



A Statistical Machine Translation Approach for Banjarese–Indonesian Language Translation

Ali Muhammad¹, Nurul Widyastuti², Novia Winda³

¹Program Studi Teknik Informatika, Universitas Sains Indonesia, Indonesia

²Kementerian Agama Indonesia, Indonesia

³Program Studi Pendidikan Bahasa dan Sastra Indonesia, Universitas PGRI Kalimantan, Indonesia

Correspondence: E-mail: ali.muhammad@lecturer.sains.ac.id

ABSTRACT

The Banjarese language has recently become a particular concern not only among regional language observers in Kalimantan but also at the national and even international levels. This is due to the fact that regional languages are increasingly being abandoned by the younger generation as a result of the hegemony of contemporary languages and other languages that are perceived as more prestigious. This phenomenon can be clearly observed in the significant decline in the number of regional language speakers. The purpose of this study is to preserve the regional language through the development of a Banjarese–Indonesian and Indonesian–Banjarese machine translation system. This research is expected to serve as a reference for other linguistic researchers who face limitations in using the Banjarese language. Automatic machine translation has been widely applied in various fields such as language recognition and vocabulary learning, sentence tree analysis, question-answering systems, text summarization, sentiment analysis, and computational linguistics. This study employs the Statistical Machine Translation (SMT) method to translate each word in a sentence into Banjarese and vice versa. The proposed method is evaluated using the Statistical Calculation, METEOR (Metric for Evaluation of Translation with Explicit Ordering) Universal Tools and human justification. The test results using the statistical calculation show that the word translation success rate using the Statistical Machine Translation method reaches 48%. Meanwhile, the evaluation of the naturalness of translated sentences using the METEOR Universal Tools with 1,000 sentence data shows that the proposed method achieves a final_score of 80. The final evaluation was conducted using human judgment involving 15 language experts. The results indicate that the proposed method is able to translate sentences based on individual words and the highest word probability available without altering words based on synonym lists. The value of naturalness evaluation using human judgment are 72.6%.

ARTICLE INFO

Article History:

Submitted/Received

13 August 2025

First Revised 25 August 2025

Accepted 28 August 2025

First Available online

29 August 2025

Publication Date

01 December 2025

Keyword:

Language Conservation,

Machine Translation,

Statistical Machine Translation.

1. INTRODUCTION

The development of Natural Language Processing (NLP) technology has progressed rapidly over the past decade and has been widely applied in various fields, including sentiment analysis, question-answering systems, machine translation, and multimodal discourse processing (Alvarez-Carmona et al., 2022; Aqlan et al., 2019). NLP has become a key technology in bridging communication across languages through the implementation of machine translation systems, whether based on statistical, neural, or modern transfer learning approaches (Lopez, 2023; Prabowo & Sanjaya, 2024). However, these advancements have not been evenly distributed, particularly for local languages that suffer from limited linguistic data resources, including the Banjarese language.

Banjarese is one of the regional languages widely spoken in South Kalimantan and plays an important role in social interaction, cultural identity, and local traditions. Nevertheless, several studies indicate a shift in the use of Banjarese, especially among the younger generation in urban areas (Hasmianti et al., 2023; Hermawan, 2022). Even in traditional community settings such as the floating market traders, the preservation of Banjarese faces significant challenges due to the dominance of the Indonesian language (Zakiah et al., 2022). This condition highlights the urgent need for language conservation efforts through digital technology approaches (Kamariah & Jamiatul, 2023), one of which is the development of machine translation systems.

A number of studies related to Banjarese language technology have been conducted, such as the development of a PC-PATR parser for Banjarese (Muhammad & Kamariah, 2020), the development of a Banjarese part-of-speech tagger based on the DevOps methodology (Muhammad & Widyastuti, 2024), as well as speech recognition for a Banjarese–Indonesian dictionary using statistical inference (Purnajaya et al., 2021). In addition, the development of machine translation applications has also been carried out using modern approaches such as the No Language Left Behind method (Pranaja & Nurhidayat, 2023). However, most of these studies remain limited in terms of the availability of parallel corpora, standardized evaluation techniques, and optimization of translation performance for low-resource languages.

Statistical Machine Translation (SMT) is one of the classical approaches in machine translation that utilizes probabilistic models based on parallel corpora (Wahyudinata et al., 2016). SMT remains relevant, especially for regional languages with limited data resources, as it is capable of learning statistically from pairs of source and target sentences. The success of SMT is highly dependent on the quality and quantity of the parallel corpora used (Larasati, 2012; Liu & Huang, 2021; Pan & Qin, 2022). Several studies also emphasize the importance of enriching corpora through semantic and multimedia approaches to further improve translation quality (Guinovart, 2019; Oliver & Álvarez, 2024; Rui & Xiuli, 2022).

In the context of translation research, the use of corpora is a vital element in building reliable translation systems (Mohammed, 2022; Shen, 2022). However, to date, the availability of Banjarese–Indonesian parallel corpora remains very limited compared to national and international languages. This limitation is further supported by studies on the development of Banjarese Neural Machine Translation, which identify the scarcity of training data as the primary challenge (Muhammad et al., 2025). Therefore, the SMT approach remains a realistic alternative for developing Banjarese–Indonesian translation systems under limited resource conditions.

In addition to technical aspects, the preservation of the Banjarese language also has strong social and cultural dimensions. The Banjarese language serves not only as a means of communication but also as a representation of the cultural values of the Banjarese

community as reflected in literature, folklore, and social systems (Muttaqin, 2019; Susilawati et al., 2021). Digitalizing the language through the development of machine translation systems is expected to support sustainable language preservation through the utilization of information technology.

Based on the above discussion, it can be concluded that the development of a Banjarese–Indonesian machine translation system based on Statistical Machine Translation is highly important, both in terms of advancing NLP technology for regional languages and in supporting the preservation of local language and culture. This study is expected to contribute to the development of an effective statistical-based translation system and to serve as a foundation for future advancements in artificial intelligence–based Banjarese translation systems.

2. METHODS

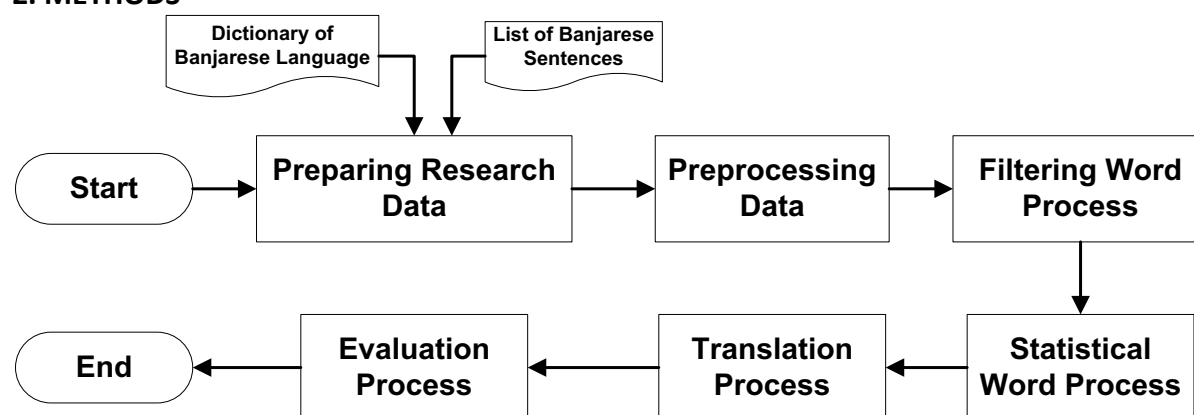


Figure 1. Proposed Methods Process

2.1. Preparing Research Data

The preparation of research data is a crucial stage in the development of the Statistical Machine Translation (SMT) system for Banjarese–Indonesian language translation. The quality of the translation model is highly dependent on the quality, consistency, and representativeness of the parallel corpus used during training. In this study, the dataset was obtained from two main sources: (1) a Banjarese–Indonesian bilingual dictionary and (2) a list of Banjarese and Indonesian sentences crawled from online mass media.

2.2. Preprocessing Data

The first data source, the bilingual dictionary, serves as the primary lexical resource for constructing the basic word-level translation pairs. All word entries were extracted digitally and manually validated to ensure correctness in spelling, meaning, and part-of-speech equivalence between Banjarese and Indonesian. Duplicate and ambiguous entries were carefully filtered to avoid noise in the training process. This dictionary-based data forms the foundation for the word alignment process in the SMT system.

The second data source was obtained through a web crawling process from several online mass media platforms that publish news articles, cultural reports, and local stories using the Banjarese language. The crawling process was conducted using automated scripts to collect raw text data. After collection, the data underwent a cleaning process to remove HTML tags, advertisements, symbols, duplicate sentences, and non-linguistic elements. Only sentences that fully used the Banjarese language and had clear syntactic structure were retained.

After the raw data collection stage, all Banjarese sentences were manually and semi-automatically translated into Indonesian to form a parallel corpus. This translation process was validated by language experts to ensure semantic equivalence and grammatical

correctness. The result of this step is a sentence-aligned Banjarese–Indonesian parallel corpus that serves as the main training and testing dataset for the SMT model.

The next stage is text normalization and preprocessing. This process includes case folding, tokenization, punctuation normalization, and standardization of spelling variations commonly found in Banjarese. Since Banjarese has multiple informal and regional orthographic variations, a normalization dictionary was developed to unify different word forms into a standard representation. This step is essential to reduce data sparsity and improve the performance of the statistical model.

Following normalization, the cleaned parallel corpus was divided into three subsets: training data, validation data, and testing data. The training dataset is used to build the language model and translation model. The validation dataset is used for parameter tuning and performance optimization. Meanwhile, the testing dataset is used exclusively for evaluating the final performance of the SMT system using automatic evaluation metrics such as METEOR and human justification evaluation.

Finally, word alignment and phrase extraction were performed using statistical alignment models. These processes generate bilingual phrase pairs and probability scores that represent the likelihood of translation between Banjarese and Indonesian. These statistical parameters become the core components of the SMT system used in this research.

Through this systematic data preparation process—starting from data acquisition, cleaning, translation, normalization, corpus alignment, to dataset partitioning—the resulting dataset is expected to be reliable and suitable for building a robust Banjarese–Indonesian Statistical Machine Translation system.

2.3. Filtering Word Process

The filtering word process is a procedure used to separate and verify sentences and words so that they purely originate from the Banjarese language, both from data sources in the form of a list of Banjarese sentences and from a Banjarese dictionary. This process filters sentences from the input list of Banjarese sentences. The initial data, which consist of a collection of sentences obtained through web crawling from online mass media, still contain mixed languages, such as Indonesian, foreign loanwords, and code-mixing. Therefore, each sentence is scanned using lexical patterns that refer to distinctive Banjarese vocabulary. Sentences dominated by Banjarese words that match entries in the Banjarese dictionary are retained, while sentences that are more dominant in Indonesian or other languages are eliminated from the dataset.

2.4. Statistical Word Process

The statistical word process aims to calculate the probability of each Banjarese word based on the list of Banjarese sentences used as the training corpus. This process forms the core of the Statistical Machine Translation (SMT) model, as it determines how likely a word appears and how it is translated into the target language based on statistical evidence.

In this process, all Banjarese sentences that have passed the filtering and normalization stages are processed to extract individual words through tokenization. Each word occurrence is then counted to obtain its frequency within the entire corpus. These frequency values represent how often each Banjarese word appears in the dataset and serve as the basis for estimating word probabilities.

After calculating word frequencies, the probability of each Banjarese word is computed using statistical formulations based on relative frequency. The probability value reflects the likelihood of a particular word appearing in a sentence compared to other words in the corpus. This probabilistic representation is crucial for the SMT model, as higher-probability words are considered more reliable during the translation process.

Furthermore, these word probabilities are integrated into the translation model to support the word alignment process between Banjarese and Indonesian. By analyzing co-occurrence patterns between Banjarese words and their Indonesian equivalents in the parallel corpus, the system learns the most probable translation pairs based on statistical evidence rather than predefined linguistic rules.

The results of the statistical word process directly influence the quality of the translation output. Words with higher probability values are prioritized during decoding, allowing the system to select the most likely translation based on the learned statistical patterns. Consequently, this process plays a critical role in improving translation accuracy, reducing ambiguity, and strengthening the overall performance of the Banjarese–Indonesian Statistical Machine Translation system developed in this study.

2.5. Translation Process

The translation process aims to translate a total of 1,000 test sentences from Indonesian into the Banjarese language using the developed Statistical Machine Translation (SMT) model. This stage is conducted to evaluate the system's ability to produce accurate, natural, and linguistically appropriate translations in Banjarese. In the initial stage, 1,000 Indonesian test sentences are selected from data that are not included in the training dataset derived from the list of Banjarese sentences. This is intended to ensure that the testing process genuinely measures the system's generalization capability on new data that have never been encountered during training. The test sentences consist of various sentence forms, including declarative, interrogative, and complex sentences.

Each test sentence is then processed and matched with its corresponding Banjarese word pairs based on the probabilistic model generated during the training phase. The system selects word translations with the highest probability values and rearranges them according to the most statistically plausible Banjarese phrase-order patterns. Furthermore, the system also considers the Banjarese language model to determine the fluency and naturalness of the word arrangement in the translation output. The language model plays an important role in ensuring that the translation results are not only semantically accurate but also conform to the grammatical structures commonly used by native Banjarese speakers. The translation results, consisting of 1,000 Banjarese sentences, are then stored as output data for subsequent evaluation.

2.6. Evaluation Process

The evaluation process is divided into two stages: evaluation using the METEOR Universal Tool and evaluation based on Human Judgment.

2.6.1. METEOR (Metric for Evaluation of Translation with Explicit Ordering) Universal Tools

METEOR (Metric for Evaluation of Translation with Explicit ORdering) is an automatic machine translation evaluation metric proposed by Banerjee and Lavie (2005) to achieve a higher correlation with human judgment compared to traditional metrics such as BLEU. METEOR evaluates translation quality not only based on exact word matches but also by considering deeper linguistic similarities, including stemming, synonym matching, and paraphrase matching. This comprehensive approach enables METEOR to better capture semantic equivalence between the system output and the reference translation.

1. Word Alignment Stage

The first stage of METEOR evaluation is word alignment between the system-generated translation (*hypothesis*) and the human reference translation (*reference*). The alignment process is performed hierarchically using several matching strategies:

1. Exact word matching,
2. Stem-based matching,
3. Synonym-based matching,
4. Paraphrase-based matching.

This multi-level alignment allows METEOR to recognize semantic matches even when the surface word forms differ.

2. Precision and Recall Computation

After word alignment is completed, METEOR computes two fundamental statistical measures: Precision (P) and Recall (R).

The Precision formula is defined as:

$$P = \frac{m}{|H|} \quad (1)$$

where:

- (m) is the number of matched words,
- (|H|) is the total number of words in the hypothesis translation.

The Recall formula is defined as:

$$R = \frac{m}{|R|} \quad (2)$$

where:

- (m) is the number of matched words,
- (|R|) is the total number of words in the reference translation.

Precision measures how many words produced by the system are correct, while Recall measures how well the system captures the content of the reference translation.

3. Harmonic Mean (F-score) Calculation

The Precision and Recall values are then combined using a weighted harmonic mean to produce the F-score. In METEOR, Recall is given higher importance than Precision.

The F-score formula is:

$$F_{mean} = \frac{10PR}{R+9P} \quad (3)$$

This formulation emphasizes translation completeness while still maintaining accuracy.

4. Fragmentation Penalty Calculation

To account for differences in word order between the hypothesis and reference translations, METEOR introduces a fragmentation penalty. This penalty is based on the number of contiguous matched word segments (*chunks*).

The Penalty formula is defined as:

$$Penalty = 0.5 * \frac{ch}{m} \quad (4)$$

where:

- (ch) is the number of matched chunks,
- (m) is the number of matched words.

A higher number of fragmented chunks indicates greater word order inconsistency and results in a higher penalty.

5. Final METEOR Score Computation

The final METEOR score is obtained by integrating the F-score with the fragmentation penalty as follows:

$$Score = Fmean * (1 - Penalty) \quad (5)$$

The METEOR score ranges from 0 to 1. A score closer to 1 indicates better translation quality and higher similarity to the human reference translation.

6. Interpretation of Results

The METEOR score provides a comprehensive evaluation of machine translation quality by simultaneously considering lexical accuracy, semantic equivalence, and sentence structure. Due to its strong correlation with human judgment, METEOR is highly suitable for evaluating the performance of Statistical Machine Translation systems, particularly for low-resource languages such as Banjarese–Indonesian translation.

Through the stages of word alignment, Precision and Recall computation, harmonic mean calculation, fragmentation penalty estimation, and final score computation, METEOR provides an objective and linguistically informed framework for translation evaluation.

2.6.2. Human Judgment

In addition to automatic evaluation using the METEOR Universal Tool, the performance of the proposed Banjarese–Indonesian Statistical Machine Translation (SMT) system is also evaluated through a Human Judgment assessment. This evaluation aims to measure the translation quality based on direct human perceptions of accuracy, fluency, and naturalness, which cannot be fully captured by automatic metrics alone.

The human evaluation process involves 15 language experts and native speakers of Banjarese. These evaluators are asked to assess the translation results of 1,000 testing sentences generated by the SMT system. Each translated sentence is compared with its corresponding reference translation, and the evaluators provide scores based on several criteria, including semantic accuracy (meaning equivalence), grammatical correctness, and naturalness of expression in the Banjarese language.

To ensure objectivity, the evaluators do not have access to the translation method used by the system during the assessment process. Each sentence is rated using a predetermined scoring scale, and the final score is obtained by calculating the average of all evaluators' ratings. This approach minimizes individual bias and ensures a more reliable assessment of translation quality.

The results of the human judgment evaluation are then compared with the METEOR automatic evaluation scores. This combined evaluation strategy provides a comprehensive performance analysis by integrating both quantitative and qualitative perspectives. The consistency between METEOR scores and human judgment results also strengthens the validity of the evaluation process.

Through this human-based evaluation, the proposed SMT system's ability to produce accurate, fluent, and culturally appropriate Banjarese translations can be thoroughly verified. The results serve as strong evidence of the system's effectiveness and practical usability in real-world Banjarese–Indonesian translation applications.

3. RESULTS AND DISCUSSION

This section discusses and analyzes the results of Meteor and human judgment.

3.1. Meteor Results

This study used samples of 10, 50, 100, 500, and 1000 sentences from the list of news sentences crawled from @hariankompas on X Digital Platform (Team, 2020-2025), translated sentences using the statistical machine translation method from the si palui corpus (Team, 2020-2025). The testing results are divided into two, namely the Percentage Results of Translated Sentences and Translation Metric Evaluation Results.

3.1.1. The Percentage Results of Translated Sentences

The percentage of translated sentences was analyzed to determine the maximum effectiveness of the proposed method for translating sentences from Indonesian to Banjar. This evaluation involved 10, 50, 100, 500, and 1,000 sentences taken from the original list of news sentences. The sentences were translated using the statistical machine translation method. The evaluation results using statistical calculation showed in the Figure 2 below.

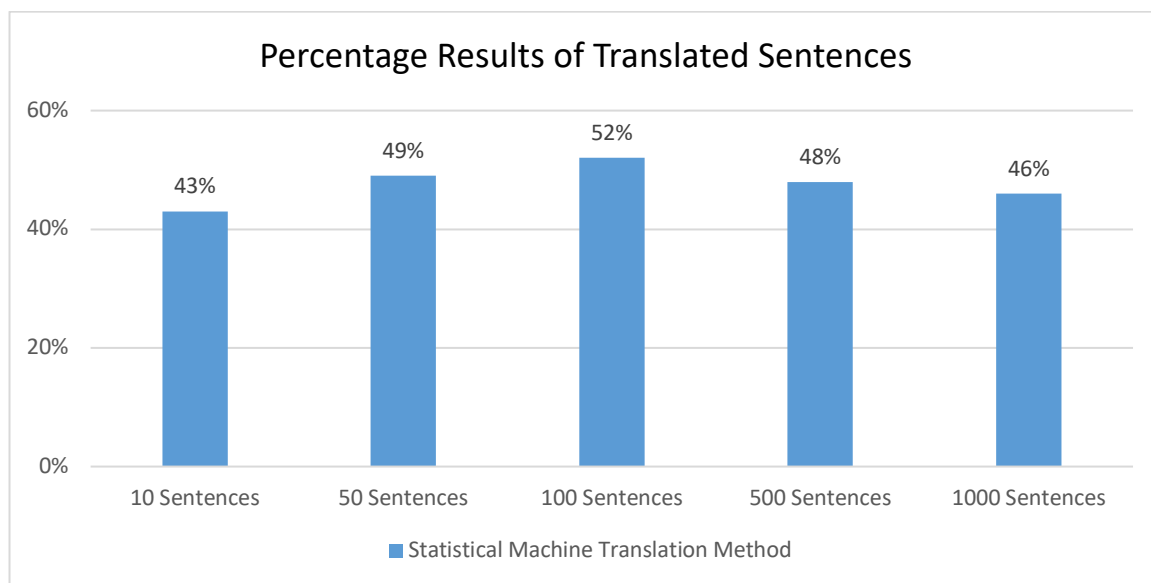


Figure 2. The Percentage Results of Translated Results

Figure 2 above shows the evaluation results of the percentage of translated sentences randomly taken from a list of news sentences. These sentences were then translated using statistical machine translation. The evaluation results show that the percentage of statistical machine translation using the statistical calculation method showed in the figure 2 above. The highest sentence translation percentage was 52% for the 100 testing sentences, while the lowest sentence translation percentage was found for the first 10 testing sentences. This condition is because the translator's dictionary did not find an equivalent word for the new trending word. The average words can be translated using statistical translation method are 48%.

3.1.2. The Naturalness Evaluation Results Using Meteor

The naturalness evaluation results of the statistical machine translation method showed better. The Meteor universal tools results showed the value of final_score are 85. This condition occurs because the statistical machine translation method translated each word from the Indonesian words to banjarese word with considering the highest value of probabilities words. The statistical machine translation method is only able to translate sentences based on the words and the highest probability of the available words without changing words based on the list of synonyms.

3.2. The Naturalness Evaluation Results Using Human Judgment

Figure 3 below shows the results of the naturalness evaluation using human judgment. The human judgment results indicate that the translation result using the statistical machine translation method achieved the highest score of 87% for translated sentence ID 8 and the lowest score for translated sentence ID 9, at 59%. This naturalness occurs because there are homographs in the translation, so the translation is out of context. Finally, the average naturalness evaluation for translated sentences was 72.6% per sentence. This occurs because there are loanwords that cannot be translated into Banjarese.

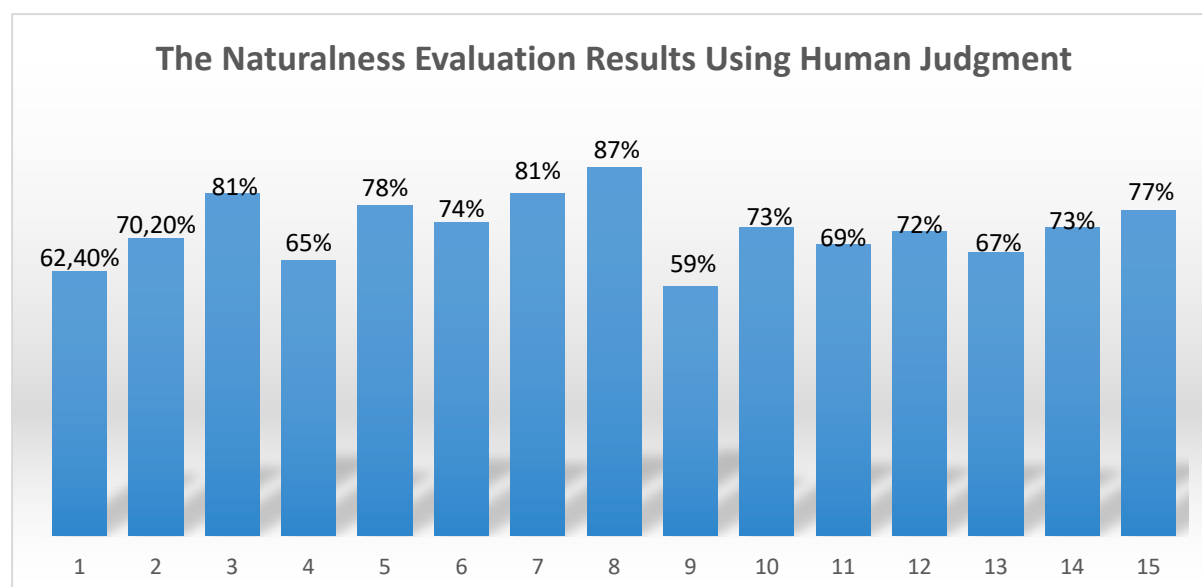


Figure 3. The Evaluation Results using Human Judgment

4. CONCLUSION

Based on the evaluation results using human judgment and the METEOR Universal Tools, the sentence translation method using statistical machine translation produces good quality translations. In several translation cases, the level of naturalness of the translated sentences generated by the statistical machine translation method demonstrates near-perfect results. This is indicated by the fact that 48% of Indonesian sentences can be successfully translated into the Banjar language. In addition, the measurement results using the METEOR Universal Tools on 1,000 testing sentences show that the final score obtained is 80. Finally the results of naturalness evaluation score using human judgment are 72.6%. This result occurs because the statistical machine translation method measures and prioritizes the highest word probability values. However, experts have assessed that this statistical machine translation method does not yet adequately consider sentence context in the translation process.

5. ACKNOWLEDGMENT

The authors would like to express their gratitude to the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia (Kemendikbudristek) for the funding support through the 2025 Beginner Lecturer Research Grant Program (PDP). We would also like to express our gratitude to the academic community of Universitas Sains Indonesia for their institutional support as a research location so that this research can be completed well.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

7. REFERENCES

- Alvarez-Carmona, M., Aranda, R., Rodríguez-Gonzalez, A. Y., Fajardo-Delgado, D., Sánchez, M. G., Pérez-Espinosa, H., et al. (2022). Natural language processing applied to tourism research: A systematic review and future research directions. *Journal of King Saud University - Computer and Information Sciences*, 34(10), 10125–10144.
- Aqlan, A. A. Q., Manjula, B., & Lakshman Naik, R. (2019). A study of sentiment analysis: Concepts, techniques, and challenges. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 28, pp. 147–162). Springer. http://dx.doi.org/10.1007/978-981-13-6459-4_16
- Barmawi, A. M., & Muhammad, A. (2019). Paraphrasing method based on contextual synonym substitution. *Jurnal ICT Research and Applications*, 13(3), 257–282.
- Barmawi, A. M., Wahyudi, B. A., & Pristi, T. (2023). Linguistic based one time password. *International Journal of Electrical Engineering and Informatics*, 15(1), 1–16.
- Fashwan, A., & Alansary, S. (2021). A morphologically annotated corpus and a morphological analyzer for Egyptian Arabic. *Procedia CIRP*, 189, 203–210. <https://doi.org/10.1016/j.procs.2021.05.084>
- Ghirrid, A. A., Sari, R. T. K., & Aldisa, R. T. (2024). Algoritma natural language processing untuk aplikasi penerjemah (Indonesia–Jawa) menggunakan metode speech processing. *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 8(3), 746–759. <https://doi.org/10.35870/jtik.v8i3.2244>
- Guinovart, X. G. (2019). Enriching parallel corpora with multimedia and lexical semantics from the CLUVI corpus to WordNet and SemCor. In John Benjamins Publishing Company (pp. 141–158).
- Hapsari, W. P., Labib, U. A., Haryanto, H., & Safitri, D. W. (2021). A literature review of Human, Organization, Technology (HOT) – Fit evaluation model. In *Proceedings of the 6th International Seminar on Science Education (ISSE 2020)* (pp. 876–883).
- Hasmianti, L., Usman, U., & Amir, J. (2023). Pergeseran penggunaan kata sapaan oleh generasi milenial Banjar di Kota Banjarmasin. *JP-BSI: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 8(2), 122.
- Hermawan, D. (2022). Pergeseran penggunaan bahasa. *Locana*, 5(1), 23–37.
- Kamariah, & Jamiatul, N. K. H. (2023). Konservasi Bahasa Banjar sebagai usaha. *Konfiks*, 10(2), 24. <https://journal.unismuh.ac.id/index.php/konfiksPermalink/DOI:https://doi.org/10.26618/jk/13118>
- Larasati, S. D. (2012). IDENTIC corpus: Morphologically enriched Indonesian-English parallel corpus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 902–906).
- Liu, B., & Huang, L. (2021). ParaMed: A parallel corpus for English–Chinese translation in the biomedical domain. *BMC Medical Informatics and Decision Making*, 21(1).
- Lopez, A. (2023). *Machine translation evaluation metrics benchmarking: From traditional MT to LLMs* (1st ed.). Universitat de Barcelona.
- Mohammed, T. A. S. (2022). The use of corpora in translation into the second language: A project-based approach. *Frontiers in Education*, 7, 1–14.
- Muhammad, A., & Kamariah, K. (2020). Pengurai kalimat Bahasa Banjar dengan menggunakan parser PC-PATR. *Jurnal Linguistik Komputasional*, 3(1), 20.

- Muhammad, A., & Widyastuti, N. (2024). Pengembangan aplikasi part-of-speech tagger Bahasa Banjar menggunakan metode pengembangan DevOps. *JIKOMTI: Jurnal Ilmiah Ilmu Komputer dan Teknologi Informasi*, 1(1).
- Muhammad, A., Winda, N., Firizkiansah, A., Setiawan, D., Dewi, S. H. F., Maulana, I. R., Ardiansyah, M. (2025). Review of Banjarnese Neural Machine Translation Development With Minimal Resources. *Journal of Software Engineering, Information and Communicaton Technology (SEICT)*, Vol 6, No 1.
- Muttaqin, A. I. (2019). Konstruksi verba gerak direksional dalam Bahasa Banjar. *PRASASTI: Journal of Linguistics*, 4(2), 99–103. <https://jurnal.uns.ac.id/pjl/article/view/34129>
- Nur, S., Assyifa, A. N., & Nurjannah, H. (2023). Pengembangan aplikasi penerjemah bahasa isyarat Indonesia (Bisindo) menggunakan metode long-short term memory. *EDUSAINTEK: Jurnal Pendidikan, Sains dan Teknologi*, 11(1), 13–30.
- Oliver, A., & Álvarez, S. (2024). LitPC: A set of tools for building parallel corpora from literary works. In *CTT 2024 – 1st Workshop on Creative Translation Technologies Proceedings* (pp. 21–31).
- Pan, B., & Qin, Q. (2022). Construction of parallel corpus for English translation teaching based on computer aided translation software. *Computer-Aided Design and Applications*, 19(s1), 70–80.
- Prabowo, A., & Indra Sanjaya, F. (2024). Penerapan metode transfer learning pada IndoBERT untuk analisis sentimen teks Bahasa Jawa Ngoko Lugu. *Simkom*, 9(2), 205–217.
- Purnajaya, A., Indriani, F., & Faisal, M. R. (2021). Pengenalan suara pada kamus Banjar–Indonesia dan Indonesia–Banjar menggunakan statistik inferensi. *Jurnal Ilmiah Informatika*, 8(1), 1–8. <https://doi.org/10.33884/jif.v8i01.1727>
- Pranaja, M. A., & Nurhidayat, A. I. (2023). Pembuatan aplikasi mesin penerjemah menggunakan metode *No Language Left Behind* dari bahasa Indonesia ke bahasa Banjar. *Jurnal Manajemen Informatika*, 12(1).
- Rui, L., & Xiuli, G. (2022). Basic research on construction of multimodal parallel corpus of tourism translation in new media era. *Academic Journal of Humanities and Social Sciences*, 5(15), 139–144.
- Satanjeev Banerjee and Alon Lavie. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization* (EvalMT '05), Association for Computational Linguistics, Stroudsburg, PA, USA, 65–72.
- Setiawan, D., Muhammad, A., & Firizkiansah, A. (2024). Pengklasifikasian dokumen teks Bahasa Indonesia berbasis vector space model dengan menggunakan metode k-nearest neighbor (k-NN) dan Euclidean distance. *JIKOMTI: Jurnal Ilmiah Ilmu Komputer dan Teknologi Informasi*, 1(1).
- Shen, N. (2022). English–Chinese corpus collection and translation wisdom algorithm implementation based on Ajax + JQuery. *International Journal of Science and Engineering Applications*, 11(12), 300–302.
- Spatioti, A. G., Kazanidis, I., & Pange, J. (2022). A comparative study of the ADDIE instructional design model in distance education. *Information*, 13(9), 1–20.
- Susilawati, E., Winda, N., & Akbari, S. (2021). Representatif model pemimpin masyarakat Banjar pada cerita rakyat Kisah Datu Wani. In *Prosiding SENSASEDA* (Vol. 1, pp. 55–64). <https://jurnal.stkipbjm.ac.id/index.php/-sensaseda/article/download/1560/786>

- Team, Si Palui., *Si Palui*, Banjarmasin Post. <https://banjarmasin.-tribunnews.com/>, 2020-2025.
- Team, Harian Kompas, X. <https://x.com/hariankompas>. 2020-2025.
- Valentino, M., Thayaparan, M., & Freitas, A. (2021). Unification-based reconstruction of multi-hop explanations for science questions. In *EACL 2021 – Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 200–211).
- Wahyudinata, T., Sujaini, H., & Nyoto, R. D. (2016). Implementasi mesin penerjemah statistik berbasis Android dengan Moses Decoder. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 3(1).
- Winda, N., & Muhammad, A. (2023). Pengembangan parsing PCPATR sebagai preservasi Bahasa dan Sastra Banjar. *Bahasa dan Sastra*, 9. <https://e-journal.my.id/onoma>
- Zakiah, Sumarlam, & Supana. (2022). Faktor pemertahanan Bahasa Banjar pada pedagang pasar terapung di Kota Banjarmasin. In *Proceedings Seminar Nasional Linguistik dan Sastra* (Vol. 4, pp. 1–9). <https://jurnal.uns.ac.id/-prosidingsemantiks>
- Zheng, X., & Wu, H. (2022). Autoregressive linguistic steganography based on BERT and consistency coding. *Security and Communication Networks*, 2022, 1–8.