

Evaluating iTaukei Translations: Key Findings

****Methodology****

- Translated 32 sentences (English ↔ iTaukei)
- Systems tested: GPT-5, Gemini, Google, Microsoft
- Evaluated using:
 - Automatic metrics (BLEU, chrF++, TER, etc.)
 - Human ratings (Adequacy, Fluency, Cohesion)

****Results****

- Gemini performed best overall
- Microsoft Translator weakest
- Adequacy > Fluency & Cohesion
- Simple sentences translated better than complex
- Automatic metrics only partly matched human ratings
 - chrF++ & Levenshtein most reliable
 - BLEU less consistent, TER least reliable

****Key Takeaways****

- Current systems convey meaning but sound unnatural
 - Complex iTaukei sentences remain challenging
 - Human evaluation is essential
 - Need fine-tuning on local/community data
- LLMs (like Gemini) show promise for future MT

Summary of the Paper

Title: Evaluating iTaukei Language Translations Generated by Leading LLMs and NMT Systems

Purpose:

This study focuses on how well modern translation systems, both large language models (LLMs) like GPT-5 and Gemini, and neural machine translation (NMT) systems like Google Translate and Microsoft Translator, translate between English and iTaukei (Fijian). iTaukei is considered a low-resource language, meaning that there are fewer digital resources available to train translation systems.

The goal was to find out:

1. How accurate and natural these systems are when translating iTaukei.
2. Which evaluation methods best capture translation quality.
3. What challenges remain for translating complex iTaukei structures.

Methodology

1. Data: Accumulation of parallel text (English ↔ iTaukei) from religious, educational, and community sources.
2. Systems Tested:
 - LLMs: GPT-5 and Gemini 1.5 Pro.
 - NMT Systems: Google Translate and Microsoft Translator.
3. Translation Tasks:
 - 32 sentences covering different sentence types (short, long, idiomatic, and domain-specific).
 - Both English → iTaukei and iTaukei → English directions.
4. Evaluation:
 - Automatic metrics: BLEU, chrF++, TER, Levenshtein ratio, and COMET (these compare machine output against human reference translations).
 - Human evaluation: Three bilingual reviewers rated each translation for adequacy (meaning preserved), fluency (how natural it sounds), and cohesion (flow and connection).
 - Statistics: Several statistical tests were used to check if differences between systems were real and meaningful.

Results

1. Overall Quality:
 - Translations generally preserved meaning (adequacy) better than they preserved fluency or cohesion.
 - In other words, the systems often conveyed the main message but sounded unnatural or awkward.
2. System Comparisons:
 - Gemini 1.5 Pro performed best in human ratings (especially fluency and cohesion).
 - Microsoft Translator performed the weakest.
 - GPT-5 and Google Translate were in between.

3. Automatic Metrics vs Human Ratings:

- Levenshtein ratio (edit similarity) and chrF++ correlated most strongly with human judgments.
- BLEU, the most commonly used metric, was less reliable.
- TER often gave misleading results (high error scores for sentences that humans still considered adequate).

4. Sentence Complexity:

- Simple sentences were translated much better than compound or complex sentences.
- For example, adequacy scores averaged 4.2/5 for simple sentences but dropped to 3.2/5 for complex ones.
- This shows that long and syntactically complicated iTaukei sentences are especially difficult for current systems.

5. Translation Direction:

- No major statistical differences were found between English → iTaukei and iTaukei → English, though Fijian-to-English performed slightly better overall.

Key Takeaways

- Current translation systems can convey meaning reasonably well but struggle with fluency, cohesion, and complex sentences.
- LLMs (like Gemini) generally perform better than traditional systems.
- Automatic metrics don't fully capture translation quality—human evaluation remains essential.
- For iTaukei MT, improvements require fine-tuning on community data and methods that handle complex sentence structures better.