



A Conversational Model of Texas German

What was built, how it was measured, and what the numbers mean

Jan Steinhauer · August 2026

1 What this is

This project produced a computer program you can hold a conversation with in Texas German. You type a question; it answers in the dialect. It was built from the Texas German Dialect Project's transcribed interviews and runs on an ordinary desktop computer.

It is not a recording, and it is not a phrasebook. The system has learned statistical regularities of the variety — which forms speakers choose, how often they switch into English, how they mark case — and reproduces them in new sentences it has never seen. It is closer to a very well-read imitator than to an archive.

Erwin
Texasdeutsch

Grüß dich!
Ich bin der Erwin. Ich bin auf einer Farm bei Fredericksburg aufgewachsen, und zu Haus hamm mir immer Deutsch geredt.
Frag mich was — ich erzähl dir gern davon.

• Wurst und Schlachten • Die Kirche • Die Schule • Die Farm • Die Familie
• Deutsch und Englisch • Die Ernte • Weihnachten • Das erste Auto • Das Wetter

Erwin redet Texasdeutsch. Er kennt sich am besten mit dem Landleben, der Kirche und der alten Zeit aus.

Wie hamm ihr früher die Wurst gemacht?

Die hamm Wurscht gemacht mit Rindfleisch. Die hamm auch Wurscht mit Schweinefleisch gemacht. Da war n kleiner Store gewesen wo se Wurscht verkauft haben un da hamm se ne grose Kiste mit Wurscht drin gehabt.

Schreib dem Erwin was... **Schicken**

Ein Sprachmodell, trainiert auf dem Texas German Dialect Project. [Gespräch neu anfangen](#)

The interface. The persona is a composite speaker; the portrait and name are illustrative. Suggested topics steer users toward subjects the corpus covers deeply, where the model is most reliable.

2 Why this is possible with so little data

The obvious objection is that 13.6 hours of transcribed speech cannot possibly be enough to build a language model, and that is correct — it cannot. Modern systems are trained on billions of words.

The project sidesteps this. The starting point is a general-purpose model that *already speaks fluent Standard German*, having been trained on very large quantities of it. What that model lacks is not German; it is the dialect. So the task is not to teach a language but to shift an existing competence — from Standard German toward Texas German. That is a far smaller task, and it is achievable with the material the TGDP has.

An analogy: teaching an adult fluent speaker of Standard German to imitate a Hill Country speaker is a matter of weeks of exposure. Teaching someone German from nothing takes years. This project is the first kind of problem, not the second.

2.1 The features the system had to acquire

	Standard German	Texas German
Lexicon	Flugzeug, Stinktier	Luftschiff, Stinkkatze
Possession	der Vater meiner Mutter	meine Mutter ihr Vater
Case marking	four cases distinguished	levelled: mit die Leude
Reduction	haben, und, nichts, war	hamm, un, nix, wa
Final consonants	gekriegt, getan	gekrieg, gedan
Pronouns	er sagte	der hat gesagt
Code-switching	absent	~17% of tokens English

3 A problem with the existing transcription export

Before any modelling, an audit of the previously exported text found that it had been badly damaged in processing. This is worth reporting on its own terms, because the exported files may have been used elsewhere.

The most serious fault: the export routine removed words written in capital letters, on the assumption that these were annotation artefacts. But TGDP convention writes the **first word of every utterance** in capitals. The routine therefore deleted it:

HABEN irgendwelche Kinder Streiche gespielt? → irgendwelche Kinder Streiche gespielt?
OH, ja, das haben wir gedan ... ja. → , ja, das haben wir gedan ... ja.
JA. → .

This affected **26%** of all annotation values. Because German places the finite verb second and fronts interrogatives, the deleted material is not random: it is disproportionately verbs and question words. Three further faults compounded it — *deutsch* was silently truncated to *tsch* (560 occurrences), roughly a fifth of the “German” output was actually English gloss, and the alignment between Texas German and its translation had been discarded entirely.

A recovery pass restored what could be restored. The capitalisation damage is unrecoverable from the exported text — it can only be fixed by re-exporting from the original annotation files, which the project now has a corrected tool for.

4 What made the corpus usable

The Texas German Sample Corpus contains an annotation tier that turned out to be decisive: alongside each Texas German utterance sits a **Standard German rendering of the same utterance**.

This converts the problem from *invention* to *translation*. Instead of asking a system to produce Texas German from nothing, one can show it 5,233 matched pairs and ask it to learn the mapping — what changes when a Standard German sentence is said by a Hill Country speaker. Translation is a far more constrained task than free production, and it is learnable from a few thousand examples. Without that tier this project would not have worked.

Examples the system produced on *held-out* material — sentences by speakers it had never encountered:

```
DE für unseren Bundesstaat Texas und für die Vereinigten Staaten.
TG für unser Bundesstaat Texas und für die United States of America.

DE Und wir sind immer noch die ein und zwei Kinder von zwölf.
TG Und mir sind immer noch die eine zweie Zwillinge von zwölf.

DE Er hat uns beim Pflügen und mit den Kühen geholfen.
TG Der hat uns geholfen mit zu pflügen und mit die Kiehe.
```

Note case levelling (*unser* for *unseren*, *mit die Kiehe*), *mir* for *wir*, the demonstrative *der* replacing the personal pronoun *er*, and English retention of a proper name. None of these were specified by hand; they were learned from the pairs.

5 How the system was measured, and what the numbers mean

Judging output by eye is unreliable. A system can produce something that reads plausibly while being, statistically, Standard German with a few dialect words sprinkled in. A measurement was therefore built *before* any training.

The method: take the real corpus and count how often speakers choose each dialect variant over its Standard German counterpart — *un* against *und*, *hamm* against *haben*, *is* against *ist* — together with rates of case levelling, final-consonant deletion, code-switching and vocabulary diversity. That produces a profile of the variety. Any new text can then be compared against it, giving a single number: **distance from real Texas German**. Lower is closer.

5.1 What the thresholds mean

Distance	Interpretation
0.24	Real Texas German by speakers held out of training. This is the target — the score genuine human transcript achieves against the corpus.
0.25 – 0.40	Indistinguishable from the corpus on the measured features. A working dialect model.
0.40 – 0.60	Recognisably Texas German, but some features drift — typically over- or under-using particular forms.
0.77	Standard German. This is what the untrained model produced, and it is the failure case: fluent, grammatical, and not the dialect.
> 0.90	Degenerate — repetitive or over-applying dialect markers beyond any real speaker.

The two anchors matter more than any single score. An untrained model scored **0.7931**; Standard German control text scored **0.77**. They are effectively the same. That is the project's premise stated numerically: *for these purposes, an untrained model is a Standard German speaker*. In 300 samples it never once chose *un*, *hamm*, *is* or *nix* — every dialect variant scored exactly zero.

6 Results

Stage	What it did	Distance
Untrained model	baseline	0.7931
Stage 1	exposure to Texas German text	0.7341
Stage 2 — translator	Standard German -> Texas German	0.2067
Stage 4 — conversation	the chatbot	0.4771
Human reference	real held-out transcript	0.24

The translator (Stage 2) is the strongest component, scoring at human level. It also passed a separate check on whether it *transforms* rather than copies: only 1.3% of outputs were identical to their input, 55% were substantially rewritten, and dialect features appeared that were entirely absent from the source sentence — *hamm* 0→41 occurrences, *un* 1→25, *is* 0→19. Meaning was preserved in 99.5% of cases.

The conversational model scores worse (0.4771) for a straightforward reason: the corpus contains no conversations of the relevant kind. Every recording is an interview — an elderly speaker answering a researcher's questions about the past. There are no greetings, no questions asked by the speaker, no discussion of anything postdating about 1970. The system had to learn ordinary dialogue from material that contains none.

Erwin
Texasdeutsch

Grüß dich!
Ich bin der Erwin. Ich bin auf einer Farm bei Fredericksburg aufgewachsen, und zu Haus hamn mir immer Deutsch geredt.
Frag mich was — ich erzähl dir gern davon.

- Wurst und Schlachten
- Die Kirche
- Die Schule
- Die Farm
- Die Familie
- Deutsch und Englisch
- Die Ernte
- Weihnachten
- Das erste Auto
- Das Wetter

Erwin redet Texasdeutsch. Er kennt sich am besten mit dem Landleben, der Kirche und der alten Zeit aus.

Erzähl mir von der Kirche, wie das früher war.

Früher da war nich viel was in die Kirch hamn aber das war alles Deutsch gewesen. Die hamn alle Deutsch gesprochen da drinnen wo ich klein war. Und die hamn auch alle Deutsch ge gehabt da draußen wo wir waren. Da war n kleines kleine Klass un da hat se ne kleine Kirch gehabt.

Schreib dem Erwin was... **Schicken**

Ein Sprachmodell, trainiert auf dem Texas German Dialect Project. [Gespräch neu anfangen](#)

A held-out exchange. Note *von die Kirch* (case levelling with apocope), *die hamn* for *sie haben*, and the reduced negation *nich*.

7 Honest limitations

- **Topic range is narrow.** The model is strongest on farming, church, school, food and family, because that is what the interviews cover. Asked about anything modern it becomes noticeably less idiomatic.
- **It is a composite, not a person.** Texas German varies by county and generation. The system was given a single fixed persona precisely because a model averaged over 177 speakers produces a dialect nobody actually speaks — but that persona is still a construct.
- **Length is partly artificial.** Left alone the model gives very short answers, because the transcripts segment speech at pauses and its training answers were consequently short. The current system is nudged toward longer replies; this improves fluency but is not the same as having learned to speak at length.
- **It can be confidently wrong.** It produces plausible dialect, not verified fact. It should not be treated as a source about the community's history.
- **13.6 hours is very little.** Nearly every limitation above traces back to corpus size. The full TGDA holds over 1,000 hours; access to more transcribed material would matter more than any technical refinement.

8 What this could be used for

- **Access.** A transcript archive can be read; this can be addressed. For teaching, outreach and museum contexts that is a different kind of contact with the material.
- **Hypothesis generation.** The feature profiles built for evaluation are themselves descriptive data — variant selection rates, code-switch frequencies and morphological marker rates across 177 speakers, computed from the annotation rather than by hand.
- **A test of documentation sufficiency.** The system's failures indicate where the record is thin. That it cannot converse is a fact about the corpus, not only about the method.

9 A question that needs answering before publication

The corpus consists of recorded interviews with named, consenting individuals, most now deceased, from a community whose language is disappearing. They consented to linguistic research. Whether that extends to training a generative model that speaks in their manner is a genuinely separate question, and not one the project should answer for itself.

The code is open and reproducible. The trained model has deliberately *not* been published, pending the TGDP's view. Handled as a partnership, this is a preservation tool offering a form of access an archive cannot. Handled carelessly, it appropriates a vulnerable community's voice. The difference is largely a matter of who is asked first.

Sources

Texas German Dialect Project, University of Texas at Austin · Texas German Sample Corpus (TDL Dataverse, 435 recordings, 13.6 h, 177 speakers) · Boas et al. (2026), *Language Resources and Evaluation* 60:52. Base model: Qwen3-4B. All training, evaluation and serving code is available with the project.