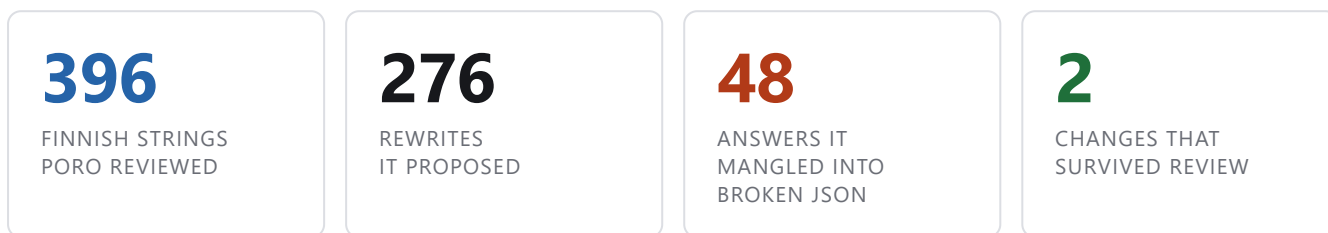


Poro vs. the original

I handed all 396 translated strings on this site to a local Finnish language model and asked it to do better than the machine translation I started with. This is what it found — and what that says about the translation it was grading.

Mikko Numminen · 21 July 2026 · mikkonumminen.dev



1 The setup

The Finnish you read on this site did not start out in Finnish. I write the source copy in English, in `en.ts`, and mirror it into Finnish and Swedish. The Finnish mirror was produced with an AI translator — competent, and it passes my own reading, but I had never sat a native-level Finnish speaker down to go over it line by line. I can't hire one for a portfolio site. So I did the next best thing I had the hardware for: I pointed **Poro** at it.

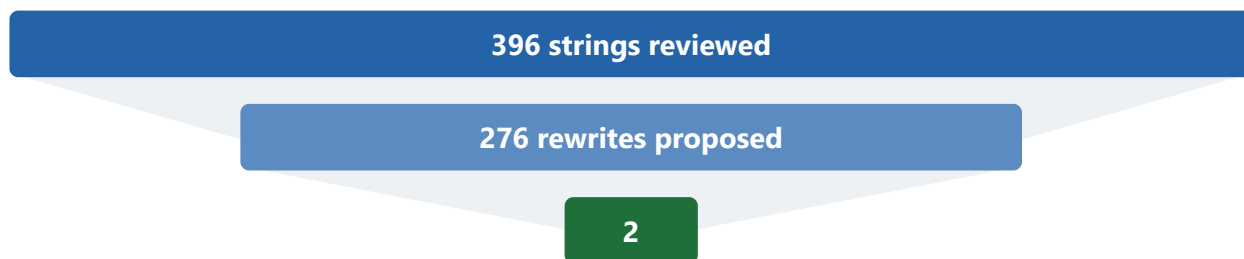
Poro-2-8B is a Finnish-first language model — the same one that answers questions in the chat on my contact page, running locally on my own GPU. If any model was going to catch a stiff, translated-sounding sentence in my copy, it was the one purpose-built for the language. The question I actually wanted answered was not "can Poro translate?" but "how good was the translation I already shipped?"



2 The method

I pulled every translated string out of `fi.ts` and paired it with its English source — 396 pairs, once I dropped the lists of technology names (Rust, TypeScript) that are identical in every language and nothing to translate. Poro read each pair and returned a verdict: leave it alone, or here is a better Finnish sentence, with a reason.

An eight-billion-parameter model is an eager editor. To stop its enthusiasm from becoming a bad pull request, I did not trust its verdicts directly. Every change it proposed went through a second, **adversarial** review — two dozen independent passes plus a skeptic whose only job was to reject anything that broke a `{placeholder}`, mistranslated a product name, drifted from the English meaning, or simply wasn't clearly better. Only the survivors reached the branch.



2 verified as genuine improvements — a 99.3% rejection rate on Poro's own suggestions

3 What Poro wanted to change

Poro proposed rewriting **276 of the 396 strings** — seventy percent of the copy on the site. Read literally, that looks like a damning verdict on the original translation. It isn't. It's what a small, eager model does: shown a sentence and asked whether it could be phrased differently, it nearly always says yes. It approved 59 strings outright, and on 61 more it never produced a usable answer — 48 came back as malformed JSON and 13 flagged a problem but offered no actual replacement.



■ 59 approved as-is ■ 276 it wanted to rewrite ■ 61 no usable answer

4 What actually held up

After the adversarial pass, exactly **two** of those 276 changes survived. Two.

`projectsPage.keyExternalDesc` **ACCURACY**

EN — *orbiting satellite — connects to an outside service*

BEFORE **kiertävä satelliitti – yhteys ulkomailmaan**

AFTER **kiertoradalla oleva satelliitti – yhteys ulkoiseen palveluun**

The original said "connection to the outside *world*" and lost the word *service*. Poro's restores it, and "satellite in orbit" is the more precise phrasing.

`contactPage.noscriptIntro` **GRAMMAR**

EN — *This page is an interactive terminal that requires JavaScript.*

BEFORE **...terminaali joka vaatii JavaScriptin.**

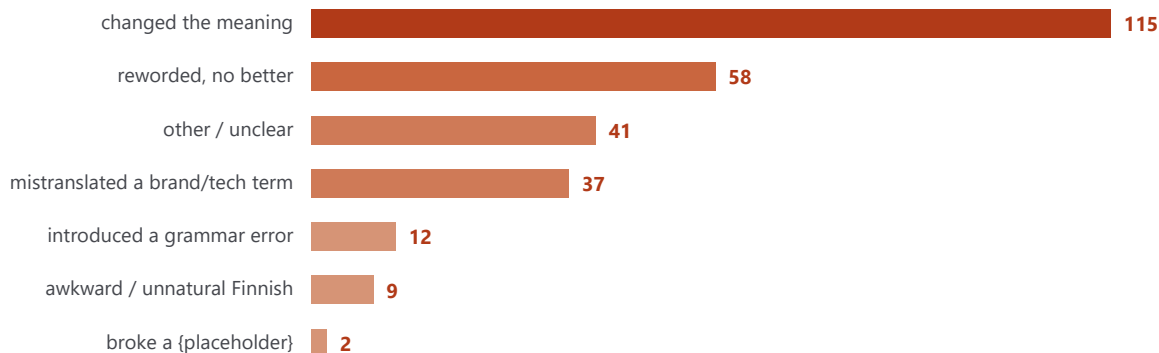
AFTER **...terminaali, joka toimii vain JavaScriptin kanssa.**

Finnish requires a comma before a relative *joka* clause; the original was missing it. A real grammar fix.

Neither is dramatic. One restores a noun the translator dropped; the other adds a comma the grammar requires. But both are real — a careful human reviewer would have made exactly these two edits — and that is the whole point. Across an entire site, a native-Finnish model given free rein to change anything found only two things worth fixing.

5 Where Poro went wrong

The 274 rejected suggestions are more interesting than the 2 that passed, because together they map out precisely how an 8B model fails at translation. I grouped the reviewers' stated reasons for rejecting each one:



Approximate grouping of 274 rejections by the reviewers' stated reason. The largest bucket, by far, is meaning drift — Poro rewording a sentence and quietly losing part of it.

The pattern is consistent, and it is the mirror image of what the original translator got right:

ENGLISH	ORIGINAL (KEPT)	PORO WANTED	WHY IT WAS REJECTED
Full-Stack Developer	Fullstack-kehittäjä	Täysipinokehityksen kehittäjä	The literal calque the rules forbid — a term of art that stays in its conventional form.
Subagents (AI sub-agents)	Subagentit	Alisopimukset	"Alisopimukset" is <i>subcontracts</i> , the legal sense. Straight meaning drift.
Project list	Projektiluettelo	Hankelista	"Hanke" is the wrong register for a developer portfolio; "projekti" is conventional.
...the seams between them are the point	...saumat niiden välillä ovat se juju	(clause dropped entirely)	Rewriting the long intro paragraph, Poro silently deleted the closing clause.

Poro reaches for the literal Finnish calque exactly where the original correctly kept an English term of art. It swaps a word for a synonym from the wrong domain. And when it rewrites a long sentence, it loses a clause on the way through. Every one of these is a mistake the original translation did not make.

6 So which translator won?

The one I started with. A purpose-built Finnish model, allowed to change anything on the site, **improved it in two small places and made it worse everywhere else it tried.**

The machine translation I'd been quietly unsure about turned out to be solid. That is the honest headline of this exercise, and it's a more reassuring one than I expected going in. Poro earned its keep not as a rewriter but as a *second reader*: it confirmed the overwhelming bulk of the work and surfaced two genuine misses a human

proofreader would also have caught. That is exactly the job you want a reviewer to do — and it's why the two edits it found are worth shipping even though they're small.

7 What this doesn't claim

Poro answered in malformed JSON 48 times out of 396 — a plain reminder that small local models are unreliable at structured output. I treated every unreadable answer as "no change" rather than guessing at what it meant, so the tooling never invented an edit. I ran each string once, at a low temperature; a second pass would phrase its suggestions differently, though I doubt it would find a third genuine fix where the first pass found two. And this is one model's read of one site's strings on one afternoon — not a benchmark of Finnish machine translation in general. What it is, is an honest before-and-after on my own work. The before held up better than the reviewer I brought in to grade it.

Method: 396 translated strings extracted from `src/i18n/locales/fi.ts` and paired with their English source; each reviewed by Llama-Poro-2-8B-Instruct (Q4_K_M) running locally via Ollama at temperature 0.2; every proposed change independently verified by a 24-agent adversarial pass before acceptance. Two changes shipped in PR #389. Generated 21 July 2026.