

Translating by Meaning, Not by Words: A Two-Stage Method for Native-Quality Machine Translation

Frits Lyneborg

FRITS AI ApS

ABSTRACT

Machine translation, including translation by strong LLMs, produces text that is grammatically correct and unmistakably foreign: calqued phrases no native speaker would write, invented compounds, English sentence structure wearing local vocabulary. English "long-press" becomes Danish "Langtryk", a word that does not exist. Users in smaller-language markets read this correctly as "the product was not made for me." The cause is not model quality. It is the instruction "translate", which tells the model to preserve the source's structure — so the calques are obedience, not failure. We give a two-stage method built on that: stage one reduces the English to plain meaning; stage two hands that meaning to a native-speaker persona who authors the copy without ever seeing the original wording. Two anti-rules remove the most reliable machine tells. We describe the plural and placeholder guards that make this safe for software strings across 26 languages and 47,000 strings, and the negative result that keeps our language list honest: for one low-resource language, no available model writes usably, and the right decision was to refuse to ship it.

1. Correct translations that read as foreign

Machine translation gives you output where nothing is *mistranslated* and none of it sounds like a native speaker [1 (#ref-1)]. The tells are consistent across engines and

target languages [2 (#ref-2)].

- **Calques** — grammatical, alien, word-for-word transfers. English “free welcome credits” becomes Danish “*Gratis velkomstkreditter*”, a phrase no Dane has written. A Dane would say “*Gratis credits*”, or reach for a different idea entirely: “*velkomstgave*”, welcome gift.
- **Invented compounds** — the surest machine tell. “Long-press” becomes “*Langtryk*”, a word that does not exist. Danish usage is either the English loanword or a plain description: “*hold fingeren nede*”.
- **Mirrored structure** — the English skeleton, clause order and rhythm, wearing local vocabulary. Each sentence defensible. The whole unmistakably translated.

We run a consumer product in 26 European languages — the 24 official EU languages plus Norwegian and Icelandic, with Ukrainian added — at roughly 1,950 interface strings each. That is about 47,000 strings under continuous change, so hand-polishing is not available to us.

2. Translation is the wrong task to ask for

Given a source text and told to translate it, a model preserves the source’s structure. Structural fidelity is what separates translation from paraphrase in its training data. The calques and the mirrored syntax are not failures of capability. They are obedience [4 (#ref-4)].

So we never ask for translation. The task is split so that no stage ever holds both the English phrasing and responsibility for the target phrasing.

Stage 1 reduces the source to meaning. A first pass rewrites the English into plain descriptive English, stripping metaphor, marketing energy and jargon while keeping the intent. “Your phone now sings — a brand new chime plays when your answer is ready” becomes “A short sound now plays when the answer is ready.” Temperature 0.2: a controlled, almost mechanical rewrite.

Stage 2 authors that meaning natively. The plain text goes to a pass framed as a person rather than a function — a native speaker of the target language, raised in the country, working as a senior copywriter, who receives a description of meaning and writes original copy for it. The prompt says so outright: the English is a description of meaning, not a template to mirror. The model may split, merge and reorder freely. Temperature 0.7, deliberately high, because the most probable continuation of a translation-shaped context is precisely the calque [3 (#ref-3)]. Escaping it takes room.

The split matters more than either half. Paraphrase-then-translate still calques, because stage 2 still hears “translate”. A single-stage “write natively” on raw marketing English imports the source’s rhetoric instead. Meaning extraction and native authorship have to be separate jobs.

3. The two anti-rules

Two rules in the stage-2 instruction remove the most reliable tells. They are the transferable core of the method.

The calque test. If the most obvious target-language equivalent of an English word is a calque or a foreign-sounding loan, write what a native would actually write instead — even where that changes the image. English “tactile” tempts Danish “*taktil*”, a word that exists and that nobody uses. The native choice is “*mærkbar*”, noticeable. The rule forces the question “what would be written here?” in place of “what does this word map to?”

Never invent words. Where an English UI or technical concept has no established native equivalent, the model must not manufacture a compound from native roots. Two moves are permitted: the English loanword, where that is genuinely how natives refer to the thing, or a short plain description of the action. The trap exists in every language, so the rule is stated generically and enforced per locale.

Each batch also carries register context — a live-chat widget is “warm, plain”; a security page is “precise, factual, lightly reassuring, a trust centre rather than marketing” — plus a consistency constraint. Where several strings share a concept, the

model picks one native root and holds it, rather than mixing synonyms across a screen.

4. Making it safe for software strings

Interface strings carry structure that must survive an aggressive re-authoring stage untouched. Three mechanisms buy that safety.

ICU plurals are translated inside-out. Strings like `{count, plural, one {# source} other {# sources}}` keep their skeleton verbatim; only the words inside each branch are authored. The model is told to *add* the plural categories the target language needs rather than mirror English's two, so Polish correctly gains `few` and `many`. A malformed plural fails the batch rather than the user.

Brand phrases bypass the model. Recurring brand-critical phrases are pre-authored natively once per locale and substituted deterministically. The model sees a placeholder token; post-processing swaps in the approved phrase. It cannot degrade what it never touches.

Leaks are ship-blockers. Models intermittently emit the placeholder token itself into output. We learned this the usual way, with a token leaking into six locales in one batch, and made detection structural: a scanner runs over every locale file after every batch, matching token patterns that cannot occur in legitimate copy while ignoring real acronyms like GDPR. Any hit fails the run. A leak can still happen. It can no longer ship.

The surrounding pipeline is deliberately boring — chunks of about 20 strings to stay inside output-token limits, failed chunks simply not written, and a resumable runner that fills exactly the gaps the last run left. One catch-up run filled 1,868 missing strings without touching an existing translation.

5. Verifying without native staff

A one-company team cannot employ 26 native reviewers, so review is delegated to the model in a different role. A reviewer persona — native speaker, senior copywriter, shown only the target text and a description of intent — rates key phrases from 1 (“obvious machine translation”) to 5 (“indistinguishable from native copy”), and must justify anything below 5 with a concrete objection and a fix [6 (#ref-6)]. Below 4 goes to a human.

This is weaker than native human review and we use it accordingly, as a regression detector rather than a certificate. It reliably catches the difference between bad and plausible. The difference between plausible and perfect still needs someone who grew up with the language.

6. Knowing when not to ship a language

The method has a boundary, and finding it is a result in its own right.

Evaluating Greenlandic (Kalaallisut), we found no available model writes it usably [5 (#ref-5)] — including the strongest European ones. Output mislabels basic interface concepts and would read as parody. No pipeline design fixes this, because re-authoring amplifies fluency, and there is no fluency here to amplify.

So we did not ship it. Greenlandic-locale users get Danish, which effectively all Greenlandic readers also read, rather than broken Kalaallisut. We record it because the pressure in localisation always runs toward claiming more languages, and a language list is only worth anything if every entry passed the same bar. Icelandic, our lowest-resource shipped language, passed — with a standing flag for extra native review.

7. Limitations

The pipeline runs on a single model family (Mistral Large; smaller models of the same family calque noticeably more), and every temperature, chunk size and artefact pattern is calibrated to it. Register decisions genuinely contested within a language, such as formal versus informal address in German, are delegated to the persona and per-batch context rather than encoded as rules, trading auditability for naturalness.

The reviewer loop shares a model family with the authoring loop and therefore shares blind spots. And the method is validated on string-sized units — interface copy, notifications, short marketing text — not on long-form documents, where discourse-level cohesion adds constraints this pipeline does not model.

The engine runs as an internal tool at FRITS AI and is not exposed as a product. If you implement the method, especially for languages we have not covered, we would like to hear how it behaves: [contact \(/contact/\)](/contact/).

8. Conclusion

Ask a model to translate and it will carry the source's structure across, calques and all. Ask it to reduce the source to plain meaning, then to author that meaning as someone who never sees the original wording, and the tells disappear. Everything else here — the anti-rules, the plural and placeholder guards, the scanner, the reviewer persona — is the engineering that makes the idea safe to run at scale.

The same discipline sets the limit. Where no model can write a language, no method makes it usable, and the honest move is not to ship it.

References

1. Gellerstam, M. — *Translationese in Swedish Novels Translated from English*. In L. Wollin & H. Lindquist (eds.), *Translation Studies in Scandinavia (SSOTT II)*, pp. 88–95. Lund: CWK Gleerup, 1986.
2. Volansky, V., Ordan, N., Wintner, S. — *On the Features of Translationese*. *Digital Scholarship in the Humanities* 30(1):98–118, 2015. doi:10.1093/llc/fqt031.
3. Raunak, V., Menezes, A., Post, M., Hassan, H. — *Do GPTs Produce Less Literal Translations?* *Proceedings of ACL 2023 (Short Papers)*, pp. 1041–1050. arXiv:2305.16806.
4. Hendy, A., et al. — *How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation*. arXiv:2302.09210, 2023.

5. Ranathunga, S., Lee, E.-S. A., Prifti Skenduli, M., Shekhar, R., Alam, M., Kaur, R. — *Neural Machine Translation for Low-Resource Languages: A Survey*. ACM Computing Surveys 55(11):1–37, 2023. doi:10.1145/3567592.
6. Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., Macherey, W. — *Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation*. Transactions of the ACL 9:1460–1474, 2021. doi:10.1162/tac1_a_00437.

How to cite

Frits Lyneborg (2026). Translating by Meaning, Not by Words: A Two-Stage Method for Native-Quality Machine Translation. FRITS AI ApS, Technical Report FRITS-TR-2026-02.
<https://frits.ai/research/meaning-first-translation/> doi:10.5281/zenodo.21274519.

```
@techreport{lyneborg2026translating,  
  title      = {Translating by Meaning, Not by Words: A Two-Stage Method for Native-Quality Ma  
  author     = {Lyneborg, Frits},  
  institution = {FRITS AI ApS},  
  number     = {FRITS-TR-2026-02},  
  year       = {2026},  
  month      = {jul},  
  doi        = {10.5281/zenodo.21274519},  
  url        = {https://frits.ai/research/meaning-first-translation/}  
}
```