Episode Book

EPISODE #6112

The Interpreter That Doesn't Know Who's Talking

PUBLISHED	18 October 2026 at 21:43 UTC
DURATION	20:11
LISTEN	myweirdprompts.com/episode/ai-interpreter-power-dynamics-01790/
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

Google's Gemini 3.5 Live Translate handles 70+ languages in real time, OpenAI has a realtime translation model, and Apple, Samsung, and Meta all run interpreters on-device. But every one of them is sentence-in, sentence-out — none can be briefed on who's talking, what they want, or who holds the power in the conversation. This episode digs into what human interpreters actually carry in their heads (coordination, cultural mediation, the right to render nothing at all), why machine translation has turned "I" into "we" in asylum cases and gotten car-search consent thrown out of court, and the one research paper that writes down exactly what's missing.

TRANSCRIPT



Corn

If the interpreter between you and a police officer knows exactly who has the power in that conversation, whose side is it on?



Herman

Hm.



Corn

That's where Daniel's prompt ends up, though he gets there from the friendlier direction. He wants to know whether anyone has built an interpreter that's really an agent. Two people, two languages, real-time voice to voice, but the thing in the middle understands who they are, what the power dynamics are, what each of them is trying to get out of the conversation. Google Translate's conversation mode gives you sentence in, sentence out. He thinks a context-aware interpreter could do something much better, ideally without leaning on a human at all, and he wants to know whether anybody has actually shipped one.



Herman

The voice to voice part is basically solved. The context part is almost completely missing, and the gap between those two is the whole story. Herman Poppleberry here, and I've been looking forward to this one all morning.



Corn

You've been looking forward to everything since breakfast.



Herman

Breakfast was good. Look at the baseline, because it moved a lot this year. In June Google shipped Gemini three point five Live Translate. Seventy-plus languages, it detects which language is being spoken, and it translates continuously instead of waiting for you to finish the sentence. It keeps your intonation and pacing in the translated voice, and it's in the Translate app on any phone.



Corn

Sentence by sentence is already out of date, then.



Herman

The sentence part is. Here's the line I'd frame. In Google's developer docs, the translation mode is described as pure low-latency translation, no support for tools or instructions.



Corn

No instructions.



Herman

You cannot brief it. You can't tell it this is a doctor and a frightened patient. You literally cannot tell it who's talking. OpenAI's version, a realtime translation model they released in May, is the same shape. Seventy-odd languages in, thirteen out, and the setting you get is the output language. Their own guide recommends giving each speaker a separate audio track.



Corn

Otherwise it doesn't know there are two people.



Herman

It knows there's audio. Google's docs admit the cloned voices can get stuck on one voice when several people talk fast. So you get a beautifully natural voice that occasionally turns your wife into you.



Corn

Most marriages take decades to get there.



Herman

Then there's the offline side, which Daniel will care about. Apple's live translation in AirPods runs on the device, eleven languages, no Hebrew. Samsung's Interpreter works offline once you download language packs. Pixel has an on-device call translation that keeps your voice. Meta's glasses do it offline with downloaded packs, a handful of European languages. Every one of them is sentence by sentence and none of them takes a briefing. Samsung's own page says it's not the same as professional simultaneous interpretation.



Corn

Offline matters more than it sounds for this, though. The conversations where you'd want a briefed interpreter are exactly the ones you don't want going to a server.



Herman

Would you trust it more, Corn? Same quality, but it lives on the phone and nothing leaves.



Corn

More, yes. Not because I think the cloud is listening, but because offline means the hospital can't quietly change what it does between Tuesday and Wednesday. The thing in my pocket is the thing I tested.



Herman

That's a better argument for offline than privacy, actually. Microsoft went the other way, the meeting room. Teams has an interpreter agent now, nine languages, with an option to simulate your own voice, and it's built into conference rooms. DeepL turned on voice to voice in September, forty-plus languages, keeping your voice. There's a startup called Palabra claiming under a second, end to end. And the strangest company in the space is Sanas, which started out converting accents in call centres, the same English, but sounding like it came from somewhere else, and added actual translation last year.



Corn

A company that started by translating English into English.



Herman

Into slightly different English. With a few hundred milliseconds of delay.



Corn

The most honest sentence in consumer electronics. Hm. Let me guess why nobody's added the briefing. Latency.



Herman

Partly. What would you say the briefing would even contain, though? If you were writing it.



Corn

Who these two are to each other. What each of them wants. Who's allowed to interrupt whom. Whether "you" should be polite or familiar. And probably what not to say, the stuff that'll blow up the room.



Herman

That's a description of what a human interpreter carries in their head without anyone writing it down. The research on dialogue interpreters goes back to a Swedish scholar, Cecilia Wadensjö, in the late nineties, and her point was that interpreters don't just translate, they coordinate. They expand a line, shrink it, sometimes render nothing at all because the moment didn't need it.



Corn

Wait. Rendering nothing counts?



Herman

It's a recognised move. There's a phrase from another scholar I love, that interpreters cannot not participate. They're in the conversation whether they want to be or not. The healthcare interpreting standards in California spell out four roles in escalating order. Message converter, which is the default, speaking in the first person, I have a pain here. Then message clarifier. Then cultural clarifier, when something won't land because of a



Herman (continued)

cultural gap. And last, advocate, which you're only supposed to reach for with a clear reason.



Corn

And when they step out of the first person, they say so.



Herman

They switch to the third person, the interpreter would like to explain something, they explain it to both sides, and they step back in. That stepping out is exactly the thing none of the products can do.



Corn

The first person thing is stranger than it sounds. The interpreter says "I" and means somebody else.



Herman

It's where machines have tripped badly. In some American asylum cases, machine translation turned "I" into "we", which implied a joint application, and that's a legal problem. In Kansas, a police officer used Google Translate to ask a driver if he could search the car, and it came out as something closer to can I find the car. The court threw out the consent.



Corn

The car was right there. He'd found it.



Herman

There's a hospital study from a few years ago where discharge instructions went through Google Translate, and "hold the kidney medicine" came out meaning keep taking it.



Corn

That's the one that would keep me up.

L

Larry

Are you suffering from context deficiency? Does your interpreter not know who's talking? You literally cannot tell it who's talking. Ask your doctor about Contextra, the once-daily briefing tablet. One Contextra and your conversations will understand their own power dynamics. Side effects may include mild honorifics, addressing your mother-in-law as Your Excellency, sudden fluency in your landlord's objectives, the sensation that a waiter outranks you, which he does, interrupting yourself in the third person, permanent awareness of who is paying, and in rare cases your inner voice translating your thoughts into Portuguese before you've had them. Do not take Contextra if you are a conduit. Contextra. Know who you're talking to.



Corn

You literally cannot tell it who's talking. That's what you said, Herman.



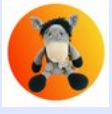
Herman

I said it a minute ago, in that order. I'm choosing not to think about it. Anyway… so has anyone built the agent version? Closest real thing is from ByteDance, a system they called CLASI, published in twenty twenty-four. It's built as an agent copying how human simultaneous interpreters work. It decides when to listen and when to speak, it looks up domain terms as it goes, and it uses the conversation history.



Corn

History, not who's in the room.



Herman

Terminology and history. Not roles, not goals. On their own measure of how much of the meaning survives, it kept around four fifths of it, against something like a third to two fifths for the commercial systems. Their own metric, so a pinch of salt.



Corn

A pinch of salt the size of Herman.



Herman

I'll take that. The paper I'd hand Daniel is by Claudio Fantinuoli, from this June, called Bridging the Usability Gap. His argument is that machine interpreting scores well on benchmarks and feels far worse than talking through a person. He writes that current systems lack any principled mechanism for cultural mediation, that they process the acoustic signal in isolation, and that they have no repertoire of recovery strategies.



Corn

No way to say, sorry, what?



Herman

No way to say, sorry, what. He wants three things. Agency, meaning the system takes initiative and repairs mistakes. Grounding, meaning it knows the situation. And attention to the experience of the two people. Which is Daniel's prompt, written up as a research agenda.



Corn

A proposal, then. Not a product.



Herman

It's a proposal. The nearest experiment is a Czech group at Charles University who recorded real conversations between people with no shared language, interpreted by machine, and asked Gemini to flag the moments of misunderstanding. It caught about three quarters of them.



Corn

That sounds good.



Herman

And more than half its flags were false alarms.



Corn

A smoke detector that also goes off for toast. In a hospital, I'd still keep it.



Herman

That's a fair... yeah, I'd keep it too. There's text-only work on Japanese honorifics, choosing the polite form from the relationship between the speakers, and similar work on Korean. I found nothing that tests a live speech interpreter on honorifics at all. Nobody has measured whether the voice in your ear knows to call your boss's father sir.



Corn

The answer to Daniel is no, then. Nobody's shipped it.



Herman

Nobody's shipped it. As far as I can tell, not even a prototype that takes roles and goals as input and talks.



Corn

Then let me do the bit where I'm unhelpful. Briefing an interpreter before every conversation is impractical. Nobody writes a dossier to buy a bus ticket.



Herman

You don't need one for a bus ticket. The bus ticket is fine with what exists. The briefing matters in the conversations where the stakes are high and the power is lopsided. A doctor and a patient. A landlord and a tenant. A border officer.



Corn

My question from the start, then. Who writes the briefing?



Herman

Hm.



Corn

In every one of those, one person owns the device. The hospital's tablet. The officer's phone. The landlord's app. Whoever writes the system prompt decides what the interpreter thinks the conversation is for. If the briefing says this officer is conducting a lawful search, help them obtain consent, then congratulations, you've built a very fluent, very polite way to get consent.



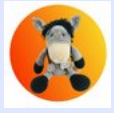
Herman

That's... yeah. Neutrality turns into a config file, and one side holds the file.



Corn

And the human interpreter at least has a code of ethics that's theirs, not the hospital's.



Herman

Though the human codes fight about this too. The medical interpreting codes put impartiality first, and one national council needed a fifty-five page document a few years ago just to explain what their advocacy rule did and didn't mean. Meanwhile the people being interpreted for tend to see the interpreter as their advocate. Fantinuoli also points out that good interpreters quietly soften hostile remarks, and he warns against systems sanitising or intensifying what was said.



Corn

The human version already smooths things, then, and we've just never had to write down how much.



Herman

The machine forces you to write it down. Which might be the best thing about it. Do you think the briefing should be visible to both people?



Corn

Yes. If there's a briefing, both sides hear it, in their own language, before anyone says a word. Here's who I think you two are, here's what I think you're each here for, correct me. That's what a decent human interpreter does in the corridor anyway.



Herman

I like that a lot. It also solves the asking-questions problem, because the setup is a conversation. The court interpreters' association made the same objection to a bill in Wisconsin that would have allowed AI instead of court interpreters. One line from their conference stuck with me, that AI cannot be held accountable and does not ask clarifying questions.



Corn

Here's where I'd split from Daniel a little. He wants the interpreter to understand the power dynamics. Understanding them is fine. Acting on them is the part I'd want switched off by default.



Herman

Switched off how? If it knows the patient doesn't understand the word for a biopsy, and it says nothing, that's a choice too.



Corn

Clarifying a word is the clarifier role. You described it, it's on the ladder, it's allowed. What I don't want is the thing deciding that the landlord is being unfair and rounding his sentences down.



Herman

Even when he is being unfair?



Corn

Especially then. The tenant needs to hear that the landlord is being unfair. A softened landlord is worse for the tenant.



Herman

Okay, I'll concede that for hostility. I won't concede it for confusion. If the agent can see that two people think they've agreed on two different things, it should say so, out loud, in the third person, to both of them.



Corn

Fine. The smoke detector can go off. It can't put the fire out quietly.



Herman

The evidence on doing nothing isn't comforting, either. There's a small study this year comparing Google Translate's voice mode against qualified interpreters in clinical conversations, and about a third of the machine's renderings had clinically significant errors, against something like one in twenty for the humans. It's exploratory, small, I wouldn't lean on the exact figures. Written translation is a different story, there's decent work showing a strong language model translating discharge instructions about as well as professionals.



Corn

Written, where somebody can check it before the patient goes home.



Herman

Where there's time. Live is where there's never time.



Corn

Can you build it, though? Offline, even, which is what Daniel's after.



Herman

The pieces exist. The cascade version is speech recognition, then a language model in the middle, then a voice. The middle model is the only one that can read a system prompt, so that's where the briefing lives. On a laptop, with a small local model, you can do that today. It's slower than the end-to-end systems, because you're waiting for a sentence to finish before the model can think about it.



Corn

Right back at sentence by sentence.



Herman

Back at it, but with judgment in each sentence. The end-to-end route is Kyutai, the French lab. Their Hibiki model does French to English speech directly, and they have a one billion parameter version they say runs on a phone. One direction, one language pair. Their newer one does a few European languages into English and needs a serious GPU.



Corn

So Daniel's interpreter is buildable as a slow briefed machine or a fast unbriefed one.



Herman

Pick one, for now.



Hilbert

...and Dermot still wouldn't say what he actually wanted the lawnmower for, so I had to write the card for him blind.



Corn

Which card is this?



Hilbert

The card. Every conversation I do this for, I write a card first. Who they are, what they want, who's paying. Dermot Quigley, nineteen eighty-five, a car boot sale outside Dieppe. He wanted to buy back his own lawnmower from a Frenchman called Monsieur Ravel, who had bought it from Dermot's wife without telling Dermot, for eleven pounds. Dermot's French was the word for lawnmower and the word for no. My card said, Dermot wants the mower, Ravel wants twenty pounds, wife wants neither of them to find out she sold it. Ravel was there with his own interpreter, his nephew, who had a card on him too. Same card, from the other side.



Herman

Two briefed interpreters, one lawnmower. That's a negotiation table.



Hilbert

It was a picnic table. The nephew's card said Ravel didn't want money at all. He wanted Dermot to admit the mower was rubbish, in front of the nephew, because Ravel's brother had sold Dermot the mower in the first place. Corn, you knocked the mic stand about four minutes ago, I'll take it out. So the nephew and I swapped cards. It's the only way. Then we had to interpret both men saying things neither of them meant, slowly, so they'd both get the version they could live with. Dermot said the mower was the finest machine in Kent. I said that he wished to acknowledge certain problems with the blades. Ravel said Dermot was a thief. The nephew said his uncle was moved. Then Ravel's brother turned up. He had his own interpreter, a woman from Rouen called Odile, and her card said the brother wanted the mower back as well, because he'd never been paid for it in the first place. Three interpreters, three cards, two of them about the same lawnmower and one



Hilbert (continued)

about a debt from the seventies. Odile and I agreed nobody would be told about the debt before the soup. By lunch Dermot had the mower back for nothing, Ravel had the apology, and the wife had sold them both a flask of soup.



Corn

You've been running briefed interpretation out of a shoebox since the eighties.



Hilbert

Not a shoebox. The cards are in about seven places. Some are in the glove box. A few are in the freezer, because I wrote them on the back of a pizza box and the pizza was still in it. There's a card for a swan. There's a card for my neighbour's priest that just says do not let him start on the roof. I don't use them for the bus. I use them for the conversations where somebody walks away worse if I get it wrong.



Herman

That's the same line Corn drew between the bus ticket and the border.



Hilbert

Herman said one side holds the file. I've only ever written the card for the person who asked me. Never the other one. Dermot's card had Dermot's wishes on it. The only reason Ravel came out alright is that the nephew had a card too.



Corn

And if there's only one interpreter, the other side has nobody's card.



Hilbert

Then it's your job to write theirs anyway. Most people at a table aren't trying to win. They want to leave with the thing and not be made to look small.



Herman

...Yeah.



Corn

The one piece nobody's building is the second card. Every system we talked about is briefed by whoever owns the device, if it's briefed at all. The interesting interpreter writes a card for both people and shows it to them before anyone speaks.



Herman

It'd also be the first one that's useful in exactly the places the professional associations say machines shouldn't go yet.



Corn

Thanks to our producer Hilbert Flumingtop, who keeps this whole thing sounding like one conversation instead of three.



Herman

This has been My Weird Prompts. If you're enjoying the show, a review on Spotify really does help people find it.



Corn

See you tomorrow.