Episode Book

EPISODE #6111

Translation Models That Know Drama From Contracts

PUBLISHED	18 October 2026 at 21:40 UTC
DURATION	21:10
LISTEN	myweirdprompts.com/episode/translation-models-register-control-36485/
HOSTS	Corn, Herman, Larry, Hilbert
SCRIPT MODEL	anthropic/claude-opus-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.10
BOOK GENERATED	19 October 2026 at 17:43 UTC

SYNOPSIS

This week the show goes out in Hebrew, Russian and French, all of it translated by a system prompt rather than a translation engine — which raised a fair question: is anyone actually selling a model built for translation, one you can point at a register and get back a script instead of a wall of text? The answer is yes, and it comes in three layers. There are open specialist models like TranslateGemma, Unbabel's Tower Plus, ByteDance's Seed X and Tencent's Hunyuan MT, which took first place in 30 of 31 language pairs at WMT — until human judges ranked it well below a general model. There are commercial engines like DeepL and Translated's Lara that turned fidelity versus fluency into a dropdown. And there's the audio layer, where dubbing tools come closest to translating for the ear. Along the way: why literary translation was the easiest material in the competition and transcribed speech the hardest, and why Hebrew and Arabic keep falling off the formality menu.

TRANSCRIPT



Corn

A joke that works in English and dies in Russian is still a dead joke, and as of this week we have a few of those to answer for. This show now goes out in Hebrew, Russian and French, with Arabic close behind, and none of it went through a translation engine. Daniel wrote a system prompt that tells a model to translate for the ear, for podcast audio, and let it rip. So his question is a fair one. Is anyone actually selling that? Not Google Translate with a fresh coat of paint, but a model built for translation, where English goes in and the Spanish that comes out has been reasoned about. Maybe one you can point at a register, drama say, and maybe one that hands you back a script instead of a wall of text.



Herman

And what language pairs these things cover, which matters more than it sounds, because the answer for French and the answer for Hebrew are wildly different. Herman Poppleberry, by the way, and yes, the products exist. There are more of them than I expected, and they come in three layers.



Corn

You've ranked them already.



Herman

I ranked them last night. The bottom layer is open models you can download that were trained specifically to translate. The middle layer is the commercial engines that quietly swapped their insides for language models and then added knobs. The top layer is the stuff built for audio, dubbing, which is the closest thing to translating for the ear as a product.



Corn

Start at the bottom, then. What does a translation-specific model have that a general one doesn't, if the general one can already translate?



Herman

Mostly that it's small and still good. Google released a family called TranslateGemma in January, built on their open Gemma model, covering fifty-five languages. The interesting result is that their twelve billion parameter translation version beats the general Gemma model that's more than twice its size. You give up a bit of everything else and you get a model that fits on a laptop and translates like a much bigger one.



Corn

A specialist who's worse at small talk. Every translator I've ever met.



Herman

Then there's Unbabel, the Portuguese company, with a family called Tower. The current one, Tower Plus, came out last year in three sizes, and the whole point of the paper was keeping the model able to chat and follow instructions while making it a translation specialist. They even built a benchmark for that, where the test is whether the model obeys a glossary or a date format while it translates.



Corn

Which is the thing a classic engine can't do.



Herman

It can't be told anything. That's the whole gap. A neural machine translation engine is a function. Sentence in, sentence out. You can't say, these two speakers are brothers, keep the teasing.



Corn

Hm. And the one you're dying to tell me about?



Herman

Tencent. Hunyuan MT, a seven billion parameter model they open-sourced in September of last year, and Tencent announced it took first place in thirty of thirty-one language pairs at WMT, which is the big annual translation competition.



Corn

Thirty of thirty-one. What happened in the other one, they got bored?



Herman

Well... that's the part I like. When the competition organisers wrote up their findings, they said the automatic scores were biased, and when humans actually judged the output, Tencent's system was considerably lower than the top ones. The system humans put on top overall was Gemini two point five Pro. A general model.



Corn

So the specialist won the exam and lost the job interview.



Herman

Partly because some of these systems get trained against the same metrics that grade them. You teach to the test, you ace the test.



Corn

We did a whole episode on that, the one about how the field's own scoring rubrics rate machine output above human translators until you put two versions side by side and ask people which they'd rather read. Same disease, different patient.



Herman

Same disease. And here's the line from that write-up I think about most. Human translations were in the winning group for only six of the fifteen language pairs.



Corn

Wait. The humans lost?



Herman

Lost, or tied with machines, depending on how you read the clusters. The English to Russian human reference landed in the third group. But the same report has a Russian test suite built from university translation exercises, and it says even the best systems still make beginner translator mistakes. Literal word order, verb forms that are wrong, stiff phrasing.



Corn

Fluent and wrong in the same sentence. The worst student in the class with the best handwriting.



Herman

Corn, which do you think was the hardest material in the whole competition? Guess.



Corn

Literature. Poetry, novels, the stuff with feelings in it.



Herman

Literary was the easiest. Speech was the hardest. Transcribed spoken language, probably because the speech recognition errors come along for the ride.



Corn

Which is the exact material we feed it. A transcript of two animals talking over each other.



Herman

Yeah. That one surprised me.



Corn

Let me guess the language part before you tell me. The open models cover French and Russian beautifully and Hebrew not at all.



Herman

You're half right, and the half depends on the model. Tower Plus covers French and Russian and has no Hebrew and no Arabic. ByteDance has one called Seed X, twenty-eight languages, Arabic yes, Hebrew no. The Tencent one actually covers all four of ours, Hebrew, Arabic, Russian and French. And there's an academic one, X ALMA, fifty languages, Hebrew and Arabic included, with little language-specific add-on modules you plug in.



Corn

So if you're translating into Hebrew, the open-source specialist is from Tencent.



Herman

With a catch. Their licence excludes the European Union, the United Kingdom and South Korea.



Corn

A translation model that doesn't do Europe. The one continent famous for needing it.



Herman

It's a licensing thing, not a capability thing, but yes, you read the licence before you build on it. For Hebrew specifically, the strongest local work is Dicta, the Israeli lab. Their language models handle Hebrew and English, and they're the people whose vowel marking we already use, but they're general models, not translation engines.



Herman

So here's one for you. If you had to redo our Hebrew edition tomorrow, would you hand it to the Tencent specialist or keep the big general model with Daniel's prompt?



Corn

Keep the general one. Hm... no, that's too quick. For a contract I'd want the specialist, because I want the same sentence back every time. For us I want the thing that understands that Herman gets carried away and Corn is the one who notices, and a seven billion parameter model trained on news and manuals has never met either of us.



Herman

That's my instinct too. Specialists win on consistency and cost, generalists win on the stuff nobody wrote a benchmark for. Teasing doesn't have a benchmark.



Corn

Give it a year.



Corn

Okay, middle layer. The commercial engines. This is where Daniel's register question lives, because a product either has a knob or it doesn't.



Herman

DeepL is the obvious one. In twenty twenty-four they replaced the guts with their own language model, built just for translation, and they've been adding controls since. There's a formality setting, more formal or less formal. There are style rules you write in plain language, and it'll even draft the rules for you from sample texts you upload. And there's a custom instructions field in the interface for developers, where you just tell it things.



Corn

Formality for which languages?



Herman

German, French, Spanish, Italian, Japanese, Russian, a handful of others. Not Hebrew, not Arabic.



Corn

Of course not. The two languages where formality would be the interesting fight.



Herman

But the cleanest answer to Daniel's question is a product called Lara, from a company called Translated. Every request, you pick one of three styles. Faithful, which keeps the structure, for legal and technical text. Fluid, which is faithful but reads naturally. And Creative, which their own description says prioritises impact and emotion over the original wording, for marketing and storytelling.



Corn

Fidelity versus fluency, finally available as a dropdown.



Herman

Two thousand years of translation theory, three radio buttons.

L

Larry

Friends, are you tired of fidelity? Are you weary of fluency? Fidelity versus fluency, finally available as a dropdown. That's what Gary told me, folks. Gary, a satisfied customer, a real man, wearing shoes. Gary was translating his grandmother's lasagne recipe into the creative register, and the lasagne got so emotional it left him. That's when Gary discovered Register Rinse, the only mouthwash with three settings. Faithful, for breath that means exactly what it says. Fluid, for breath that flows. And Creative, for breath that prioritises impact and emotion over the original wording. Today Brenda, sorry, Gary,

L

Larry (continued)

Brenda is a new woman. She gargles in French now. She doesn't speak French. Register Rinse. Now with idioms. Do not swallow the idioms.



Corn

That was my line. Word for word.



Herman

You said it about a minute ago. I don't love that he was listening. Anyway... the big clouds have all done some version of the Lara move, just less elegantly. Microsoft's Azure translator now lets you pick, per request, a classic engine or a GPT model, and if you pick the GPT one you can set the tone, formal, informal or neutral, and the gender, and give it up to five example translations to copy.



Corn

Set the gender of a translation. That's a sentence that would've confused everyone ten years ago.



Herman

It matters enormously in Hebrew, actually. Every verb is gendered, so if you're addressing a woman, the machine has to know. Google's version is called Adaptive Translation, where you hand it example translations and it imitates the style. Amazon's is a formal and informal switch, and here's the detail I love, it silently ignores the setting for languages that don't support it.



Corn

The most honest product decision in the whole layer. It just doesn't tell you.



Herman

And Alibaba sells one called Qwen MT, ninety-two languages, with what they call domain prompting, where you describe the field, legal, marketing, and it adjusts the style. Which is a register knob by another name.



Corn

Here's what I don't get, though. Every one of these is a menu. Formal, informal, faithful, creative. Daniel asked about drama. Nobody's selling a drama button.



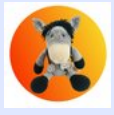
Herman

Nobody's selling a drama button. The closest is the free-text field, DeepL's custom instructions, Lara's styles, where you're back to writing a prompt. Which means the... no. Which means the frontier product, for anything past a few registers, is still a system prompt.



Corn

Daniel's method, with a company logo on the front.



Herman

More or less. What do you think the menu is actually for, though? Why would DeepL ship three settings instead of a text box?



Corn

Because a menu can be tested and a text box can't. If you sell to a legal department, you need to promise that formal means formal every single time. My text box says make the hosts sound like brothers who've shared a flat too long, and nobody in compliance can sign off on that.



Herman

That's probably it. Predictability is the product.



Corn

Top layer. The audio people.



Herman

This is the part that touches what we do. ElevenLabs has dubbing in ninety-plus languages, and the dubbing editor they built, with a timeline and a track per speaker and an editable transcript and translation, is now in maintenance mode.



Corn

The one product that looked like a script editor, and they parked it.



Herman

You can still export subtitles and a separate audio track per speaker, which is the nearest thing I found to a structured script coming out the other end. YouTube has auto dubbing open to every creator in twenty-seven languages, with an expressive mode in eight that tries to carry tone. Papercup, the British dubbing startup, had its technology bought by the language services giant RWS last year, so it lives on inside a bigger company. And DeepL launched voice to voice this year, real-time spoken translation, and in September added models that keep tone and emotion.



Corn

What about a script, though? Speaker labels, turns, timings. The thing Daniel actually means when he says structured output.



Herman

I couldn't find a product that documents a timed, speaker-labelled dubbing script as something it hands you. Not as a first-class output. The research is closer than the products. There's a demo paper from a big language conference last year where the



Herman (continued)

model predicts how many sounds the translated line can have, from how long the original line runs, and keeps rewriting until it fits. They got the dubbed speech lining up with the original far better without losing quality.



Corn

That's lip-sync from the text side. Translate for the stopwatch.



Herman

Isochrony is the term. It matters because Spanish and French tend to run longer than English, so a faithful line overruns the speaker's mouth.



Corn

So when we translate the show, does length matter for us? Nobody's watching Herman's lips.



Herman

My lips are very watchable.



Corn

Nobody's watching them.



Herman

For audio-only it matters less, but the flip side matters a lot. Our pipeline sends a model batches of turns as a list, twenty-four at a time, and checks it gets the same number back. Every number gets written out as words, because there's no number normaliser outside English. For Russian, the model puts a stress mark on every word longer than one syllable, because the voice engine needs it to stress correctly. For Arabic it writes all the vowel marks. A translation engine can't be told any of that.



Corn

Which is the real answer to Daniel. The specialised models exist, the register knobs exist, but the moment you need the output shaped for a machine downstream, you're writing the instructions yourself.



Herman

Which you can, now. That's the change.



Corn

There's one thing in our own prompt I'd want a product to copy. Carry jokes over as jokes, drop English-only wordplay rather than explain it.



Herman

That's the idiom problem in one line. There's a study on Chinese idioms where even the best system got them wrong more than a quarter of the time. And the automatic metrics barely agreed with what humans thought.



Corn

A metric that can't tell when a joke died.



Hilbert

Register. You've both said register about eleven times and neither of you means the same thing.



Herman

We mean the level of language, Hilbert. Formal, casual.



Hilbert

A register is a book. You write things in it. My cousin Ambrose keeps one. He lives in a lighthouse off the Portuguese coast that he rents from the navy for eleven euros a month, and he translates my letters to them.



Corn

Why are you writing to the Portuguese navy?



Hilbert

About the light. It comes on at the wrong time. Four minutes early, every night, since August. Ambrose doesn't do formal or casual. He does three registers, and he decides which one based on the weather. Calm sea, my letter goes out as a legal notice. Choppy, it goes as a sea shanty. Fog, he writes it as a recipe. The navy has answered every one of them in the same register, so I now have a shanty from a rear admiral called Teodoro Pires, forty-one verses, about the lamp timer, and a recipe for sardines that, if you read it carefully, accepts liability.



Herman

That's a reasonably strong legal position, if the recipe is signed.



Hilbert

It's signed. Ambrose charges by the word. That's the part I wanted to talk about, not your dropdowns. Eighty cents a word, Portuguese, which is about what it was when I last paid anyone by the word. So every letter comes back longer than I sent it. Ninety words go out from me, three hundred and twelve come back. The shanties are the worst. A chorus repeats.



Corn

He's billing you for the chorus.



Hilbert

Every chorus. Last week Ambrose decided the navy's shanty deserved a reply in the same register, so he wrote one on my behalf without asking, forty-two verses, one more than the admiral, as a matter of rank. I was billed for all forty-two. The navy has now sent a choir. Not a letter about a choir. The choir. Nine men in a launch, anchored off the lighthouse, singing the reply to the reply, and Ambrose is translating them live from the gallery and charging me by the word for that as well. He says live is a separate register. He's added a surcharge for wind.



Corn

At some point the cheapest option is to fix the timer yourself.



Hilbert

The timer belongs to the navy. Touching it is an act of war. Ambrose looked it up. He charged me for looking it up.



Hilbert

I seal the envelopes myself, one lick, which with my tongue covers the flap and most of the stamp. He sends them back resealed with a little drawing of a lighthouse on the back. Every time. Different weather in the drawing. He didn't have to do that. I had a telescope, so I could check the light from the hill behind my sister's place. The navy took it in September. Teodoro Pires said, in a verse, that it was interfering with shipping.



Herman

A telescope interfering with shipping is technically very hard.



Hilbert

He didn't explain it. It was in verse. The lamp is still four minutes early.



Corn

Here's the thing about paying by the word, though. You've given him a reason to make it longer and nothing that tells him how long it should be. Everyone who dubs anything has your problem the other way round. They've got a fixed amount of mouth time and a translation that wants to run long, so the research we were just talking about puts the length limit inside the translation. The model knows how long the line is allowed to be before it starts.



Herman

You'd pay Ambrose per letter and give him a word ceiling. The same budget the dubbing model gets.



Hilbert

Ninety words, then. He'll use longer ones. Portuguese has the room.



Corn

So where does that leave Daniel? Dedicated translation models are real. Google, Unbabel, Tencent, ByteDance all ship them, and the coverage for French and Russian is generous and for Hebrew and Arabic it's patchy.



Herman

And the commercial engines give you register as a short menu, three styles at best, which is enough for a contract and not enough for a comedy.



Corn

The thing worth keeping is where the control lives. The menus exist because a company can promise what a menu does. Everything past the menu, drama, two brothers teasing, a joke that has to land in Russian, is still written in plain language by somebody who knows the show.



Herman

The specialised models make that cheaper and more reliable. They don't make it unnecessary.



Corn

What I'd watch for is the Ambrose fix showing up as a feature. A translation model that takes the length budget along with the register. Here's the tone, here's how many seconds you've got. Somebody will build that, and it'll be the first product that actually translates for the ear instead of for the page.



Herman

And until then, if you want to hear how the system prompt version came out, the Hebrew, Russian and French editions are on the site under languages, with the hosts' cloned voices, and Arabic is coming.



Corn

Thanks as always to our producer Hilbert Flumingtop for keeping all of this running.



Herman

This has been My Weird Prompts. If you want the translated editions, they're at my weird prompts dot com.



Corn

See you tomorrow.