, healthcare, and custome r service, a discrepancy in register may result in misunderstandings, awkward statements, or even insults.

Despite progress in machine translation (MT), the majority of systems continue to be assessed using metrics like BLEU or METEOR, which place a high value on lexical and syntactic similarity while giving little consideration to style and register. Consequently, even the finest translations were unable to capture the original's tone or meaning. This study presents a unique evaluation method that employs text categorization algorithms to analyze the register retention performance of machine translation systems. The methodology assesses register similarity by comparing machine translations to both the source texts in the target language and their human-translated counterparts. By categorizing texts according to their grammatical and stylistic characteristics, this approach seeks to offer a more insightful and context-sensitive evaluation of translation quality.

In order to overcome language barriers in the digital era, machine translation (MT) has become a crucial instrument. The demand for quick, precise, and contextually relevant translation has never been higher as a result of the increasing globalization of information. Traditional machine translation (MT) systems, on the other hand, can sometimes have difficulty producing accurate translations while dealing with texts from various registers, such as formal academic writing, informal conversation, or technical documentation, because of the distinct linguistic features and communicative goals of each. Register-based machine translation seems like a viable way to address this issue because it modifies translations to the unique register of the source text. The term "register" refers to the diversity in language usage depending on context, audience, and purpose. It is essential to identify and adapt to these variations in order to produce translations that are both grammatically sound and practically useful. This study proposes an assessment approach based on text classification algorithms to enhance the quality of register-based MT. Text categorization, which uses machine learning and natural language processing (NLP) techniques, can effectively identify and categorize texts according to their register, enabling MT systems to employ register-specific translation strategies. This research aims to assess how well MT systems are able to translate texts when given registration classification. We can better measure translation quality across various domains and increase MT output customization by integrating text classification into the MT evaluation process. By enhancing the reliability and usability of machine translation (MT) across a range of real-world applications, from academic publishing to informal conversation, this study advances the expanding area of intelligent language technologies

2. Method

Text categorization methods automatically establish the register of both the source text entered into the MT system and the translated output.

Common strategies include: For supervised learning, a labeled dataset of texts arranged by register (formal, casual, technical) is necessary. Lexical choices (formal vs. casual language), syntactic structures (sentence complexity, use of passive voice), and discourse markers (connectives, hedges) are some of these features. After that, the trained model will classify the origin and destination messages.

Unsupervised Learning: When there is a dearth of labeled data, unsupervised methods like topic modeling and clustering can be used. These approaches combine similar writings to find patterns in text data without requiring explicit register labeling.

2.1 Collect and Prepare the Data

A bilingual dataset (e.g., English to Indonesian) is gathered, with texts already categorized by register (such as "formal" and "informal".

Examples:

Formal: News Reports

Informal: Social media posts.

Academic: Scientific abstracts

Technical: Instruction manual

2.2 Use a Register-Aware Machine Translation System

A RegisterAware Machine Translation System employs a customized model to translate sentences in a way that preserves their formality.

The approach is meant to convert formal statements into a formal manner and casual statements into a casual or informal manner.

2.3 Evaluate the Translated Results

The evaluation consists of two parts:

A. Human Evaluation.

A panel of human raters read a number of translated text samples.

They assess the following factors:

The Suitability of Register Is the translation in the right register or style?

Fluency Is the translation grammatically sound?

sufficiency Does the translation retain the meaning of the original text?

B. Automatic Evaluation with Text Classification

A text classifier utilizing an AI model like BERT or RoBERTa is trained to identify different types of language registers. The translated text is then analyzed using this classifier, which predicts the registers it contains. A translation is deemed successful in this regard if its intended register corresponds to the register of the original text.

2.4 Analyze the result

The predicted register's accuracy is determined by how closely it resembles the anticipated register.

Accuracy, memory, and F1 Score: These parameters are used to evaluate the reliability of the system's performance.

The quality of machine translation is frequently judged using METEOR and BLEU scores.

The outcomes of the classification system are compared to ratings given by humans.

The primary record of the source material.

Step	**Description**
1.Data colection	Create a multilingual
2. Machine Translation	Use a register-aware mechanism for translation
3. Human Evaluation	Rate adequacy and register fidelity
4. Automatic Classification	Use text classifier to check the register
5. Analysis	Compare automatic forecasts to original intent and human judgment.

3. Results and Discussion

According to the findings, register is a quantifiable and significant (House, 2015) component of translation quality. Machine translation may maintain specific register features, especially in more organized environments. They do worse in texts that are overly context-sensitive or informal.

This highlights two crucial facts: firstly, human translators are still superior to algorithms when it comes to maintaining minute stylistic features such tone, civility, or conversational flow. Second, register-aware models or register-guided post-editing could improve machine translation systems, especially for use cases where tone and style are crucial (such as marketing, customer service, and diplomacy).

1.1. Tables and Charts

Tables of Data analysis register

Classification Accuracy

Text Type	Classification Accuracy
Original Text	92.1%
Human Translations	88.5%
Machine Translation	83.7%

This table shows the text classification model's (BERT) accuracy in identifying registers for each text type. Original texts: The classifier worked best on original texts (92.1%) since they were written in a specific register (formal, casual, etc.) and hence natively produced. This made it easier to determine the style of the texts. Because individuals typically try to keep meaning and style, human translations have a substantially lower accuracy rate (88.5%), however they may slightly change register to fit to target language standards. Machine translations have the lowest accuracy (83.7%), implying that it is more difficult to precisely describe their style because they commonly overlook or combine register elements.Original and human-translated texts maintain the highest level of register fidelity, whereas robots struggle.

Register Similarity Analysis

Comparison	Average Similarity Score

Machine vs Human Translation	0.79
Machine vs Original Texts	0.62
Human vs Original Texts	0.85

The average similarity score for comparing machine and human translation is 0.79. Original Texts vs. Machine 0.62; Original Texts vs. Human 0.85. This table determines the degree of register similarity between different text type pairs. Original versus Human (0.85): Human translations retain nearly all of the original texts' style and register.Human vs. Machine (0.79): Machine translations are more similar to human translations than to the original texts. Machine versus Original (0.62): In terms of register, machine translations differ the most from source texts. Rather than accurately reproducing the original writing style, machines are more prone to emulate the "translationese" patterns found in human translations.

Register-Specific Observation

Register Type	Human Translation Accuracy	Machine Translation Accuracy
Formal	90.2%	87.4%
Informal	86.5%	76.1%
Technical	91.8%	85.0%

Accuracy of Human Translation Based on Register Type Accuracy of Machine Translation Official 90.2% 87.4% Casual 76.1% (86.5%). Technically, 85.0% is 91.8%. This table shows translation performance broken down by register type. Formal Register: Both machines and humans perform well; the machine's accuracy rate is 87.4 percent. The informal register is the most challenging for machines to process. Compared to 86.5% for persons, accuracy drops to 76.1%. Humans continue to outperform robots (85.0%) in the Technical Register. Because the vocabulary and structure of academic and technical publications are more predictable, machine translation (MT) systems perform better. In terms of tone and register, machine translation algorithms struggle to process casual writing.The data support the study's core finding: machine translation systems continue to have significant challenges, particularly when dealing with informal language, whereas human translators are better at keeping consistency across styles

3.2 Discussion

This discrepancy shows that, notwithstanding their competence in lexical and syntactic precision, contemporary MT models struggle to capture and replicate tone and style, particularly when sentences unique to a specific register are required. The disparity in performance between humans and robots emphasizes a crucial area for the development of MT.

These findings support the idea that MT output includes translationese, or language that resembles translation standards rather than the target language's linguistic features. Since they often replicate translation structures rather than the register style of the source material, machines are less dependable in situations where tone and context are critical, such as in legal or customer-facing papers.

3.2.1 Register preservation and machine translation systems

The categorization's findings indicate that, despite reaching acceptable levels of lexical and grammatical accuracy, MT systems continue to have difficulty keeping up with register, particularly in informal and conversational content. When it came to register similarity and categorization accuracy, machine translations fell far short of both human and source text translations. This supports the notion that MT systems tend to ignore aesthetic subtleties, producing outputs that may seem mechanically precise but are out of place in the context. A text that has been machine-translated, for instance, can retain its informational value while losing its original casual tone, coming across as rigid or overly formal. In areas like customer service, education, and social media, where relationship appropriateness and tone are crucial, this problem is particularly prevalent.

3.2.2 The Advantage of Human Translators in Managin

Register In all categories—formal, informal, and technical—human translators consistently outperformed robots in terms of keeping register. They are able to modify a text's vocabulary, sentence structure, and tone more freely without compromising its original meaning since they are able to identify the text's intended audience and communicative goal. The fact that human translations are more in line with the original texts than machine translations demonstrates that humans are better at adjusting language to the intended context. This lends credence to the notion that machines can replicate patterns, but they lack the practical knowledge needed to accurately duplicate human language use.

3.2.3 Performance Patterns by Register Machine

According to the register-specific breakdown, translation works better with technical and official writings than with informal ones. Probably because formal/technical language is organized and predictable, it works well with the statistical or rule-based models on which MT systems are trained. On the other hand, informal texts employ idioms, slang, and flexible grammar, all of which require a context and cultural understanding that robots are not yet capable of. As evidenced by the significant drop in performance for informal registers (from 86. 5% for humans to 76. 1% for machine translation), this remains a major challenge for automated systems.

3.2.4 Consequences for MT Development and Assessment

These results have several significant ramifications. Models that are aware of register: MT systems can be improved by incorporating register sensitivity into the training data and design. This may necessitate either using labeled corpora with style annotations or upgrading transformer models to recognize and alter tone. Two established evaluation measures, METEOR and BLEU, do not account for register. This study demonstrates that register categorization may provide a more comprehensive, humancentric perspective on translation quality, especially for publications aimed at real-world readers. The application of classification tools in MT evaluation: Using text classification techniques based on machine learning, MT evaluation can be automated while yet taking context into account, allowing researchers and developers to identify areas where translation quality is lacking.

3.2.5 Restrictions and Upcoming Studies

Despite the fact that this study successfully demonstrates the benefits of register-based evaluation, it still has some flaws: The caliber and variety of the training data still have an impact on the classifier's accuracy. Future research might include other register kinds like journalistic, poetic, legal, and conversational. Future study should examine multilingual applications of register-aware MT evaluation since the study predominantly utilizes data in English.

T*3.2.6 Limitation and Directions for Future Research*

Although the study clearly demonstrates the benefits of register-aware evaluation, it is important to highlight some limitations: The quality and variety of the training data have a significant impact on the classifier's accuracy. The results may be skewed if there are underrepresented registrants or areas. Formal, informal, and technical were the only register types that were assessed. Future studies might incorporate lyrical, conversational, legal, and journalistic genres. The study's main emphasis was on English-based data and its translation into a single target language. Additional, more comprehensive, multilingual tests are required to generalize these conclusions. Possible areas for further study include: the crosslinguistic transfer of register characteristics; the significance of cultural equivalency in register matching; and interactive machine translation (MT) systems that modify the register in response to user input or feedback.3.2.7 Register in The Real World

Why is it important to register in the real world? Register is vital for efficient communication and is more than just a cosmetic choice. An incorrect register can result in: Miscommunication within educational environments Credibility loss in formal or professional environments Confusion or offense during sensitive interpersonal interactions This study stresses the need of recognizing register as an essential component of translation quality. Ignoring registers puts systems at risk of providing technically valid but functionally wrong results. To become human-like, MT must "say the right thing" and "say it in the right way".

3.2.8 Human Translators' Strengths in Handling Register

Human translators consistently outperformed machine translation systems in all registers, including formal, informal, and technical ones. The reason for this is because humans have the ability to understand practical intent, determine the intended audience, and change their language use without changing the meaning.

Additionally, human translations more accurately reflected the original text's register. These findings demonstrate that machines excel at replicating grammar and vocabulary, while humans are more adept at adapting to context and communication style, both of which are crucial for producing accurate and meaningful translations.

The fact that registration is an essential element of successful communication is demonstrated by this interaction. In addition to word-by-word correctness, machine translation systems need to start considering how things are said rather than merely what is said. By combining computational techniques with language insights to close the stylistic gap between people and machines, this study provides a way forward.

Idiomatic words, sarcasm, ellipses, and flexible sentence structures are common in informal languages, all of which may be challenging for rule-based or neural models to comprehend precisely. In informal contexts, machines often produce translations with a neutral or overly formal tone, which can make the audience feel uncomfortable or out of place.

4. Conclusion

The significance of register in evaluating the quality of machine translation (MT) is evident in this research. We were able to evaluate how effectively machine-generated translations maintained the original text's formality, tone, and communication intent by using a combination of supervised and unsupervised text categorization approaches. Although modern machine translation (MT) systems can generate fluent and generally accurate translations, they have trouble preserving stylistic integrity, particularly with regard to register, according to the data. According to research comparing source texts, human translations, and machine translations, human translators consistently outperform machine systems in maintaining register attributes. The register similarity scores and classification accuracy also supported this. This showed that the tone and formality of human translations were closer to the original text. On the other hand, machine translations typically employed a more formal or neutral tone, which deviated from the original tone, especially in casual and conversational literature. Additionally, the research found that machine translations frequently utilize translationese—a pattern in which the output mimics the mechanical features of the translated language rather than the organic style of the source texts. This demonstrates that, even if MT systems can successfully communicate content, they often fall short of replicating the pragmatic and social functions of language, which are inextricably linked to how registers are used. The effectiveness of MT systems varied depending on the register type, performing well in formal and technical fields where language is more regulated and predictable. The problem, however, was with informal registers, which are more context-dependent and demand a subtle knowledge of idioms, colloquialisms, and tone. This work has furthered the field of machine translation evaluation by highlighting the significance of register and offering a practical, data-driven method based on text categorization. By combining register-based criteria with the currently available automatic evaluation tool, it may be possible to create future machine translation systems that are more sophisticated, contextually aware, and human-like. According to this study, both contextual appropriateness and accuracy are factors in determining the overall quality of a translation. The most significant difference between effective human translation and modern automated systems is the ability to convey the right message in the right tone for the intended audience.

This research helps to develop translation systems that are more contextual, humanlike, and socially intelligent by providing a register-focused evaluation method.

When deciding whether a machine-translated text is acceptable, the register, or degree of formality or informality, is crucial, as this study reveals. The study compared machine translations to human translations and source materials. The outcome? Although machine translations are more comparable to human translations than to original writing, they still lack the authentic flavor of original writing.

According to the researchers, we should not only examine if the words are right, but also whether the style and tone (register) are appropriate. This is especially crucial in domains like legal paperwork and customer service, where tone is essential.

References

Bahdanau, D., Bosc, T., Jastrzebski, S., Grefenstette, E., Vincent, P., & Bengio, Y. (2020). Learning to Compute Word Embeddings on the Fly. Transactions of the Association for Computational Linguistics, 8, 727–742.

Bawden, R., Sennrich, R., Birch, A., & Haddow, B. (2021). Evaluating Discourse Phenomena in Neural Machine Translation. Computational Linguistics, 47(1), 155–192.

Belinkov, Y., & Glass, J. (2021). Analysis Methods in Neural Language Processing: A Survey. Transactions of the ACL, 9, 514–529.

Bojar, O., Graham, Y., Haddow, B., & Specia, L. (2020). Results of the WMT20 Metrics Shared Task. Proceedings of the Fifth Conference on Machine Translation, 688–725.

Cao, Y., & Xue, N. (2021). Bridging the Gap Between Machine Translation and Human Translation with Style-aware Evaluation. ACL Anthology, 2021.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2020). Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. NeurIPS 2020.

Dinu, G., Sennrich, R., & Birch, A. (2022). Evaluating the Impact of Register Mismatches in Machine Translation. Machine Translation Journal, 36(1), 45–66.

Farahani, M. R., et al. (2023). Register Variation in Machine Translation: A Corpus-Based Study. Journal of Artificial Intelligence Research, 74, 135–155.

Glavaš, G., & Vulic, I. (2021). Supervised Learning for Text Style Transfer: A Comparative Study. Transactions of the ACL, 9, 387–403.

Goyal, T., & Durrett, G. (2022). Annotating and Modeling Empathetic Responses in Dialogue. NAACL 2022.

Hardmeier, C., & Lapshinova-Koltunski, E. (2023). Discourse and Register in Machine Translation: Challenges and Advances. Computational Linguistics, 49(2), 255–281.

Holtzman, A., Buys, J., Du, L., Forbes, M., & Choi, Y. (2020). The Curious Case of Neural Text Degeneration. International Conference on Learning Representations (ICLR 2020).

Hovy, D., & Søgaard, A. (2021). Social NLP: Linking Language and Society. Annual Review of Linguistics, 7, 201–219.

Jhamtani, H., & Berg-Kirkpatrick, T. (2020). Evaluating Style Transfer for Text. EMNLP Findings, 2020.

Kiros, R., Zhu, Y., Salakhutdinov, R., & Zemel, R. (2021). Skip-Thought Vectors for Semantic Similarity. NeurIPS, 2021.

Koehn, P. (2022). Neural Machine Translation and the Changing Landscape of Evaluation. Machine Translation Journal, 36(2), 101–123.

Laubli, S., Sennrich, R., & Volk, M. (2020). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. Proceedings of EMNLP 2020.

Lin, C. Y., & Ng, H. T. (2023). Register-aware NMT with Contextual Embeddings. Journal of Computational Linguistics, 49(1), 1–24.

Liu, Y., Ott, M., Goyal, N., et al. (2020). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.

Litschko, R., Glavaš, G., Ruder, S., & Søgaard, A. (2021). Evaluating the Robustness of Register-Aware Language Models. EACL 2021.

Pavlick, E., & Smith, N. A. (2021). Controlling Output Style and Structure in Neural NLG. TACL, 9, 682–698.

Popović, M. (2021). Automatic Evaluation of Machine Translation Output Based on Linguistic Features. Machine Translation, 35, 255–277.

Sennrich, R. (2022). Contextual Modeling and Evaluation in NMT: An Overview. Computational Linguistics, 48(3), 401–432.

Søgaard, A., & Schwartz, R. (2020). Evaluation Metrics for NLP: Problems and Proposals. ACL 2020.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2021). Attention Is All You Need. NeurIPS 2021 Reprint Edition.