Translation Quality Analysis and Error Categorization

General Statistics

- Total sentences evaluated: 155
- Score 3 (correct translation): 66
- Score 2 (minor mistake): 36
- Score 1 (incorrect translation): 53
- Percentage of correct translations (score = 3): ~42.6%
- Percentage of minor mistakes (score = 2): ~23.2%
- Percentage of incorrect translation (score = 1): ~34.2%

Error Categories

Category	Description	Examples	Linguistic Explanation
Lexical errors	Wrong word choice	*орехи → orchids*, *свинья → pigs*	Failure to resolve lexical ambiguity
Grammatical errors	Incorrect grammar or structure	*Is she working?* instead of *Is it working?*	Errors in syntactic parsing or morphology
Semantic distortions	Critical loss or change of meaning	*six degrees of separation* instead of *six university degrees*	Pragmatic misinterpretation
Omissions / truncation	Missing parts of a sentence	*with raisins* without *and walnuts*	Misalignment or premature truncation
Idiomatic & cultural issues	Misrepresentation of culturally specific expressions	*You can't crush it with one finger*	Lack of idiom and cultural adaptation. Still the meaning of the correct translation is not the same as in Russian.
Named entity errors	Distortion or hallucination of names	*Vladimir the Great* instead of *Vladimir Petrovich Serbsky*	Named entity recognition (NER) failures

Linguistic Explanation of Mistakes

- **Lexical ambiguity** in Russian often leads to incorrect word selection.
- **Rich morphology** (cases, gender, aspect) makes parsing more error-prone.
- **Free word order** in Russian complicates syntactic analysis.

- **Cultural concepts and idioms** are not easily transferred or understood by general-purpose models.
- **Domain-specific vocabulary** (culinary, personal narratives) is underrepresented in pretrained corpora.

Model Limitations

- **Low translation stability**: Even basic sentences are often mistranslated.
- **Inconsistency**: Similar structures receive different translations.
- **Frequent failure on short phrases**: Simple replies like “Hello”, “OK”, “Thank you” are mistranslated or missing.
- **Hallucinations**: Extra or unrelated content is sometimes added.
- **Lack of context awareness**: Sentences are translated in isolation, without discourse-level understanding.

Overall Conclusion

The evaluation shows that the machine translation model used in this project demonstrates limited accuracy and consistency when translating from Russian to English. Only 66 sentences out of 156 were translated correctly. The most common error types included **lexical mistranslations**, **omissions**, **grammatical issues**, and **named entity distortions**.

Several linguistic and technical factors contributed to these problems:

- **Russian is a morphologically rich and syntactically flexible language**, which makes automatic translation particularly difficult. Word order is more fluid than in English, and meaning is often conveyed through word endings and inflection. This increases ambiguity and demands more from translation models.
- The source material consisted of **conversational, informal language**, which is more variable and less structured than formal or technical text. Informal speech often includes ellipses, idioms, and cultural references that are difficult to capture with standard models.
- The translations were done **sentence by sentence**, without contextual information. This limited the model’s ability to preserve discourse flow and resolve pronouns or referents.

Model Training and Comparison

As part of this project, we trained **two separate machine translation models** using different architectures and hyperparameters: **mBART** and **MarianMT**.

- Both models were trained using a large dataset consisting of approximately **300,000 parallel sentence pairs** from the **OPUS corpus**, specifically the **TED Talks domain**, which includes clean and high-quality aligned translations.
- After evaluation, **mBART outperformed MarianMT** in both sentence fluency and translation accuracy. However, even the best-performing model still failed to produce high-quality translations for a large portion of the data.

This is primarily because **conversational speech is significantly more complex for machine translation** compared to formal or business language. Spoken language often includes **incomplete constructions, slang, idiomatic expressions, emotional nuance, and culturally specific elements**, making it less predictable and much more variable. Unlike formal language, where sentence structure is clearer and more standardized, **conversational style requires the model to grasp deeper context, tone, and speaker intention**. As a result, even with a well-prepared training corpus, translation quality tends to be lower when the model is faced with such unstable and informal utterances.