

Methodology

1. Language Pair Selection

Language pairs were selected based on the objectives of the study. Examples included:

- Fijian ↔ English
 - English ↔ Fijian
-

2. Sample Data Preparation

A diverse set of sentences was curated for translation. The dataset included:

- Short **and long** sentences
 - **Idiomatic expressions**
 - **Domain-specific terminology (e.g., legal, medical)**
-

3. Translation Execution

Each selected sentence was translated using various machine translation systems.

3.1 Large Language Models (LLMs)

LLMs were prompted using the format:

"Translate the following sentence to [target language]: [sentence]"

3.2 Neural Machine Translation Systems (NMTs)

Translation was performed using APIs and open-source toolkits such as Google Translate. Preprocessing and tokenization steps were standardized across systems where applicable.

4. Automatic Evaluation

Machine-generated outputs were compared to human reference translations using standard automatic evaluation metrics:

- BLEU
- CHRF++
- TER
- COMET

These metrics provided quantitative assessments of translation accuracy and fluency.

5. Human Evaluation

Where feasible, bilingual speakers assessed the translations. Each output was rated based on:

- **Fluency** (grammatical correctness and naturalness) on a 1–5 scale
- **Adequacy** (faithfulness to the source meaning) on a 1–5 scale
- **Cohesion/Discourse** (for long texts), evaluated through qualitative feedback or an extended scale

Multiple evaluators were used to ensure consistency and reduce subjective bias.

6. Performance Quantification

For each model or system, the following performance indicators were calculated:

- Average BLEU, CHRF++, and/or COMET scores
 - Mean human evaluation scores (if applicable)
 - Translation speed (tokens/sec or words/sec), recorded under consistent hardware and environmental conditions
-

7. Reporting

Results were compiled into tables and **visualizations** to highlight:

- Comparative performance across systems
 - Specific strengths and weaknesses
 - Observations in domain-specific or low-resource contexts
 - Recommendations for model tuning, deployment, or future research
-