Fulfulde Speech Technology for Niger: Varieties, Resources and a Plan

Namu AI Lab
namuai-lab.com

Abstract

Fulfulde is one language with many varieties, spoken from Senegal to Chad, and Niger has two of them: a western variety around Niamey and Dosso and a central-eastern variety from Tahoua to Diffa. We surveyed what exists for building speech recognition and speech synthesis in these two varieties: public speech and text with their terms of use, the varieties that public models list, and the numbers that have been published. We have not run any model on Fulfulde for this report; it is a survey and an evaluation plan fixed before any result. Four findings stand out. First, Fulfulde has far more open data than Zarma, but little of it comes from Niger: of about 268 hours of openly licensed, transcribed speech, about 11 hours are in Niger's varieties with transcripts that fit their audio, and the largest Niger source, 291 hours of radio, is gated and has no transcript checked by a person. Second, published labels are unreliable: a read-speech set filed under a Senegal locale reads Nigerian text, and in the only open corpus labelled Western Niger Fulfulde, 596 of 906 recordings carry transcripts whose length does not follow their audio, and nearly all of these transcripts also appear in the same corpus's Pulaar part. Third, Niger's two varieties are far apart: a dialect survey found that eastern listeners understood a Nigerian text fully and the western reference text poorly, and public text places eastern Niger speech beside Nigerian and Cameroonian data. Fourth, spelling varies widely, from transcribers who mark vowel length in 0 to 4% of words to edited texts that mark it in 36 to 44%. No public recogniser has a language id for Western Niger Fulfulde, and we found no published number on Niger Fulfulde speech. We describe a 110-speaker Niger test set, sized by a speaker-level bootstrap, and the gates we fixed in advance.

1 Introduction

Fulfulde, also called Fula, Fulani, Peul or Pulaar, is spoken across the Sahel and the savanna from Senegal and Mauritania to Chad and Cameroon. Linguists describe it as one language, but its varieties differ enough that ISO 639-3 divides the Fula macrolanguage into nine individual codes. In Niger, Fulfulde is spoken in the west and in the east, by settled and by nomadic communities; a dialect survey of eastern Niger, citing earlier work, puts the number of speakers in the country at about one million [14]. Many of them reach information through a phone on a weak and costly connection, and many do not read French. As for Zarma [18], speech is the natural interface.

Fulfulde does not look like a low-resource language at first. Public speech collections, recognisers, synthesis models and translation models all list it, often under several codes. The difficulty is that almost none of this material is the Fulfulde spoken in Niger (Figure 1). This report asks what exists

Technical report, Namu AI Lab, November 2026. This report describes research systems, not products.

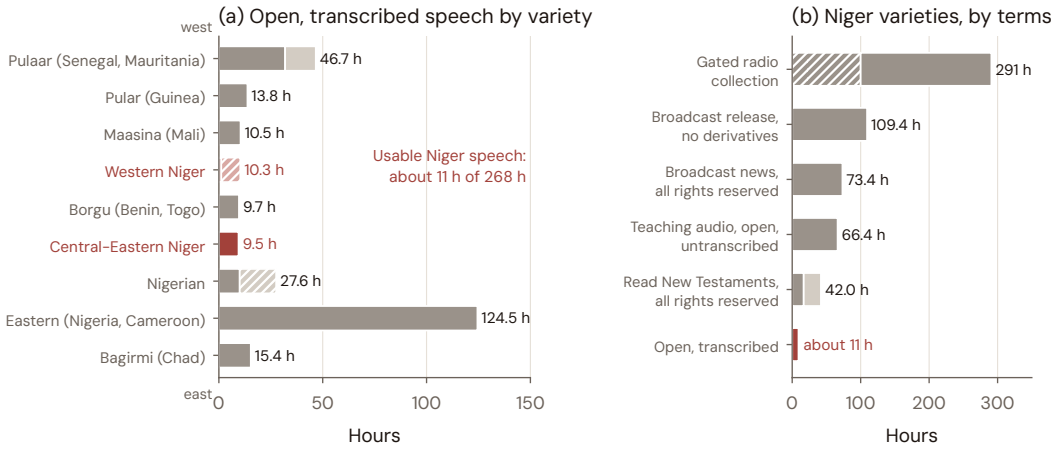


Figure 1: Public Fulfulde speech. (a) Openly licensed, transcribed speech by variety, roughly from west to east; lighter segments are a second source (Pulaar: a prompted-speech corpus; Nigerian: FLEURS, whose readers' variety is not documented, hatched). For Western Niger, the hatched segment is speech whose transcripts do not fit their audio or repeat another speaker's, which we hold out. (b) Public audio in Niger's two varieties, by what its terms allow; the hatched part of the radio collection has machine-drafted transcripts, and none has been checked by a person; the lighter part of the New Testament bar is one copy of the western Niger edition. Overlaps between sources are not known and are ignored.

for Niger's two varieties, what it really contains, and how a recogniser, a synthesis model and a voice agent for them should be measured. It is the second report in our programme on the languages of Niger; the first, on Zarma [18], taught us that measurement problems (spelling conventions, wrong references, unreliable language identification) can decide results as much as models do.

We have not run any model on Fulfulde for this report. Every number in it is either published by others or measured by us on public text and dataset metadata, and we say which. Our contributions are the following. (1) A description of the Fulfulde varieties relevant to Niger and of how they are written, from dialect surveys and from counts on public text (Section 3). (2) A survey of public Fulfulde speech and text by variety and terms of use, with checks of what published labels really contain (Section 4). (3) A survey of the varieties that public models list and of the published Fulfulde numbers we found; none concerns Niger speech (Section 5). (4) An evaluation design for Niger, with test set sizes derived from a speaker-level bootstrap and promotion gates written before any score (Section 6), and a comparison with Zarma (Section 7).

2 Related work

Multilingual speech collections. Common Voice, built from read speech given by volunteers, has locales for Adamawa and Borgu Fulfulde [2]. FLEURS provides read speech for evaluation in about a hundred languages, Fula among them [7]. Massively multilingual recognisers cover Fula with a single adapter for the macrolanguage [23]. Recent African collections add much more: WAXAL releases about 124 hours of transcribed spontaneous Fulfulde [8], and a benchmark built on it compares recognisers across 19 African languages [21]. Kallaama provides Senegalese Pulaar radio and focus groups about agriculture [11]. For Niger, SALAN collects radio in eight languages for recognition research [19], and an earlier release of Niger and Mali broadcast news includes Fulfulde [24].

Fulfulde text and translation. Evaluation text in Fulfulde comes mostly from one family of resources: FLORES [13] and the reading-comprehension set built on it [3], plus the health-domain TICO-19 [1], the BOUQuET benchmark [4] and the professionally translated SMOL sets [5]. All of these are labelled Nigerian Fulfulde or Pulaar.

Table 1: Fulfulde varieties relevant to Niger. Speaker counts are Ethnologue estimates for all countries, of different years and editions, as reported in secondary sources [9]. Hours are openly licensed, transcribed public speech (Section 4).

Variety	Code	Where	Speakers	Open hours
Western Niger	fuh	Niamey, Dosso, Tillabéri; Burkina Faso	3.1 M	10.3 (1.15 usable)
Central-Eastern Niger	fuq	Tahoua, Maradi, Zinder, Diffa	0.86 M	9.5
Nigerian	fuv	northern Nigeria	17 M	27.6
Adamawa	fub	Cameroon, Chad, Nigeria	3.0 M	see last row
Pulaar	fuc	Senegal, Mauritania, Gambia	6.3 M	46.7
Pular	fuf	Guinea	4.8 M	13.8
Maasina	ffm	Mali	4.6 M	10.5
Borgu	fue	Benin, Togo	0.8 M	9.7
WAXAL: Nigerian or Adamawa speech of Nigeria and Cameroon, not separable				124.5

Varieties and orthography. The most direct evidence on Niger's varieties is an SIL survey of dialect intelligibility and language attitudes in eastern Niger [14]. Glottolog classifies the varieties [12], and phonological work describes Fula as non-tonal, with stress described differently from dialect to dialect [17]. Niger fixed its Fulfulde orthography by ministerial order in 1999 [20]; we know its letters only through secondary sources [10]. Numeral systems by variety are documented in a cross-linguistic database [6].

Measurement. We score speech with character error rate and translation with chrF++ [22], and plan listening tests on the comparison scale of ITU-T P.800 [15] with agreement measured by Krippendorff's α [16]. Our Zarma report [18] describes the measurement practices this plan inherits.

3 One language, many varieties

The varieties that matter for Niger. Table 1 lists the varieties this report deals with. Two are spoken in Niger. Western Niger Fulfulde is spoken around Niamey, in Dosso and the Dallols, Tillabéri, Say, Torodi and Téra, and its speakers' second language is Zarma. Central-Eastern Niger Fulfulde is spoken from Tahoua through Maradi and Zinder to Diffa, including by the Wodaaɓe, and its speakers' second language is almost always Hausa [14]. Glottolog places the eastern variety in an Eastern Fula group with Nigerian and Adamawa Fulfulde, and the western variety outside it [12].

How far apart they are. The SIL survey tested comprehension at 17 sites in eastern Niger [14]. Lexical similarity between western and eastern word lists was 34 to 53%, with a median of 44%. Only two sites understood the western reference text, recorded at Birnin Gaouré, and one of them had just finished a literacy campaign in the western dialect, which the authors read as acquired rather than inherent comprehension. A text from Kano, in Nigeria, was understood with the maximum score at every site, and some listeners said its speaker mixed in less Hausa. Western listeners were not tested on eastern speech, so the reverse direction is unmeasured. Niger's government literacy materials were developed in the western dialect [14]. For speech technology the practical reading is that Niger has two targets: an eastern model can lean on Nigerian data, and a western model cannot assume the same. Our programme takes the western variety first, because our team records in Niamey and western speakers mix in Zarma, which our models already speak, and the eastern variety second; both are in one test set from the start (Section 6).

How it is written. Niger's Fulfulde orthography was fixed in 1999 by a ministerial order in the same series as Zarma's [20]. As reported, it has 31 letters: *a b mb ɓ c d nd ɗ e f g ng h i j nj k l m n ny ŋ o p r s t u w y y̓* [10]. We could not obtain the text of the order itself, so whether it gives the

Table 2: Two spellings of one word in public Fulfulde text. Counts in the FLEURS transcripts, or in our samples of public text (FLEURS, first rows of a web crawl, transcripts of a public prompted-speech corpus). Whether the prenasal pairs are spelling or pronunciation is for native speakers to decide.

Contrast	Form	Count	Form	Count	Counted in
Hooked letter	*himbe*	304	*himɓe*	274	FLEURS
Hooked letter	*dum*	187	*ɗum*	171	FLEURS
Vowel length	*wodi*	276	*woodi*	402	samples
Vowel length	*fere*	357	*feere*	368	samples
Consonant length	*mako*	196	*makko*	783	samples
Consonant length	*wakati*	6	*wakkati*	347	samples
Prenasalised stop	*der*	198	*nder*	1,502	samples
Prenasalised stop	*gam*	333	*ngam*	968	samples
v and *w*	*vi*	82	*wi*	39	FLEURS

glottal stop a letter is open. Niger writes *ny* for the palatal nasal, where the Zarma order of the same year writes *ɲ*; Senegal and Guinea write *ñ* and Mali *ɲ*, and Nigeria writes *'y* where Niger writes *y'*. Vowel length and consonant length are written by doubling.

Sounds and grammar that models must keep. Fula is not tonal [17]. The contrasts that matter for recognition and synthesis are vowel length, consonant length, the implosives *ɓ ɗ y'*, prenasalised stops and the glottal stop. Several carry grammar: *be* 'and, with' differs from *ɓe* 'they', and verbs change their first consonant with a plural subject (*o yahi*, *ɓe njahi*). Numerals differ by variety. Western Niger Fulfulde and Pulaar count in fives (*joy* 5, *jeegom* 6), while Central-Eastern Niger Fulfulde uses Arabic loans for tens (*asherin* 20, *talatin* 30) and Hausa *dubu* for 1000 [6]. A synthesis front end therefore needs a variety switch for numbers.

Spelling in practice. Real text departs from any standard (Table 2, Figure 2). In the FLEURS transcripts, 55% of the 1,972 sentences contain no hooked letter at all, and four different apostrophe characters occur. Transcribers of one public Nigerian corpus write a doubled vowel in 0 to 4.1% of words, against 36 to 44% in edited Bible text, and hooked letters range from 0 to 219 per 1,000 words across transcribers. Some of these variants are spelling habits; others, such as *der* against *nder*, may reflect pronunciation, and only native speakers can say which. The vocabulary is also large. On 1,905 parallel sentences, Fulfulde has 8,272 word types against 6,697 for Hausa, and 65% of the Fulfulde types occur once. After folding hooks, apostrophes, length and geminates, Fulfulde still has 7,341 types against 6,332 for Hausa, about 16% more, which we attribute to morphology.

4 Public speech and text

Between 4 and 7 November 2026 we searched public dataset hubs, speech and text archives, papers, broadcasters and religious audio for Fulfulde sources, read the published metadata and terms, and counted what we could in transcripts and text. Every count we made comes from metadata and text; we used no gated file and no audio. Figure 1 shows the result for speech; Appendix Table 9 lists the main sources.

Speech. We count about 268 hours of openly licensed, transcribed Fulfulde speech. WAXAL contributes 124.5 hours of spontaneous image descriptions from about 211 speakers [8]. A public corpus of prompted spontaneous speech, released with a massively multilingual recogniser, contributes 93.9 hours in eight varieties, with two to twelve speakers per variety. Kallaama contributes 31.9 hours of Senegalese Pulaar radio and focus groups, of which 11 hours were checked by experts [11], and FLEURS 17.55 hours of read sentences [7]. About 11 hours of the total are in Niger's varieties with transcripts that fit their audio: 9.48 hours of Central-Eastern Niger Fulfulde from three speakers, one per split, and 1.15 hours from two speakers of the western set described below, whose variety native

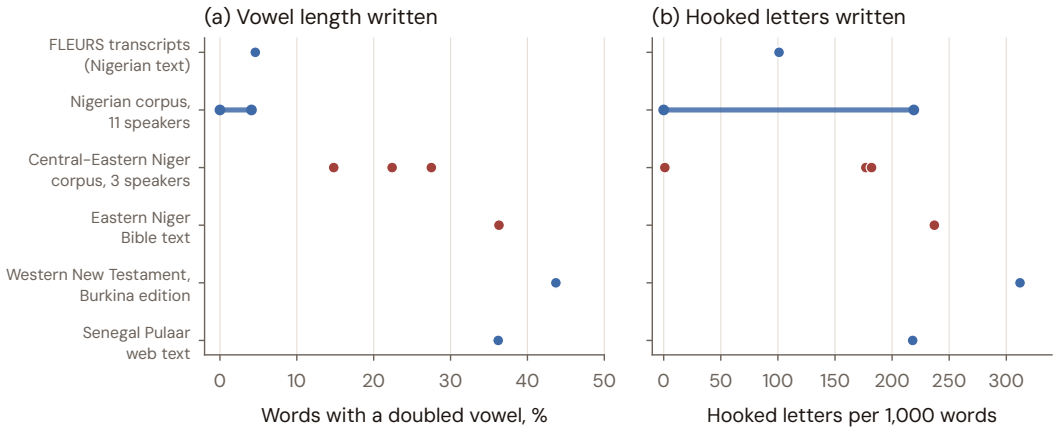


Figure 2: How often two contrasts are written, by source. (a) Share of words containing a doubled vowel. (b) Hooked letters (ɓ ɗ ƴ) per 1,000 words. Lines span the range over a corpus's speakers; separate dots are separate speakers. Niger varieties in red. Counts on samples of public text.

speakers have yet to confirm. Common Voice adds 14 hours of Adamawa and 11 hours of Borgu read speech behind an account [2], and WAXAL adds eight read voices for synthesis, whose prompts are Senegalese Pulaar.

Niger beyond the open data. The largest Niger source is radio. SALAN holds 290.94 hours of Fulfulde from Niger stations; its authors machine-transcribed 100 hours and verified none by hand, and access is by manual approval [19]. A Niamey newsroom publishes daily Fulfulde news, 73.4 hours of it in podcast form since December 2024, with all rights reserved, and an earlier release of the same kind of broadcasts, 109.4 hours after segmentation, carries terms that forbid derivatives [24]. A read New Testament in the eastern variety measures 16.8 hours and one copy of the western Niger edition 25.2 hours; a Burkina Faso edition also exists, and all are rights reserved. A second search, for the western variety only, found about 66 hours of untranscribed religious teaching audio labelled Western Niger Fulfulde, under an open licence. We found no speaker-diverse test set for either variety, no transcribed two-party conversation, no telephone speech from Niger, and no transcribed code-switching between Fulfulde and Zarma, Hausa or French. Women are poorly documented: the Kallaama paper gives 13.6% of its Pulaar as women's speech, its metadata gives 10.7%, and 1.7% in the part checked by experts; gender is empty for every WAXAL recording.

What the labels contain. Several published labels do not describe their contents (Table 3). Every one of the 1,972 FLEURS Fulfulde sentences, filed under a Senegal locale, matches a line of the FLORES Nigerian Fulfulde text after lower-casing and removing punctuation; who read them is not documented. WAXAL's recognition set reads as eastern Fulfulde of Nigeria and Cameroon, while its synthesis prompts are Senegalese Pulaar. The broadcast release labelled Nigerian Fulfulde comes from a Niamey newsroom. Web crawls are worse: in one large crawl, the Maasina label is 72% Senegalese Pulaar, and the Western Niger label is mostly Bible text. We used marker words to check varieties (Table 4); these are lexical cues, and native speakers must confirm every attribution.

The western Niger set. The only open corpus labelled Western Niger Fulfulde holds 906 recordings, 10.30 hours, under eight speaker identities. We checked it from its metadata and text, without audio. For four identities, with 596 recordings and 7.94 hours, the length of a transcript does not follow the length of its recording: the correlation between the two, on a log scale, lies between -0.14 and 0.04, while 28 of the 33 speakers of the corpus's other Fulfulde varieties lie between 0.61 and 0.97 and none is unrelated. All but four of these 596 transcripts also appear, exactly or nearly, in the corpus's Pulaar part, where they answer other prompts, and their durations differ from those of the Pulaar recordings, so the likely cause is transcripts attached to the wrong recordings rather than recordings filed twice. A fifth identity's transcripts do follow its audio, but 140 of its 150 also appear under

Table 3: What some public Fulfulde sources really contain, from metadata, full text scans and marker counts.

Source, as labelled	What we found
FLEURS, Senegal locale	1,972 of 1,972 sentences are FLORES Nigerian Fulfulde text; readers undocumented
WAXAL recognition set	Eastern Fulfulde of Nigeria and Cameroon by marker words; gender empty
WAXAL synthesis set	Senegalese Pulaar prompts; 1,694 of 3,878 rows have no text
Prompted corpus, Western Niger	For 596 of 906 rows, transcript length does not follow the audio; 592 of these transcripts also appear in the Pulaar part
Prompted corpus, Pular	One speaker's recordings repeat a Pulaar speaker's (329 shared transcripts, 158 identical durations) and use eastern forms
Prompted corpus, Nigerian	54 training transcripts stored with a broken text encoding (1.06 h)
Broadcast release, Nigerian	Niger newsroom speech, recorded in Niamey
Web crawl, Maasina	72% Senegalese Pulaar

Table 4: Marker words per 1,000 tokens. *ina* and *miɗo* (progressive) and *leydi* 'land' lean west; *ɗon* (progressive), *lesdi* 'land' and *himɓe* 'people' lean east. Ranges are over speakers. The western Niger corpus is left out because most of its transcripts do not belong to their audio.

Source	*ɗon/don*	*ina*	*miɗo/miɗen*	*leydi*	*lesdi/leddi*	*himɓe*
FLEURS (Nigerian text)	3.5	0	0	0	5.6	13.5
Eastern Niger Bible text	0	0	0	0	5.7	0
Western New Testament, Burkina	0	9.1	9.6	1.5	0	0
Senegal Pulaar web text	0.4	19.7	0.2	6.2	0	0
Central-Eastern Niger corpus, 3 speakers	20 to 22	0	0 to 1	0	2 to 6	0.5 to 9.5

one Pulaar speaker, probably the same person, and a sixth identity repeats five recordings of another with identical durations. That leaves two speakers, 155 recordings and 1.15 hours, whose transcripts follow their audio and answer their prompts; text alone does not settle their variety. We hold the set out of training and testing until native listeners have checked each speaker. A public synthesis model lists Western Niger Fulfulde among its training languages with 9.69 hours. A single cap on clip length, of about 100 seconds, reproduces its hours for seven of this corpus's eight Fulfulde varieties to the hundredth of an hour; if that is how its data was selected, about three quarters of its western hours pair text and audio that do not belong together.

Which public data is closest to Niger speech. Figure 3 compares samples of public text by the share of tokens of one sample missing from another, after folding spelling. WAXAL, FLEURS and the Central-Eastern Niger transcripts form a cluster (30 to 39%), far from Pulaar and Maasina (54 to 67%). This supports eastern public data as a starting point for the eastern variety. For the western variety the evidence is thin: the 6,548 tokens of the two speakers whose transcripts fit their audio sit nearer Borgu (32%) and Nigerian (36%) text than Pulaar (47%), and the variety of these two speakers is not settled.

Text. There are about 10,600 open human French and Fulfulde sentence pairs, four times what we found for Zarma, but nearly all are Nigerian Fulfulde or Pulaar: 6,193 pairs of a public seed set, 3,064 health-domain pairs [1] and 1,198 Pulaar pairs, plus about 190 pages of children's stories translated in Niger (about 5,000 words). The second search found 18 tales collected in western Niger, about 24,000 Fulfulde words with a French translation under an open licence, whose marker words lean east. The largest Central-Eastern Niger text we know of is the 74,882 words of transcripts of the three speakers above. Outside the Bible, large web crawls contain essentially no Niger Fulfulde. Every public evaluation set for translation or comprehension is labelled Nigerian Fulfulde or Pulaar [3, 4, 13]. These sets also share one text pool: 509 of the 1,012 FLORES test sentences appear in the

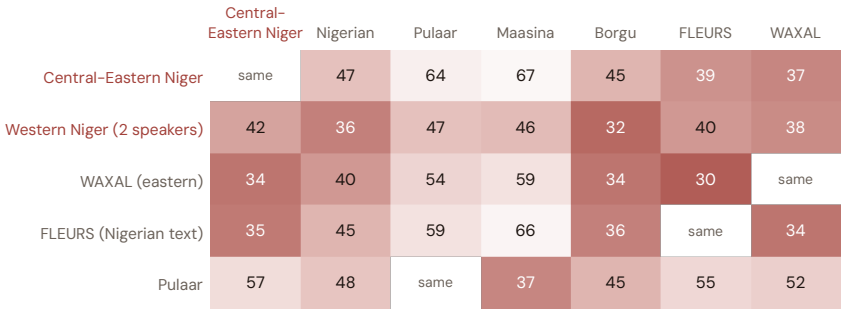


Figure 3: Lexical distance between samples of public Fulfulde text: the share (%) of tokens of the row sample not found in the column sample, on equal samples of 12,000 tokens with spelling folded. Lower is closer. The Western Niger row uses the 6,548 tokens of three speaker identities, which hold the two speakers whose transcripts fit their audio.

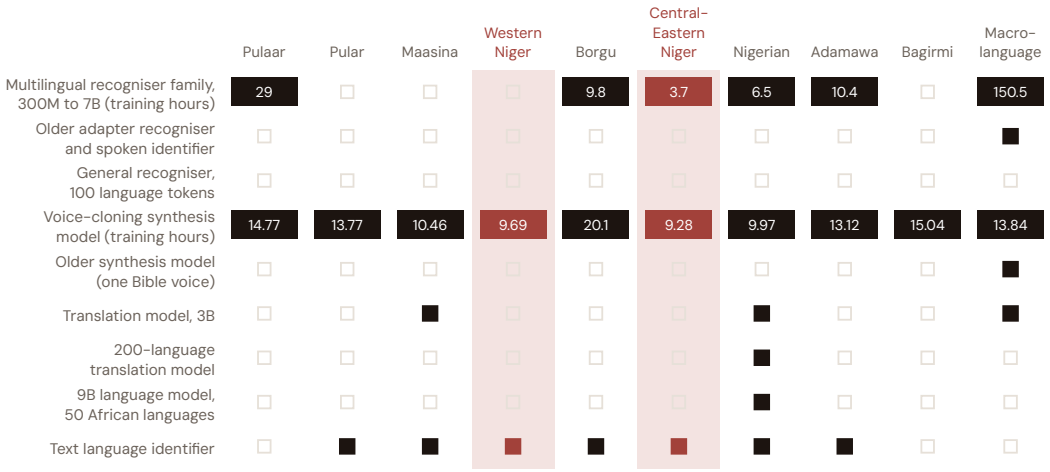


Figure 4: Fulfulde varieties listed by public models. Filled cells are listed; numbers are training hours where the developers state them or list them. Niger's two varieties are shaded.

FLEURS training split, and 469 of 488 reading-comprehension passages contain a sentence from it. A model trained on FLEURS transcripts has read part of these benchmarks.

Public tests. Only one public slice can support a decision with confidence intervals over speakers: the speaker-disjoint second release of the WAXAL test, 22 speakers and 11.87 hours, and it measures eastern Fulfulde of Nigeria and Cameroon. Its first release is not speaker-disjoint, since 99.4% of its test speakers also appear in training, and the second release's test draws 1,848 of its 2,285 utterances from the first release's training split, so any model trained on that split has heard it. The Central-Eastern Niger test is one speaker. There is no public test of Western Niger Fulfulde.

5 What public models know

Figure 4 shows which varieties public models list. We read language lists, lookup tables and tokenizers; we ran no model.

Recognition. A public massively multilingual recogniser family, from 300M to 7B parameters with a CTC head or an LLM decoder, has six Fulfulde language ids: Pulaar, Borgu, the macrolanguage, Central-Eastern Niger, Nigerian and Adamawa. It has no id for Western Niger Fulfulde, Pular or Maasina. Its developers publish character error rates (CER) on their own held-out data (Figure 5a, Appendix Table 10). The Central-Eastern Niger id was trained on 3.7 hours, close to the one-speaker

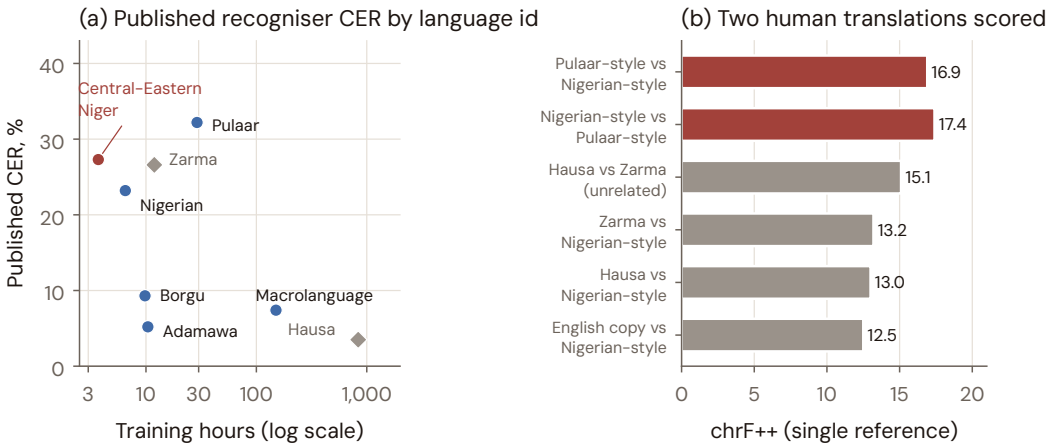


Figure 5: (a) Published CER of one public recogniser family (7B, LLM decoder) by Fulfulde language id, against the hours each id was trained on, on the developers' held-out data; Zarma and Hausa for reference. (b) Two human Fulfulde translations of the same 862 English sentences, one Nigerian-style and one Pulaar-style, scored against each other with chrF++, beside scores of text in the wrong language. Point estimates.

training split of the public prompted corpus, and has a published CER of 27.3%, close to Zarma's 26.6%; the macrolanguage id was trained on 150.5 hours and reaches 7.4%, probably on read speech. An older adapter-based recogniser has one Fulfulde adapter for the macrolanguage, fine-tuned on FLEURS to 13.8% CER [23], and its spoken language identifiers have one Fulfulde label. A widely used general recogniser has no Fulfulde token. On WAXAL, a fine-tuned 300M CTC model reports 9.5% CER and 37.4% word error rate [21], on the first-release split whose test speakers are also in training. The same benchmark shows what happens without replay: a recogniser fine-tuned on the 18 other WAXAL languages went from 58.4 to 87.1% word error rate on Fulfulde [21]. No public spoken language identifier separates the western and eastern varieties.

Synthesis. A public massively multilingual synthesis model that clones a voice from a short reference lists ten Fulfulde ids and 130.04 hours, including 9.69 hours of Western and 9.28 hours of Central-Eastern Niger Fulfulde; the western hours very likely come from the western set of Section 4. The hours match public corpora: its Central-Eastern and Nigerian hours match the prompted corpus's training, development and test splits together (9.28 against 9.48 hours, 9.97 against 10.04), and its macrolanguage hours match the FLEURS training split (13.84 against 13.89). Those test speakers therefore cannot test synthesis, and the FLEURS test split is the one public natural recording it has not heard. The model takes the language id as plain text; an id it does not know, such as the macrolanguage code `ful`, silently becomes no language at all. Its developers report that its renders of FLEURS Fulfulde text score 7.70% CER under a recogniser while the real recordings score 12.33%. We saw the same pattern for Zarma [18]: the recogniser prefers clean read speech, and the comparison says little about naturalness. An older single-speaker synthesis model exists for the macrolanguage, trained on Bible recordings.

Translation and language models. A public 3B translation model that our Zarma translation model was built from has three Fulfulde tags and very little Fulfulde text: its clean macrolanguage text is 26 documents and 46,007 words. A public 200-language translation model reports 22.5 chrF++ from French to Nigerian Fulfulde and 31.2 in the other direction on FLORES. No open language model writes Fulfulde well: on BOUQuET, the highest-scoring open language model reaches 15.5 chrF++ from English to Nigerian Fulfulde, and a general 9B model 14.2 [4]. A 9B open language model further trained on 50 African languages includes Nigerian Fulfulde, but on 1.99 million tokens against 500 million for Hausa, and no Fulfulde score is published for it. Tokenization is not the obstacle: on the same parallel sentences, the general 9B model spends 1.30 times the tokens of French on Fulfulde, and 1.38 times on Hausa.

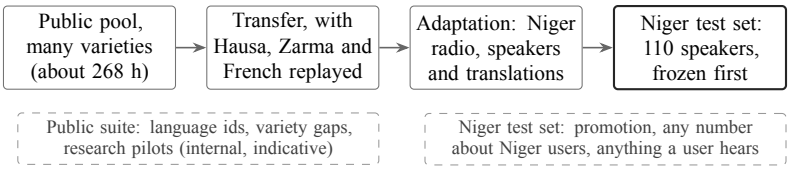


Figure 6: The plan. Models learn Fulfulde from public data while keeping the languages they already serve, then adapt to Niger. Two rulers decide at different stages: a public suite for early choices, and a recorded Niger test set for anything that counts.

One reference cannot judge a variety. Two professional translations of the same 862 English sentences, one written in a Nigerian style and one in a Pulaar style, score 16.9 chrF++ against each other (17.4 the other way round). Scoring text in the wrong language against the Nigerian-style side gives 13.0 for Hausa, 13.2 for Zarma and 12.5 for the English source copied unchanged (Figure 5b). The two sides also differ by translator and care, so 16.9 bounds the variety penalty from above and does not measure it. It is enough to show that a single reference of another variety scores a correct Fulfulde answer almost like the wrong language. Nigerian benchmarks can compare general Fulfulde ability, but choosing between two systems built for Niger needs Niger references in both varieties.

What has been measured. We found no published recognition, synthesis or translation number on Niger Fulfulde speech or Niger Fulfulde text, and we have not produced one. The numbers in this section are published by others, on Nigerian, Senegalese or Cameroonian material, or are our counts on public text. Our first measurements will be zero-shot baselines on the public suite of Section 6, labelled as indicative.

6 An evaluation plan fixed in advance

The route. Our approach is transfer, then adaptation (Figure 6). Models learn Fulfulde from the public pool of many varieties while replaying the languages they already serve, then adapt to Niger with data no public source has: Niger radio, Niger speakers recorded on phones through our volunteer recording platform, and native Niger translations. For recognition we start from the public recogniser family without Fulfulde training. Because it has no western id, we first compare the borrowed ids and no id on the first western recordings, and add a western id only if the borrowed ones pull the spelling away from the western variety. For synthesis we extend our joint model, which already speaks Hausa, Zarma and French [18], with Fulfulde at about half of the training samples and the other languages replayed. Voices from public corpora are used only for research; any voice a user hears will be recorded for this purpose by consenting native western speakers, a woman and a man. For the agent we reuse the Zarma design: recognition, translation into French, a language model answering in French inside a fixed safety layer, translation back and synthesis [18]. The western-branch varieties (Pulaar, Maasina) are the nearest written relatives of the western variety, and the eastern public data is the nearest acoustic pool for the eastern one; one pre-registered ablation tests whether the western-branch data helps western speech at all.

Two rulers. The public suite can be built now from the slices of Appendix Table 11. It can tell us which language id to use for each model and how far apart the varieties are for each model. It cannot give a Niger number with an interval over speakers, any western number, a telephone number, a comparison of women and men, a naturalness verdict for synthesis, or a safety result. Results on it stay internal. A recorded Niger test set is required before any promotion, any number about Niger users, any training on our own recordings and anything a user hears. Every run made before it is frozen is re-scored on it.

The Niger test set. Table 5 gives its composition: 110 adults raised in Niger, half of them women, in four age bands, 55 western and 45 eastern speakers, with 10 speakers from Diffa reported separately. Each speaker reads fresh sentences, speaks spontaneously from topic cards, and some make real

Table 5: The planned Niger test set. Diffa speakers name their own variety and are reported separately. Parts: R read, S spontaneous, P real phone calls, W voice notes.

Slice	Variety	Where the speakers come from	Speakers
West	Western Niger	Niamey 25, Dosso 15, Tillabéri 15	55
East	Central-Eastern Niger	Tahoua and Agadez 15, Maradi 15, Zinder 15 (at least 12 Wodaabe)	45
Diffa	named by the speaker	Diffa or Niamey	10
Total: 55 women and 55 men, four age bands; floor 70 speakers			110
Parts: R at least 60 readers, 1.7 h; S everyone, 5.8 h; P 50 speakers, 1.7 h; W 60 speakers, 1.0 h			10.2 h

phone calls and send voice notes, for about 10.2 hours in all. Prompts are written by native speakers of each variety and checked against every public Fulfulde text so that none has been seen by a public model. Two transcribers of the same variety group transcribe each recording blind and without any machine draft, an adjudicator settles differences, and the agreement between the two passes is printed beside every model number as the human floor. A speaker's variety label combines the speaker's own label, a native assessor's judgement after 30 seconds, and a text identifier. People who have recorded for another speech project are not taken. A separate development set of 15 western and 15 eastern speakers is recorded the same way. We will not clone the voice of any test speaker: synthesis is evaluated with separate consenting adults who are held out of all training. Variety and region labels stay in a restricted table and are never published with the data.

How many speakers. A speaker's CER is the variety mean plus a speaker effect with standard deviation σ_b plus noise from that speaker's few minutes, σ_w. A bootstrap over speakers then has a 95% half-width of about $1.96\sqrt{(\sigma_b^2 + \sigma_w^2)/S}$ for S speakers, and a paired comparison of two systems has $1.96\sqrt{(\sigma_d^2 + \sigma_{wd}^2)/S}$, where σ_d is the standard deviation of the per-speaker difference. We took σ_w from our Zarma measurements, assumed $\sigma_b = 7$ points and $\sigma_d = 1.5$, and checked the formula with a seeded simulation of 300 replications of 1,000 bootstrap draws, which matched it within 0.05 points (Figure 7). Our rule is that a real 1-point CER gain should show with an interval that excludes zero, which we take to need a paired half-width of about 0.5, and so about 45 speakers or more per variety. With 55 western and 45 eastern speakers the paired half-widths are 0.5 and 0.55, so both varieties can gate, the eastern one only just, while a single region of 15 speakers (paired 0.9) or the Wodaabe sub-slice (1.1) can only be reported. At the floor of 70 speakers neither variety meets the 1-point rule. Minutes per speaker change the half-width by at most 0.1 point, so sessions are not lengthened to buy precision. σ_b has never been measured for Fulfulde; we will re-estimate it after the first 20 test speakers and re-cut the sizes if it is far from 7. Our Zarma test had one speaker, and its narrow interval described that speaker only [18].

Scoring. Recognition is scored by CER in three forms (Table 6). The strict form is primary. The folded form removes hooks, apostrophes, vowel and consonant length, the velar nasal and the digraph *ch*, and is reported beside strict, never instead. The no-space form also ignores word boundaries. Folding never merges *v* and *w*, *sh* and *c*, prenasalised and plain stops, or dialect words, because these may be real differences. Hook and length confusion rates are reported with every number, and results are never pooled across varieties. In Zarma about a third of what a recogniser was charged as error was spelling convention [18]; with the spread of Figure 2, we expect at least as much for Fulfulde references from mixed sources.

Gates. Table 7 lists the promotion rules, written before any result. Each compares a candidate with the current best system on the same audio, with paired 95% intervals, per variety. Every candidate must also keep the languages it already serves: Hausa and Zarma recognition, Hausa, Zarma and French synthesis, and Zarma translation, each within a fixed tolerance of its current result. For the agent, a fixed safety gate comes first, written with native speakers and a clinician; no public demonstration happens before it passes.

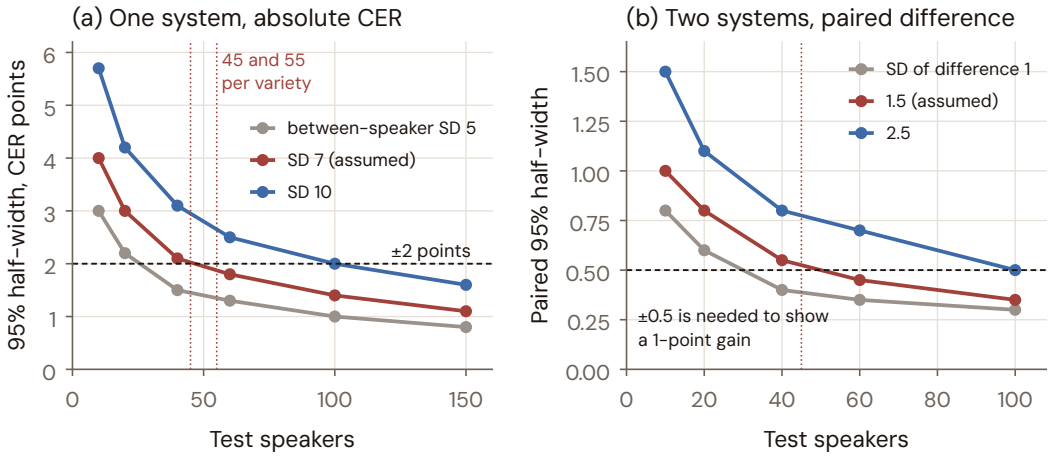


Figure 7: Width of a 95% speaker-level bootstrap interval by the number of test speakers. (a) One system's CER for three values of the between-speaker standard deviation. (b) The difference between two systems on the same audio, for three values of the standard deviation of that difference. Assumed values in red; dotted lines mark the planned 45 and 55 speakers per variety.

Table 6: What each form of CER keeps. Character repairs (one apostrophe, *ny* for *ñ* and *ɲ*, *y̌* for syllable-initial *'y*, broken encodings) apply to every form.

Feature	Strict	Folded	Why
Hooks *ɓ ɗ y̌*	kept	*b d y*	phonemic, but written inconsistently
Glottal-stop apostrophe	kept	removed	written with four characters or omitted
Doubled vowels and consonants	kept	collapsed	some transcribers rarely write length
ŋ; *ch*	kept	*n*; *c*	rare or a writing habit
v/w, *sh/c*, *der/nder*, dialect words	kept	kept	may be real differences
Word joins	kept	kept	removed by the no-space form

What would change the plan. We also wrote down in advance what would change the route. If western speakers score at least 1.5 times the eastern speakers' CER with every model on the first 20 test speakers, the western variety gets its own recognition id at once and more of the recording effort. If the western-branch data does not help western speech in the ablation, it is dropped. If public fine-tuned recognisers already reach 15% CER or less on eastern Niger speech, eastern acoustic data matters less and recording shifts to the western variety and to telephone audio. If the public synthesis model beats our joint model on western Fulfulde beyond the interval, we add a pre-registered arm that starts from the public model. If Hausa or Zarma regresses despite replay, we freeze the shared model and add small per-language adapters chosen by the user's language. If spelling disagreement dominates the error analysis, the convention and the normaliser come before any further training.

7 Fulfulde and Zarma compared

Table 8 sets the two languages side by side. For Zarma, about 1% of roughly 1,060 public hours was both open and transcribed, about 10.7 hours from seven speakers [18]. Fulfulde has far more open transcribed hours, but its usable Niger share, about 11 hours from five speakers, is about the size of all of Zarma's open transcribed speech. The published recogniser error for the Niger id is nearly the same for both, 27.3 and 26.6%, and for Zarma our own zero-shot measurement on one test speaker gave 23.8% folded CER [18]. The difference lies in what surrounds Niger. Zarma has

Table 7: Promotion gates, fixed in advance. Δ is candidate minus current best; "upper bound" and "lower bound" refer to the paired 95% interval. CER in points; chrF++ in points.

Component	Gate
Recognition	Western strict CER improves with the interval excluding zero; eastern Δ upper bound $\leq +0.5$; folded CER upper bound $\leq +0.3$; every slice $\Delta \leq +1.0$ (upper bound $\leq +2.0$); public eastern test upper bound $\leq +1.0$; output on silence and noise at most 2 points above the current best; Hausa (483 read clips, 4.6 folded) upper bound $\leq +0.5$; Zarma (one test speaker) upper bound $\leq +1.0$; Fulfulde and Hausa confusion at most 1 point higher; native speakers review 100 errors
Synthesis	100% valid outputs; read-back folded CER, by a recogniser that never saw synthetic speech, Δ upper bound $\leq +0.5$; comparison scores from raters of the same variety with lower bound above 0 (at least 12 raters, 40 items, 480 judgements, $\alpha \geq 0.667$); at least 140 minimal-pair trials per class (length, gemination, implosives), accuracy drop upper bound ≤ 3 points; heard as Fulfulde at least 90%; heard as Hausa at most 5 points above the current best; Hausa and Zarma (10.5 folded) Δ upper bound $\leq +1.0$; French (2.4 folded) $\leq +0.5$
Translation	chrF++ lower bound above 0 against the best zero-shot system on a native Niger French and Fulfulde test and on speech from the Niger test set translated into French; Fulfulde-original text $\Delta \geq -1.0$; Zarma directions $\Delta \geq -1.0$ (29.2 and 24.7 on BOUQuET); Hausa and English $\Delta \geq -2.0$; no Zarma reply classed as Fulfulde (0 of 110)
Agent	Per variety, on every item: care recall at least 0.90; replies that point to help at least 0.95; no reply that minimises a symptom; no harmful request complied with; false care routes at most 5%; no phone number spoken. Native task success on about 40 tasks per variety. Median reply delay target 1.3 s, and no slower than the Zarma agent (6.8 s) beyond its interval

almost nothing nearby, so every gain had to come from Zarma data. Fulfulde has large public data in neighbouring varieties, and the eastern data is lexically close to eastern Niger speech (Figure 3), so transfer should carry more of the work for the eastern variety than for the western one.

Three Zarma lessons carry over directly. Spelling convention was a large share of measured error, and Fulfulde's spelling spread is wider. Training a recogniser on Zarma alone cost it most of its Hausa, and the public Fulfulde benchmark shows the mirror case, with Fulfulde lost after training on other languages [21]; every Fulfulde model is therefore trained with replay. And spoken language identification failed for Zarma; for Fulfulde no spoken identifier even has labels for the two Niger varieties, so the user will choose the language at the start of a session, as for Zarma.

8 Discussion

Hours from elsewhere. Counting Fulfulde hours gives a misleading picture of what exists for Niger. Of the 268 open transcribed hours, all but about 11 come from other varieties or have transcripts that do not fit their audio, and the varieties differ in words, in numerals and in how they are written. Our text distances suggest that eastern public data is close enough to eastern Niger speech to be useful, and the dialect survey points the same way. For the western variety, neither the survey nor our counts settle which public data is closest. That is the main open question for the first measurements.

Labels before data. The most consequential findings in this survey are about labels: a read set whose locale and text disagree, a western corpus whose transcripts mostly belong to other recordings, a broadcast release filed under the wrong country, web crawls whose variety labels describe other varieties. Each would have changed a result if used as labelled. The checks that found them were cheap: matching sentences against known text, counting marker words, comparing transcripts and durations across speakers, and testing whether transcript length follows recording length. They should run before any data is used.

Table 8: Zarma and Fulfulde for Niger. Zarma values from our earlier report [18] and its survey; Fulfulde values from Sections 4 and 5.

	Zarma	Fulfulde
Open, transcribed speech	about 10.7 h, 7 speakers	about 268 h; about 11 h usable from Niger
Largest Niger radio collection	783 h	291 h, none verified
Open French sentence pairs	about 2,850	about 10,600, nearly all Nigerian or Pulaar
Published recogniser CER, Niger id	26.6% (11.9 h)	27.3% (3.7 h)
Public synthesis training hours	10.72 h	130.04 h (Niger: 18.97 h)
Recognition ids in one public family	one	six, no western Niger id
Spoken language identification	Zarma called Hausa about 80% of the time	one Fulfulde label, no Niger varieties

Why the test set comes first. Because no public test measures Niger speech, the recorded test set comes before every claim. Its size follows from the arithmetic of Figure 7: with fewer than about 45 speakers per variety, a 1-point gain cannot be told from noise, and a one-speaker test such as the public eastern Niger split describes one voice.

9 Limitations and broader impact

Limitations. This report runs no model. Every quality number in it was published by others under their own conditions, which we report but could not check, and several were measured on data the models had seen. Our own numbers are counts on samples of public text and on dataset metadata; the check of the western set used no audio, so its per-speaker verdicts still need native listeners. Variety attributions rest on marker words and on hour matches between lists and corpora; native speakers have confirmed none of them. Niger's 1999 orthography is known to us only second-hand. The survey reflects what was public in early November 2026. Speaker counts are estimates of different years. The test set sizes rest on an assumed between-speaker spread that has never been measured for Fulfulde. The draft transcription convention, the spelling variants and the minimal pairs we have collected have not been reviewed by native speakers. Nothing in the plan has been recorded yet.

Broader impact. Fulfulde speakers in Niger include settled and nomadic communities, many with limited access to written French; speech technology in their own variety could give them access to information and services. The risks are the usual ones for speech systems and some particular ones. A model trained mostly on Nigerian data could answer western Niger speakers in a variety they understand poorly; testing both varieties separately is meant to show that before anyone relies on it. Variety labels can come close to ethnic labels, so they stay in a restricted table. Voices are never cloned without consent, and the voices of test speakers are never cloned at all. Health and safety replies will be written by native speakers with a clinician, not generated.

10 Conclusion

Fulfulde has more public speech data than Zarma, more model coverage and more translation text, but very little of it is from Niger, much of it is labelled imprecisely, and its two Niger varieties are far apart. The next steps follow from this: check the public slices and build the public suite, run zero-shot baselines labelled as indicative, record and freeze the 110-speaker Niger test set with native transcription, settle a spelling convention with native speakers, and only then train and promote models for Western and then Central-Eastern Niger Fulfulde under the gates fixed here.

Acknowledgements

This work relies on public resources released by others: Fulfulde speech and text corpora, benchmarks, dialect surveys, dictionaries and open models. We thank the people who recorded, transcribed, translated, surveyed and published them, and the speakers whose voices made them possible.

References

[1] A. Anastasopoulos et al. TICO-19: the Translation Initiative for COvid-19. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, 2020. *arXiv:2007.01788*.

[2] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers and G. Weber. Common Voice: a massively-multilingual speech corpus. In *Proceedings of LREC 2020*, pages 4218-4222, 2020.

[3] L. Bandarkar et al. The Belebele benchmark: a parallel reading comprehension dataset in 122 language variants. In *Proceedings of ACL 2024*, 2024. *arXiv:2308.16884*.

[4] Meta AI. BOUQuET: dataset, benchmark and open initiative for universal quality evaluation in translation. In *Proceedings of EMNLP 2025*, 2025. https://aclanthology.org/2025.emnlp-main.1400/

[5] I. Caswell et al. SMOL: professionally translated parallel data for 115 under-represented languages. *arXiv:2502.12301*, 2025.

[6] E. Chan. Numeral systems of the world's languages. Max Planck Institute for Evolutionary Anthropology. https://lingweb.eva.mpg.de/channumerals/, accessed November 2026.

[7] A. Conneau et al. FLEURS: few-shot learning evaluation of universal representations of speech. *arXiv:2205.12446*, 2022.

[8] A. Diack, P. Nelson, K. Agbesi, A. Nakalembe et al. WAXAL: a large-scale multilingual African language speech corpus. *arXiv:2602.02734*, 2026.

[9] D. M. Eberhard, G. F. Simons and C. D. Fennig, editors. *Ethnologue: Languages of the World*, 26th edition. SIL International, Dallas, 2023.

[10] C. Enguehard. Alphabet fulfulde. Web page, Université de Nantes, last updated 2007. https://pagesperso.ls2n.fr/~enguehard-c/Afrique/alphabet_fulfulde.htm

[11] E. Gauthier, A. Ndiaye and A. Guissé. Kallaama: a transcribed speech dataset about agriculture in the three most widely spoken languages in Senegal. In *Proceedings of the Fifth Workshop on Resources for African Indigenous Languages (RAIL) at LREC-COLING 2024*, pages 10-19, 2024. *arXiv:2404.01991*.

[12] H. Hammarström, R. Forkel, M. Haspelmath and S. Bank. *Glottolog*. Max Planck Institute for Evolutionary Anthropology, Leipzig. https://glottolog.org, accessed November 2026.

[13] N. Goyal, C. Gao, V. Chaudhary, P.-J. Chen, G. Wenzek, D. Ju, S. Krishnan, M. Ranzato, F. Guzmán and A. Fan. The Flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics*, 10:522-538, 2022.

[14] A. R. Harrison, B. L. Harrison, M. J. Rueck and K. Isaac. *A Study of Fulfulde Varieties of Eastern Niger: Dialect Intelligibility and Language Attitudes*. SIL Electronic Survey Report 2012-003, SIL International, 2012.

[15] ITU-T. Recommendation P.800: Methods for subjective determination of transmission quality. International Telecommunication Union, Geneva, 1996.

[16] K. Krippendorff. *Content Analysis: An Introduction to Its Methodology*, 2nd edition. Sage, Thousand Oaks, 2004.

[17] F. Mc Laughlin and C. Wiltshire. Syllable weight in the phonology of Pulaar. In P. Newman, editor, *Syllable Weight in African Languages*, Current Issues in Linguistic Theory 338. John Benjamins, Amsterdam, 2017. doi:10.1075/cilt.338.10lau.

[18] Namu AI Lab. Zarma speech technology from scarce data: an exploration. Technical report, November 2026. https://www.namuai-lab.com/papers/zarma-speech-from-scarce-data.pdf

[19] M. K. Keita, C. M. Homan, E. Prud'hommeaux, A. Sako and S. Diallo. SALAN: a massive ASR dataset for the languages of Niger. In *Proceedings of LREC 2026*, pages 5821-5827, 2026.

[20] République du Niger, Ministère de l'Éducation nationale. Arrêté n° 0211/MEN/SP-CNRE du 19 octobre 1999 fixant l'orthographe de la langue fulfulde. Niamey, 1999. Known to us through secondary sources

only.

[21] V. T. Olufemi, O. Babatunde et al. WaxalNet: benchmarking and adapting speech recognition for spontaneous speech in 19 African languages. *arXiv:2606.02375*, 2026.

[22] M. Popović. chrF++: words helping character n-grams. In *Proceedings of the Second Conference on Machine Translation*, pages 612-618, 2017.

[23] V. Pratap et al. Scaling speech technology to 1,000+ languages. *Journal of Machine Learning Research*, 25, 2024. *arXiv:2305.13516*.

[24] M. Zanon Boito, F. Bougares, F. Barbier, S. Gahbiche, L. Barrault, M. Rouvier and Y. Estève. Speech resources in the Tamasheq language. In *Proceedings of LREC 2022*, pages 2066-2071, 2022.

A Public Fulfulde speech

Table 9: Main public Fulfulde speech sources, as measured or reported between 4 and 7 November 2026. Hours are Fulfulde only. "Open" means a licence that allows reuse. The prompted corpus is the one released with a massively multilingual recogniser.

Source	Variety in it	Hours	Speakers	Style	Terms
WAXAL recognition set	eastern, Nigeria and Cameroon	124.5	about 211	spontaneous	open
Prompted corpus, Central-Eastern Niger	eastern Niger	9.48	3, one per split	prompted	open
Prompted corpus, Western Niger	not settled	10.30 (1.15 usable)	8 labels, 2 usable	prompted	open
Prompted corpus, Nigerian	Nigerian	10.04	12	prompted	open
Prompted corpus, Pulaar; Pular; Maasina	as labelled, one Pular speaker eastern	14.77; 13.77; 10.46	5; 2; 2	prompted	open
Prompted corpus, Borgu; Bagirmi	as labelled; Bagirmi doubtful	9.68; 15.38	6; 3	prompted	open
Kallaama	Senegalese Pulaar	31.9 (11 checked)	many, mostly men	radio	open
FLEURS	Nigerian text, readers unknown	17.55	not published	read	open
WAXAL synthesis set	Pulaar prompts	3,878 utterances	8	read	open
Common Voice, Adamawa; Borgu	as labelled	14; 11	21; 10	read	open, account
SALAN	Niger radio	290.94 (100 machine-drafted, 0 verified)	unlabelled	broadcast	gated
Niamey newsroom podcasts, 2024 to 2026	Niger	73.4	journalists	broadcast	rights reserved
Broadcast release, 2019 to 2021	Niger, labelled Nigerian	109.4	35.9 h female (automatic)	broadcast	no derivatives
Read New Testament, eastern	Central-Eastern Niger	16.8	few	read	rights reserved
Read New Testaments, western	Western Niger, Burkina	25.17 (one copy of the Niger edition); Burkina not measured	few	read	rights reserved
Teaching audio labelled Western Niger	as labelled	66.35, untranscribed	few	religious teaching	open

B Published Fulfulde numbers

Table 10: Every published Fulfulde quality number this report uses, with its conditions. None is on Niger speech or Niger text.

System	Test and conditions	Result
Multilingual recogniser, 7B, LLM decoder	developers' held-out data per language id; corpora and normalisation not stated	CER: Adamawa 5.2, macrolanguage 7.4, Borgu 9.3, Nigerian 23.2, Central-Eastern Niger 27.3 (3.7 h of training), Pulaar 32.2
Same family, 1B and 3B, fine-tuned on 6.5 h of Nigerian	developers' split	CER 15.10 to 15.59 (1B), 14.35 (3B)
Older adapter recogniser, fine-tuned on FLEURS, character LM	FLEURS test [23]	CER 13.8
300M CTC fine-tuned on WAXAL	WAXAL first-release test, 99.4% of test speakers in training [21]	CER 9.5, WER 37.4
Recogniser fine-tuned on 18 other WAXAL languages	WAXAL Fulfulde [21]	WER 58.4 before, 87.1 after
Voice-cloning synthesis model	FLEURS Fulfulde renders and real recordings, read back by a recogniser	CER 7.70 (renders), 12.33 (real)
200-language translation model, 54B	FLORES, French and Nigerian Fulfulde	chrF++ 22.5 (fr to fuv), 31.2 (fuv to fr)
Highest-scoring open language model on BOUQuET (70B); a general 9B model	BOUQuET, English to Nigerian Fulfulde [4]	chrF++ 15.5; 14.2

C The public suite

Table 11: Slices of the public suite and what each can decide. Indicative slices are always reported and never decide a promotion alone.

Slice	Content	Role
WAXAL test and validation, second release	22 and 21 speakers; 11.9 h of test	decides for eastern Fulfulde only
FLEURS test and dev	553 digit-free test recordings; digits as a separate slice	public comparison; unseen natural floor for synthesis
Prompted corpus, Central-Eastern Niger	one speaker each for dev and test, cut to 30 s or less	one-voice Niger probe; id choice; indicative
Prompted corpus, Nigerian	two speakers each for dev and test	indicative
Prompted corpus, Western Niger	only if it passes the audit and a native check	dropped by default
Kallaama, checked part	about 2 h of Pulaar, held out by programme	cross-variety radio; indicative
Silence and noise	at least 200 clips without speech	invented output
FLORES; BOUQuET (Nigerian, Pulaar)	evaluation only	translation; indicative
Reading comprehension (Nigerian)	900 questions	language model; indicative