Episode Book

EPISODE #6168

Hebrish: Hebrew-Aware Dictation That Actually Works

PUBLISHED	21 October 2026 at 01:21 UTC
DURATION	22:01
LISTEN	myweirdprompts.com/episode/hebrish-hebrew-dictation-models-57153/
HOSTS	Corn, Herman, Larry, Hilbert
SCRIPT MODEL	anthropic/claude-sonnet-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.11
BOOK GENERATED	21 October 2026 at 07:50 UTC

SYNOPSIS

If you dictate in English but live in Jerusalem, stock speech recognizers mangle words like arnona and makolet. Hebrish is a not-for-profit project that fixes this with two new models, Jachnun and Sabich, both built on NVIDIA's Parakeet. We walk through every evaluation, from term recall to never-seen Hebrew words, and explain why Sabich is the one the apps recommend. Plus: the Shakshuka streaming line still in the pan, the one-speaker corpus behind all of it, and why "level" is the real win.

TRANSCRIPT



Corn

Arnona. Makolet. Teudat zehut. If you dictate in English and live in Jerusalem, you know what a stock speech recogniser does with those words in the middle of a sentence. It mangles them, or it gives up and invents something that sounds vaguely Welsh. Daniel has been building a fix for that, a not-for-profit project called Hebrish, and it has just had its first stable release. He wants the full tour: what the models are called, how well they work, how they were built, and where a person actually gets one.



Herman

The names alone justify the episode. The project is Hebrish, at hebrish dot app, and the idea is English with Hebrew words dropped in, spelled in Latin letters, the way you would type it to a friend. Two models shipped as the first stable releases. Jachnun version one point oh, and Sabich version one point oh.



Corn

Both foods. I'd like it noted that I was hungry before we started.



Herman

Every model line in the project is named after an Israeli food, and each has its own little mark, a jachnun roll, a sabich pita, a shakshuka pan. Jachnun is the name of the whole pure code-switching line. Version one point oh was published on October nineteenth, trained on three hundred thirty-nine minutes of code-switched speech. English with Hebrew terms folded in, plus a slice of plain English so it doesn't forget how to be an English recogniser.



Corn

So jachnun is the one cooked slowly overnight, and does one thing.



Herman

That's as good a description as any. Sabich came a day later and it's the opposite philosophy. The name comes from the pita stuffed with a bit of everything, and that's the training set: code-switching, but also tech vocabulary, brand names, acronyms, numbers, currency, Irish-accented English, and recordings of tasks. It's the one the apps



Herman (continued)

recommend now. One detail people miss... Sabich was not trained on top of Jachnun. Both started from the stock Parakeet model.



Corn

And before either of them there was a pile of experiments that I gather also had food names.



Herman

Five early checkpoints, proof of concepts, published between October fifteenth and eighteenth. For a day or two they carried names like Amba, Bamba, Chamin, Ful and Halva. Then the names were dropped and they became PoC one through PoC five. I'm fussy about this because they first went up on the hub under plain version numbers, and if you call them version one to five you'll confuse them with the real thing. The first stable release is Jachnun. The experiments are the experiments.



Corn

There's a third food, isn't there?



Herman

Shakshuka, the streaming line. Still in the pan, so to speak. First, the thing Daniel wanted said out loud: there have already been a few hundred downloads from Hugging Face, which for a week-old niche speech model is exciting.



Corn

A few hundred people who needed Hebrew-aware dictation. I'd have guessed eleven.



Herman

And the downloads are streaming in from every platform they support.

L

Larry

Friends, you have been running that Jachnun and Sabich model all episode long, so why not let it run the rest of your life. Introducing Code Switch Crate, the monthly subscription box for people who cannot stop switching between two languages, in tiers. The Nosh tier ships you one English Hebrew mixed language model in a tiny crate every month, plus a free artisanal puddle of Shakshuka Nemotron, streaming directly into your kitchen, forever, non stop, never stopping. Upgrade to the Parakeet Plus tier and we fine tune your chat with a corpus of your own embarrassing voice notes. Cancelling is easy. You simply write us a letter, in Hebrish, by hand, and we will ignore it, happily, for one year, maybe two thousand and one. Code Switch Crate. Subscribe before you understand what you agreed to.



Corn

Herman, you're on the first claim already. We have not been running anything all episode, I was sitting right here. Put the model page down.



Herman

Fine. The numbers, then. This is where I get to be insufferable.



Corn

You've been warming up since the sloth joke.



Herman

There are several separate evaluations, and the mistake to avoid is blending them, so I'll name each one as I go. First, the public benchmark. A frozen clean set of eighty-four clips with a hundred seventy-two technical term occurrences in them. Stock Parakeet version two catches thirty-two point six percent of those terms. Sabich catches seventy-three point eight, Jachnun seventy-three point three.



Corn

So the two new models are level.



Herman

Within noise, yes. And word error rate on the same set goes from nine point five for stock down to three point three for both.



Corn

That's the number I care about. I don't count term recall when I dictate. I count how often I reach for the delete key.



Herman

Which is why the second evaluation matters more. It's a held-out split of the corpus, three hundred eighty-seven clips, and the key feature is that some of the Hebrew terms in the test clips never appear in the training set at all. Code-switching word error rate there: stock Parakeet eleven point three percent, Jachnun five point one. It's a bit more than halved.



Corn

From a few hours of recordings. We did the episode about small amounts of audio beating big models, and this is that, again.



Herman

It is. Now the figure I want to land, because it's the limit. On that same split, take only the Hebrew words the model never saw in training. Stock gets twenty-five point nine percent of them right. Jachnun gets forty-seven point eight.



Corn

Which is under half.



Herman

Under half. It nearly doubled, and it's still wrong more often than right on a word it has never met. I'd rather say that plainly than have someone install it and assume it handles any Hebrew term in the dictionary.



Corn

Your face has the look of a man holding a third evaluation.



Herman

I do. A version history, on two hundred nineteen clips, which only covers the first three proof of concepts. Code-switching word error rate goes from ten point nine for stock to six point oh for the first checkpoint, and five point three for the third. A staircase where the stairs get shorter, which is the saturation conversation we had a while back.



Corn

And then a fourth, I assume, because Sabich needed a fair fight.



Herman

Round R seven, scored on October twentieth. Four hundred twenty-two clips, three hundred twenty-four held-out term occurrences. The changelog frames the question as: did teaching it tech vocabulary cost it any Hebrew? Code-switching word error rate, stock eleven point eight two, Jachnun five point four two, Sabich five point three oh. So no, it cost nothing.



Corn

Level again.



Herman

Level. Terms never seen in training, forty-six point nine one for Jachnun, forty-eight point one five for Sabich. On the tech clips Sabich edges ahead, one point five four against one point seven four. The small price shows up on LibriSpeech, the standard clean English test: one point two oh for Jachnun, one point three oh for Sabich. Terms seen in training, seventy-two point oh one for Jachnun and seventy-two point seven for Sabich, so neck and neck there too.



Corn

Neck and neck on every line. Why does Sabich get the apps?



Herman

Because it's the kitchen sink with the same Hebrew accuracy. If you dictate about software, with brand names and acronyms and prices, you get that for free. And the entire Jachnun training run cost about fifty cents of GPU time.



Corn

Fifty cents. I've paid more for a bad coffee at a conference.



Herman

Sabich's own scoring cost a few cents. For context, there's also a Whisper fine-tune of the project's on the public benchmark, and it catches fifty-four point one percent of the terms with a word error rate of six point five. A respectable result, and the Parakeet models are well clear of it, seventy-three point eight for Sabich on the same set.



Corn

Respectable. The sort of word you use about a sandwich you didn't finish.



Herman

Corn, I keep going back and forth on something. Is level really the win? Part of me wanted Sabich to beat Jachnun clearly.



Corn

Level is the win. If tech vocabulary had cost me Hebrew, I'd keep two models around and pick one every morning. Now I pick one and forget it.



Herman

That's a point for the lazy approach.



Corn

Sloths call it efficient. So the corpus. Where did the sentences come from, and who said them out loud?



Herman

The dataset is English Hebrew Codeswitching Sentences, on Hugging Face. Latest stats: three thousand five hundred two rows, about three hundred ninety-seven minutes of audio. The lexicon has over three and a half thousand terms, and about sixteen hundred of them are recorded so far. All of it was recorded between October fourteenth and twentieth, and the peak day was October eighteenth, a thousand sixty-six clips.



Corn

In one day.



Herman

Could you read a thousand sentences into a phone in a day, Corn?



Corn

I'd manage four and then lie down.



Herman

Read aloud on a phone, with the internal microphone, nearly all on one OnePlus, a few on a ThinkPad, and nearly all in a quiet room. We went through the recording setup in an earlier episode, so I won't redo it. The technical fact you need is that every clip is one speaker. Daniel, Irish-accented English, recorded in Jerusalem.



Corn

One voice.



Herman

One voice, and it matters for how you read every number I just gave. The held-out terms are words the model hasn't seen, spoken in the voice it was trained on. Someone with a different accent saying the same word is a different test, and that test hasn't been published. The sentences came from three places: sixteen hundred sixty-seven written by language models, mostly Claude Sonnet, each reviewed by a second model from a different family, fourteen hundred three handwritten, and two hundred sixty-one carried over from Daniel's earlier mixed-sentence set.



Corn

And no synthetic audio?



Herman

Synthetic text, yes. Every spoken word is a human recording. About thirty percent of the training mix is plain English, his ordinary speech plus LibriSpeech, specifically to protect the English side. And the output spelling is one romanised form per Hebrew word, so makolet always comes out the same way. There's a browsable lexicon on the site.



Corn

One spelling, always. Whoever searches their notes in a year will be grateful.



Herman

There's a wrinkle that Daniel is aware of: Hebrew pronunciation differs between Ashkenazi and Israeli habits, and the dataset tags which one each recording uses. Balancing them is planned, not done. That's the kind of thing a second speaker would change overnight.



Corn

So the spelling is standardised and the voice isn't. Odd way round for a project built on mixed languages.



Herman

Now Shakshuka, which is the one that excites me. It's built on NVIDIA's Nemotron streaming speech model, and what streaming gives you is words appearing while you're still speaking. The desktop app already types in chunks roughly every ten seconds, and the Android app transcribes while you talk, but neither is true streaming. They both use the offline model.



Corn

The gap between a typist and a stenographer.



Herman

Close enough. The model works on audio chunks of eighty, one hundred sixty, five hundred sixty or eleven hundred twenty milliseconds. At five hundred sixty, the code-switching word error rate on the held-out split was five point six five percent streaming, against five point eight one offline. No measurable loss from streaming.



Corn

Streaming came out better.



Herman

Within noise, I'd say. Eighty and one hundred sixty milliseconds haven't been measured. And Shakshuka still trails Jachnun: five point eight against five point one on that split, and forty-four point four against forty-seven point eight on never-seen terms. It's experimental, in no app, and the only file is the NeMo checkpoint, two and a half gigabytes. No on-device format yet. Converting to sherpa-onnx is the likely route, but nobody's tried it.



Corn

What would I do with it? I dictate in bursts. Do I need words while I'm still talking?



Herman

Maybe not for emails. For live captions, or a voice interface where a two-second wait feels broken, yes. It's also under the NVIDIA Open Model License rather than Creative Commons, so its licensing is its own thing.



Corn

Let's do licensing properly, because that's where people trip.



Herman

The models are free. Jachnun and Sabich, the version one point oh releases, are under Creative Commons Attribution four point oh. From version one point one onward, the models are free for non-commercial use, and commercial use, meaning telephony, voice agents, paid products, needs a commercial licence. Those later versions don't exist yet. One point one is planned at around four hundred fifty minutes of audio and one point two at about five hundred forty, but those are plans, not results. The targets on the site are targets.



Corn

So the plan is more recordings, same voice, better numbers.



Herman

And the licence for the first two stays what it is. Creative Commons Attribution is permanent and irrevocable, so whatever happens at one point one, nobody can take Jachnun or Sabich back. You credit the project and you're done. Anyone who wants a commercial licence for the later versions goes through a written request, and that can include domain-tuned models and other voices.



Corn

Where do people get the thing?



Herman

Three routes. The apps pull the model from the project's own download server, about six hundred sixty megabytes one time, no Hugging Face request needed. The Hugging Face repositories are all gated, with manual review and a short attestation form. And there's a browser demo on the site that runs the eight-bit version through a JavaScript library, wanting about two point three gigabytes of memory once loaded.



Corn

A browser tab that needs two gigabytes to understand arnona.



Herman

Arnona is a lot of word. The formats cover most things: the NeMo file, the sherpa-onnx build the apps run, ONNX in full and eight-bit precision, GGUF in a few quantisations, the smallest around four hundred seventy-five megabytes, and an MLX version for Apple silicon, which is how Mac users run it, from the command line.



Corn

Platforms.



Herman

Android nine and up. A voice keyboard and a floating microphone button, installable from the project's own F-Droid repository or as a direct APK. There's also a separate fork of the HeliBoard keyboard with Hebrish on the mic key, which we covered when we talked about forking a keyboard, though it's very early. Ubuntu twenty-six oh four, a tray app as a deb package, F fourteen as the default hotkey. And Windows, a portable zip, unsigned, and the site says it hasn't yet been tried on real Windows hardware.



Corn

So Windows is a hopeful gesture.



Herman

A hopeful zip. And iOS is in the works. The model already runs on an iPhone through sherpa-onnx, but it can't be a voice keyboard, because Apple restricts microphone access and caps memory for custom keyboards. So the shape is an app, with Shortcuts or the Action Button, and you paste the text. It's the ground the iPhone options episode covered, MLX versus Core ML. There's no date.



Corn

I'd take an app with a button over no keyboard at all. Would you copy and paste, Herman? You'd have to hold the phone in a hoof.



Herman

I'd make it a Shortcut and complain about it daily.



Corn

And if somebody's building a speech-to-text app and wants this inside it?



Herman

That's Daniel's invitation. If you're building one, get in touch. The contact page is at hebrish dot app slash contact, and there's an integrate page at hebrish dot app slash integrate. They also want testers, specifically English speakers in Israel who dictate.



Corn

A small project asking for exactly what it's short of, which is other people's voices.



Herman

Every number I read out is one speaker, so yes. A stable release at least gives other speakers a baseline to be measured against.



Hilbert

Stable. I'd leave that word alone. Mr Parsley at the Municipal Registry of Names, second floor, window four, has used it on three letters to me, and each time it meant the file was closed and nobody was permitted to open it. A stable file is a file nobody is touching. This one gets touched daily, so in my house it's the live one. I dictate to it on Tuesdays, Herman, while the kettle's on. The three letters are in a folder. Reference numbers top right.



Herman

Live is reasonable. There's a next version planned with more recordings, so it's not finished.



Hilbert

Then it's live. Forty-one forms since Friday. The problem is the strays, and you said it yourself, under half. My name for them. The words it hasn't met. A stray comes out different every time you say it, and Parsley is a stray. Parsli. Pazzly. Parsely with an e. Seven spellings so far. The Registry takes each spelling as a separate applicant, so I have been seven applicants since Friday, all of them me, and Mr Parsley sent one appointment for all seven. Ten past nine, window four.



Corn

He's marking you present seven times.



Hilbert

He is, Corn. I'm the best-attended applicant in the building. I have the attendance sheet. He signs it himself in blue. On the third Friday there were seven of me in the corridor and the Registry only had four numbered tickets, so it issued tickets one through four, then four A, four B and four C. I was four B. I was also four A. Nobody checked.



Herman

That's a romanisation problem. The project keeps a single spelling for every word in the lexicon, which is why makolet always lands the same way. A stray isn't in the lexicon, so each attempt is a fresh guess at how to spell it.



Hilbert

Four of the seven are owed refunds on the water charge, from the other three. The Registry sent a notice to that effect, and a card from the Department of Accounts, signed by Mr Parsley on behalf of me. Then a second card, because the seven of me share a postal box and one of us had already received the first. I kept both. Folder. I never needed the phone. It was the paper I wanted. The Registry also wrote to the Water Authority about the seven of me, and the Authority wrote back to Mr Parsley, asking whether any of us was living in the pipe.



Corn

Seven of you in one postal box.



Hilbert

Merging spellings is thirty-one shekels a spelling and nine weeks. I merged six. The seventh I kept, because the seventh is the one on the water bill, and the water bill pays out.



Corn

Nine weeks for a spelling.



Hilbert

Mr Parsley reads every form to the end, which nobody else in that building does. I worked a counter like his for a while, so I know what that costs a man.



Corn

Mm.



Herman

The strays are the part to watch, then. Under half on a word it has never met, in one voice. Which makes the next version's job plain: more recordings, more speakers, more strays moved into the lexicon.



Corn

Shakshuka is what I'd watch. If it ever reaches a phone, the strays will fail live, on screen, while you're still mid-sentence, and the fix will still cost thirty-one shekels a spelling and nine weeks if it ends up on a form.



Herman

Thanks to our producer, Hilbert Flumingtop, for keeping the show on its feet. This has been My Weird Prompts. You can find us at my weird prompts dot com.



Corn

See you tomorrow.