ENLIVEN-1000: A Comprehensive Revitalization Framework for 1000+ Endangered Languages via Broad-Coverage LID and LLM-Augmented MT

Anonymous submission

Abstract

We present **ENLIVEN-1000**, a unified framework for endangered and low-resource language revitalization that integrates broad-coverage language identification (LID), machine translation (MT), and LLM-generated synthetic data—aimed at expanding safe, equitable NLP support for communities historically excluded from mainstream tools. We compile a text corpus for 1154 languages (1069 endangered or low resource) from public sources and train a fastText-based LID model covering this vast set. The LID system achieves high detection quality with F1 $\approx$ 0.99 and FPR $\approx 3\times10^{-6}$, substantially broadening reliable coverage beyond existing solutions. Focusing on five diverse endangered languages—Carpathian Romani, Chuj, Sunwar, Kapingamarangi, and Inuktitut—we fine-tune a 600M-parameter NLLB-200 model for translation. Our fine-tuned models outperform zero-shot baselines and even proxy models trained on related, high-resource languages, in both directions (endangered$\rightarrow$English and English$\rightarrow$endangered). We further use GPT-4o to generate synthetic parallel data, demonstrating that augmenting limited real data with LLM-generated text yields substantial MT improvements. These results illustrate a practical path toward scaling NLP support to hundreds of under-resourced languages. We discuss implications for language revitalization and ethical considerations in working with endangered language communities.

1 Introduction

Thousands of human languages are endangered, with many at risk of extinction by the end of this century (Daily Sabah 2023; UN DESA 2022). Out of over 7,000 living languages, roughly 40–50% are already endangered (Ethnologue 2023). UNESCO predicts that without intervention, we may lose half of today's languages (or more) by 2100 (Wikipedia 2025). The loss of a language signifies the disappearance of unique cultural knowledge and identity. Recognizing this crisis, global efforts have increased: UNESCO launched an *Atlas of the World's Languages in Danger* and designated 2022–2032 as the *International Decade of Indigenous Languages* to mobilize revitalization efforts (UN DESA 2022). Collaborative platforms like the Endangered Languages Project (ELP) also connect communities and linguists in documenting and sharing threatened languages.

However, most endangered languages remain underserved by language technology. Fewer than 100 languages have robust support in widely used tools (e.g., web translators or voice assistants). Even large multilingual models and services cover at best a few hundred languages, leaving the vast majority of endangered languages beyond their scope. This technological gap poses a barrier to revitalization: digital tools (such as text analysis, translation, and educational software) could greatly aid communities, but those tools require data and models for each language.

In this paper, we introduce **ENLIVEN-1000**, a framework for delivering NLP support to 1069 endangered and low-resource languages through three core components: (i) a broad-coverage language identification (LID) system that reliably recognizes and disambiguates more than a thousand languages; (ii) machine translation (MT) systems for those languages; and (iii) a large language model (LLM)–driven synthetic data pipeline that mitigates data scarcity. Together, identification, translation, and augmentation form an end-to-end workflow for corpus creation, information access, and cross-lingual communication in low-resource settings.

Our approach is inspired by recent advances in massively multilingual NLP. On the LID side, reliable identification is "paramount" for building trustworthy datasets (Caswell and Ng 2020). We curate **1,174** languages in total from diverse sources; of these, **1,154** have sufficient and usable data for training and evaluation, which we use to train a fastText-based classifier (Joulin et al. 2017). The remaining languages are included only as candidate labels at inference time to mitigate "out-of-model cousin" errors, and we therefore omit per-language F1/FPR for them (Caswell and Ng 2020; Kreutzer et al. 2022a). On the MT side, we leverage the No Language Left Behind NLLB-200 model from Meta AI (NLLB Team et al. 2022) and fine-tune it on small language-specific corpora. To expand coverage, we generate synthetic parallel data with GPT-4o using few-shot prompts augmented by bilingual lexicons (OpenAI 2023; Yong, Menghini, and Bach 2024), consistent with evidence that synthetic data improves MT quality (Sennrich, Haddow, and Birch 2016; de Gibert et al. 2025).

Contributions.

- **Corpus & LID:** A multilingual corpus spanning 1154 languages and a high-precision fastText-based LID model.
- **MT systems:** Fine-tuned NLLB-200 models for five en-

dangered languages, outperforming zero-shot and proxy baselines.

- **Data scaling case study (Inuktitut):** A controlled scaling study at 31k/62k/93k/124k sentence pairs demonstrating consistent improvements in both directions, underscoring the value of additional real bitext.
- **Synthetic data:** A lexicon-guided GPT-4o few-shot prompting pipeline that boosts MT under data scarcity.
- **Evaluation across data regimes:** We evaluate MT models trained on real-only, synthetic-only, and mixed training regimes, showing that synthetic data can boost performance, especially for the weaker language.
- **Scaling & ethics:** An analysis of how to scale to all 1000+ languages, with guidance for community-centered, ethical deployment.

2 Language Identification (LID)

A first step toward inclusivity in NLP is robust language identification (LID) for low-resource languages. Effective LID allows us to discover, filter, and curate text in target languages from the web or archives, which is crucial for building corpora. Unfortunately, most existing LID systems have limited coverage—many support only a few hundred languages or fewer (Google Research 2016; NLLB Team et al. 2022; Thoma 2018). For example, Google's compact language detector (CLD3) covers roughly $\sim$100 languages (Google Research 2016); fastText-LID and the NLLB LID model target $\sim$176 and $\sim$200 languages, respectively (Facebook AI Research 2022; NLLB Team et al. 2022). This leaves thousands of languages "in the dark," including the majority of endangered languages.

Endangerment Status. Because our goal is revitalization for underserved languages, we adopt an inclusive policy on endangerment. Using standard sources (Ethnologue, Glottolog, Endangered Languages Project, UNESCO), and mapping to the Expanded Graded Intergenerational Disruption Scale (EGIDS), we include languages rated **EGIDS 5–9** (Developing $\rightarrow$ Critically Endangered) as well as **EGIDS 10** (Extinct). By contrast, **EGIDS 0–4** (International $\rightarrow$ Educational) denote *institutional* support via schooling and broad functional use; we treat these as *institutional/non-endangered*.

Out-of-Model Cousin Errors. A pervasive failure mode in LID is the *out-of-model cousin* error: when a true language is *absent* from the label set, the classifier systematically assigns it to a *closely related* ("cousin") language that *is* in the model, inflating false positives and contaminating downstream corpora (Caswell and Ng 2020; Kreutzer et al. 2022b). To mitigate this, we emphasize broad coverage during inference. Concretely, we consider **1,174** languages in total; a small subset could not be processed reliably or lacked sufficient data for training/evaluation. We nevertheless *retain* these languages as candidate labels at inference time to reduce out-of-model cousin confusions, but we *do not* report per-language F1/FPR for them. This design follows best practice observed in prior large-scale audits: keep the label space wide at inference to avoid systematic misclassification, while restricting quantitative reporting to languages with adequate evidence (Caswell and Ng 2020; Kreutzer et al. 2022b).

Data Compilation. From our inventory, **1,154** languages have sufficient data for training/evaluation; among these, **85** are institutional (EGIDS 0–4) and **1,069** are endangered/low-resource (EGIDS 5–10). We therefore define two evaluation sets: **Full-LID** (*all 1,154* usable languages) and **Endangered-LID** (*the 1,069* endangered/low-resource languages). Unless otherwise noted, metrics (macro-F1, FPR) are reported on these splits; the remaining (non-evaluable) languages are included only as inference-time labels to guard against out-of-model cousin errors. Text data were aggregated from a variety of open sources. A major source was multilingual Bible translations (e.g., the Parallel Bible Corpus), which provide aligned verses in hundreds of languages, often including very low-resource ones (Mayer and Cysouw 2014). While such translations can be stylistically formal or archaic, they offer a starting point for many languages with otherwise scant digital presence. We also used Tatoeba (a crowd-sourced collection of translated sentences) and other public corpora such as UDHR translations and web-scraped datasets to enrich coverage (Tiedemann 2020; Unicode Consortium 2007; Ortiz Suárez, Romary, and Sagot 2020; Wenzek, Lacroix, and Bennoit 2020). Each text line is tagged with its ISO 639-3 code; we verified labels where possible, since mislabeling degrades LID performance (Caswell and Ng 2020).

LID Model. We adopt a supervised `fastText` classifier (Joulin et al. 2017) with character n-grams for short-text LID, an approach that scales well to thousands of labels while remaining fast and lightweight (Facebook AI Research 2022; Kargaran et al. 2023). In line with prior low-resource LID work, we favor fastText because it (i) is open-source and easy to deploy across platforms; (ii) delivers calibrated confidence scores that enable thresholded filtering (high-precision corpus creation); (iii) handles broad label sets efficiently, which helps curb *out-of-model cousin* errors when we keep a wide candidate set at inference; and (iv) supports granularity control (e.g., aggregating probabilities for macrolanguages vs. varieties) (Kargaran et al. 2023). Our pipeline (full details in the Appendix) performs Unicode normalization, cross-/within-language de-duplication, stratified splitting (85/5/10 with caps), light temperature upsampling for low-resource classes, and training with standard fastText hyperparameters, followed by model quantization for compact deployment.

Evaluation. We report per-language and macro-averaged Precision, Recall, F1, and false positive rate (FPR) under a thresholded decision rule: predict the top label if $\max_l P(l \mid s) \geq \theta$, otherwise output UNDETERMINED. By varying the confidence threshold θ, we control the precision–recall tradeoff: higher θ yields fewer but more reliable labels (higher precision, lower recall), while lower θ yields more coverage at the cost of additional false positives (higher recall, lower precision). This protocol emphasizes high-precision filtering for corpus construction while

Metric	Full (1,128) θ=0.3	Full (1,128) θ=0.5	Endangered (1,044) θ=0.3	Endangered (1,044) θ=0.5
Precision	0.9935	0.9942	0.9909	0.9919
Recall	0.9931	0.9924	0.9910	0.9904
F1	0.9931	0.9931	0.9908	0.9910
FPR↓	0.000004	0.000003	0.000004	0.000004

Table 1: Macro-averaged LID performance at two confidence thresholds. Metrics are computed over languages present in each test split.

improving recall on low-resource classes via light upsampling and calibrated thresholds.

Key Results. As shown in Table 1, the model attains near-perfect macro precision/recall/F1 on both **Full-LID** and **Endangered-LID**. Raising the decision threshold from θ=0.3 to 0.5 yields a modest precision gain and pushes false positives down to only a few per million negatives, with essentially unchanged F1 and a slight drop in recall. This stable trade-off across both evaluation sets lets us operate in a high-precision regime for large-scale corpus filtering while preserving strong recall for low-resource languages. Our LID model greatly outperforms prior open LID systems on low-resource languages (Kargaran et al. 2023; Thoma 2018).

Granularity and Coverage. Some languages in our set are extremely closely related or mutually intelligible dialects, which can challenge fine-grained identification (e.g., closely related Zapotec varieties); occasional confusions are expected and consistent with prior analyses (Caswell and Ng 2020). Nonetheless, the broad coverage and near-perfect performance of our LID (1154 languages) far exceeds that of widely used detectors (Google Research 2016; Facebook AI Research 2022; NLLB Team et al. 2022). In practice, this “opens the pipeline” for endangered languages: one can reliably detect and isolate text in a given endangered language from large multilingual collections. As highlighted by large-scale corpus work (Wenzek, Lacroix, and Bennoit 2020; Ortiz Suárez, Romary, and Sagot 2020; Kreutzer et al. 2022b), robust LID and filtering are essential for high-quality datasets in low-resource NLP.

3 Machine Translation (MT)

With a strong LID in hand, we turn to the challenge of machine translation (MT) for endangered languages. MT can facilitate many revitalization activities: it helps speakers translate materials (stories, educational content) into or out of their language, and can assist linguists in deciphering texts. However, building MT systems for endangered languages is daunting due to scarce parallel data. We explore a solution by leveraging a pre-trained multilingual MT model (Meta AI’s NLLB-200) and fine-tuning it on small, language-specific datasets—a strategy shown to be effective for low-resource MT via multilingual pretraining and transfer learning (NLLB Team et al. 2022; Aharoni, Johnson, and Firat 2019; Arivazhagan et al. 2019; Zoph et al. 2016; Neubig and Hu 2018). We focus on five languages as case studies, chosen for diversity in geography, family, and script:

- **Carpathian Romani** (ISO 639-3: `rmc`)—An Indo-European (Romani) language spoken in Central/Eastern Europe. Tt uses Latin-based orthographies that vary by country. It is classified as ”At risk” by The Catalogue of Endangered Languages (ElCat) (Campbell et al. 2022).
- **Chuj** (`cac`)—A Mayan language of Guatemala and Mexico. It is spoken mainly in Guatemala’s Huehuetenango highlands—with communities also in nearby Chiapas, Mexico. It is classified as ”Vulnerable” by ElCat.
- **Sunwar** (`suz`)—A Sino-Tibetan language of eastern Nepal (Kiranti branch). It is classified as ”Threatened” by ElCat. Written in Devanagari.
- **Kapingamarangi** (`kpg`)—A Polynesian outlier language spoken on Kapingamarangi atoll (FSM). It is classified as ”Threatened” by ElCat. It represents the Austronesian family and poses interesting challenges due to extremely limited textual data.
- **Inuktitut** (Eastern Canadian Inuktitut, `iku`)—An Eskimo–Aleut language spoken in Arctic Canada. It is classified as ”Vulnerable” by ElCat. It uses Canadian Aboriginal Syllabics; we include it because more data are available (e.g., government documents), enabling scaling experiments.

Corpora. For each language, we obtained a small parallel corpus of ∼31,000 sentence pairs (language↔English). Sources included aligned segments from Bible translations (Mayer and Cysouw 2014), community-preserved material from OPUS (Tiedemann 2012), and public government translations for Inuktitut (Joanis et al. 2020). All data were cleaned and deduplicated; we standardized scripts where helpful (e.g., transliterating Kapingamarangi to Latin when unsupported), consistent with practices in multilingual NMT (Aharoni, Johnson, and Firat 2019). We held out 1,000 sentences per language for validation/testing, using the remaining ∼30k for training.

Model and Fine-Tuning. We used **NLLB-200** as the base system (NLLB Team et al. 2022), specifically the 600M-parameter variant for a balance of capacity and efficiency. Although NLLB-200 was trained many-to-many over 200 languages, our five targets are all absent from its supported languages. We fine-tuned *separately* for each pair and direction (X→En and En→X). We initialized from NLLB and continued training on the small parallel corpus for a few thousand updates. We used a learning rate of 5×10^{-5} with linear decay and validated frequently to avoid overfitting (early stopping typically within 5–6 epochs). Following multilingual NMT practice, we employed a *target-language tag* to steer generation: for X→English, we prepended `<2eng>` to the source; for English→X, we prepended `<2xxx>` (the target code) to the English source (Johnson et al. 2017; NLLB Team et al. 2022). When NLLB lacked a predefined code (e.g., Sunwar, Kapingamarangi), we used a related code or an unused token as a proxy and updated the model’s dictionary accordingly (Aharoni, Johnson, and Firat 2019).

Code	Language	System (NLLB-200-600M)	BLEU	spBLEU	ChrF	TER ↓	COMET
rmc	Carpathian Romani	**Fine-tuned**	**33.99**	**35.14**	**53.68**	**56.35**	**0.666**
		Zero-shot baseline	0.40	0.39	18.65	127.21	0.271
		Best proxy: `slk_Latn` → eng_Latn	1.25	1.39	18.87	180.79	0.330
		Δ	+32.74	+33.75	+34.81	+70.86	+0.34
cac	Chuj	**Fine-tuned**	**11.42**	**11.92**	**33.54**	**108.47**	**0.559**
		Zero-shot baseline	0.33	0.22	16.87	171.91	0.256
		Best proxy (COMET): `por_Latn` → eng_Latn	0.17	0.23	13.43	260.20	0.292
		Best proxy (ChrF): `grn_Latn` → eng_Latn	0.16	0.13	13.72	139.24	0.237
		Δ	+11.09	+11.69	+16.67	+30.77	+0.27
suz	Sunwar	**Fine-tuned**	**13.90**	**14.69**	**35.26**	**93.45**	**0.645**
		Zero-shot baseline	0.27	0.26	17.26	209.71	0.368
		Best proxy (COMET): `mai_Deva` → eng_Latn	0.43	0.52	13.29	290.85	0.369
		Best proxy (ChrF): `dzo_Tibt` → eng_Latn	0.33	0.33	13.61	223.89	0.349
		Δ	+13.47	+14.17	+18.00	+116.26	+0.28
kpg	Kapingamarangi	**Fine-tuned**	**28.14**	**28.90**	**47.53**	**66.74**	**0.668**
		Zero-shot baseline	0.25	0.38	16.49	169.75	0.285
		Best proxy (COMET): `fij_Latn` → eng_Latn	0.52	0.60	13.38	262.06	0.318
		Best proxy (ChrF): `smo_Latn` → eng_Latn	0.79	0.81	14.62	225.31	0.316
		Δ	+27.35	+28.09	+31.04	+103.01	+0.35
iku	Inuktitut	**Fine-tuned**	**10.92**	**11.25**	**24.10**	130.80	**0.544**
		Zero-shot baseline	1.67	0.15	2.91	104.61	0.414
		Best proxy (COMET): `tur_Latn` → eng_Latn	1.76	0.32	4.06	**101.37**	0.437
		Best proxy (ChrF): `fin_Latn` → eng_Latn	1.54	0.50	5.97	108.23	0.435
		Δ	+9.16	+10.75	+18.13	-29.43	+0.11

Table 2: Endangered → English translation results with NLLB-200-600M (fine-tuned on ∼31k sentence pairs per language). For each language, the bottom Δ row reports the **absolute improvement of the fine-tuned model over the second-best system in that block**, metric by metric.

Evaluation Setup. We report five complementary metrics, each capturing a different facet of translation quality. We report **BLEU** and **spBLEU** computed case-insensitively with consistent tokenization using SacreBLEU (Papineni et al. 2002; Post 2018); spBLEU applies standardized segmentation to reduce tokenizer variance. We include **ChrF** as a character n-gram F-score (Popović 2015), **TER** as the normalized edit operations required to match the reference (Snover et al. 2006), and **COMET** as a reference-based neural quality estimator aligned with human judgments (Rei et al. 2020).

For each language and direction, we compare:

1. **Zero-Shot Baseline**—off-the-shelf NLLB-200 without fine-tuning (for unsupported languages, effectively near-zero).
2. **Best Proxy Fine-Tuned Baseline**—we screen a panel of ∼5 typologically or geographically related high-resource proxies, fine-tune one model per proxy, and report the best dev-set performer. *Example (Sunwar):* Nepali, Central Tibetan, Maithili, Hindi, Dzongkha, Burmese.
3. **ENLIVEN Fine-Tuned**—our model fine-tuned on the true language's ∼31k parallel pairs.

MT Analysis. Table 2 shows **consistent improvements across all metrics** after fine-tuning on ∼31k pairs per language. BLEU and spBLEU jump from near-zero to clearly usable ranges, indicating better n-gram fluency; ChrF rises substantially, reflecting stronger character-level fidelity and robustness to orthographic variation; COMET increases across the board, confirming gains in semantic adequacy and overall quality; and TER **drops sharply**, meaning far fewer edits would be needed in a post-editing workflow. The Δ rows make this explicit: the fine-tuned system outperforms **both** the zero-shot baseline and the best proxy for language and metric besides one, with the largest gains where the off-the-shelf model struggles most.

We also observe a **script effect**: Latin-script languages show the biggest jumps, Devanagari is close behind, and Canadian Aboriginal Syllabics (Inuktitut) still benefits meaningfully—particularly on COMET and ChrF—even when word-overlap metrics (BLEU/spBLEU) rise more modestly. This pattern aligns with expected tokenization and pretraining coverage differences. Overall, a modest amount of language-specific supervision turns an unreliable zero-shot model into a **reliable endangered→English translator** across multiple quality axes, suitable for corpus growth, data mining, and downstream applications.

Translating *into* endangered languages (English→X; see Appendix B for the full table) is harder, and BLEU scores are lower on average than X→English, but fine-tuning still makes the task tractable. We obtain **12.72** BLEU for *English→Romani* (vs. 0.12), **12.47** for *English→Kapingamarangi* (vs. 0.22), **3.25** for *English→Chuj* (vs. 0.02), **2.79** for *English→Sunwar* (vs. 0.01), and **5.00**

Size	BLEU	spBLEU	ChrF	TER↓	COMET
31K	10.92	11.25	24.10	130.80	0.544
62K	11.86	10.57	24.80	**109.79**	0.551
93K	9.04	8.61	24.51	138.79	0.560
124K	**14.79**	**15.41**	**30.36**	123.13	**0.586**

Table 3: Inuktitut→English with NLLB-600M fine-tuned on one language and increasing data (31K–124K).

for *English→Inuktitut* (vs. 0.68). While one metric (spBLEU) for *English→Inuktitut* is slightly lower than baseline—likely an artifact of segmentation—COMET and TER improve substantially (COMET +0.17; TER drops from 439.40 to 97.85).

Overall, across both directions and all five languages, **ENLIVEN** fine-tuning reliably surpasses zero-shot and proxy baselines. A practical takeaway is that collecting on the order of 30k authentic pairs—via community engagement or targeted mining—suffices to produce usable systems when combined with a pretrained multilingual backbone.

Qualitative Analysis. To concretely illustrate the improvements, consider Sunwar: the zero-shot NLLB (leveraging incidental Hindi/Nepali knowledge) yielded mostly garbled outputs for Sunwar, often mixing Nepali words. The fine-tuned model, by contrast, produces fluent Sunwar closely matching reference translations. For a test sentence meaning "You are strangers in the land" the baseline produced unreadable output, whereas the fine-tuned system generated a similarly-correct Sunwar sentence . This transformation from unusable to reasonably accurate translation is replicated across the five case studies.

Case Study for Data Scaling (Inuktitut). We also investigated data scaling using Inuktitut, where more bitext is available. We created training sets of 31k, 62k, 93k, and 124k Inuktitut–English pairs by adding sentences from government and news sources, and fine-tuned models on each. Table 3 shows a clear upward trend: moving from 31k to 124k steadily strengthens performance for English→Inuktitut and significantly improves Inuktitut→English as well (table in Appendix B), consistent with long-standing observations that additional in-domain parallel data improves NMT quality (Koehn and Knowles 2017). We note a small non-monotonic dip at 93k, likely due to mixed-domain additions and a fixed training schedule; performance rebounds at 124k. This suggests that whenever additional real parallel data can be obtained, it should be incorporated. Nevertheless, for most of the 1000+ languages we target, such large corpora do not exist, motivating synthetic data generation in the next section.

4 Synthetic Data Experiments

While real parallel data is the gold standard for MT, it is scarce for endangered languages. A promising solution is to generate synthetic parallel data using a powerful language model, effectively leveraging high-resource language data (like English texts) to create translations in the low-resource language (Sennrich, Haddow, and Birch 2016; de Gibert et al. 2025). Recent studies have shown that large LLMs (such as GPT-4o or Claude) can produce surprisingly good translations for languages where they were not explicitly trained (OpenAI 2023; Anthropic 2024). We investigate this approach for two of our languages: Carpathian Romani (rmc) and Inuktitut (iku), chosen as the best- and worst-performing languages in our earlier MT results. Romani had the highest baseline translation quality after fine-tuning (likely due to its linguistic similarity to other languages the model knows, and possibly simpler morphology), whereas Inuktitut remained the most challenging (due to complex morphology and polysynthesis). By testing on both, we gauge the effect of synthetic data in favorable vs. difficult conditions.

Synthetic Data Generation. We generate 1:1 synthetic parallel corpora for *Carpathian Romani* (rmc) and *Inuktitut* (iku) using a lexicon-guided few-shot prompting strategy (Yong, Menghini, and Bach 2024). Concretely, we start from the English side of our real bitext (31k sentences per language) and construct prompts that interleave (i) nearest-neighbor exemplars retrieved by character n-gram similarity from the seed corpus, and (ii) compact bilingual token/phrase lexicons automatically induced from the same seed (token co-occurrence and shallow alignments). A frontier LLM (GPT-4o) translates each English sentence into the target language under style/orthography constraints. For Inuktitut, we request romanized outputs and apply a deterministic romanization→syllabics mapping post hoc to ensure script conformity.

Quality control and de-duplication. We employ a two-stage filter: (1) a semantic gate using LaBSE cosine similarity on small held-out checks (with retry-based selection) to reject low-adequacy generations, and (2) near-duplicate removal with a token-level Levenshtein similarity threshold to discard trivially paraphrastic or repeated items. We also exclude the current source sentence from few-shot blocks (leak-safe exemplars) and enforce basic well-formedness constraints. The pipeline is implemented with strict rate limiting, timeout-aware retries, and periodic checkpoints to enable safe resume.

Downstream MT. Because reference-free quality estimators are unavailable for rmc and iku in our setting, we evaluate the utility of synthetic data via downstream MT performance rather than standalone reference-free metrics. We trained MT models under four regimes for each language (rmc and iku): (1) Real-only 31k – using the original 31k real parallel sentences (same fine-tuned model from early MT); (2) 100% Synthetic 31k – using only the 31k GPT-4o generated pairs for training; (3) Concat-62k – concatenating the real and synthetic data, giving 62k training pairs which mirrors a realistic deployment pattern; (4) 50/50 @31k – mixing real and synthetic but capping the total at 31k (randomly sampled 15.5k from real and 15.5k from synthetic to keep training data size constant at 31k, to isolate data quality effects). We fine-tuned NLLB-200 models for each setting, using the same hyperparameters as before. We then evaluated

Code	Language	System (NLLB-600M)	BLEU	spBLEU	ChrF	TER ↓	COMET
rmc	Carpathian Romani	Real-only (31K)	**33.99**	35.14	53.68	56.35	0.666
		100% Synthetic (31K)	10.87	14.14	35.62	77.30	0.551
		Concat-62K (31K + 31K)	26.86	**38.61**	**56.38**	**52.73**	**0.683**
		50/50 @31K (real/synth balanced)	10.87	14.14	35.62	77.30	0.551
iku	Inuktitut	Real-only (31K)	**10.92**	**11.25**	**24.10**	130.80	**0.544**
		100% Synthetic (31K)	0.75	0.20	4.25	99.56	0.420
		Concat-62K (31K + 31K)	5.47	10.02	23.49	**97.30**	0.507
		50/50 @31K (real/synth balanced)	4.21	7.07	21.87	102.39	0.491

Table 4: Synthetic augmentation regimes for *rmc*→En and *iku*→En with NLLB-600M fine-tuned per setting. "Real-only" trains on 31K human bitext; "100% Synthetic" uses 31K GPT-4o outputs; "Concat-62K" mixes 31K real + 31K synthetic; "50/50 @31K" balances real/synthetic while keeping the total at 31K. Best per-language scores are bolded.

all models on the same real test set (consisting of genuine translations) and compared metrics for MT.

Results. Table 4 compares real-only, synthetic-only, fixed-budget mixes, and simple concatenation for *rmc* and *iku* (into English). A few themes emerge.

Carpathian Romani (rmc). Synthetic sentences act as strong *augmenters*. Even when trained in isolation they approach the real-only system, and when concatenated with real data they yield the best overall profile: adequacy/fluency improve jointly (ChrF/COMET), while TER drops, indicating fewer edits are needed to reach references. The gains plausibly come from added lexical and constructional coverage that the smaller human set lacks (e.g., rarer inflections and multiword expressions). At a fixed budget (50/50), however, replacing half of the reliable human data with noisier synthetic examples can slightly blunt improvements—suggesting synthetic is most valuable as an *add-on* rather than a substitute for truly low-resource languages.

Inuktitut (iku). Synthetic-only training is weak, and even concatenation produces uneven gains: robustness (TER) improves more consistently than overlap-based or semantic metrics (BLEU/ChrF/COMET). Several factors likely contribute: (i) orthography handling (romanization ↔ syllabics conversion) introduces normalization noise; (ii) domain/style mismatch between GPT outputs and government/news references reduces n-gram overlap (hurting BLEU) even when translations are usable; (iii) morphology and compounding create sparse tokenizations that penalize overlap metrics; and (iv) higher LLM noise for *iku* increases the burden on the model to denoise. In short, synthetic data helps *volume* and robustness, but its intrinsic quality matters more for *iku* than for *rmc*.

Metric-level diagnostics. We observe expected divergences across metrics: BLEU/spBLEU (token/segmented overlap) can undervalue acceptable paraphrases and morphological variants that ChrF tolerates; COMET better tracks adequacy/meaning preservation but may be less calibrated for truly low-resource scripts; TER often captures practical editability gains even when overlap metrics move modestly. Reading these together yields a more faithful picture of data utility than any single metric.

Takeaways. (1) LLM-augmented training is best used as an *add-on*, not a replacement: when synthetic quality is imperfect, concatenating synthetic with real data outperforms fixed-budget mixes that substitute synthetic for human data and can even surpass the real-only 31k setting—primarily because of increased training size. (2) To raise the floor of synthetic quality, invest in script-aware prompting and post-processing (e.g., romanization/syllabics normalization) plus light filtering, and assess impact with multiple metrics (ChrF, COMET, BLEU, TER), which is especially important for morphologically rich or script-divergent languages. (3) Crucially, these gains appear only when the synthetic data is of reasonable quality: in Inuktitut, where synthetic pairs were noisier, concatenating to 62k with synthetic *reduced* scores relative to the real-only 31k model, showing that added scale cannot compensate for low-quality synthetic data and that augmentation should be applied additively only when the synthetic component is sufficiently good. These conclusions resonate partially with de Gibert et al. (2025), who report improvements from LLM-generated data even when noisy (de Gibert et al. 2025).

Future Outlook. If general-purpose LLMs continue to improve in multilingual grounding, script fidelity, and controllability, we foresee a workflow in which our LID→MT→synthesis pipeline becomes a virtuous loop rather than a one-off bootstrap. LID discovers and filters monolingual text; stronger LLMs then translate or co-generate parallel data directly in community-preferred orthographies and dialects, attach uncertainty and style metadata, self-check via reference-free signals (e.g., round-trip consistency, contrastive probes, QE-like scores), and adapt with small preference datasets from speakers. MT fine-tuned on this higher-quality synthetic+real mix harvests even more usable text, further raising LLM promptability and data quality in the next round. Under community consent and safeguards, this trajectory makes the "translator assistant" role ever more realistic: LLMs augment (not replace) human expertise, turning tiny corpora into sustainable pipelines that can extend practical MT coverage to languages that currently have none.

5 Discussion and Implications

Our study demonstrates a practical framework to scale NLP support to hundreds of endangered languages. By combining broad LID for data acquisition with LLM-augmented MT training, we address two key bottlenecks: identifying data in the target language and generating enough training material to learn translations. The success on our five case-study languages suggests that similar methods can be applied to many others. In fact, our LID dataset (1154 languages) and project title "ENLIVEN-1000" indicate the ambition to cover over 1000 endangered languages – nearly an order of magnitude more than the languages supported by current translation systems (for context, Google Translate covers 133 languages, and the largest research MT models cover 200) (Google Support 2025; NLLB Team et al. 2022; Fan et al. 2021). Achieving this scale will require careful engineering and community collaboration, but the components presented here lay the groundwork (Caswell and Ng 2020; Sennrich, Haddow, and Birch 2016).

Generalizability: The languages we chose for MT experiments were diverse in family and structure, yet for each we found that a pre-trained model could be adapted with minimal data to produce understandable translations. This bodes well for other languages in our collection. Many endangered languages have at least some related language or similar typology represented in multilingual models (e.g., if a model knows some Algonquian languages, it might help with another Algonquian language). Our results imply that fine- tuning will be effective as long as a small authentic corpus can be obtained. In scenarios where there is sparse parallel data, our synthetic data approach provides a pathway: one can help a starter corpus using LLM-based augmentation and then train an MT model. This approach can be applied iteratively and at scale – for example, generating 10k–20k sentences for each of hundreds of languages and training separate models, or even training one massive multilingual MT model with synthetic data for all those languages (which could leverage transfer across them). We anticipate that integrating human feedback (through community speakers reviewing and correcting some LLM outputs) will further enhance quality, in a loop that involves native speakers in the technology development (Johnson et al. 2017; Aharoni, Johnson, and Firat 2019; Arivazhagan et al. 2019; de Gibert et al. 2025; Nekoto et al. 2020; Bird 2020).

Resource Scaling and Quality: One interesting implication from our Inuktitut data scaling is that more data clearly improves performance, but the returns can diminish if the data is of lower quality (as seen with synthetic vs real). This underlines a trade-off: quantity vs. quality. For many endangered languages, we might be forced to rely on imperfect data (machine-generated or scraped from noisy sources). Our LID model's high precision can ensure that whatever data we gather is at least in the correct language, reducing one kind of noise (Kreutzer et al. 2022b; Wenzek, Lacroix, and Bennoit 2020; Ortiz Suárez, Romary, and Sagot 2020). As for MT training, techniques like filtering synthetic outputs with back-translation or using agreement between multiple LLMs could improve quality. We view the use of synthetic data not as a one-time shortcut, but as part of a larger pipeline where initial synthetic translations enable a rough MT system, which can then be refined with any new real data that becomes available (for instance, through crowdsourcing translations of some evaluation texts, or leveraging indigenous language media) (Sennrich, Haddow, and Birch 2016; Wang et al. 2022).

Towards 1000+ Languages: Scaling to 1000 languages also raises engineering challenges: maintaining separate models for each vs. a single unified model, dealing with different scripts and tokenization issues, and evaluating performance when human expertise for many of these languages is limited. Nonetheless, the upside is tremendous – providing even rudimentary translation capability for 1000 endangered languages would dramatically broaden digital inclusion. Consider an indigenous community that currently has no software or online content in their language: an MT system could help translate public health information, educational materials, or allow elders to record stories in their language and translate them for younger generations or outsiders. It also aids linguists in creating bilingual texts and dictionaries. By bridging these languages to a lingua franca (like English), we effectively open a channel of communication that was absent. This contributes to language documentation as well – automatic translation can speed up the processing of field data and creation of translated texts for archival (NLLB Team et al. 2022; Fan et al. 2021).

Our work aligns with the vision of "no language left behind" in NLP and provides a concrete step in that direction focused on endangered languages. Moreover, it complements broader efforts in language technology for low-resource languages, such as speech recognition/synthesis and OCR. A natural next step is to incorporate speech-to-text and text-to-speech for these languages, enabling voice-based translation systems that could be used by non-literate speakers. Another direction is to expand the domain coverage: our experiments used fairly formal texts (Bible, government proceedings). Community use cases might require handling informal speech or dialectal variations; achieving that will require collecting or generating data in those registers, and perhaps fine-tuning models specifically for conversational language.

6 Conclusion

ENLIVEN-1000 shows that a pragmatic combination of broad-coverage LID, targeted fine-tuning of a multilingual MT backbone, and carefully curated LLM-generated data can extend usable NLP support to hundreds of endangered and low-resource languages without bespoke infrastructure for each. Beyond headline scores, the key contribution is a fully-deployable workflow: LID reliably allows for text discovery and filtering; MT adaptation turns small, authentic corpora into functional translation systems in both directions; and synthetic augmentation acts as an *additive* accelerant when quality is high enough, with multi-metric evaluation (BLEU, spBLEU, ChrF, TER, COMET) guiding decisions about when and how to integrate it. Looking forward, we see three immediate fronts for impact: (i) community-centered scaling (consent, orthography/script preferences, dialect labels) and human-in-the-loop post-editing; (ii) robustness work for script-divergent, morphologically rich languages (prompting, normalization, tokenization) alongside reference-free quality estimation; and (iii) modality expansion to speech (ASR/TTS) and OCR to widen access and documentation. As general-purpose LLMs improve in multilingual grounding and controllability, this LID→MT→synthetic loop can become a virtuous, repeatable pipeline—amplifying, rather than replacing, community expertise—to sustainably broaden language technologies for those most at risk of digital exclusion. Our hope is that this work sparks further research and funding for underrepresented languages – aligning with UNESCO's calls to preserve linguistic diversity as part of our common heritage (UN DESA 2022).

Ethical Statement and Limitations

Working with endangered languages requires careful ethical consideration. These languages often belong to indigenous or marginalized communities who have historically faced oppression, and their linguistic data may carry cultural sensitivities (Bird 2020; Bender and Friedman 2018). We approached the ENLIVEN-1000 project with a mindset of respect, transparency, and inclusivity. We acknowledge the following ethical aspects:

Community Engagement and Consent: Whenever possible, development of language technology for an endangered language should involve the speech community. In this study, we did *not* acquire direct engagement with native speakers due to limited resources and time, and our study exclusively leveraged publicly available texts (e.g., religious translations, governmental proceedings, and other publicly released corpora). We recognize that true revitalization impact comes from community buy-in. In future work, we plan to reach out to native speakers for guidance on our selected languages. For example, we can consult Romani community members about the appropriate orthography and dialect labeling, or discuss with Inuit representatives about the use of Inuktitut syllabics vs romanization. In addition, as we expand to more languages, we plan to establish partnerships with community organizations and obtain consent especially if using any non-public data. We also note that not all communities may want automatic translation of their language; some might fear misuse or loss of control. In such cases, we would refrain from deploying models publicly without community approval (Nekoto et al. 2020; Carroll et al. 2020; United Nations 2007).

Data Transparency and Quality: We only used data from open sources or those with clear usage permissions. Still, much of the data (like Bible translations) was created by missionaries or outside linguists, not the communities themselves. These texts might contain biases or archaic language (Mayer and Cysouw 2014; Hovy and Spruit 2016). We disclose the provenance of our datasets and caution that the MT systems' outputs may mirror the style of the training data. Users should be informed that these tools are assistive and not a replacement for learning the language or consulting fluent speakers. We also took care to not include any sacred or private texts that we thought communities might not want in a training dataset (beyond widely circulated scripture translations). All source data and model outputs will be open-sourced, in line with ethical open research practices, so that community members can inspect and audit what has been done with "their" language (Gebru et al. 2021; Bender and Friedman 2018).

Bias and Hallucination in LLM Outputs: For the synthetic data generated by GPT-4, we acknowledge that an LLM might introduce subtle biases or outright errors. GPT-4 is a product of predominantly English internet data; when it produces text in an endangered language, it might inadvertently inject stereotypes (for example, always translating certain words in a way that aligns with colonial-era texts). We attempted to mitigate this by manually checking a sample of outputs and by instructing the model clearly, but some problematic content could slip through. This is an area for improvement – possibly employing multiple LLMs or human reviewers to cross-verify synthetic data. In any case, we emphasize that caution must be used when interpreting translations. Critical decisions (e.g., legal, medical) should not be made solely on automated translations without human verification, especially given the lower reliability in low-resource settings (Bender et al. 2021; Ji et al. 2023; Weidinger et al. 2021).

Cultural Considerations: Language is deeply tied to culture. Automated systems might misinterpret idioms or output phrases that are culturally offensive if used in the wrong context. We did not specifically filter training data for such content beyond basic profanity filters. It's possible the MT outputs could occasionally produce disrespectful or nonsensical phrases (especially the LLM-synthetic portions). As part of ethical deployment, we advocate for a human-in-the-loop approach in sensitive contexts. That is, use MT to assist, but have native speakers or bilingual experts review and correct important translations. Our ultimate aim is to help preserve and promote the use of these languages, and we strongly believe this must be done with the communities' leadership. Technology can amplify revitalization efforts, but it cannot succeed without the human element – the desire of people to keep their languages alive (Bird 2020; Nekoto et al. 2020).

In conclusion, we commit to ongoing dialogue with language communities and experts. We will treat any community feedback (e.g., requests to remove data or adjust how a language is represented) with utmost seriousness. This project is driven by a passion for linguistic diversity, and we strive to ensure it remains a force for good in supporting the world's endangered languages, aligning with ethical standards and the rights of indigenous peoples (United Nations 2007; Carroll et al. 2020).

References

Aharoni, R.; Johnson, M.; and Firat, O. 2019. Massively Multilingual Neural Machine Translation. In *Proceedings of NAACL-HLT*.

Anthropic. 2024. Claude 3 Family: Capabilities and Evaluation. https://www.anthropic.com/. Accessed: 2025-11-05.

Arivazhagan, N.; Bapna, A.; Firat, O.; et al. 2019. Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges. In *Proceedings of ACL*.

Bender, E. M.; and Friedman, B. 2018. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, 6: 587–604.

Bender, E. M.; Gebru, T.; McMillan-Major, A.; and Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of FAccT*, 610–623.

Bird, S. 2020. Decolonising Speech and Language Technology. *ACL 2020 Workshop on NLP for Indigenous Languages*.

Campbell, L.; Lee, N. H.; Okura, E.; Simpson, S.; and Ueki, K. 2022. The Catalogue of Endangered Languages (ElCat). Database available at http://endangeredlanguages.com/userquery/download/, accessed 2022-08-28.

Carroll, S. R.; Garba, I.; Figueroa-Rodríguez, O.; et al. 2020. The CARE Principles for Indigenous Data Governance. https://www.gida-global.org/care. Global Indigenous Data Alliance. Accessed: 2025-11-05.

Caswell, I.; and Ng, Y. 2020. Language ID in the Wild: Unexpected Challenges on the Path to a Thousand-Language Web Text Corpus. In *Proceedings of EMNLP*.

Daily Sabah. 2023. 3,000 languages may go extinct by end of 21st century: UNESCO. https://www.dailysabah.com/arts/3000-languages-may-go-extinct-by-end-of-21st-century-unesco/news#:~:text=Around%203%2C000%20of%20the%20over,to%20cultural%20and%20linguistic%20diversity. Accessed: 2025-11-05.

de Gibert, O.; Attieh, J.; Vahtola, T.; et al. 2025. Scaling Low-Resource MT via Synthetic Data Generation with LLMs. *arXiv preprint arXiv:2505.14423*.

Ethnologue. 2023. How many languages are there in the world? https://www.ethnologue.com/insights/how-many-languages/#:~:text=languages%20every%20day,half%20of%20the%20world%E2%80%99s%20population. Accessed: 2025-11-05.

Facebook AI Research. 2022. fastText Language Identification (LID-176). https://fasttext.cc/docs/en/language-identification.html. Accessed: 2025-11-05.

Fan, A.; Bhosale, S.; Schwenk, H.; et al. 2021. Beyond English-Centric Multilingual Machine Translation. In *Proceedings of JMLR (M2M-100 paper; also arXiv:2010.11125)*.

Gebru, T.; Morgenstern, J.; Vecchione, B.; et al. 2021. Datasheets for Datasets. *Communications of the ACM*, 64(12): 86–92.

Google Research. 2016. CLD3: Compact Language Detector v3. https://github.com/google/cld3. Accessed: 2025-11-05.

Google Support. 2025. Languages you can use with Google Translate. https://support.google.com/translate/answer/6350850. Accessed: 2025-11-05.

Hovy, D.; and Spruit, S. L. 2016. The Social Impact of Natural Language Processing. In *Proceedings of ACL*, 591–598.

Ji, Z.; Lee, N.; Frieske, R.; et al. 2023. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12): 1–38.

Joanis, E.; Knowles, R.; Kuhn, R.; Larkin, S.; Littell, P.; Lo, C.-k.; Stewart, D.; and Micher, J. 2020. The Nunavut Hansard Inuktitut–English Parallel Corpus 3.0 with Preliminary Machine Translation Results. In *Proceedings of LREC*, 2562–2572.

Johnson, M.; Schuster, M.; Le, Q. V.; et al. 2017. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. In *Transactions of the ACL*, volume 5, 339–351.

Joulin, A.; Grave, E.; Bojanowski, P.; and Mikolov, T. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of EACL*, 427–431.

Kargaran, A. H.; Imani, A.; Yvon, F.; and Schuetze, H. 2023. GlotLID: Language Identification for Low-Resource Languages. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, 6155–6218. Singapore: Association for Computational Linguistics.

Koehn, P.; and Knowles, R. 2017. Six Challenges for Neural Machine Translation. In *Proceedings of the First Workshop on Neural Machine Translation*, 28–39.

Kreutzer, J.; Caswell, I.; Wang, L.; Wahab, A.; van Esch, D.; Ulzii-Orshikh, N.; Tapo, A.; Subramani, N.; Sokolov, A.; Sikasote, C.; Setyawan, M.; Sarin, S.; Samb, S.; Sagot, B.; Rivera, C.; Rios, A.; Papadimitriou, I.; Osei, S.; Ortiz Suarez, P.; Orife, I.; Ogueji, K.; Rubungo, A. N.; Nguyen, T. Q.; Müller, M.; Müller, A.; Muhammad, S. H.; Muhammad, N.; Mnyakeni, A.; Mirzakhalov, J.; Matangira, T.; Leong, C.; Lawson, N.; Kudugunta, S.; Jernite, Y.; Jenny, M.; Firat, O.; Dossou, B. F. P.; Dlamini, S.; de Silva, N.; Çabuk Ballı, S.; Biderman, S.; Battisti, A.; Baruwa, A.; Bapna, A.; Baljekar, P.; Azime, I. A.; Awokoya, A.; Ataman, D.; Ahia, O.; Ahia, O.; Agrawal, S.; and Adeyemi, M. 2022a. Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets. *Transactions of the Association for Computational Linguistics*, 10: 50–72.

Kreutzer, J.; Caswell, I.; Wang, L.; et al. 2022b. Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets. *Transactions of the Association for Computational Linguistics*, 10: 50–72.

Mayer, T.; and Cysouw, M. 2014. Creating a Massively Parallel Bible Corpus. In *Proceedings of LREC*, 3158–3163.

Nekoto, W.; Marivate, V.; Matsila, T.; et al. 2020. Participatory Research for Low-Resourced Machine Translation: A Case Study in African Languages (Masakhane). In *Findings of EMNLP*.

Neubig, G.; and Hu, J. 2018. Rapid Adaptation of Neural Machine Translation to New Languages. In *Proceedings of EMNLP*, 875–880.

NLLB Team; Costa-jussà, M. R.; Cross, J.; et al. 2022. No Language Left Behind: Scaling Human-Centered Machine Translation. *arXiv preprint arXiv:2207.04672*.

OpenAI. 2023. GPT-4 Technical Report. https://arxiv.org/abs/2303.08774.

Ortiz Suárez, P. J.; Romary, L.; and Sagot, B. 2020. A Monolingual Corpus for Multilingual Language Modeling. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*, 1218–1228.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of ACL*, 311–318.

Popović, M. 2015. chrF: Character n-gram F-score for Automatic MT Evaluation. In *Proceedings of WMT*, 392–395.

Post, M. 2018. A Call for Clarity in Reporting BLEU Scores. In *Proceedings of WMT*, 186–191.

Rei, R.; Stewart, C.; Farinha, A. C.; and Lavie, A. 2020. COMET: A Neural Framework for MT Evaluation. In *Proceedings of EMNLP*, 2685–2702.

Sennrich, R.; Haddow, B.; and Birch, A. 2016. Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of ACL*.

Snover, M.; Dorr, B.; Schwartz, R.; Micciulla, L.; and Makhoul, J. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of AMTA*, 223–231.

Thoma, M. 2018. WiLI-2018 – A Benchmark Dataset for Written Language Identification. *arXiv preprint arXiv:1801.07779*.

Tiedemann, J. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of LREC*, 2214–2218.

Tiedemann, J. 2020. The Tatoeba Translation Challenge – Realistic Data Sets for Low Resource and Multilingual MT. In *Proceedings of WMT*, 1174–1182.

UN DESA. 2022. International Decade of Indigenous Languages (2022–2032). https://social.desa.un.org/issues/indigenous-peoples/international-decade-of-indigenous-languages-2022-2032#:~:text=Optimistic%20estimates%20suggest%20that%20at,the%20end%20of%20this%20century. Accessed: 2025-11-05.

Unicode Consortium. 2007. The Universal Declaration of Human Rights (UDHR) in Unicode. https://unicode.org/udhr. Accessed: 2025-11-05.

United Nations. 2007. United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html. Accessed: 2025-11-05.

Wang, X.; Wei, J.; Schuurmans, D.; et al. 2022. Self-Consistency Improves Chain of Thought Reasoning in Language Models. In *arXiv preprint arXiv:2203.11171*.

Weidinger, L.; Uesato, J.; Rauh, M.; et al. 2021. Ethical and Social Risks of Harm from Language Models. arXiv:2112.04359.

Wenzek, G.; Lacroix, M.-A.; and Bennoit, O. 2020. CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data. In *Proceedings of LREC*.

Wikipedia. 2025. Lists of endangered languages — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/Lists_of_endangered_languages#:~:text=Lists%20of%20endangered%20languages%20are,lists%20are%20organized%20by%20region. Accessed: 2025-11-05.

Yong, Z.-X.; Menghini, C.; and Bach, S. H. 2024. LexC-Gen: Generating Data for Extremely Low-Resource Languages with Large Language Models and Bilingual Lexicons. *arXiv preprint arXiv:2402.14086*.

Zoph, B.; Yuret, D.; May, J.; and Knight, K. 2016. Transfer Learning for Low-Resource Neural Machine Translation. *arXiv preprint arXiv:1604.02201*.

A LID Training and Evaluation Details

Preprocessing and De-duplication

Normalization. Each line is normalized with Unicode NFKC, control/format characters are removed, whitespace is collapsed, and lines with fewer than 8 or more than 512 visible characters are discarded. **Two-pass de-duplication.** *Pass 1* computes a SHA-1 hash over every normalized line in every language file and flags hashes that occur in more than one language (cross-language conflicts). *Pass 2* writes per-language cleaned files that (i) drop any cross-language conflicting hashes and (ii) drop within-language duplicates (set semantics).

Stratified Splits

For each language ℓ, lines are shuffled (fixed seed) and split into **train/valid/test** at **85/5/10**. We cap validation and test at ≤ 1000 instances per language (excess is returned to train). The resulting files are concatenated into fastText format with `__label__<iso>` prefixes.

Temperature Upsampling (Train Only)

To mitigate class imbalance, we apply light *temperature upsampling* on the training file. Let p_ℓ be the empirical proportion of language ℓ in train. We draw with weights

$$w_\ell \;\propto\; p_\ell^{1/T} \qquad (default 1/T = 0.3,\; T \approx 3.33),$$

and sample to reach a target of $\approx 2\times$ the original train size. This increases exposure to low-resource classes without overwhelming high-resource ones.

Model and Optimization

We train a supervised fastText classifier (Joulin et al. 2017) over character n-grams with the following settings:

`lr`	0.8
`epoch`	2
`dim`	256
`wordNgrams`	1
`minn`/`maxn`	2/5
`loss`	`softmax`
`bucket`	10^6
`minCount`/`minCountLabel`	1000/0

Training uses multi-threading and saves both the standard `.bin` and a quantized `.ftz` model; quantization employs `qnorm` with `dsub`=2.

Thresholded Inference and Metrics

At inference, predict $\hat{l}(s) = \arg\max_{l \in B} P(l \mid s)$ if $\max_{l \in B} P(l \mid s) \geq \theta$, otherwise return UNDETERMINED, where B is the base label set (all model labels or only languages present in the benchmark). We report per-language Precision, Recall, F1, and False Positive Rate (FPR), plus macro-averages over languages present in the test set; results are given at $\theta \in \{0.5, 0.3\}$, and we also log top-1/top-3 accuracy on dev/test.

Aggregation and Reporting

For each language ℓ, we aggregate TP_ℓ, FP_ℓ, and FN_ℓ under the decision rule; UNDETERMINED counts toward FN_ℓ for the true label. We compute $\mathrm{Prec}_\ell = \frac{\mathrm{TP}_\ell}{\mathrm{TP}_\ell + \mathrm{FP}_\ell}$, $\mathrm{Rec}_\ell = \frac{\mathrm{TP}_\ell}{\mathrm{TP}_\ell + \mathrm{FN}_\ell}$, $\mathrm{F1}_\ell = \frac{2\,\mathrm{Prec}_\ell\,\mathrm{Rec}_\ell}{\mathrm{Prec}_\ell + \mathrm{Rec}_\ell}$, and (inline to avoid overflow) $\mathrm{FPR}_\ell = \frac{\mathrm{FP}_\ell}{\mathrm{FP}_\ell + \mathrm{TN}_\ell}$. Macro-averages are arithmetic means over languages included in the benchmark split. We export per-language CSVs (metrics, support, and UNDETERMINED counts) and print best/worst languages by F1 for diagnostics.

Reproducibility Notes

We fix the random seed to 13 for shuffling and sampling, and use a single pass of upsampling with the parameters above. The decision threshold(s) and base set B are explicitly reported with each result. All preprocessing, splits, model artifacts, and per-language metrics are saved to disk to facilitate exact reruns.

B Additional MT Results

Synthetic Generation Pipeline

- **Inputs/target:** Load a parallel CSV (`SRC_COL`, `TGT_COL`); set `TARGET_PAIRS`=62k.
- **Chat wrapper:** SDK-compatible chat completions (accepts `max_tokens`/`max_output_tokens`), with RPM/TPM limits, timeouts, retries/backoff.
- **Validation (LaBSE):** Small pre-generation check; require mean cosine & hit-rate thresholds or abort/override.
- **Retrieval & hints:** Char n-gram NN over sources; prompts include nearest examples + compact token/phrase lexicon (SimAlign optional).
- **Diverse seeding:** Bin by length/punctuation; round-robin to form a global seed order covering the corpus.
- **Batched creation:** Ask for k new pairs in strict format `LRL - English`; normalize (NFC) and parse.
- **Dedup/filter:** Maintain a *seen* set; drop near-duplicates via token-level Levenshtein (sliding window); discard malformed rows.
- **Checkpoints/resume:** Every 2k pairs/5 rounds save partial CSV + JSON state (seed pointer/order, rounds); resumable after interrupts.
- **Finalize:** Write the 62k synthetic pairs to CSV and store a brief validation report for audit.

Endangered→English Tables

In the main paper, we reported translation quality for the endangered→English direction for our representative languages. For completeness, this appendix presents the complementary English→endangered results for the same languages, model settings, and training regimes.

Table 5 mirrors the endangered→English results by reporting English→endangered scores for five languages (Carpathian Romani, Chuj, Sunwar, Kapingamarangi, and Inuktitut) with NLLB-200-600M fine-tuned on roughly 31K sentence pairs per language. For each language, we compare the fine-tuned system to the zero-shot baseline across BLEU, spBLEU, ChrF, TER, and COMET, and summarize the absolute metric-wise improvement in the bottom Δ row (for TER, positive Δ indicates a reduction in edits).

Table 6 provides the English→Inuktitut counterpart to our data-scaling analysis, varying the amount of supervised bitext from 31K to 124K sentence pairs while keeping the NLLB-600M architecture fixed.

Finally, Table 7 is the English→endangered analogue of our synthetic augmentation study: it evaluates "Real-only", "100% Synthetic", "Concat-62K", and "50/50 @31K" regimes for English→Carpathian Romani and English→Inuktitut, using the same metrics and experimental conditions as in the endangered→English direction.

After Table 7, the remaining appendix tables report our full LID results, covering all 1,154 languages at both confidence thresholds $\theta = 0.3$ and $\theta = 0.5$.

Code	Language	System (NLLB-200-600M)	BLEU	spBLEU	ChrF	TER ↓	COMET
rmc	Carpathian Romani	**Fine-tuned**	**12.72**	**18.07**	**37.98**	**81.33**	**0.536**
		Zero-shot baseline	0.12	0.12	12.62	130.95	0.271
		Δ	+12.60	+17.95	+25.36	+49.62	+0.27
cac	Chuj	**Fine-tuned**	**3.25**	**8.09**	**23.69**	136.08	**0.511**
		Zero-shot baseline	0.02	0.04	9.40	**111.20**	0.212
		Δ	+3.23	+8.05	+14.29	-24.88	+0.30
suz	Sunwar	**Fine-tuned**	**2.79**	**10.42**	**23.57**	176.25	**0.500**
		Zero-shot baseline	0.01	0.00	0.16	**118.65**	0.167
		Δ	+2.78	+10.42	+23.41	-57.60	+0.33
kpg	Kapingamarangi	**Fine-tuned**	**12.47**	**16.53**	**39.50**	**89.80**	**0.664**
		Zero-shot baseline	0.22	0.28	9.91	112.05	0.222
		Δ	+12.25	+16.25	+29.59	+22.25	+0.44
iku	Inuktitut	**Fine-tuned**	**5.00**	0.17	**2.20**	**97.85**	**0.491**
		Zero-shot baseline	0.68	**0.21**	0.88	439.40	0.325
		Δ	+4.32	-0.04	+1.32	+341.55	+0.17

Table 5: English $\rightarrow$ Endangered translation results with NLLB-200-600M (fine-tuned on $\sim$31k sentence pairs per language). For each language, the bottom Δ row reports the **absolute improvement of the fine-tuned model over the zero-shot baseline**, metric by metric (for TER, positive Δ denotes a reduction in edits).

Size	**BLEU**	**spBLEU**	**ChrF**	**TER↓**	**COMET**
31K	**5.00**	0.17	2.20	97.85	0.491
62K	4.40	**0.19**	2.55	**97.70**	0.489
93K	2.33	0.04	1.62	99.06	0.481
124K	1.51	0.08	**2.65**	97.94	**0.576**

Table 6: English$\rightarrow$Inuktitut with NLLB-600M fine-tuned on one language and increasing data (31K–124K).

Code	Language	System (NLLB-600M)	BLEU	spBLEU	ChrF	TER ↓	COMET
rmc	Carpathian Romani	Real-only (31K)	12.72	18.07	37.98	81.33	0.536
		100% Synthetic (31K)	5.27	9.27	27.29	93.72	0.497
		Concat-62K (31K + 31K)	**16.33**	**23.86**	**42.40**	**73.47**	**0.556**
		50/50 @31K (real/synth balanced)	5.27	9.27	27.29	93.72	0.497
iku	Inuktitut	Real-only (31K)	**5.00**	**0.17**	**2.20**	**97.85**	**0.491**
		100% Synthetic (31K)	0.40	0.00	1.21	98.56	0.462
		Concat-62K (31K + 31K)	1.21	0.01	1.82	98.16	0.472
		50/50 @31K (real/synth balanced)	1.06	0.01	1.74	98.42	0.471

Table 7: Synthetic augmentation regimes for *En→rmc* and *En→iku* with NLLB-600M fine-tuned per setting. “Real-only” trains on 31K human bitext; “100% Synthetic” uses 31K GPT-4o outputs; “Concat-62K” mixes 31K real + 31K synthetic; “50/50 @31K” balances real/synthetic while keeping the total at 31K. Best per-language scores are bolded.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
aai	Arifama-Miniafia	3470	795	0.9949874686716790	0.9987421383647800	0.9968612680477090	4.39811760566478E-06	0.9949874686716790	0.9987421383647800	0.9968612680477090	4.39811760566478E-06
aak	Ankave	1500	774	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
aau	Abau	7000	789	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
aaz	Amarasi	50k	902	0.9988888888888890	0.9966740576496670	0.9977802441731410	1.09965877588184E-06	0.9988888888888890	0.9966740576496670	0.9977802441731410	1.09965877588184E-06
abx	Inabaknon	26k	785	0.9949238578680200	0.9987261146496820	0.9968213604577240	4.3980692476003E-06	0.9974554707379140	0.9987261146496820	0.9980903882877150	2.19903462380015E-06
aby	Aneme Wake	1700	704	0.9971631205673760	0.9985795454545450	0.9978708303761530	2.19883879323329E-06	0.9985795454545450	0.9985795454545450	0.9985795454545450	1.09941939661665E-06
acu	Achuar-Shiwiar	5443	1000	0.9920477137176940	0.998	0.9950149551345960	8.79821836078194E-06	0.9920239282153540	0.995	0.9935097353969050	8.79821836078194E-06
adz	Adzera	28,900	370	0.9946236559139790	1.0	0.9973045822102430	2.1980316626461E-06	0.9973045822102430	1.0	0.9986504723346830	1.09901583132305E-06
aer	Eastern Arrernte	3820	894	0.9988826815642460	1.0	0.9994410285075460	1.09964910197156E-06	0.9988826815642460	1.0	0.9994410285075460	1.09964910197156E-06
aey	Amele	5300	904	1.0	0.9988938053097350	0.9994465965688990	0.0	1.0	0.9988938053097350	0.9994465965688990	0.0
agd	Agarabi	27k	792	0.9962216624685140	0.9987373737373740	0.9974779319041620	3.29857732360033E-06	0.9962216624685140	0.9987373737373740	0.9974779319041620	3.29857732360033E-06
agg	Angor	1270	776	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
agm	Angaataha (Angaatiha)	2500	747	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
agn	Agutaynen	15k	784	0.9849056603773590	0.9987244897959180	0.9917669411019630	1.3194193235557E-05	0.9886363636363640	0.9987244897959180	0.9936548223350250	9.89564492666777E-06
agr	Awajún	55k	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06
agt	Central Cagayan Agta	780	753	0.99734395750332	0.99734395750332	0.99734395750332	2.19895725446993E-06	0.99734395750332	0.99734395750332	0.99734395750332	2.19895725446993E-06
agu	Awakateko	16k	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
aia	Arosi	6750	780	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
aii	Assyrian Neo-Aramaic	500k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ake	Akawaio	6k	793	0.9987389659520810	0.9987389659520810	0.9987389659520810	1.0995269834917E-06	0.9987389659520810	0.9987389659520810	0.9987389659520810	1.0995269834917E-06
alp	Alune	25k	782	0.9974457215836530	0.9987212276214830	0.9980830670926520	2.19902737019416E-06	0.9987212276214830	0.9987212276214830	0.9987212276214830	1.09951368509708E-06
alq	Algonquin	2680	784	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06
aly	Alyawarr	1800	818	0.9987789987789990	1.0	0.9993891264508250	1.09955720831221E-06	0.9987789987789990	1.0	0.9993891264508250	1.09955720831221E-06
ame	Yanesha' (Amuesha)	8k	748	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
amf	Hamer-Banna	42.8k	780	0.9974424552429670	1.0	0.998719590268886	2.19902253448342E-06	0.998719590268886	1.0	0.9993593850096090	1.09951126724171E-06
amk	Ambai	10k	784	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
amm	Ama (Sawiyanu)	1.4k	781	0.9987212276214830	1.0	0.999360204734485	1.09951247616807E-06	0.9987212276214830	1.0	0.999360204734485	1.09951247616807E-06
amo	Amo	25k	795	0.9962311557788940	0.9974842767295600	0.9968573224387180	3.29858820424858E-06	0.9962216624685140	0.9949685534591200	0.9955947136563880	3.29858820424858E-06
amp	Alamblak	1.2k	768	0.9922480620155040	1.0	0.9961089494163420	6.59698056199677E-06	0.9948186528497410	1.0	0.9974025974025970	4.39798704133118E-06
amr	Amarakaeri (Harakmbut)	1206	757	0.9986807387862800	1.0	0.9993399339933990	1.09948346266924E-06	0.9986807387862800	1.0	0.9993399339933990	1.09948346266924E-06
amu	Guerrero Amuzgo	23k	793	0.9987389659520810	0.9987389659520810	0.9987389659520810	1.0995269834917E-06	1.0	0.9987389659520810	0.9993690851735020	0.0
amx	Anmatyerre	1220	402	1.0	0.9975124378109450	0.9987546699875470	0.0	1.0	0.9975124378109450	0.9987546699875470	0.0
anh	Nend	2,000	63	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
anv	Denya	11,200	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06
aoi	Anindilyakwa	1,500	339	1.0	0.9970501474926250	0.9985228951255540	0.0	1.0	0.9970501474926250	0.9985228951255540	0.0
aom	Ömie	1200	782	0.998719590268886	0.9974424552429670	0.9980806142034550	1.09951368509708E-06	1.0	0.9974424552429670	0.998719590268886	0.0
aon	Bumbita Arapesh	2353	8	0.0	0.0	0.0	2.0872996604293E-05	0.0	0.0	0.0	2.0872996604293E-05
apb	Sa'a	12k	775	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ape	Bukiyip (Mountain Arapesh)	16k	8	0.2	0.25	0.22222222222222200	8.78863014917601E-06	0.2	0.25	0.22222222222222200	8.78863014917601E-06
apn	Apinagé	1.5k	689	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
apr	Arop-Lokep	4,5k	783	0.9987244897959180	1.0	0.9993618379068280	1.09951489402875E-06	0.9987244897959180	1.0	0.9993618379068280	1.09951489402875E-06
apu	Apurinã	2780	792	0.9962264150943400	1.0	0.998109640831758	3.29857732360033E-06	0.9962264150943400	1.0	0.998109640831758	3.29857732360033E-06
apw	Western Apache	¡14k	791	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
apz	Safeyoka	2390	770	1.0	1.0	1.0	0.0	1.0	0.9987012987012990	0.9993502274204030	0.0
arb	Standard Arabic	420m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
are	Western Arrarnta	1k	788	0.998730964467005	0.998730964467005	0.998730964467005	1.099520938727E-06	0.998730964467005	0.998730964467005	0.998730964467005	1.099520938727E-06
arl	Arabela	50	775	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
arn	Mapudungun	100k	791	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
arp	Arapaho	100k	114	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
asm	Assamese	15.3m	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	1.0	1.0	1.0	0.0
aso	Dano	30k	757	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ata	Pele-Ata	2k	942	0.9989361702127660	0.9968152866242040	0.9978746014877790	1.09970714798649E-06	0.9989361702127660	0.9968152866242040	0.9978746014877790	1.09970714798649E-06
atb	Zaiwa	155k	790	1.0	0.9962025316455700	0.9980976537729870	0.0	1.0	0.9962025316455700	0.9980976537729870	0.0
atd	Ata Manobo	27k	784	0.9836888331242160	1.0	0.9917773561037320	1.42937093385201E-05	0.98989898989899	1.0	0.9949238578680200	8.79612882370469E-06
atg	Ivbie North-Okpela-Arhe	5k	794	0.9987421383647800	1.0	0.9993706733794840	1.09952819245262E-06	0.9987421383647800	1.0	0.9993706733794840	1.09952819245262E-06
att	Pamplona Atta	1k	795	0.9937421777221530	0.9987421383647800	0.9962358845671270	5.49764700708097E-06	0.9962358845671270	0.9987421383647800	0.9974874371859300	3.29858820424858E-06
auc	Waorani (Huaorani)	1,2k	785	0.9987277353689570	1.0	0.9993634627625720	1.09951731190008E-06	0.9987277353689570	1.0	0.9993634627625720	1.09951731190008E-06
aui	Anuki	890	407	0.995085995085995	0.995085995085995	0.995085995085995	2.19812104612977E-06	0.995085995085995	0.995085995085995	0.995085995085995	2.19812104612977E-06
auy	Awiyaana	11k	767	0.9960886571056060	0.9960886571056060	0.9960886571056060	3.298486654323E-06	0.9973890339425590	0.9960886571056060	0.9967384213959560	2.198991102882E-06
avt	Au	¡10k	787	0.998730964467005	1.0	0.9993650793650790	1.09951972978203E-06	0.998730964467005	1.0	0.9993650793650790	1.09951972978203E-06
awb	Awa (Samsak)	12k	787	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
awk	Awabakal	0	114	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
awx	Awara	1900	163	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
azb	South Azerbaijani	13.8m	1000	0.997	0.997	0.997	3.29933188529323E-06	0.996996996996997	0.996	0.9964982491245620	3.29933188529323E-06
azg	San Pedro Amuzgos Amuzgo	4k	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
azz	Highland Puebla Nahuatl	227k	786	0.9924242424242420	1.0	0.9961977186311790	6.59711112503835E-06	0.9949367088607600	1.0	0.9974619289340100	4.3980740833589E-06
bao	Waimaha (Tucanoan; Colombia/Brazil)	3k	793	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06
bba	Baatonum (Borgu)	1.26m	1000	1.0	0.985	0.9924433249370280	0.0	1.0	0.984	0.9919354838709680	0.0
bbb	Barai	2k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bbr	Girawa	3.6k	767	0.9960988296488950	0.9986962190352020	0.9973958333333330	3.298486654323E-06	1.0	0.9986962190352020	0.9993476842791910	0.0
bch	Bariai	3,400	1000	0.9950149551345960	0.998	0.9965052421367950	5.49888647548871E-06	0.9960079840319360	0.998	0.997002997002997	4.39910918039097E-06
bco	Kaluli	3,100	913	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bdd	Bunama	4k	786	0.9974587039390090	0.9987277353689570	0.9980928162746340	2.19903704167945E-06	0.9987277353689570	0.9987277353689570	0.9987277353689570	1.09951852083972E-06
bea	Beaver (Danezaa Zaage)	300	68	0.9855072463768120	1.0	0.9927007299270070	1.09865118593902E-06	0.9855072463768120	1.0	0.9927007299270070	1.09865118593902E-06
bef	Benabena	45k	784	0.9961880559085130	1.0	0.9980903882877150	3.29854830888926E-06	0.9961880559085130	1.0	0.9980903882877150	3.29854830888926E-06
beo	Beami	5k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
beu	Blagar	11k	297	0.9966329966329970	0.9966329966329970	0.9966329966329970	1.09892766638314E-06	1.0	0.9966329966329970	0.9983136593591910	0.0
bgs	Tagabawa	43k	995	0.9979939819458380	1.0	0.998995983935743	2.19954249516101E-06	0.9979919678714860	0.9989949748743720	0.9984932194876950	2.19954249516101E-06
bgt	Bughotu	4048	782	0.9987212276214830	0.9987212276214830	0.9987212276214830	1.09951368509708E-06	0.9987212276214830	0.9987212276214830	0.9987212276214830	1.09951368509708E-06
bhg	Binandere	7k	547	1.0	1.0	1.0	0.0	1.0	0.9981718464351010	0.9990850869167430	0.0
bhl	Bimin	2k	951	0.9979013641133260	1.0	0.9989495798319330	2.19943606459304E-06	0.9979013641133260	1.0	0.9989495798319330	2.19943606459304E-06
big	Biangai	7.4k	775	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bjk	Barok	7k	786	0.9974587039390090	0.9987277353689570	0.9980928162746340	2.19903704167945E-06	1.0	0.9974554707379140	0.9987261146496820	0.0
bjp	Fanamaket	2,2k	791	0.9949685534591200	1.0	0.9974779319041620	4.39809826231138E-06	0.9949685534591200	1.0	0.9974779319041620	4.39809826231138E-06
bjr	Binumarien	1,2k	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
bjv	Bedjond	36k	792	0.9912390488110140	1.0	0.9956002514142050	7.69668042173411E-06	0.9912390488110140	1.0	0.9956002514142050	7.69668042173411E-06
bjz	Baruga	2.5k	754	1.0	1.0	1.0	0.0	1.0	0.9986737400530500	0.9993364299933640	0.0
bkd	Binukid	100k	777	0.9987146529562980	1.0	0.9993569131832800	1.09950764047859E-06	0.9987146529562980	1.0	0.9993569131832800	1.09950764047859E-06
bki	Baki	350	776	0.9987080103359170	0.9961340206185570	0.9974193548387100	1.09950643156287E-06	0.9987080103359170	0.9961340206185570	0.9974193548387100	1.09950643156287E-06
bkq	Bakairí	950	775	0.9974259974259980	1.0	0.9987113402061860	2.19901044529962E-06	0.9974259974259980	1.0	0.9987113402061860	2.19901044529962E-06
bkx	Baikeno (Uab Meto)	20k	63	0.9545454545454550	1.0	0.9767441860465120	3.2959354524001E-06	0.9545454545454550	1.0	0.9767441860465120	3.2959354524001E-06
bla	Blackfoot (Siksiká)	4600	105	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
blw	Balangao	14k	787	0.9961977186311790	0.9987293519695040	0.9974619289340100	3.29855918934609E-06	0.9987293519695040	0.9987293519695040	0.9987293519695040	1.09951972978203E-06
blz	Balantak	30k	792	1.0	0.9974747474747480	0.9987357774968400	0.0	1.0	0.9974747474747480	0.9987357774968400	0.0
bmh	Kein (Keinim)	1.8k	768	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bmk	Ghayavi	2.8k	65	0.9411764705882350	0.9846153846153850	0.962406015037594	4.39459025939069E-06	0.9545454545454550	0.9692307692307690	0.9618320610687020	3.29594269454302E-06
bmr	Muinane	150	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bmu	Somba-Siawari	10k	954	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bnp	Bola	14k	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.999	0.999	0.999	1.09977729509774E-06
boa	Bora	¡2,500	760	0.9986859395532200	1.0	0.9993425378040760	1.09948708927285E-06	1.0	1.0	1.0	0.0
boj	Anjam	3,800	763	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bon	Bine	2k	789	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
box	Buamu (Bwa)	270k	789	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06
bpr	Koronadal Blaan	60k	790	0.951310861423221	0.9645569620253170	0.9578881206788190	4.28814109083712E-05	0.9536340852130330	0.9632911392405060	0.9584382871536530	4.06823641951214E-05
bps	Sarangani Blaan	75k	1000	0.9671641791044780	0.972	0.9695760598503740	3.62926507382255E-05	0.9681274900398410	0.972	0.9700598802395210	3.51928734431278E-05
bqc	Boko (Benin)	150k	1000	0.9989939637826960	0.993	0.995987963891675	1.09977729509774E-06	0.9989939637826960	0.993	0.995987963891675	1.09977729509774E-06
bqp	Busa-Bisã	110k	1000	0.9698149951314510	0.996	0.9827331031080410	3.409309614803E-05	0.9707602339181290	0.996	0.983218163869694	3.29933188529323E-05
bre	Breton	107k	373	0.9973262032085560	1.0	0.998661311914324	1.09901945484239E-06	1.0	1.0	1.0	0.0
bsj	Bangwinji	14k	781	1.0	0.9961587708066580	0.9980756895445800	0.0	1.0	0.9961587708066580	0.9980756895445800	0.0
bsn	Barasana-Eduria	1.9k	741	0.9986486486486490	0.9973009446693660	0.9979743416610400	1.09946412118733E-06	0.9986486486486490	0.9973009446693660	0.9979743416610400	1.09946412118733E-06

Table 8: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
bsp	Baga Sitemu	47k	448	0.9845814977973570	0.9977678571428570	0.9911308203991130	7.69377035414425E-06	0.991130820399113	0.9977678571428570	0.9944382647385980	4.39644020236814E-06
bss	Akoose	100k	790	1.0	1.0	1.0	0.0	1.0	0.9987341772151900	0.999366687777074	0.0
buk	Bukawa	9690	777	0.9961538461538460	1.0	0.9980732177263970	3.29852292143578E-06	0.9961538461538460	1.0	0.9980732177263970	3.29852292143578E-06
bus	Bokobaru (Bokobaru)	70k	1000	0.9930139720558880	0.995	0.994005994005994	7.6984410656842E-06	0.9930139720558880	0.995	0.994005994005994	7.6984410656842E-06
bvd	Baeggu (Baegu)	5.9k	789	0.9874529485570890	0.9974651457541190	0.9924337957124840	1.09952214767462E-05	0.989937106918239	0.9974651457541190	0.9936868686868690	8.79617718139697E-06
bvr	Burarra	800	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
bxh	Buhutu	1.4k	419	0.9976190476190480	1.0	0.99880810488677	1.09907501846446E-06	0.9976190476190480	1.0	0.99880810488677	1.09907501846446E-06
byx	Qaqet	15k	782	0.9961783439490450	1.0	0.998085513720485	3.29854105529125E-06	0.9961783439490450	1.0	0.998085513720485	3.29854105529125E-06
bzd	Bribri	6k	971	0.9989711934156380	1.0	0.9994853319608850	1.09974222042353E-06	0.9989711934156380	1.0	0.9994853319608850	1.09974222042353E-06
bzh	Mapos Buang	10.5k	793	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06
caa	Ch'orti'	55,250	975	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cab	Garifuna	195,800	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06
cac	Chuj	51,200	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
caf	Southern Carrier (Dakelh)	1,039	793	0.9912060301507540	0.9949558638083230	0.9930774071743240	7.69668888444191E-06	0.9924528301886790	0.9949558638083230	0.9937027707808560	6.59716190095021E-06
cak	Kaqchikel	450,900	790	0.9974715549936790	0.9987341772151900	0.9981024667931690	2.19904671324981E-06	0.9987341772151900	0.9987341772151900	0.9987341772151900	1.0995233566249E-06
cao	Chácobo	550	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cap	Chipaya	1200	788	0.998730964467005	0.998730964467005	0.998730964467005	1.099520938727E-06	0.998730964467005	0.998730964467005	0.998730964467005	1.099520938727E-06
car	Carib (Kari'nja)	8,600	789	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06
cav	Cavineña	601	766	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cax	Chiquitano	5000	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cbc	Carapana	650	773	0.9961290322580650	0.9987063389391980	0.9974160206718350	3.29850841449497E-06	0.9961290322580650	0.9987063389391980	0.9974160206718350	3.29850841449497E-06
cbi	Chachi (Cha'palaa)	5,465	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
cbk	Cavite Chavacano	359k	792	0.9962168978562420	0.9974747474747480	0.9968454258675080	3.29857732360033E-06	0.9962168978562420	0.9974747474747480	0.9968454258675080	3.29857732360033E-06
cbr	Kakataibo-Kashibo (Cashibo-Cacataibo)	2191	930	0.9989258861439310	1.0	0.999462654486835	1.09969263590826E-06	0.9989258861439310	1.0	0.999462654486835	1.09969263590826E-06
cbs	Kashinawa (Huni Kuin)	5,457	739	0.9986468200270640	0.9986468200270640	0.9986468200270640	1.09946170354994E-06	1.0	0.9986468200270640	0.9993229519295870	0.0
cbt	Shawi (Chayahuita)	14k	914	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cbu	Kandozi-Chapra	1,586	1000	0.9960159362549800	1.0	0.998003992015968	4.39910918039097E-06	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06
cbv	Cacua	400	764	1.0	0.9986910994764400	0.9993451211525870	0.0	1.0	0.9986910994764400	0.9993451211525870	0.0
cco	Comaltepec Chinantec	3,200	785	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ceb	Cebuano	20m	1000	0.9891089108910890	0.999	0.9940298507462690	1.20975502460752E-05	0.9910714285714290	0.999	0.9950199203187250	9.89799565587969E-06
cek	Eastern Khumi Chin	50k	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
cgc	Kagayanen	30k	782	0.988621997471555	1.0	0.9942784488239030	9.89562316587373E-06	0.9936467598475220	1.0	0.9968132568514980	5.49756842548541E-06
cha	Chamorro	58k	724	0.9986206896551720	1.0	0.9993098688750860	1.09944357160841E-06	0.9986206896551720	1.0	0.9993098688750860	1.09944357160841E-06
chd	Highland Oaxaca Chontal	3.6k	784	0.9987244897959180	0.9987244897959180	0.9987244897959180	1.09951610296309E-06	1.0	0.9987244897959180	0.9993618379068280	0.0
chf	Tabasco Chontal (Yokot'an)	38k	793	1.0	0.9962168978562420	0.998104864181933	0.0	1.0	0.9962168978562420	0.998104864181933	0.0
chk	Chuukese	45.9k	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.997	0.99849774661993	0.0
chq	Quiotepec Chinantec	10,440	789	0.9974715549936790	1.0	0.9987341772151900	2.19904429534924E-06	0.9974715549936790	1.0	0.9987341772151900	2.19904429534924E-06
chz	Ojitlán Chinantec	37,900	792	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06	1.0	1.0	1.0	0.0
cjo	Pajonal Ashéninka	20k	763	0.9794079794079790	0.9973787680209700	0.9883116883116880	1.75918514544063E-05	0.9844760672703750	0.9973787680209700	0.9908854166666670	1.31938885908047E-05
cjv	Chuave	26k	784	1.0	0.9987244897959180	0.9993618379068280	0.0	1.0	0.9987244897959180	0.9993618379068280	0.0
ckb	Central Kurdish (Sorani)	7mm	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cle	Lealao Chinantec	3.2k	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
clu	Caluyanun	30k	785	0.966707768187423	0.9987261146496820	0.9824561403508770	2.9686967421302E-05	0.9679012345679010	0.9987261146496820	0.9830721003134800	2.8587450109402E-05
cme	Cerma (San)	53.6k	789	0.9886934673366830	0.9974651457541190	0.9930599369085170	9.89569932907159E-06	0.9911838790931990	0.9974651457541190	0.994314592545799	7.69665503372234E-06
cni	Asháninka	89,516	770	0.9961190168175940	1.0	0.9980557355800390	3.29849753437309E-06	0.9974059662775620	0.9987012987012990	0.9980532121998700	2.19899835624873E-06
cnl	Lalana Chinantec	10,700	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cnt	Tepetotutla Chinantec	2k	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cof	Tsafiki (Colorado)	1,872	749	0.9986648865153540	0.9986648865153540	0.9986648865153540	1.09947379184322E-06	1.0	0.9986648865153540	0.9993319973279890	0.0
con	Cofán (A'i)	1,5k	938	0.997872340425532	1.0	0.9989350372736960	2.19940462116905E-06	0.9989350372736950	1.0	0.9994672349493870	1.09970231058452E-06
cot	Caquinte	500	776	0.9987113402061860	0.9987113402061860	0.9987113402061860	1.09950643156287E-06	0.9987113402061860	0.9987113402061860	0.9987113402061860	1.09950643156287E-06
cpa	Palantla Chinantec	30k	794	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cpb	Ucayali-Yurúa Ashéninka	10k	789	0.9836065573770490	0.9885931558935360	0.9860935524652340	1.42937879197701E-05	0.9835858585858590	0.9873257287705960	0.9854522454142950	1.42937879197701E-05
cpc	Ajyíninka Apurucayali	20k	789	0.9528415961305930	0.9987325728770600	0.9752475247524750	4.28813637593102E-05	0.9551515151515150	0.9987325728770600	0.9764560099132590	4.0682319463961E-05
cpu	Pichis Ashéninka	17k	790	0.9885786802030460	0.9860759493670890	0.9873257287705960	9.89571020962413E-06	0.9910941475826970	0.9860759493670890	0.9885786802030460	7.69666349637432E-06
cpy	South Ucayali Ashéninka	7870	1000	0.993963782696177	0.988	0.9909729187562690	6.59866377058646E-06	0.993963782696177	0.988	0.9909729187562690	6.59866377058646E-06
crn	El Nayar Cora	15k	765	0.9986945169712790	1.0	0.9993468321358590	1.09949313366538E-06	0.9986945169712790	1.0	0.9993468321358590	1.09949313366538E-06
crx	Babine (Northern Carrier)	¡300	946	0.9957627118644070	0.9936575052854120	0.9947089947089950	4.39884794172406E-06	0.9968186638388120	0.9936575052854120	0.9952355743779780	3.29913595629305E-06
cso	Sochiapam Chinantec	3,590	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
csy	Siyin Chin	10k	795	0.9987389659520810	0.9962264150943400	0.9974811083123430	1.09952940141619E-06	1.0	0.9962264150943400	0.998109640831758	0.0
cta	Tataltepec Chatino	540	786	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ctp	Western Highland Chatino	12k	788	0.9987325728770600	1.0	0.9993658845909960	1.099520938727E-06	1.0	1.0	1.0	0.0
ctu	Chol	133.9k	834	0.9976047904191620	0.9988009592326140	0.9982025164769320	2.19915310613883E-06	0.9976019184652280	0.9976019184652280	0.9976019184652280	2.19915310613883E-06
cub	Cubeo	6260	770	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cuc	Usila Chinantec	7410	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cui	Cuiba	2,830	761	0.9986876640419950	1.0	0.9993434011818780	1.09948829814604E-06	1.0	1.0	1.0	0.0
cuk	San Blas Kuna	57.1k	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cut	Teutila Cuicatec	3690	783	1.0	0.9987228607918260	0.9993610223642170	0.0	1.0	0.9987228607918260	0.9993610223642170	0.0
cux	Tepeuxila Cuicatec	4.2k	795	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cwe	Kwere	150k	1000	0.980980980980981	0.98	0.9804902451225610	2.08957686068571E-05	0.9819458375125380	0.979	0.9804707060590890	1.97959913117594E-05
cya	Nopala Chatino	6k	794	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dgc	Casiguran Dumagat Agta	610	780	0.9885931558935360	1.0	0.994263862332696	9.8956014051754E-06	0.9911054637865310	1.0	0.9955328653477980	7.69657887069198E-06
dgr	Tlicho (Dogrib)	2,470	936	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dgz	Daga	9k	780	0.9910941475826970	0.9987179487179490	0.9948914431673050	7.69657887069198E-06	0.9910941475826970	0.9987179487179490	0.9948914431673050	7.69657887069198E-06
dhg	Dhangu-Djangu	350	61	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dif	Dieri	1-2	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06
dik	Southwestern Dinka	1.3mm	788	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dji	Djinang	120	88	0.9887640449438200	1.0	0.9943502824858760	1.09867532715805E-06	0.9887640449438200	1.0	0.9943502824858760	1.09867532715805E-06
djr	Djambarrpuyu	2760	741	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dob	Dobu	10k	785	0.9924146649810370	1.0	0.9961928934010150	6.59710387140046E-06	0.9936708860759490	1.0	0.9968253968253970	5.49758655950038E-06
dop	Lukpa	160k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dov	Dombe	5430	793	0.9778051787916150	1.0	0.9887780548628430	1.97914857028506E-05	0.9789864029666250	0.9987389659520810	0.9887640449438200	1.86919587193589E-05
dwr	Dawro	1mm	779	0.9910941475826970	1.0	0.9955271565495210	7.69657040822609E-06	0.9936224489795920	1.0	0.9968010236724250	5.49755029159007E-06
dww	Dawawa	2.5k	787	0.9987293519695040	0.9987293519695040	0.9987293519695040	1.09951972978203E-06	0.9987293519695040	0.9987293519695040	0.9987293519695040	1.09951972978203E-06
dwy	Dhuwaya	3650	62	0.9841269841269840	1.0	0.9920000000000000	1.09864394378019E-06	0.9841269841269840	1.0	0.9920000000000000	1.09864394378019E-06
ebk	Eastern Bontok	6.2k	791	0.9875	0.9987357774968390	0.9930861093651790	1.09952456557784E-05	0.9899749373433580	0.9987357774968390	0.9943360604153560	8.79619652462275E-06
eko	Koti	100k	632	0.9968454258675080	1.0	0.9984202211690360	2.19866475089678E-06	0.9984202211690360	1.0	0.9992094861660080	1.09933237544839E-06
emi	Mussau-Emira	5k	804	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
emp	Northern Embera (Emberá)	60.2k	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
enq	Enga	432k	756	1.0	0.9973544973544970	0.9986754966887420	0.0	1.0	0.9973544973544970	0.9986754966887420	0.0
eri	Ogea	4.2k	790	0.9949622166246850	1.0	0.9974747474747480	4.39809342649961E-06	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06
ese	Ese Ejja	725	768	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
esk	Northwest Alaska Inupiatun	3k	783	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
etr	Edolo	1.3k	420	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ewe	Ewe	7.5m	1000	0.996011964107677	0.999	0.9975037443834250	4.39910918039097E-06	0.997002997002997	0.998	0.9975012493753120	3.29933188529323E-06
faa	Fasu	1,200	777	0.9961538461538460	1.0	0.9980732177263970	3.29852292143578E-06	0.9961538461538460	1.0	0.9980732177263970	3.29852292143578E-06
fai	Faiwol	4,500	794	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06	1.0	1.0	1.0	0.0
far	Fataleka	14k	785	0.9948979591836740	0.9936305732484080	0.994263862332696	4.3980692476003E-06	0.9948979591836740	0.9936305732484080	0.994263862332696	4.3980692476003E-06
ffm	Maasina Fulfulde	1.6m	786	0.9910025706940870	0.9809160305343510	0.9859335038363170	7.69662964587807E-06	0.9922680412371130	0.9796437659033080	0.9859154929577470	6.59711112503835E-06
for	Fore	30k	789	0.9974651457541190	0.9974651457541190	0.9974651457541190	2.19904429534924E-06	0.998730964467005	0.9974651457541190	0.9980976537729870	1.09952214767462E-06
fue	Borgu Fulfulde	900k	466	0.9624217118997910	0.9892703862660940	0.9756613756613760	1.9784372324301E-05	0.9663865546218490	0.9871244635193130	0.9766454352441610	1.7586108732712E-05
fuf	Pular (Futa Jalon)	4.3m	789	0.9911167512690360	0.9898605830164770	0.9904882688649340	7.69665503372234E-06	0.9911167512690360	0.9898605830164770	0.9904882688649340	7.69665503372234E-06
fuh	Western Niger Fulfulde	3.1m	789	0.965559655596556	0.9949302915082380	0.9800249687890140	3.07866201348894E-05	0.969097651421508	0.9936628643852980	0.981226533166458	2.74880536918655E-05
gah	Gahuku (Alekano)	25k	788	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
gai	Borei (Warembori?)	1202	785	0.9987277353689570	1.0	0.9993634627625720	1.09951731190008E-06	0.9987277353689570	1.0	0.9993634627625720	1.09951731190008E-06
gam	Kandawo	4k	780	0.9936305732484080	1.0	0.9968051118210860	5.49755633620856E-06	0.9948979591836740	1.0	0.9974424552429670	4.39804506896684E-06
gaw	Nobonob (Garus)	2.4k	783	0.9987228607918260	0.9987228607918260	0.9987228607918260	1.09951489402875E-06	0.9987228607918260	0.9987228607918260	0.9987228607918260	1.09951489402875E-06
gdn	Umanakaina (Gaina)	2.8k	767	0.9986962190352020	0.9986962190352020	0.9986962190352020	1.099495551441E-06	0.9986962190352020	0.9986962190352020	0.9986962190352020	1.099495551441E-06
gdr	Wipi	3.5k	787	0.9974651457541190	1.0	0.9987309644670050	2.19903945956406E-06	0.9974651457541190	1.0	0.9987309644670050	2.19903945956406E-06
geb	Kire	1,9k	780	0.998719590268886	1.0	0.9993593850096090	1.09951126724171E-06	0.998719590268886	1.0	0.9993593850096090	1.09951126724171E-06
gfk	Patpatar	7,000	767	0.9922279792746110	0.9986962190352020	0.99545159194282	6.59697330864599E-06	0.9922279792746110	0.9986962190352020	0.99545159194282	6.59697330864599E-06

Table 9: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
ghs	Guhu-Samane	13k	745	0.9986559139784950	0.9973154362416110	0.9979852249832100	1.09946895649401E-06	1.0	0.9973154362416110	0.9986559139784950	0.0
glk	Gilaki	2.8m	88	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
gmv	Gamo	2.1m	780	0.9873257287705960	0.9987179487179490	0.9929891650732950	1.09951126724171E-05	0.98856416772554	0.9974358974358970	0.9929802169751120	9.8956014051754E-06
gng	Ngangam	250k	790	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06
gnn	Gumatj	450	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
gnw	Western Bolivian Guaraní	9.7k	780	0.9948320413436690	0.9871794871794870	0.9909909909909910	4.39804506896684E-06	0.9948253557567920	0.985897435897436	0.9903412749517070	4.39804506896684E-06
gof	Gofa	392k	790	0.9923664122137400	0.9873417721518990	0.9898477157360410	6.59714013974942E-06	0.9961685823754790	0.9873417721518990	0.9917355371900830	3.29857006987471E-06
gub	Guajajára (Tenetehára)	15k	1000	1.0	1.0	1.0	0.0	1.0	0.999	0.9994997498749370	0.0
guh	Guahibo (Hiwi)	34.2k	738	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
gui	Eastern Bolivian Guaraní	33,670	780	0.9847715736040610	0.9948717948717950	0.9897959183673470	1.31941352069005E-05	0.9860228716645490	0.9948717948717950	0.990427568602425	1.20946239396588E-05
gum	Guambiano (Misak)	23,500	946	0.9989429175475690	0.9989429175475690	0.9989429175475690	1.09971198543102E-06	0.9989429175475690	0.9989429175475690	0.9989429175475690	1.09971198543102E-06
gun	Mbyá Guaraní	25,450	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
guo	Guayabero (Jiw)	1,118	776	0.9974293059125960	1.0	0.9987129987129990	2.19901286312574E-06	0.9987129987129990	1.0	0.9993560849967800	1.09950643156287E-06
gux	Gourmanchéma	1.5m	794	0.9974779319041620	0.9962216624685140	0.9968494013862630	2.19905638490524E-06	0.9974747474747480	0.9949622166246850	0.9962168978562420	2.19905638490524E-06
gvc	Guanano (Wanana)	2,045	774	0.9974160206718350	0.9974160206718350	0.9974160206718350	2.1990080274788E-06	0.9987063389391980	0.9974160206718350	0.998060762766645	1.0995040137394E-06
gvf	Golin	51.1k	789	1.0	0.9987325728770600	0.9993658845909960	0.0	1.0	0.9987325728770600	0.9993658845909960	0.0
gvn	Kuku-Yalanji	240	807	0.9987608426270140	0.9987608426270140	0.9987608426270140	1.09954390918647E-06	0.9987608426270140	0.9987608426270140	0.9987608426270140	1.09954390918647E-06
gvs	Gumawana	470	1000	0.9950149551345960	0.998	0.9965052421367950	5.49888647548871E-06	0.996003996003996	0.997	0.9965017491254370	4.39910918039097E-06
gwi	Gwich'in	800	783	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
gym	Ngäbere (Guaymí)	5,090	774	0.9974226804123710	1.0	0.9987096774193550	2.1990080274788E-06	0.9974226804123710	1.0	0.9987096774193550	2.1990080274788E-06
gyr	Guarayu	5k	787	1.0	1.0	1.0	0.0	1.0	0.9987293519695040	0.9993642720915450	0.0
haw	Hawaiian	2k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
hbo	Ancient Hebrew	0 (n/a)	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
hch	Wixárika (Huichol)	20k	791	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
heb	Hebrew	10m	768	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
heg	Helong	14k	903	0.9955849889624720	0.9988925802879290	0.9972360420121610	4.39863994053039E-06	0.9966814159292040	0.9977851605758580	0.9972329828444940	3.29897995539779E-06
hix	Hixkaryána	600	759	0.9986824769433470	0.9986824769433470	0.9986824769433470	1.09948588040232E-06	0.9986824769433470	0.9986824769433470	0.9986824769433470	1.09948588040232E-06
hla	Halia	25k	770	0.9961190168175940	1.0	0.9980557355800390	3.29849753437309E-06	0.9987029831387810	1.0	0.9993510707332900	1.09949917812436E-06
hlt	Matu Chin	40k	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
hmo	Hiri Motu	95k	1000	0.9950199203187250	0.999	0.9970059880239520	5.49888647548871E-06	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06
hns	Caribbean Hindustani	300k	782	0.9974457215836530	0.9987212276214830	0.9980830670926520	2.19902737019416E-06	0.9974457215836530	0.9987212276214830	0.9980830670926520	2.19902737019416E-06
hop	Hopi	5,260	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
hot	Malei	2.3k	764	0.9960886571056060	1.0	0.9980404964075770	3.29847577434468E-06	0.9960886571056060	1.0	0.9980404964075770	3.29847577434468E-06
hto	Minica Huitoto	6,800	791	1.0	0.9974715549936790	0.9987341772151900	0.0	1.0	0.9974715549936790	0.9987341772151900	0.0
hub	Wampis (Huambisa)	9330	772	0.9974059662775620	0.9961139896373060	0.9967595593000650	2.19900319185313E-06	0.9974059662775620	0.9961139896373060	0.9967595593000650	2.19900319185313E-06
hui	Huli	80k	1000	0.9960159362549800	1.0	0.998003992015968	4.39910918039097E-06	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06
hun	Hungarian	13m	794	0.9924812030075190	0.9974811083123430	0.9949748743718590	6.59716915471571E-06	0.9937264742785450	0.9974811083123430	0.9956002514142050	5.49764096226309E-06
hus	Huastec (Wastek)	121,750	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
huu	Murui Huitoto	1,130	766	0.9960988296488950	1.0	0.9980456026058630	3.29848302765558E-06	0.9960988296488950	1.0	0.9980456026058630	3.29848302765558E-06
huv	San Mateo del Mar Huave	12k	795	0.9974905897114180	1.0	0.9987437185929650	2.19905880283239E-06	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06
hvn	Hawu	100k	63	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ian	Iatmul	46k	765	0.9973856209150330	0.9973856209150330	0.9973856209150330	2.19898626733076E-06	0.9986910994764400	0.9973856209150330	0.9980379332897320	1.09949313366538E-06
ign	Ignaciano (Mojeño-Ignaciano)	4.5k	736	0.9986431478968790	1.0	0.9993211133740670	1.09945807711379E-06	0.9986431478968790	1.0	0.9993211133740670	1.09945807711379E-06
ikk	Ika	100k	795	1.0	0.9987421383647800	0.9993706733794840	0.0	1.0	0.9974842767295600	0.9987405541561710	0.0
ikw	Ikwere	2mm	793	0.9974747474747480	0.9962168978562420	0.9968454258675080	2.1990539669834E-06	0.9987357774968390	0.9962168978562420	0.9974747474747470	1.0995269834917E-06
ilo	Ilocano	11m	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06
imo	Imbongu	42.5k	765	0.9947984395318600	1.0	0.9973924380704040	4.39797253466152E-06	0.99609375	1.0	0.9980430528375730	3.29847940099614E-06
inb	Inga (Jungle Kichwa)	11,200	783	0.9987228607918260	0.9987228607918260	0.9987228607918260	1.09951489402875E-06	0.9987228607918260	0.9987228607918260	0.9987228607918260	1.09951489402875E-06
ind	Indonesian	275m	1000	0.9282945736434110	0.958	0.9429133858267720	8.1383519837233E-05	0.9300291545189500	0.957	0.9433218334154760	7.91839652470375E-05
ino	Inoke-Yate	10k	743	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
iou	Tuma-Irumu	2,200	1000	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
ipi	Ipili	26k	584	0.9982905982905980	1.0	0.9991445680068440	1.09927436898903E-06	0.9982905982905980	1.0	0.9991445680068440	1.09927436898903E-06
isn	Isanzu	32,400	794	0.9790123456790120	0.9987405541561710	0.9887780548628430	1.86919792716945E-05	0.9814126394052050	0.9974811083123430	0.9893816364772020	1.64929228867893E-05
ita	Italian	68m	1000	1.0	0.994	0.9969909729187560	0.0	1.0	0.994	0.9969909729187560	0.0
iws	Sepik Iwam	2,500	781	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ixl	Ixil	69k	792	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06
jac	Eastern Jacalteco (Popti')	50k	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
jae	Yabem	2k	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.998	0.998998998998999	0.0
jao	Yanyuwa	52	71	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
jic	Tol (Jicaque)	550	789	0.9974715549936790	1.0	0.9987341772151900	2.19904429534924E-06	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06
jid	Bu (Jidda)	6k	794	0.9937027707808560	0.9937027707808560	0.9937027707808560	5.49764096226309E-06	0.9936948297604040	0.9924433249370280	0.9930686830497800	5.49764096226309E-06
jiv	Shuar	50k	779	0.9987080103359170	0.9922978177150190	0.9954925949774630	1.09951005831801E-06	0.9987080103359170	0.9922978177150190	0.9954925949774630	1.09951005831801E-06
jni	Janji	1,150	794	0.9924337957124840	0.9911838790931990	0.9918084436042850	6.59716915471571E-06	0.9924337957124840	0.9911838790931990	0.9918084436042850	6.59716915471571E-06
jpn	Japanese	130m	3	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
jvn	Caribbean Javanese	77k	773	0.9987080103359170	1.0	0.9993535875888820	1.09950280483166E-06	1.0	1.0	1.0	0.0
kan	Kannada	59m	1000	0.9910802775024780	1.0	0.9955201592832260	9.89799565587969E-06	0.9920634920634920	1.0	0.9960159362549800	8.79821836078194E-06
kaq	Capanahua	400	791	0.9974779319041620	1.0	0.9987373737373740	2.19904913115569E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
kbc	Kadiwéu	1590	782	0.9987228607918260	1.0	0.9993610223642170	1.09951368509708E-06	0.9987228607918260	1.0	0.9993610223642170	1.09951368509708E-06
kbh	Camsá (Kamsá/Biatá)	4773	783	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kbm	Iwal	1900	778	0.9961538461538460	0.9987146529562980	0.9974326059050070	3.29852654819092E-06	0.9987146529562980	0.9987146529562980	0.9987146529562980	1.09950884939697E-06
kdc	Kutu	45k	791	0.94625	0.9570164348925410	0.951602765556254	4.72795563198473E-05	0.9521410579345090	0.9557522123893810	0.9539432176656150	4.17819334919581E-05
kde	Makonde	2.4mm	791	0.9924528301886790	0.9974715549936790	0.9949558638083230	6.59714739346707E-06	0.9937027707808560	0.9974715549936790	0.9955835962145110	5.49762282788922E-06
kdl	Tsikimba	100k	790	0.9813664596273290	1.0	0.9905956112852670	1.64928503493735E-05	0.9838107098381070	1.0	0.9918392969240430	1.42938036361237E-05
kek	Q'eqchi'	423,500	794	1.0	0.9987405541561710	0.9993698802772530	0.0	1.0	0.9987405541561710	0.9993698802772530	0.0
ken	Kenyang (Manyu)	65k	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kew	West Kewapi	45k	936	0.9936102236421730	0.9967948717948720	0.9952	6.598199351397E-06	0.9936102236421730	0.9967948717948720	0.9952	6.598199351397E-06
kgf	Kube	7.5k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kgk	Kaiwá	18,510	753	0.9947089947089950	0.99867197875166	0.9966865473823720	4.39791450893986E-06	0.9960264900662250	0.99867197875166	0.9973474801061010	3.2984358817049E-06
kgp	Kaingang	18,500	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
khs	Kasua	600	775	0.9974259974259980	1.0	0.9987113402061860	2.19901044529962E-06	0.9987113402061860	1.0	0.9993552546744040	1.09950522264981E-06
khz	Keapara	19,400	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06
kik	Kikuyu (Gikuyu)	8.1m	1000	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06
kiw	Northeast Kiwai	4.4k	114	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kiz	Kisi	10.2k	792	0.9911949685534590	0.9949494949494950	0.9930686830497800	7.69668042173411E-06	0.9936948297604040	0.9949494949494950	0.9943217665615140	5.49762887266722E-06
kje	Kisar	20k	782	0.9974457215836530	0.9987212276214830	0.9980830670926520	2.19902737019416E-06	0.9974457215836530	0.9987212276214830	0.9980830670926520	2.19902737019416E-06
kjn	Kunjen (Oykangand)	20	1	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kjs	East Kewa	45k	789	0.9924146649810370	0.9949302915082380	0.9936708860759490	6.59713288604772E-06	0.9936708860759490	0.9949302915082380	0.9943001899936670	5.4976107383731E-06
kkc	Odoodee	490	568	0.993006993006993	1.0	0.9964912280701760	4.39702013945149E-06	0.9947460595446580	1.0	0.9973661106233540	3.29776510458862E-06
kkl	Kosarek Yale	2300	342	0.9941690962099130	0.9970760233918130	0.9956204379562040	2.19796402592279E-06	0.9970760233918130	0.9970760233918130	0.9970760233918130	1.09898201296139E-06
klt	Nukna	1000	376	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
klv	Maskelynes	1,100	789	0.9974651457541190	0.9974651457541190	0.9974651457541190	2.19904429534924E-06	0.998730964467005	0.9974651457541190	0.9980976537729870	1.09952214767462E-06
kmg	Kâte	6,130	615	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kmk	Limos Kalinga	20k	790	0.9949622166246850	1.0	0.9974747474747480	4.39809342649961E-06	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06
kmo	Kwoma	2,925	784	0.9961832061068700	0.9987244897959180	0.9974522292993630	3.29854830888926E-06	0.9974522292993630	0.9987244897959180	0.9980879541108990	2.19903220592617E-06
kms	Kamasau	700	784	0.9987228607918260	0.9974489795918370	0.998085513720485	1.09951610296309E-06	0.9987228607918260	0.9974489795918370	0.998085513720485	1.09951610296309E-06
kmu	Kanite	9.3k	696	0.6186666666666670	1.0	0.7644151565074140	0.0004716467728476580	0.6275924256086570	1.0	0.771191135734072	0.0004540562172169760
kne	Kankanaey	150k	1000	0.995995995995996	0.995	0.9954977488744370	4.39910918039097E-06	0.9979939819458380	0.995	0.9964947421131700	2.19955459019549E-06
knf	Mankanya	65k	944	1.0	0.9989406779661020	0.9994700582935880	0.0	1.0	0.9989406779661020	0.9994700582935880	0.0
knj	Akateko	20k	792	0.9987341772151900	0.9962121212121210	0.9974715549936790	1.09952577453344E-06	0.9987341772151900	0.9962121212121210	0.9974715549936790	1.09952577453344E-06
kos	Kosraean	8,570	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
kpf	Komba	15k	785	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kpg	Kapingamarangi	3k	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
kpj	Karajá	3.6k	761	0.9973753280839900	0.9986859395532200	0.9980302035456340	2.19897659629209E-06	0.9986859395532200	0.9986859395532200	0.9986859395532200	1.09948829814604E-06
kpr	Korafe-Yegha	3630	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
kpw	Kobon	10k	772	0.9987046632124350	0.9987046632124350	0.9987046632124350	1.09950159592657E-06	0.9987046632124350	0.9987046632124350	0.9987046632124350	1.09950159592657E-06
kpx	Mountain Koiali	4k	779	0.9974358974358970	0.9987163029525030	0.9980756895445800	2.19902011663603E-06	0.9974358974358970	0.9987163029525030	0.9980756895445800	2.19902011663603E-06
kqa	Mum	3290	104	0.9904761904761910	1.0	0.9952153110047850	1.09869464089715E-06	0.9904761904761910	1.0	0.9952153110047850	1.09869464089715E-06
kqc	Doromu-Koki	1.5k	799	1.0	0.9987484355444310	0.9993738259236070	0.0	1.0	0.9974968710888610	0.9987468671679200	0.0
kqf	Kakabai	900	474	1.0	0.9978902953586500	0.9989440337909190	0.0	1.0	0.9978902953586500	0.9989440337909190	0.0
kql	Kyenele (Miyak)	548	67	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0

Table 10: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
kqw	Kandas	480	309	0.9871794871794870	0.9967637540453080	0.9919484702093400	4.39576863311376E-06	0.9967637540453080	0.9967637540453080	0.9967637540453080	1.09894215827844E-06
ksd	Kuanua (Tolai)	61k	1000	0.999	0.999	0.999	1.09977729509774E-06	0.999	0.999	0.999	1.09977729509774E-06
ksj	Uare	1.3k	420	0.9976133651551310	0.9952380952380950	0.9964243146603100	1.09907622643168E-06	1.0	0.9952380952380950	0.9976133651551310	0.0
ksr	Borong	2070	954	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ktm	Kurti	3k	242	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kto	Kuot	2.4k	779	0.9987179487179490	1.0	0.9993585631815270	1.09951005831801E-06	1.0	1.0	1.0	0.0
kud	'Auhelawa	1.2k	789	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06	1.0	0.9987325728770600	0.9993658845909960	0.0
kue	Kuman	115k	794	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06	0.9987421383647800	1.0	0.9993706733794840	1.09952819245262E-06
kup	Kunimaipa	8.2k	749	1.0	1.0	1.0	0.0	1.0	0.9986648865153540	0.9993319973279890	0.0
kvg	Kuni-Boazi	4.5k	801	0.9987531172069830	1.0	0.9993761696818470	1.09953665525348E-06	0.9987515605493130	0.9987515605493130	0.9987515605493130	1.09953665525348E-06
kvn	Border Kuna (Guna)	700	791	0.9974779319041620	1.0	0.9987373737373740	2.19904913115569E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
kwd	Kwaio	13.2k	787	0.998730964467005	1.0	0.9993650793650790	1.09951972978203E-06	1.0	1.0	1.0	0.0
kwf	Kwara'ae	32.4k	932	0.9978540772532190	0.9978540772532190	0.9978540772532190	2.19939010912274E-06	0.9978540772532190	0.9978540772532190	0.9978540772532190	2.19939010912274E-06
kwi	Awa-Cuaiquer (Awa Pit)	17,464	791	1.0	0.9987357774968390	0.9993674889310560	0.0	1.0	0.9987357774968390	0.9993674889310560	0.0
kwj	Kwanga	7,520	793	1.0	0.9974779319041620	0.9987373737373740	0.0	1.0	0.9974779319041620	0.9987373737373740	0.0
kyc	Kyaka	15.4k	1000	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
kyf	Kouya	10.1k	785	0.9924146649810370	1.0	0.9961928934010150	6.59710387140046E-06	0.9924146649810370	1.0	0.9961928934010150	6.59710387140046E-06
kyg	Keyagana	12.3k	950	0.9958071278826000	1.0	0.9978991596638660	4.3988672916724E-06	0.9958071278826000	1.0	0.9978991596638660	4.3988672916724E-06
kyq	Kenga	40k	790	0.9987357774968390	1.0	0.9993674889310560	1.0995233566249E-06	0.9987357774968390	1.0	0.9993674889310560	1.0995233566249E-06
kyz	Kayabí	1,000	727	1.0	0.9986244841815680	0.9993117687543020	0.0	1.0	0.9986244841815680	0.9993117687543020	0.0
kze	Kosena	2,000	766	0.9947984395318600	0.9986945169712790	0.996742671009772	4.39797737020744E-06	0.99609375	0.9986945169712790	0.9973924380704040	3.29848302765558E-06
lac	Lacandón	1,000	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
lbb	Label	140	788	0.9949367088607600	0.9974619289340100	0.9961977186311790	4.39808375490799E-06	0.9961977186311790	0.9974619289340100	0.9968294229549780	3.29856281618099E-06
lbk	Central Bontok (Bontoc)	26k	938	0.9978609625668450	0.9946695095948830	0.996262680192205	2.19940462116905E-06	0.9989293361884370	0.9946695095948830	0.9967948717948720	1.09970231058452E-06
lcm	Tungag (Lavongai)	12k	954	0.9979013641133260	0.9968553459119500	0.9973780807551130	2.19944332089548E-06	1.0	0.9968553459119500	0.9984251968503940	0.0
leu	Kara	5k	786	0.9987212276214830	0.9936386768447840	0.996173469387755	1.09951852083972E-06	0.9987212276214830	0.9936386768447840	0.996173469387755	1.09951852083972E-06
lex	Luang	18k	961	0.995846313603323	0.9979188345473470	0.9968814968814970	4.3989205049081E-06	0.9968814968814970	0.9979188345473470	0.9973998959958400	3.29919037868107E-06
lgl	Wala	6,980	787	0.9974651457541190	1.0	0.9987309644670050	2.19903945956406E-06	1.0	1.0	1.0	0.0
lid	Nyindrou	4.2k	950	0.9978925184404640	0.9968421052631580	0.9973670352817270	2.1994336458362E-06	0.9978925184404640	0.9968421052631580	0.9973670352817270	2.1994336458362E-06
lif	Limbu	334k	783	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
lin	Lingala	2.04mm	1000	0.9940357852882700	1.0	0.9970089730807580	6.59866377058646E-06	0.9950248756218910	1.0	0.9975062344139650	5.49888647548871E-06
llg	Lole	20k	900	0.9856035437430790	0.9888888888888890	0.9872434830837490	1.42955326460481E-05	0.9899888765294770	0.9888888888888890	0.989438576987215	9.89690721649485E-06
lug	Ganda (Luganda)	4.13mm	1000	0.9891196834817010	1.0	0.9945300845350570	1.20975502460752E-05	0.9920634920634920	1.0	0.9960159362549800	8.79821836078194E-06
luo	Dholuo	4.27mm	1000	0.999	0.999	0.999	1.09977729509774E-06	0.999	0.999	0.999	1.09977729509774E-06
lww	Lewo	2.2k	777	0.9948783610755440	1.0	0.9974326059050070	4.39803056191438E-06	0.9948783610755440	1.0	0.9974326059050070	4.39803056191438E-06
maa	San Jerónimo Tecóatl Mazatec	21,100	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
maj	Jalapa de Díaz Mazatec	24,200	788	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mam	Mam	842,000	793	0.9974811083123430	0.9987389659520810	0.9981096408317580	2.1990539669834E-06	0.9974811083123430	0.9987389659520810	0.9981096408317580	2.1990539669834E-06
maq	Chiquihuitlán Mazatec	169,400	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mav	Sateré-Mawé	7,134	742	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
maz	Central Mazahua	350k	782	0.9974457215836530	0.9987212276214830	0.9980830670926520	2.19902737019416E-06	0.998719590268886	0.9974424552429670	0.9980806142034550	1.09951368509708E-06
mbb	Western Bukidnon Manobo	19k	1000	0.9950149551345960	0.998	0.9965052421367950	5.49888647548871E-06	0.998	0.998	0.998	2.19955459019549E-06
mbc	Makushi	29,100	784	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mbh	Mangseng	2,500	789	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06
mbj	Nadëb	300	784	0.9961880559085130	1.0	0.9980903882877150	3.29854830888926E-06	0.9961880559085130	1.0	0.9980903882877150	3.29854830888926E-06
mbl	Maxakalí	800	790	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mbs	Sarangani Manobo	58k	932	0.9957264957264960	1.0	0.9978586723768740	4.39878021824548E-06	0.9989281886388000	1.0	0.9994638069705100	1.09969505456137E-06
mbt	Matigsalug Manobo	30k	789	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mca	Maca (Enxet-Maca)	1,500	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mcb	Machiguenga (Matsigenka)	10,100	766	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mcd	Sharanahua	438	735	1.0	0.998639455782313	0.9993192648059900	0.0	1.0	0.998639455782313	0.9993192648059900	0.0
mcf	Matsés (Mayoruna)	4,000	776	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mco	Coatlán Mixe	5,000	784	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06
mcp	Makaa	80k	792	1.0	0.9974747474747480	0.9987357774968400	0.0	1.0	0.9974747474747480	0.9987357774968400	0.0
mcq	Ese	10k	792	0.9949748743718590	1.0	0.9974811083123430	4.39810309813378E-06	0.9962216624685140	0.9987373737373740	0.9974779319041620	3.29857732360033E-06
mcr	Menya	20k	761	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mdy	Maale (Malé)	53,800	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
med	Melpa	130k	750	0.9960159362549800	1.0	0.998003992015968	3.29842500206152E-06	0.9960159362549800	1.0	0.998003992015968	3.29842500206152E-06
mee	Mengen	8,400	787	0.9974651457541190	1.0	0.9987309644670050	2.19903945956406E-06	0.9987293519695040	0.9987293519695040	0.9987293519695040	1.09951972978203E-06
mek	Mekeo	19,000	777	1.0	0.9987129987129990	0.9993560849967800	0.0	1.0	0.9987129987129990	0.9993560849967800	0.0
meq	Merey	10k	790	0.9974715549936790	0.9987341772151900	0.9981024667931690	2.19904671324981E-06	0.9987341772151900	0.9987341772151900	0.9987341772151900	1.0995233566249E-06
met	Mato	1,250	937	0.9989316239316240	0.9978655282817500	0.998398291510945	1.09970110124068E-06	0.9989304812834220	0.9967982924226250	0.9978632478632480	1.09970110124068E-06
meu	Motu	39k	1000	0.9969909729187560	0.994	0.9954932398597900	3.29933188529323E-06	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06
mgc	Morokodo	3,400	382	0.9973753280839900	0.9947643979057590	0.9960681520314550	1.09903032554377E-06	1.0	0.9947643979057590	0.9973753280839900	0.0
mgh	Makhuwa-Meetto	2.4m	794	1.0	0.9987405541561710	0.9993698802772530	0.0	1.0	0.9974811083123430	0.9987389659520810	0.0
mgw	Matumbi	72k	215	0.963963963963964	0.9953488372093020	0.9794050343249430	8.79062918928422E-06	0.9683257918552040	0.9953488372093020	0.981651376146789	7.6918005406237E-06
mib	Atatláhuca Mixtec	12,000	788	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mic	Mi'kmaq (Micmac)	8,145	789	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06	1.0	1.0	1.0	0.0
mie	Ocotepec Mixtec	6,500	790	0.9974747474747480	1.0	0.9987357774968400	2.19904671324981E-06	1.0	1.0	1.0	0.0
mig	San Miguel El Grande Mixtec	5,600	794	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mih	Chayuco Mixtec	10k	790	0.9987357774968390	1.0	0.9993674889310560	1.0995233566249E-06	0.9987357774968390	1.0	0.9993674889310560	1.0995233566249E-06
mil	Peñoles Mixtec	7.3k	788	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mio	Pinotepa Nacional Mixtec	20k	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mir	Isthmus Mixe	20k	742	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mit	Southern Puebla Mixtec	3,180	777	1.0	1.0	1.0	0.0	1.0	0.9987129987129990	0.9993560849967800	0.0
miz	Coatzospan Mixtec	2.1k	795	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06
mjc	San Juan Colorado Mixtec	7.82k	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mkj	Mokilese	1,230	602	0.9983416252072970	1.0	0.9991701244813280	1.09929612069392E-06	0.9983416252072970	1.0	0.9991701244813280	1.09929612069392E-06
mkl	Mokole	44k	787	1.0	1.0	1.0	0.0	1.0	0.9987293519695040	0.9993642720915450	0.0
mkn	Kupang Malay	200k	929	0.9646569646569650	0.9989235737351990	0.9814912744579590	3.73895085039138E-05	0.9666666666666670	0.9989235737351990	0.9825304393859190	3.51901256507424E-05
mks	Silacayoapan Mixtec	18.7k	795	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mle	Manambu	2,800	725	0.9986225895316810	1.0	0.9993108201240520	1.0994447803859E-06	0.9986225895316810	1.0	0.9993108201240520	1.0994447803859E-06
mlp	Bargam	4k	772	0.9961290322580650	1.0	0.9980607627666450	3.2985047877797E-06	0.9974160206718350	1.0	0.9987063389391980	2.19900319185313E-06
mmo	Mangga Buang	1,500	794	0.9949874686716790	1.0	0.9974874371859300	4.39811276981047E-06	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06
mmx	Madak	4k	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
mna	Mbula	4.5k	1000	0.9920398009950250	0.997	0.9945137157107230	8.79821836078194E-06	0.9930278884462150	0.997	0.9950099800399200	7.6984410656842E-06
mop	Mopán Maya	11,800	789	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06	1.0	0.9987325728770600	0.9993658845909960	0.0
mox	Molima	4,010	782	1.0	0.9948849104859340	0.9974358974358970	0.0	1.0	0.9948849104859340	0.9974358974358970	0.0
mph	Maung	240	58	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mpj	Martu Wangka	720	504	0.9960474308300400	1.0	0.998019801980198	2.19835541031754E-06	0.9960474308300400	1.0	0.998019801980198	2.19835541031754E-06
mpm	Yosondúa Mixtec	2k	788	0.998730964467005	0.998730964467005	0.998730964467005	1.099520938727E-06	0.998730964467005	0.998730964467005	0.998730964467005	1.099520938727E-06
mpp	Migabac	4k	784	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06	1.0	1.0	1.0	0.0
mps	Dadibi	13k	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06
mpt	Mian	1,750	763	0.9986910994764400	1.0	0.9993451211525870	1.09949071590039E-06	0.9986910994764400	1.0	0.9993451211525870	1.09949071590039E-06
mqb	Mbuko	13k	787	0.9936868686868690	1.0	0.9968334388853710	5.49759864891016E-06	0.9949367088607600	0.9987293519695040	0.9968294229549780	4.39807891912812E-06
mqj	Mamasa	89k	786	0.9961734693877550	0.9936386768447840	0.9949044585987260	3.29855556251917E-06	0.998719590268886	0.9923664122137400	0.9955328653477980	1.09951852083972E-06
msb	Masbatenyo	600k	790	0.986232790988736	0.9974683544303800	0.9918187539332910	1.20947569228739E-05	0.9886934673366830	0.9962025316455700	0.9924337957124840	9.89571020962413E-06
msc	Sankaran Maninka	127k	293	0.9932203389830510	1.0	0.9965986394557820	2.19784567167263E-06	0.9932203389830510	1.0	0.9965986394557820	2.19784567167263E-06
msk	Mansaka	57,800	775	0.9532595325953260	1.0	0.9760705289672540	4.17811984606927E-05	0.9591584158415840	1.0	0.9791535060012630	3.62836723474437E-05
msm	Agusan Manobo	81k	777	0.9847328244274810	0.9961389961389960	0.9904030710172740	1.31940916857431E-05	0.989769820971867	0.9961389961389960	0.9929441949967930	8.79606112382875E-06
msy	Aruamu	5,350	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
mti	Maiwa	41k	790	0.9962073324905180	0.9974683544303800	0.9968374446552820	3.29857006987471E-06	0.9962073324905180	0.9974683544303800	0.9968374446552820	3.29857006987471E-06
mto	Totontepec Mixe	5,200	788	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mux	Bo-Ung	30k	745	0.9959893048128340	1.0	0.9979906229068990	3.29840686948204E-06	0.9986595174262740	1.0	0.9993293091884640	1.09946895649401E-06
muy	Muyang	30k	788	0.9974651457541190	0.998730964467005	0.9980976537729870	2.19904187745399E-06	0.9974651457541190	0.998730964467005	0.9980976537729870	2.19904187745399E-06
mva	Manam	7,950	794	0.9974842767295600	0.9987405541561710	0.9981120201384520	2.19905638490524E-06	0.9974842767295600	0.9987405541561710	0.9981120201384520	2.19905638490524E-06
mvn	Minaveha	2,300	778	0.9987163029525030	1.0	0.9993577392421320	1.09950884939697E-06	1.0	1.0	1.0	0.0
mwc	Are	1,230	440	0.9909706546275400	0.9977272727272730	0.9943374858437150	4.39640154533514E-06	0.9909706546275400	0.9977272727272730	0.9943374858437150	4.39640154533514E-06
mwe	Mwera (Nyasa)	953k	794	0.9987341772151900	0.9937027707808560	0.9962121212121210	1.09952819245262E-06	1.0	0.9937027707808560	0.9968414403032220	0.0
mwf	Murrinh-Patha	1,430	400	0.9950248756218910	1.0	0.9975062344139650	2.1981041351834E-06	1.0	1.0	1.0	0.0

Table 11: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
mwp	Kala Lagaw Ya	700	575	0.9948096885813150	1.0	0.997398091934085	3.29779048037815E-06	0.9982638888888890	1.0	0.999131190269331	1.09926349345938E-06
mxb	Tezoatlán Mixtec	5,080	793	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06
mxp	Tlahuitoltepec Mixe	5k	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mxq	Juquila Mixe	8k	792	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06	1.0	1.0	1.0	0.0
mxt	Jamiltepec Mixtec	10k	791	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06	1.0	1.0	1.0	0.0
mya	Burmese	43m	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
myk	Mamara Senoufo	738k	805	0.9877300613496930	1.0	0.9938271604938270	1.09954149119817E-05	0.990159901599016	1.0	0.9950556242274410	8.79633192958536E-06
myu	Mundurukú	10,100	759	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
myw	Muyuw (Murua)	4,700	671	0.9985096870342770	0.9985096870342770	0.9985096870342770	1.09937951020444E-06	0.9985096870342770	0.9985096870342770	0.9985096870342770	1.09937951020444E-06
myy	Macuna	1,110	782	0.9961734693877550	0.9987212276214830	0.9974457215836530	3.29854105529125E-06	0.9961734693877550	0.9987212276214830	0.9974457215836530	3.29854105529125E-06
mzz	Maiadomu	730	65	0.9848484848484850	1.0	0.9923664122137400	1.09864756484767E-06	1.0	1.0	1.0	0.0
nab	Southern Nambikuára	720	682	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
naf	Nabak	16k	767	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nak	Nakanai	13k	786	0.9987293519695040	1.0	0.9993642720915450	1.09951852083972E-06	0.9987293519695040	1.0	0.9993642720915450	1.09951852083972E-06
nas	Naasioi	45k	798	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nay	Ngarrindjeri	0	39	0.975	1.0	0.9873417721518990	1.09861618305582E-06	1.0	1.0	1.0	0.0
nbq	Nggem	4,400	369	0.994579945799458	0.994579945799458	0.994579945799458	2.19802924697716E-06	0.994579945799458	0.994579945799458	0.994579945799458	2.19802924697716E-06
nca	Iyo	6,900	786	0.9949367088607600	1.0	0.9974619289340100	4.3980740833589E-06	0.9961977186311790	1.0	0.9980952380952380	3.29855556251917E-06
nch	Central Huasteca Nahuatl	200k	1000	0.9851924975320830	0.998	0.9915548931942370	1.64966594264661E-05	0.9871414441147380	0.998	0.9925410243659870	1.42971048362707E-05
ncj	Northern Puebla Nahuatl	40k	792	0.9949622166246850	0.9974747474747480	0.9962168978562420	4.39810309813378E-06	0.9949558638083230	0.9962121212121210	0.995583596214511	4.39810309813378E-06
ncl	Michoacán Nahuatl	4.7k	780	0.9987163029525030	0.9974358974358970	0.9980756895445800	1.09951126724171E-06	0.9987163029525030	0.9974358974358970	0.9980756895445800	1.09951126724171E-06
ncu	Chumburung	69k	763	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ndg	Ndengereko	72k	791	0.9750312109862670	0.9873577749683940	0.9811557788944720	2.19904913115569E-05	0.982367758186398	0.9860935524652340	0.9842271293375390	1.53933439180898E-05
ndj	Ndamba	55k	795	0.9874843554443050	0.9924528301886790	0.9899623588456710	1.09952940141619E-05	0.9899623588456710	0.9924528301886790	0.9912060301507540	8.79623521132955E-06
ngp	Ngulu	132k	794	0.9824120603015080	0.9848866498740550	0.9836477987421380	1.53933946943367E-05	0.9848866498740550	0.9848866498740550	0.9848866498740550	1.31943383094314E-05
ngu	Guerrero Nahuatl	150k	791	0.9974715549936790	0.9974715549936790	0.9974715549936790	2.19904913115569E-06	0.9987341772151900	0.9974715549936790	0.9981024667931690	1.09952456557784E-06
nhe	Eastern Huasteca Nahuatl	450k	772	0.9738903394255870	0.966321243523316	0.9700910273081920	2.19900319185313E-05	0.9738562091503270	0.9650259067357510	0.9694209499024070	2.19900319185313E-05
nhg	Tetelcingo Nahuatl	3,500	790	1.0	0.9949367088607600	0.9974619289340100	0.0	1.0	0.9949367088607600	0.9974619289340100	0.0
nhi	Zacatlán-Ahuacatlán-Tepetzintla Nahuatl	17,100	790	0.9974715549936790	0.9987341772151900	0.9981024667931690	2.19904671324981E-06	0.9974683544303800	0.9974683544303800	0.9974683544303800	2.19904671324981E-06
nho	Takuu	1,750	791	0.9669926650366750	1.0	0.9832193909260410	2.96871632706018E-05	0.9669926650366750	1.0	0.9832193909260410	2.96871632706018E-05
nhr	Naro	14k	790	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06	0.9974747474747480	1.0	0.9987357774968400	2.19904671324981E-06
nhu	Noone	25k	792	0.9924812030075190	1.0	0.9962264150943400	6.59715464720066E-06	0.9924812030075190	1.0	0.9962264150943400	6.59715464720066E-06
nhw	Western Huasteca Nahuatl	400k	1000	0.9719157472417250	0.969	0.9704556835252880	3.07937642627368E-05	0.9748490945674040	0.969	0.9719157472417250	2.74944323774436E-05
nhy	Northern Oaxaca Nahuatl	12k	792	0.9974683544303800	0.9949494949494950	0.9962073324905180	2.19905154906689E-06	0.9974683544303800	0.9949494949494950	0.9962073324905180	2.19905154906689E-06
nif	Nek	2k	371	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nii	Nii	12k	778	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nin	Ninzo	35k	794	0.99125	0.9987405541561710	0.9949811794228360	7.69669734716833E-06	0.9924906132665830	0.9987405541561710	0.9956057752667920	6.59716915471571E-06
nko	Nkonya	28k	782	0.991128010139417	1.0	0.9955442393380010	7.69659579567957E-06	0.9949109414758270	1.0	0.9974489795918370	4.39805474038833E-06
nlg	Gela	11,876	1000	0.9979959919839680	0.996	0.9969969969969970	2.19955459019549E-06	0.9979919678714860	0.994	0.9959919839679360	2.19955459019549E-06
nmw	Rifao	10k	122	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nna	Nyangumarta	260	209	0.9952380952380950	1.0	0.9976133651551310	1.0988214041619E-06	0.9952380952380950	1.0	0.9976133651551310	1.0988214041619E-06
nnq	Ngindo	220k	793	0.982521847690387	0.9924337957124840	0.9874529485570890	1.53933777688838E-05	0.9849812265331660	0.9924337957124840	0.9886934673366830	1.31943238019004E-05
noa	Waunana (Noanamá)	8,177	784	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nop	Numanggang	2,300	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
not	Nomatsiguenga	6,500	778	0.9974326059050060	0.9987146529562980	0.9980732177263970	2.19901769879395E-06	0.9987146529562980	0.9987146529562980	0.9987146529562980	1.09950884939697E-06
nou	Ewage-Notu	12,900	775	0.9987113402061860	1.0	0.9993552546744040	1.09950522264981E-06	0.9987096774193550	0.9987096774193550	0.9987096774193550	1.09950522264981E-06
npi	Nepali	20m	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
npl	Southeastern Puebla Nahuatl	135k	794	0.9924906132665830	0.9987405541561710	0.9956057752667920	6.59716915471571E-06	0.9962311557788940	0.9987405541561710	0.9974842767295600	3.29858457735786E-06
nsn	Nehan	6,500	788	0.9987325728770600	1.0	0.9993658845909960	1.099520938727E-06	0.9987325728770600	1.0	0.9993658845909960	1.099520938727E-06
nss	Nali	2,900	688	0.998546511627907	0.998546511627907	0.998546511627907	1.09940005738868E-06	0.998546511627907	0.998546511627907	0.998546511627907	1.09940005738868E-06
ntj	Ngaanyatjarra	990	1000	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06
ntp	Northern Tepehuan	6,200	772	1.0	0.9987046632124350	0.9993519118600130	0.0	1.0	0.9987046632124350	0.9993519118600130	0.0
ntu	Natügu	4,280	995	0.996993987975952	1.0	0.9984947315604620	3.29931374274151E-06	0.996993987975952	1.0	0.9984947315604620	3.29931374274151E-06
nuy	Nunggubuyu (Wubuy)	360	795	1.0	0.9987421383647800	0.9993706733794840	0.0	1.0	0.9974842767295600	0.9987405541561710	0.0
nvm	Namiae	3,500	724	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nwi	Southwest Tanna	4,500	783	1.0	1.0	1.0	0.0	1.0	0.9987228607918260	0.9993610223642170	0.0
nya	Chewa (Nyanja)	15m	1000	0.9930209371884350	0.996	0.9945082376435350	7.6984410656842E-06	0.9940119760479040	0.996	0.995004995004995	6.59866377058646E-06
nys	Noongar	0	116	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
obo	Obo Manobo	60k	788	0.9924433249370280	1.0	0.9962073324905180	6.59712563236198E-06	0.9924337957124840	0.998730964467005	0.9955724225173940	6.59712563236198E-06
okv	Orokaiva	47k	777	1.0	0.9987129987129990	0.9993560849967800	0.0	1.0	0.9974259974259980	0.9987113402061860	0.0
ong	Olo	13,700	786	0.9974619289340100	1.0	0.9987293519695050	2.19903704167945E-06	0.9974619289340100	1.0	0.9987293519695050	2.19903704167945E-06
ons	Ono	5,500	794	1.0	0.9987405541561710	0.9993698802772530	0.0	1.0	0.9987405541561710	0.9993698802772530	0.0
ood	Tohono O'odham	9,600	752	1.0	0.9960106382978720	0.9980013324450370	0.0	1.0	0.9960106382978720	0.9980013324450370	0.0
opm	Oksapmin	8,000	759	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ory	Odia (Oriya)	50m	1000	0.999	0.999	0.999	1.09977729509774E-06	0.999	0.999	0.999	1.09977729509774E-06
ote	Mezquital Otomí	100k	794	0.9974811083123430	0.9974811083123430	0.9974811083123430	2.19905638490524E-06	0.9974811083123430	0.9974811083123430	0.9974811083123430	2.19905638490524E-06
otm	Eastern Highland Otomí	70k	787	1.0	0.9987293519695040	0.9993642720915450	0.0	1.0	0.9987293519695040	0.9993642720915450	0.0
otn	Tenango Otomí	16k	786	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
otq	Querétaro Otomí	33k	794	1.0	0.9987405541561710	0.9993698802772530	0.0	1.0	0.9987405541561710	0.9993698802772530	0.0
ots	Estado de México Otomí	10k	789	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06
pab	Parecís	1,800	793	0.9974811083123430	0.9987389659520810	0.9981096408317580	2.1990539669834E-06	0.9987389659520810	0.9987389659520810	0.9987389659520810	1.0995269834917E-06
pad	Paumarí	290	755	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
pah	Tenharim (Kagwahiva)	350	749	0.9986648865153540	0.9986648865153540	0.9986648865153540	1.09947379184322E-06	0.9986648865153540	0.9986648865153540	0.9986648865153540	1.09947379184322E-06
pan	Punjabi (Eastern)	37m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
pao	Northern Paiute	700	698	0.997134670487106	0.997134670487106	0.997134670487106	2.19882428865286E-06	0.9985652797704450	0.997134670487106	0.9978494623655910	1.09941214432643E-06
pes	Iranian Persian	83m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
pib	Yine (Piro)	3,490	789	0.9974683544303800	0.9987325728770600	0.9981000633312220	2.19904429534924E-06	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06
pio	Piapoco	6,380	767	0.9948119325551230	1.0	0.9973992197659300	4.397982205764E-06	0.9961038961038960	1.0	0.9980481457384520	3.298486654323E-06
pir	Piratapuyo	1,070	770	0.9948253557567920	0.9987012987012990	0.9967595593000650	4.39799671249746E-06	0.9948253557567920	0.9987012987012990	0.9967595593000650	4.39799671249746E-06
piu	Pintupi-Luritja	390	655	0.9954407294832830	1.0	0.9977151561309980	3.29808051713903E-06	0.9954407294832830	1.0	0.9977151561309980	3.29808051713903E-06
pjt	Pitjantjatjara	2120	1000	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06
pls	San Marcos Tlacoyalco Popoloca	11k	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
plu	Palikúr	1,420	787	0.9974619289340100	0.9987293519695040	0.998095238095238	2.19903945956406E-06	0.9974619289340100	0.9987293519695040	0.998095238095238	2.19903945956406E-06
pma	Paama	6k	1000	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06
poe	San Juan Atzingo Popoloca	5k	794	0.9987421383647800	1.0	0.9993706733794840	1.09952819245262E-06	1.0	1.0	1.0	0.0
poh	Poqomchi'	92k	767	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
poi	Highland Popoluca (Sierra Popoluca)	30k	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
pol	Polish	43m	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.995	0.9974937343358400	0.0
poy	Pogolo	185k	795	0.9949811794228360	0.9974842767295600	0.9962311557788940	4.39811760566478E-06	0.9987405541561710	0.9974842767295600	0.9981120201384520	1.09952940141619E-06
ppo	Folopa	3k	766	0.9973958333333330	1.0	0.998696219035202	2.19898868510372E-06	1.0	1.0	1.0	0.0
prf	Paranan	16,700	793	0.9924906132665830	1.0	0.9962311557788940	6.59716190095021E-06	0.9924812030075190	0.9987389659520810	0.9956002514142050	6.59716190095021E-06
pri	Paicî	6,500	761	1.0	0.9986859395532200	0.9993425378040760	0.0	1.0	0.9986859395532200	0.9993425378040760	0.0
ptp	Patep	1,700	792	1.0	0.9987373737373740	0.9993682880606440	0.0	1.0	0.9987373737373740	0.9993682880606440	0.0
ptu	Bambam	42,100	927	0.9914255091103970	0.9978425026968720	0.9946236559139790	8.79751206358842E-06	0.9935553168635880	0.9978425026968720	0.9956942949407970	6.59813404769131E-06
pwg	Gapapaiwa	3k	943	0.9957492029755580	0.9936373276776250	0.994692144373673	4.39883342937453E-06	0.9968085106382980	0.9936373276776250	0.9952203929899100	3.2991250720309E-06
qub	Huallaga Huánuco Quechua	24k	1000	0.9632139399806390	0.995	0.9788489916379730	4.17915372137142E-05	0.9688412852969820	0.995	0.9817464232856440	3.51928734431278E-05
quc	K'iche'	1.1m	795	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
quf	Lambayeque Quechua	20k	778	0.9974326059050060	0.9987146529562980	0.9980732177263970	2.19901769879395E-06	0.9974326059050060	0.9987146529562980	0.9980732177263970	2.19901769879395E-06
quh	South Bolivian Quechua	2.7m	788	0.9702602230483270	0.9936548223350250	0.9818181818181820	2.63885025294479E-05	0.9726708074534160	0.9936548223350250	0.9830508474576270	2.41894606519939E-05
qul	North Bolivian Quechua	116k	790	0.9948186528497410	0.9721518987341770	0.9833546734955190	4.39809342649961E-06	0.9948119325551230	0.970886075949367	0.9827033952594490	4.39809342649961E-06
qup	Southern Pastaza Quechua	1550	764	0.9947780678851180	0.9973821989528800	0.996078431372549	4.39796769912623E-06	0.9947780678851180	0.9973821989528800	0.996078431372549	4.39796769912623E-06
qvc	Cajamarca Quechua	30k	778	1.0	0.9987146529562980	0.9993569131832800	0.0	1.0	0.9987146529562980	0.9993569131832800	0.0
qve	Eastern Apurímac Quechua	200k	790	0.9974747474747480	1.0	0.9987357774968400	2.19904671324981E-06	0.9974747474747480	1.0	0.9987357774968400	2.19904671324981E-06
qvh	Huamalíes-Dos de Mayo Huánuco Quechua	72,400	726	0.9676113360323890	0.987603305785124	0.9775051124744380	2.63867037399854E-05	0.9715061058344640	0.9862258953168040	0.9788106630211890	2.30883657724872E-05
qvm	Margos-Yarowilca-Lauricocha Quechua	83,400	725	0.9806629834254140	0.9793103448275860	0.9799861973775020	1.53922269254027E-05	0.9806629834254140	0.9793103448275860	0.9799861973775020	1.53922269254027E-05
qvn	North Junín Quechua	60k	794	0.9073226544622430	0.9987405541561710	0.9508393285371700	8.90617835886621E-05	0.9146482122260670	0.9987405541561710	0.954846478025286	8.13650862414938E-05
qvs	San Martín Quechua	15k	821	0.9975669099756690	0.9987819732034100	0.9981740718198420	2.19912167080468E-06	0.998780487804878	0.997563946406821	0.9981718464351010	1.09956083540234E-06
qvw	Huaylla Wanca Quechua	250k	758	0.9934469200524250	1.0	0.9967126890203810	5.49742335767226E-06	0.994750656167979	1.0	0.9973684210526320	4.39793868613781E-06
qvz	Northern Pastaza Quichua	14k	780	0.9974424552429670	1.0	0.998719590268886	2.19902253448342E-06	0.9974424552429670	1.0	0.998719590268886	2.19902253448342E-06
qwh	Huaylas Ancash Quechua	336k	783	0.9987228607918260	0.9987228607918260	0.9987228607918260	1.09951489402875E-06	0.9987228607918260	0.9987228607918260	0.9987228607918260	1.09951489402875E-06

Table 12: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
qxh	Panao Quechua	64k	770	0.9758576874205850	0.9974025974025970	0.9865125240847780	2.08904843843629E-05	0.9795918367346940	0.9974025974025970	0.9884169884169880	1.75919868499898E-05
qxn	Northern Conchucos Ancash Quechua	250k	792	0.9726708074534160	0.9886363636363640	0.9805886036318100	2.41895670397358E-05	0.975093399750934	0.9886363636363640	0.9818181818181820	2.19905154906689E-05
qxo	Southern Conchucos Ancash Quechua	250k	739	0.9613848202396800	0.9769959404600810	0.9691275167785240	3.18843894029483E-05	0.963903743315508	0.9756427604871450	0.9697377269670480	2.96854659958484E-05
rai	Ramoaaina	10,300	783	0.9974457215836530	0.9974457215836530	0.9974457215836530	2.19902978805751E-06	0.9987212276214830	0.9974457215836530	0.9980830670926520	1.09951489402875E-06
reg	Kara	100k	793	0.9850931677018630	1.0	0.9924906132665830	1.31943238019004E-05	0.9863013698630140	0.9987389659520810	0.9924812030075190	1.20947968184087E-05
rgu	Ringgou (Rikou)	12,000	817	0.9938875305623470	0.9951040391676870	0.9944954128440370	5.49777999643744E-06	0.9938650306748470	0.9914320685434520	0.9926470588235290	5.49777999643744E-06
rkb	Rikbaktsá	910	765	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
rmc	Carpathian Romani	472,470	114	0.991304347826087	1.0	0.9956331877729260	1.09870671232892E-06	0.9912280701754390	0.9912280701754390	0.9912280701754390	1.09870671232892E-06
rmy	Vlax Romani	885,970	228	0.991304347826087	1.0	0.9956331877729260	2.19768869080388E-06	0.9956331877729260	1.0	0.9978118161925600	1.09884434540194E-06
roo	Rotokas	4,320	786	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
rop	Kriol	7,500	1000	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
row	Dela-Oenale	7,000	902	0.9944258639910810	0.9889135254988910	0.9916620344635910	5.49829387940922E-06	0.9955357142857140	0.9889135254988910	0.9922135706340380	4.39863510352738E-06
rro	Waima	15k	794	0.9987405541561710	0.9987405541561710	0.9987405541561710	1.09952819245262E-06	0.9987405541561710	0.9987405541561710	0.9987405541561710	1.09952819245262E-06
ruf	Luguru	692k	794	0.992462311557789	0.9949622166246850	0.9937106918238990	6.59716915471571E-06	0.9962168978562420	0.9949622166246850	0.9955891619407690	3.29858457735786E-06
rug	Roviana	9,870	1000	0.9950149551345960	0.998	0.9965052421367950	5.49888647548871E-06	0.996003996003996	0.997	0.9965017491254370	4.39910918039097E-06
sab	Buglere (Bocotá)	2,500	724	1.0	1.0	1.0	0.0	1.0	0.9986187845303870	0.9993089149965450	0.0
sbe	Saliba	2,500	788	0.9961977186311790	0.9974619289340100	0.9968294229549780	3.29856281618099E-06	0.9961977186311790	0.9974619289340100	0.9968294229549780	3.29856281618099E-06
sbk	Safwa	158,000	792	0.9788557213930350	0.9936868686868690	0.9862155388471180	1.86919381670685E-05	0.9800747198007470	0.9936868686868690	0.986833855799373	1.75924123925351E-05
sbs	Subiya (Kuhane)	35,090	792	0.9738805970149250	0.9886363636363640	0.981203007518797	2.30900412652023E-05	0.9811320754716980	0.9848484848484850	0.9829867674858220	1.64928866180017E-05
seh	Sena	2.8m	196	0.9848484848484850	0.9948979591836740	0.9898477157360410	3.29641712422768E-06	0.9898477157360410	0.9948979591836740	0.9923664122137400	2.19761141615178E-06
sey	Secoya (Paicoca)	1,174	791	1.0	1.0	1.0	0.0	1.0	0.9987357774968390	0.9993674889310560	0.0
sgb	Mag-antsi Ayta	4,200	785	0.9727385377943000	1.0	0.9861809045226130	2.41893808618017E-05	0.9800249687890140	1.0	0.9899117276166460	1.75922769904012E-05
sgz	Sursurunga	3,000	773	0.9987080103359170	1.0	0.9993535875888820	1.09950280483166E-06	0.9987080103359170	1.0	0.9993535875888820	1.09950280483166E-06
shj	Shatt	30k	115	1.0	0.991304347826087	0.9956331877729260	0.0	1.0	0.991304347826087	0.9956331877729260	0.0
shp	Shipibo-Konibo	26k	976	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
sim	Mende	3,180	778	1.0	0.9987146529562980	0.9993569131832800	0.0	1.0	0.9974293059125960	0.9987129987129990	0.0
sja	Epena (Saija)	3,550	786	0.9987277353689570	0.9987277353689570	0.9987277353689570	1.09951852083972E-06	1.0	0.9974554707379140	0.9987261146496820	0.0
sll	Salt-Yui (Yui)	6,500	786	1.0	0.9987277353689570	0.9993634627625720	0.0	1.0	0.9987277353689570	0.9993634627625720	0.0
smk	Bolinao	50k	787	0.9874371859296480	0.9987293519695040	0.9930511686670880	1.09951972978203E-05	0.9899244332493700	0.9987293519695040	0.9943074003795070	8.79615783825625E-06
snc	Sinaugoro	15k	792	0.9974747474747480	0.9974747474747480	0.9974747474747480	2.19905154906689E-06	0.9987357774968390	0.9974747474747480	0.9981048641819330	1.09952577453344E-06
snn	Siona	200	786	0.9987293519695040	1.0	0.9993642720915450	1.09951852083972E-06	1.0	1.0	1.0	0.0
snp	Siane	29k	783	0.9987244897959180	1.0	0.9993618379068280	1.09951489402875E-06	1.0	1.0	1.0	0.0
snx	Sam	780	66	1.0	0.9696969696969700	0.9846153846153850	0.0	1.0	0.9696969696969700	0.9846153846153850	0.0
sny	Saniyo-Hiyewe	644	776	0.9974293059125960	1.0	0.9987129987129990	2.19901286312574E-06	0.9974293059125960	1.0	0.9987129987129990	2.19901286312574E-06
som	Somali	24m	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	0.999	0.999	0.999	1.09977729509774E-06
soq	Kanasi	2,400	1000	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
soy	Miyobe	8,700	792	0.9949748743718590	1.0	0.9974811083123430	4.39810309813378E-06	0.9949748743718590	1.0	0.9974811083123430	4.39810309813378E-06
spl	Selepet	7,200	687	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
spm	Akukem (Sepen)	409	162	0.9938650306748470	1.0	0.9969230769230770	1.09876465889401E-06	0.9938650306748470	1.0	0.9969230769230770	1.09876465889401E-06
spp	Supyire Sénoufo	350k	944	0.996832101372756	1.0	0.9984135378106820	3.2991287001103E-06	0.9989417989417990	1.0	0.9994706193753310	1.09970956670343E-06
sps	Saposa	3,100	814	0.9975460122699390	0.9987714987714990	0.9981583793738490	2.19910474445853E-06	0.9975460122699390	0.9987714987714990	0.9981583793738490	2.19910474445853E-06
spy	Sabaot	279,000	780	0.9987179487179490	0.9987179487179490	0.9987179487179490	1.09951126724171E-06	1.0	0.9987179487179490	0.9993585631815270	0.0
sri	Siriano	217	780	0.9987179487179490	0.9987179487179490	0.9987179487179490	1.09951126724171E-06	0.9987179487179490	0.9987179487179490	0.9987179487179490	1.09951126724171E-06
srm	Saramaccan	23k	774	0.9987096774193550	1.0	0.9993544222078760	1.0995040137394E-06	0.9987096774193550	1.0	0.9993544222078760	1.0995040137394E-06
srn	Sranan Tongo	519,600	789	0.9962121212121210	1.0	0.9981024667931690	3.29856644302386E-06	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06
srq	Sirionó	400	781	0.998719590268886	0.998719590268886	0.998719590268886	1.09951247616807E-06	1.0	0.998719590268886	0.9993593850096090	0.0
ssd	Siroi	1,310	1000	0.996	0.996	0.996	4.39910918039097E-06	0.996	0.996	0.996	4.39910918039097E-06
ssg	Seimat	1,400	789	0.9924242424242420	0.9961977186311790	0.9943074003795070	6.59713288604772E-06	0.9961977186311790	0.9961977186311790	0.9961977186311790	3.29856644302386E-06
ssx	Samberigi	3,130	940	1.0	0.9968085106382980	0.9984017048481620	0.0	1.0	0.9968085106382980	0.9984017048481620	0.0
stp	Southeastern Tepehuan	36,543	785	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
sua	Sulka	2,500	738	0.9986468200270640	1.0	0.9993229519295870	1.09946049473523E-06	1.0	1.0	1.0	0.0
sue	Suena	3,000	788	0.9962073324905180	1.0	0.9981000633312220	3.29856281618099E-06	0.9974683544303800	1.0	0.9987325728770600	2.19904187745399E-06
sus	Susu	2.4m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
suz	Sunwar	32,708	1000	1.0	1.0	1.0	0.0	1.0	0.999	0.9994997498749370	0.0
swh	Swahili	200m	1000	0.9850894632206760	0.991	0.9880358923230310	1.64966594264661E-05	0.988011988011988	0.989	0.9885057471264370	1.31973275411729E-05
swp	Suau	20k	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
sxb	Suba	174k	790	0.9875	1.0	0.9937106918238990	1.0995233566249E-05	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06
tac	Lowland Tarahumara (Western Tarahumara)	39,800	785	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
taj	Eastern Tamang	759k	788	0.986232790988736	1.0	0.9930686830497800	1.2094730325997E-05	0.9886934673366830	0.998730964467005	0.9936868686868690	9.89568844854297E-06
tam	Tamil	80m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tav	Tatuyo	331	756	0.9973579920739760	0.9986772486772490	0.9980171844018510	2.19896450761337E-06	0.9986754966887420	0.9973544973544970	0.9980145598941100	1.09948225380668E-06
taw	Tai (Tay)	900	767	0.9986979166666670	1.0	0.9993485342019550	1.099495551441E-06	0.9986979166666670	1.0	0.9993485342019550	1.099495551441E-06
tbc	Takia	40k	783	0.9974522292993630	1.0	0.9987244897959180	2.19902978805751E-06	0.9987244897959180	1.0	0.9993618379068280	1.09951489402875E-06
tbf	Mandara	4,000	792	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06
tbl	Tboli	95,300	107	0.9724770642201840	0.9906542056074770	0.9814814814814820	3.29609478689649E-06	0.9814814814814820	0.9906542056074770	0.986046511627907	2.19739652459766E-06
tbo	Tawala	20k	789	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06
tbz	Ditammari	227k	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tca	Ticuna	50,102	769	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tcs	Torres Strait Creole (Yumplatok)	6,000	955	0.9989528795811520	0.9989528795811520	0.9989528795811520	1.0997228698368E-06	0.9989528795811520	0.9989528795811520	0.9989528795811520	1.0997228698368E-06
tcz	Thado Chin	115,045	1000	0.997002997002997	0.998	0.9975012493753120	3.29933188529323E-06	1.0	0.998	0.998998998998999	0.0
tdt	Tetun Dili	944k	783	0.9949109414758270	0.9987228607918260	0.9968132568514980	4.39805957611502E-06	0.9949109414758270	0.9987228607918260	0.9968132568514980	4.39805957611502E-06
tee	Huehuetla Tepehua	3k	792	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06	1.0	1.0	1.0	0.0
tel	Telugu	97m	1000	0.9727626459143970	1.0	0.9861932938856020	3.07937642627368E-05	0.9727626459143970	1.0	0.9861932938856020	3.07937642627368E-05
ter	Terêna	19k	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tet	Tetum	1m	902	0.9988851727982160	0.9933481152993350	0.9961089494163420	1.09965877588184E-06	0.9988839285714290	0.9922394678492240	0.9955506117908790	1.09965877588184E-06
tew	Tewa	1,300	482	0.9958677685950410	1.0	0.9979296066252590	2.19830225117142E-06	0.9979296066252590	1.0	0.9989637305699480	1.09915112558571E-06
tfr	Teribe (Naso)	3,005	904	0.9977924944812360	1.0	0.9988950276243090	2.19932238877202E-06	0.9977900552486190	0.9988938053097350	0.9983416252072970	2.19932238877202E-06
tgk	Tajik	8m	432	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tgl	Tagalog	90m	1000	0.9871159563924680	0.996	0.9915380786460930	1.42971048362707E-05	0.9880834160873880	0.995	0.9915296462381660	1.31973275411729E-05
tgo	Sudest	1,693	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
tgp	Tangoa	370	793	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06
tha	Thai	70m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
thd	Kuuk Thaayorre	150	17	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tif	Tifal	2,800	738	0.9959514170040490	1.0	0.9979716024340770	3.2983814842057E-06	0.9986468200270640	1.0	0.9993229519295870	1.09946049473523E-06
tim	Timbe	11k	778	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tiw	Tiwi	1,830	105	1.0	0.9904761904761910	0.9952153110047850	0.0	1.0	0.9904761904761910	0.9952153110047850	0.0
tiy	Tiruray (Teduray)	50k	790	0.9974715549936790	0.9987341772151900	0.9981024667931690	2.19904671324981E-06	0.9974715549936790	0.9987341772151900	0.9981024667931690	2.19904671324981E-06
tke	Takwane	181k	115	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tku	Upper Necaxa Totonac	3,400	791	0.9936868686868690	0.9949431099873580	0.9943145925457990	5.49762282788922E-06	0.9936788874841970	0.9936788874841970	0.9936788874841970	5.49762282788922E-06
tlf	Telefol	4,500	743	0.9986541049798120	0.9986541049798120	0.9986541049798120	1.09946653883536E-06	0.9986541049798120	0.9986541049798120	0.9986541049798120	1.09946653883536E-06
tmd	Haruai	100	772	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tna	Tacana	1,820	768	0.9948186528497410	1.0	0.9974025974025970	4.39798704133118E-06	0.9974025974025970	1.0	0.9986996098829650	2.19899352066559E-06
tnc	Tanimuca-Retuarã	1952	298	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tnk	Kwamera	3,500	787	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tnn	North Tanna	5,000	785	0.9974587039390090	1.0	0.9987277353689570	2.19903462380015E-06	0.9974587039390090	1.0	0.9987277353689570	2.19903462380015E-06
tnp	Whitesands	7,500	785	1.0	0.9987261146496820	0.9993626513703000	0.0	1.0	0.9987261146496820	0.9993626513703000	0.0
toc	Coyutla Totonac	48,100	785	0.9974587039390090	1.0	0.9987277353689570	2.19903462380015E-06	0.9974587039390090	1.0	0.9987277353689570	2.19903462380015E-06
tod	Toma (Loma)	144,000	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
tof	Gizrra	1,050	887	0.9977477477477480	0.9988726042841040	0.9983098591549300	2.19928127487937E-06	0.9988726042841040	0.9988726042841040	0.9988726042841040	1.09964063743968E-06
toj	Tojol-ab'al	36,000	784	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ton	Tongan	126k	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
too	Xicotepec de Juárez Totonac	3,000	792	0.9949431099873580	0.9936868686868690	0.9943145925457990	4.39810309813378E-06	0.9949431099873580	0.9936868686868690	0.9943145925457990	4.39810309813378E-06
top	Papantla Totonac	80,000	787	0.9987293519695040	0.9987293519695040	0.9987293519695040	1.09951972978203E-06	0.9987293519695040	0.9987293519695040	0.9987293519695040	1.09951972978203E-06
tos	Highland Totonac	120,000	777	0.9987146529562980	1.0	0.9993569131832800	1.09950764047859E-06	0.9987146529562980	1.0	0.9993569131832800	1.09950764047859E-06
tpa	Taupota	1,280	66	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tpi	Tok Pisin	4m	1000	0.999	0.999	0.999	1.09977729509774E-06	1.0	0.999	0.9994997498749370	0.0
tpt	Tlachichilco Tepehua	3,000	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06
tpz	Tinputz	3,900	784	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06	1.0	1.0	1.0	0.0
trc	Copala Triqui	25,000	783	0.9987244897959180	1.0	0.9993618379068280	1.09951489402875E-06	0.9987244897959180	1.0	0.9993618379068280	1.09951489402875E-06
tsw	Tsishingini	100k	790	0.98625	0.9987341772151900	0.9924528301886790	1.20947569228739E-05	0.9887218045112780	0.9987341772151900	0.9937027707808560	9.89571020962413E-06

Table 13: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
ttc	Tektiteko	5,900	795	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tte	Bwanabwana	2,400	704	0.9929378531073450	0.9985795454545450	0.9957507082152970	5.49709698308323E-06	0.9929278642149930	0.9971590909090910	0.9950389794472010	5.49709698308323E-06
tue	Tuyuca	1442	779	0.9974391805377720	1.0	0.9987179487179490	2.19902011663603E-06	0.9987163029525030	0.9987163029525030	0.9987163029525030	1.09951005831801E-06
tuf	Central Tunebo (U'wa)	3,550	786	0.9974619289340100	1.0	0.9987293519695050	2.19903704167945E-06	0.9974619289340100	1.0	0.9987293519695050	2.19903704167945E-06
tuo	Tucano	13,996	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
tur	Turkish	90m	1000	0.9930348258706470	0.998	0.9955112219451370	7.6984410656842E-06	0.9950149551345960	0.998	0.9965052421367950	5.49888647548871E-06
tvk	Southeast Ambrym (Vatlongos)	3,700	800	1.0	0.99875	0.9993746091307070	0.0	1.0	0.9975	0.9987484355444310	0.0
twi	Twi	20m	1000	0.992992992992993	0.992	0.9924962481240620	7.6984410656842E-06	0.9949748743718590	0.99	0.9924812030075190	5.49888647548871E-06
txq	Tii	20,000	898	0.9844961240310080	0.9899777282850780	0.9872293170460860	1.5395155144676E-05	0.9877641824249170	0.9888641425389760	0.988313856427379	1.20961933279597E-05
txu	Kayapó (Mebêngôkre)	7,100	750	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tzj	Tz'utujil	83,800	782	0.9974489795918370	1.0	0.9987228607918260	2.19902737019416E-06	0.9974489795918370	1.0	0.9987228607918260	2.19902737019416E-06
tzo	Tzotzil	260k	776	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ubr	Ubir	2,560	942	0.9957627118644070	0.9978768577494690	0.9968186638388120	4.39882859194597E-06	0.997874601487779	0.9968152866242040	0.997344662772172	2.19941429597298E-06
udu	Uduk	22k	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	1.0	1.0	1.0	0.0
uig	Uyghur	13m	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	0.999	0.999	0.999	1.09977729509774E-06
ukr	Ukrainian	51m	962	0.998960498960499	0.998960498960499	0.998960498960499	1.0997313356347E-06	0.998960498960499	0.998960498960499	0.998960498960499	1.0997313356347E-06
uli	Ulithian	3,075	790	0.9962168978562420	1.0	0.998104864181933	3.29857006987471E-06	0.9974747474747480	1.0	0.9987357774968400	2.19904671324981E-06
ulk	Meriam Mir	320	377	0.9973474801061010	0.9973474801061010	0.9973474801061010	1.09902428623868E-06	1.0	0.9973474801061010	0.99867197875166	0.0
upv	Uripiv-Wala-Rano-Atchin	9,000	791	0.9962216624685140	1.0	0.9981072555205050	3.29857369673353E-06	0.9962216624685140	1.0	0.9981072555205050	3.29857369673353E-06
ura	Urarina	3,000	772	0.9987046632124350	0.9987046632124350	0.9987046632124350	1.09950159592657E-06	0.9987046632124350	0.9987046632124350	0.9987046632124350	1.09950159592657E-06
urb	Urubú-Ka'apor (Ka'apor)	800	767	0.9960988296488950	0.9986962190352020	0.9973958333333330	3.298486654323E-06	1.0	0.9986962190352020	0.9993476842791910	0.0
uri	Urim	3,740	249	0.996	1.0	0.9979959919839680	1.09886970262388E-06	0.996	1.0	0.9979959919839680	1.09886970262388E-06
urt	Urat	6,280	782	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
urw	Sob (Sop)	2,500	118	0.9915254237288140	0.9915254237288140	0.9915254237288140	1.0987115409759E-06	0.9915254237288140	0.9915254237288140	0.9915254237288140	1.0987115409759E-06
usa	Usarufa	1,300	812	0.996319018404908	1.0	0.9981561155500920	3.29864986261123E-06	0.996319018404908	1.0	0.9981561155500920	3.29864986261123E-06
usp	Uspanteko	3,000	789	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06	0.9987325728770600	0.9987325728770600	0.9987325728770600	1.09952214767462E-06
uvh	Uri	2,500	749	0.9986666666666670	1.0	0.9993328885923950	1.09947379184322E-06	0.9986666666666670	1.0	0.9993328885923950	1.09947379184322E-06
uvl	Lote	6,000	790	0.9987341772151900	0.9987341772151900	0.9987341772151900	1.0995233566249E-06	1.0	0.9974683544303800	0.9987325728770600	0.0
vid	Vidunda	32,000	1000	0.9919839679358720	0.99	0.990990990990991	8.79821836078194E-06	0.9939698492462310	0.989	0.9914786967418550	6.59866377058646E-06
vie	Vietnamese	86m	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
viv	Iduna	5,560	744	1.0	0.9973118279569890	0.9986541049798120	0.0	1.0	0.9973118279569890	0.9986541049798120	0.0
vmy	Ayautla Mazatec	3,700	794	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06
waj	Waffa	1,990	783	0.9974522292993630	1.0	0.9987244897959180	2.19902978805751E-06	0.9974522292993630	1.0	0.9987244897959180	2.19902978805751E-06
wal	Wolaytta	2.7m	1000	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06
wap	Wapishana	10,900	788	0.9987325728770600	1.0	0.9993658845909960	1.099520938727E-06	0.9987325728770600	1.0	0.9993658845909960	1.099520938727E-06
wat	Kaninuwa	360	259	0.9810606060606060	1.0	0.9904397705544930	5.49440888951403E-06	0.9810606060606060	1.0	0.9904397705544930	5.49440888951403E-06
wbi	Vwanji	28,000	793	0.9874686716791980	0.9936948297604040	0.9905719673161530	1.0995269834917E-05	0.9924433249370280	0.9936948297604040	0.9930686830497800	6.59716190095021E-06
wbp	Warlpiri	2,670	932	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
wed	Wedau	2,200	408	0.9975490196078430	0.9975490196078430	0.9975490196078430	1.09906173100025E-06	1.0	0.9975490196078430	0.9987730061349690	0.0
wer	Weri	4,160	785	1.0	0.9987261146496820	0.9993626513703000	0.0	1.0	0.9987261146496820	0.9993626513703000	0.0
wim	Wik-Mungkan	840	748	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
wiu	Wiru	15,300	781	0.9974457215836530	1.0	0.9987212276214830	2.19902495233613E-06	0.9974457215836530	1.0	0.9987212276214830	2.19902495233613E-06
wiv	Vitu	7,000	792	0.9937106918238990	0.9974747474747480	0.9955891619407690	5.49762887266722E-06	0.9949622166246850	0.9974747474747480	0.9962168978562420	4.39810309813378E-06
wmt	Walmajarri	860	97	0.9797979797979800	1.0	0.989795918367347	2.19737238210548E-06	0.9797979797979800	1.0	0.989795918367347	2.19737238210548E-06
wmw	Mwani	167,150	793	0.9862155388471180	0.9924337957124840	0.9893148962916400	1.20947968184087E-05	0.9899244332493700	0.9911727616645650	0.9905482041587900	8.79621586793362E-06
wnc	Wantoat	8,200	1000	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06
wnu	Usan	1,400	772	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
wol	Wolof	18m	1000	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06	0.9979879275653920	0.992	0.9949849548645940	2.19955459019549E-06
wos	Hanga Hundi	7,200	782	0.9974489795918370	1.0	0.9987228607918260	2.19902737019416E-06	0.9987228607918260	1.0	0.9993610223642170	1.09951368509708E-06
wrk	Garrwa	110	411	0.9975728155339810	1.0	0.9987849331713240	1.09906535482226E-06	1.0	1.0	1.0	0.0
wro	Worrorra	20	182	1.0	0.9945054945054950	0.9972451790633610	0.0	1.0	0.9945054945054950	0.9972451790633610	0.0
wrs	Waris	2,500	895	0.9988839285714290	1.0	0.9994416527079840	1.09965031120104E-06	0.9988839285714290	1.0	0.9994416527079840	1.09965031120104E-06
wsk	Waskia	20,000	1000	0.9960159362549800	1.0	0.998003992015968	4.39910918039097E-06	0.9970089730807580	1.0	0.9985022466300550	3.29933188529323E-06
wuv	Wuvulu-Aua	1,600	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
xav	Xavánte	13,303	699	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
xed	Hdi	29,000	794	0.9962264150943400	0.9974811083123430	0.9968533668974200	3.29858457735786E-06	0.9962264150943400	0.9974811083123430	0.9968533668974200	3.29858457735786E-06
xla	Kamula	800	961	0.998960498960499	1.0	0.999479979199168	1.09973012622702E-06	0.998960498960499	1.0	0.999479979199168	1.09973012622702E-06
xnn	Northern Kankanay	70,000	772	0.9834183673469390	0.9987046632124350	0.9910025706940870	1.42935207470454E-05	0.9834183673469390	0.9987046632124350	0.9910025706940870	1.42935207470454E-05
xon	Konkomba	920,000	786	0.9886792452830190	1.0	0.9943074003795070	9.89566668755752E-06	0.9924242424242420	1.0	0.9961977186311790	6.59711112503835E-06
xsi	Sio	3,500	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	1.0	0.999	0.9994997498749370	0.0
xtd	Diuxi-Tilantongo Mixtec	4,220	788	1.0	0.9974619289340100	0.9987293519695050	0.0	1.0	0.9974619289340100	0.9987293519695050	0.0
xtm	Magdalena Peñasco Mixtec	7,350	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
yaa	Yaminahua	1,390	756	0.9986789960369880	1.0	0.9993390614672840	1.09948225380668E-06	0.9986789960369880	1.0	0.9993390614672840	1.09948225380668E-06
yad	Yagua	4,297	741	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
yal	Yalunka	181k	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
yap	Yapese	6,590	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
yaq	Yaqui	16,000	793	0.993734335839599	1.0	0.9968573224387180	5.49763491745851E-06	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06
yby	Yaweyuha	4,600	787	0.9974587039390090	0.9974587039390090	0.9974587039390090	2.19903945956406E-06	0.9987277353689570	0.9974587039390090	0.9980928162746340	1.09951972978203E-06
ycn	Yucuna	770	777	1.0	1.0	1.0	0.0	1.0	0.9987129987129990	0.9993560849967800	0.0
yka	Yakan	106,000	800	0.9937888198757760	1.0	0.9968847352024920	5.49767723136975E-06	0.9937888198757760	1.0	0.9968847352024920	5.49767723136975E-06
yle	Yélî Dnye	6,000	756	0.994729907773386	0.9986772486772490	0.9966996699669970	4.39792901522673E-06	0.996042216358839	0.9986772486772490	0.9973579920739760	3.29844676142005E-06
yml	Iamalele	2,800	727	0.9986244841815680	0.9986244841815680	0.9986244841815680	1.09944719794887E-06	0.9986225895316810	0.9972489683631360	0.9979353062629040	1.09944719794887E-06
yon	Yongkom	4,000	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06
yor	Yoruba	50m	1000	0.999	0.999	0.999	1.09977729509774E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
yrb	Yareba	1,200	711	0.9985955056179780	1.0	0.9992972593113140	1.09942785774283E-06	0.9985915492957750	0.9971870604782000	0.9978888106966920	1.09942785774283E-06
yre	Yaouré	40,000	791	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
yuj	Karkar-Yuri	1,000	784	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06	0.9987261146496820	1.0	0.9993626513703000	1.09951610296309E-06
yut	Yopno	9,000	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
yuw	Yau	140	784	1.0	0.9987244897959180	0.9993618379068280	0.0	1.0	0.9987244897959180	0.9993618379068280	0.0
yva	Yawa (Kalou)	820	778	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zaa	Sierra de Juárez Zapotec	12,900	788	0.9974683544303800	1.0	0.9987325728770600	2.19904187745399E-06	0.9974683544303800	1.0	0.9987325728770600	2.19904187745399E-06
zab	Western Tlacolula Valley Zapotec	29,200	790	0.9987341772151900	0.9987341772151900	0.9987341772151900	1.0995233566249E-06	0.9987325728770600	0.9974683544303800	0.9981000633312220	1.0995233566249E-06
zac	Ocotlán Zapotec	15,000	794	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zad	Cajonos Zapotec	5,000	793	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06	1.0	1.0	1.0	0.0
zai	Isthmus Zapotec (Juchitán)	85,000	791	0.9987341772151900	0.9974715549936790	0.9981024667931690	1.09952456557784E-06	0.9987341772151900	0.9974715549936790	0.9981024667931690	1.09952456557784E-06
zam	Miahuatlán Zapotec	1,000	786	0.9949367088607600	1.0	0.9974619289340100	4.3980740833589E-06	0.9961977186311790	1.0	0.9980952380952380	3.29855556251917E-06
zao	Ozolotepec Zapotec	6,500	793	0.987546699875467	1.0	0.993734335839599	1.0995269834917E-05	0.99125	1.0	0.9956057752667920	7.69668888444191E-06
zap	Zapotec (macrolanguage)	490,000	785	0.9974587039390090	1.0	0.9987277353689570	2.19903462380015E-06	0.9987261146496820	0.9987261146496820	0.9987261146496820	1.09951731190008E-06
zar	Rincón Zapotec	5,000	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
zas	Santo Domingo Albarradas Zapotec	5,500	792	0.9987373737373740	0.9987373737373740	0.9987373737373740	1.09952577453344E-06	0.9987373737373740	0.9987373737373740	0.9987373737373740	1.09952577453344E-06
zat	Tabaa Zapotec	2,000	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zav	Yatzachi Zapotec	2,500	782	1.0	0.9974424552429670	0.998719590268886	0.0	1.0	0.9974424552429670	0.998719590268886	0.0
zaw	Mitla Zapotec	19,500	795	0.9962406015037590	1.0	0.9981167608286250	3.29858820424858E-06	0.9962406015037590	1.0	0.9981167608286250	3.29858820424858E-06
zca	Coatecas Altas Zapotec	4,880	793	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06
zga	Kinga	150,000	787	0.9949302915082380	0.9974587039390090	0.9961928934010150	4.39807891912812E-06	0.9949302915082380	0.9974587039390090	0.9961928934010150	4.39807891912812E-06
zia	Zia	4,500	788	0.9961928934010150	0.9961928934010150	0.9961928934010150	3.29856281618099E-06	0.9974587039390090	0.9961928934010150	0.996825396825397	2.19904187745399E-06
ziw	Zigula	380,000	793	0.9823899371069180	0.9848675914249690	0.9836272040302270	1.53933777688838E-05	0.9848484848484850	0.9836065573770490	0.9842271293375400	1.31943238019004E-05
zlm	Malay (Standard Malay)	18m	794	0.9517601043024770	0.9193954659949620	0.9352978859705320	4.06825431207469E-05	0.954248366013072	0.9193954659949620	0.9364977549711350	3.84834867358417E-05
zos	Zoque, Francisco León	20,000	795	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06	1.0	1.0	1.0	0.0
zpc	Choapan Zapotec	12,000	791	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
zpl	Lachixío Zapotec	6,500	795	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06
zpm	Mixtepec Zapotec	9,000	769	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zpo	Amatlán Zapotec	10,000	790	1.0	0.9949367088607600	0.9974619289340100	0.0	1.0	0.9949367088607600	0.9974619289340100	0.0
zpq	Zoogocho Zapotec	1,000	787	0.9962025316455700	1.0	0.9980976537729870	3.29855918934609E-06	0.9962025316455700	1.0	0.9980976537729870	3.29855918934609E-06
zpu	Yalálag Zapotec	3,500	793	0.9962311557788940	1.0	0.9981120201384520	3.29858095047511E-06	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06
zpv	Chichicapan Zapotec	2,720	794	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zpz	Texmelucan Zapotec	4,100	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zsr	Southern Rincón Zapotec	12,000	791	0.9974715549936790	0.9974715549936790	0.9974715549936790	2.19904913115569E-06	0.9974683544303800	0.9962073324905180	0.9968374446552820	2.19904913115569E-06
ztq	Quioquitani-Quierí Zapotec	4,000	789	0.9987341772151900	1.0	0.999366687777074	1.09952214767462E-06	1.0	1.0	1.0	0.0
zty	Yatee Zapotec	5,000	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06

Table 14: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
acf	Saint Lucian Creole French	357,000	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06
bel	Belarusian	10m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ben	Bengali	270m	1000	0.9960159362549800	1.0	0.998003992015968	4.39910918039097E-06	0.9960159362549800	1.0	0.998003992015968	4.39910918039097E-06
bzj	Belize Kriol English	200,000	779	1.0	1.0	1.0	0.0	1.0	0.9987163029525030	0.9993577392421320	0.0
ces	Czech	12m	791	0.9801734820322180	1.0	0.9899874843554440	1.75923930492455E-05	0.9801734820322180	1.0	0.9899874843554440	1.75923930492455E-05
cop	Coptic	0	792	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
cth	Thaiphum Chin	1,000	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	0.999	0.999	0.999	1.09977729509774E-06
daa	Dangaléat	60,000	784	0.9961880559085130	1.0	0.9980903882877150	3.29854830888926E-06	0.9961880559085130	1.0	0.9980903882877150	3.29854830888926E-06
dad	Marik	3,500	792	0.9962264150943400	1.0	0.998109640831758	3.29857732360033E-06	0.9962264150943400	1.0	0.998109640831758	3.29857732360033E-06
dah	Gwahatike	1,570	768	0.9961089494163420	1.0	0.9980506822612090	3.29849028099839E-06	0.9961089494163420	1.0	0.9980506822612090	3.29849028099839E-06
dan	Danish	5.6m	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.998	0.998998998998999	0.0
ded	Dedua	6,500	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
deu	German	155m	1000	0.996011964107677	0.999	0.9975037443834250	4.39910918039097E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
djk	Aukan (Ndyuka)	67,000	786	0.9961928934010150	0.9987277353689570	0.9974587039390090	3.29855556251917E-06	0.9961928934010150	0.9987277353689570	0.9974587039390090	3.29855556251917E-06
eng	English	1.46bn	1000	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06	0.997002997002997	0.998	0.9975012493753120	3.29933188529323E-06
epo	Esperanto	100,000	1000	0.997002997002997	0.998	0.9975012493753120	3.29933188529323E-06	0.997	0.997	0.997	3.29933188529323E-06
fra	French	321m	1000	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06	0.9979959919839680	0.996	0.9969969969969970	2.19955459019549E-06
grc	Ancient Greek	0	791	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
guj	Gujarati	62m	1000	0.9950248756218910	1.0	0.9975062344139650	5.49888647548871E-06	0.9950248756218910	1.0	0.9975062344139650	5.49888647548871E-06
gul	Gullah (Sea Island Creole English)	300	794	0.9962358845671270	1.0	0.9981143934632310	3.29858457735786E-06	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06
hat	Haitian Creole	13m	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
hau	Hausa	94m	796	0.980246913580247	0.9974874371859300	0.9887920298879200	1.75924897661188E-05	0.9814585908529050	0.9974874371859300	0.9894080996884740	1.64929591557364E-05
hin	Hindi	610m	1000	0.9784735812133070	1.0	0.9891196834817010	2.41951004921503E-05	0.9803729146221790	0.999	0.9895988112927190	2.19955459019549E-05
hrv	Croatian	6.4m	1000	0.9737373737373740	0.964	0.9688442211055280	2.85942096725413E-05	0.974721941354904	0.964	0.9693313222724990	2.74944323774436E-05
kbq	Kamano	63,000	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	1.0	1.0	1.0	0.0
lat	Latin	0	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
lit	Lithuanian	3.1m	703	0.9957507082152970	1.0	0.9978708303761530	3.2982545636849E-06	0.9971631205673760	1.0	0.9985795454545450	2.19883637578993E-06
mal	Malayalam	39m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mar	Marathi	84m	1000	0.9910802775024780	1.0	0.9955201592832260	9.89799565587969E-06	0.9930486593843100	1.0	0.9965122072745390	7.6984410656842E-06
mau	Huautla Mazatec	74,600	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mpx	Misima-Panaeati	18,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.997	0.99849774661993	0.0
nfa	Dhao	5,000	903	0.9988938053097350	1.0	0.9994465965688990	1.0996599851326E-06	0.9988925802879290	0.9988925802879290	0.9988925802879290	1.0996599851326E-06
nld	Dutch	25m	1000	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06
por	Portuguese	279m	786	0.9949367088607600	1.0	0.9974619289340100	4.3980740833589E-06	0.9961977186311790	1.0	0.9980952380952380	3.29855556251917E-06
ron	Romanian	28m	114	0.991304347826087	1.0	0.9956331877729260	1.09870671232892E-06	0.991304347826087	1.0	0.9956331877729260	1.09870671232892E-06
rus	Russian	260m	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
san	Sanskrit	24,821	791	0.9949685534591200	1.0	0.9974779319041620	4.39809826231138E-06	0.9987373737373740	1.0	0.9993682880606440	1.09952456557784E-06
spa	Spanish	600m	1000	0.9930348258706470	0.998	0.9955112219451370	7.6984410656842E-06	0.9940239043824700	0.998	0.9960079840319360	6.59866377058646E-06
srp	Serbian	12m	1000	0.9635108481262330	0.977	0.9702085402184710	4.06917599186165E-05	0.9654150197628460	0.977	0.9711729622266400	3.8492205328421E-05
swe	Swedish	11m	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.997	0.99849774661993	0.0
urd	Urdu	246m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
zaj	Zaramo	293,000	792	0.947565543071161	0.9583333333333330	0.9529190207156310	4.61800825304046E-05	0.94875	0.9583333333333330	0.9535175879396980	4.50805567558712E-05
zyp	Zyphe Chin	21,000	793	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mgj	Abureni	4k	563	1.0	0.9928952042628780	0.9964349376114080	0.0	1.0	0.9928952042628780	0.9964349376114080	0.0
aca	Achagua	250	515	1.0	0.9883495145631070	0.9941406250000000	0.0	1.0	0.9883495145631070	0.9941406250000000	0.0
adj	Adioukrou	160k	1000	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06	1.0	0.99	0.9949748743718590	0.0
adt	Adnyamathanha	262	2	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
duo	Agta, Dupaninan	1,400	1000	0.9979654120040690	0.981	0.9894099848714070	2.19955459019549E-06	0.9989785495403470	0.978	0.988377968671046	1.09977729509774E-06
due	Agta, Umiray Dumaget	3,000	844	0.99644128113879	0.995260663507109	0.995850622406639	3.29876593166496E-06	0.99644128113879	0.995260663507109	0.995850622406639	3.29876593166496E-06
esg	Aheri Gondi	150,000	1000	0.9881305637982200	0.999	0.9935355544505220	1.31973275411729E-05	0.9900891972249750	0.999	0.9945246391239420	1.09977729509774E-05
ahr	Ahirani	1.6 million	174	0.9885714285714290	0.9942528735632180	0.991404011461318	2.19755829298067E-06	0.9942528735632180	0.9942528735632180	0.9942528735632180	1.09877914649033E-06
alw	Alaba-K'abeena	204,000	906	1.0	0.9900662251655630	0.9950083194675540	0.0	1.0	0.9889624724061810	0.9944506104328520	0.0
mim	Alacatlatzala Mixtec	34,000	1000	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06
omb	Ambae, East	5,000	122	0.9918032786885250	0.9918032786885250	0.9918032786885250	1.09871636966532E-06	0.9918032786885250	0.9918032786885250	0.9918032786885250	1.09871636966532E-06
abc	Ambala Ayta	2,000	187	0.9887640449438200	0.9411764705882350	0.9643835616438360	2.19758968363499E-06	0.9943502824858760	0.9411764705882350	0.9670329670329670	1.09879484181749E-06
amh	Amharic	32 million	776	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
akh	Angal Heneng	91,000	1000	0.999	0.999	0.999	1.09977729509774E-06	0.999	0.999	0.999	1.09977729509774E-06
cko	Anufo	180,000	943	0.9968017057569300	0.9915164369034990	0.9941520467836260	3.2991250720309E-06	0.9978609625668450	0.9893955461293740	0.9936102236421720	2.19941671468727E-06
any	Anyin	1.45 million	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.994	0.9969909729187560	0.0
apy	Apalaí	450	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
app	Apma	7,800	10	0.9090909090909090	1.0	0.9523809523809520	1.09858118240293E-06	0.9090909090909090	1.0	0.9523809523809520	1.09858118240293E-06
arq	Arabic, Algerian	42 million	497	0.9959919839679360	1.0	0.9979919678714860	2.1983384957649E-06	0.9959919839679360	1.0	0.9979919678714860	2.1983384957649E-06
aeb	Arabic, Tunisian	13 million	854	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
luc	Aringa	495,000	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06
prq	Ashéninka Perené	7,200	1000	1.0	0.957	0.9780275932549820	0.0	1.0	0.954	0.9764585465711360	0.0
auu	Auye	5, 350	1000	0.997	0.997	0.997	3.29933188529323E-06	0.996996996996997	0.996	0.9964982491245620	3.29933188529323E-06
avu	Avokaya	40,000	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
abp	Ayta, Abellen	4,400	1000	0.9979695431472080	0.983	0.9904282115869020	2.19955459019549E-06	0.9989816700611000	0.981	0.9899091826437940	1.09977729509774E-06
blx	Ayta, Mag-Indi	5,000	1000	0.9782393669634030	0.989	0.9835902536051720	2.41951004921503E-05	0.9820538384845460	0.985	0.9835247129306040	1.97959913117594E-05
miy	Ayutla Mixtec	5,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
bfd	Bafut	105,000	1000	0.9989775051124740	0.977	0.9878665318503540	1.09977729509774E-06	0.9989775051124740	0.977	0.9878665318503540	1.09977729509774E-06
bsw	Baiso	3 500	286	0.9761092150170650	1.0	0.9879101899827290	7.69240067737083E-06	0.979381443298969	0.9965034965034970	0.9878682842287700	6.59348629488928E-06
bcc	Southern Balochi	8.8 million	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.997	0.99849774661993	0.0
bmo	Bambalang	29,000	875	0.9988532110091740	0.9954285714285710	0.9971379507727530	1.09962612711678E-06	0.9988532110091740	0.9954285714285710	0.9971379507727530	1.09962612711678E-06
bcw	Bana	23,000	1000	0.9960079840319360	0.998	0.997002997002997	4.39910918039097E-06	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06
bqj	Bandial	16,000	1000	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06
bjn	Banjar	3.5 million	361	0.9887005649717510	0.9695290858725760	0.9790209790209790	4.39601984363357E-06	0.9914772727272730	0.9667590027700830	0.9789621318373070	3.29701488272518E-06
bno	Bantoanon	75,000	1000	0.9948979591836740	0.975	0.984848484848485	5.49888647548871E-06	0.9948979591836740	0.975	0.984848484848485	5.49888647548871E-06
bpx	Bareli, Palya	10,000	107	0.8916666666666670	1.0	0.9427312775330400	1.42830774098848E-05	0.9067796610169490	1.0	0.9511111111111110	1.20856808852871E-05
buj	Basa-Gurmana	5,700	135	1.0	0.9703703703703700	0.9849624060150380	0.0	1.0	0.9703703703703700	0.9849624060150380	0.0
ifb	Batad Ifugao	43,000	1000	0.9829659318637280	0.981	0.981981981981982	1.86962140166616E-05	0.9839518555667000	0.981	0.9824737105658490	1.75964367215639E-05
bkw	Bekwel	13,000	1000	1.0	0.991	0.9954796584630840	0.0	1.0	0.988	0.993963782696177	0.0
bhd	Bhadrawahi	120,000	792	0.9987373737373740	0.9987373737373740	0.9987373737373740	1.09952577453344E-06	1.0	0.9987373737373740	0.9993682880606440	0.0
bht	Bhattiyali	24,000	785	0.9949238578680200	0.9987261146496820	0.9968213604577240	4.3980692476003E-06	0.9974554707379140	0.9987261146496820	0.9980903882877150	2.19903462380015E-06
bhi	Bhilali	1.2 million	1000	0.9939455095862770	0.985	0.9894525364138620	6.59866377058646E-06	0.9959514170040490	0.984	0.9899396378269620	4.39910918039097E-06
bhu	Bhunjia	12,000	921	0.9902491874322860	0.992399565689468	0.9913232104121470	9.89713576890848E-06	0.9923830250272030	0.990228013029316	0.9913043478260870	7.69777226470659E-06
bim	Bimoba	200,000	1000	0.9979879275653920	0.992	0.9949849548645940	2.19955459019549E-06	1.0	0.99	0.9949748743718590	0.0
biv	Birifor, Southern	231,000	860	0.9988331388564760	0.9953488372093020	0.9970879440885260	1.09960798975165E-06	1.0	0.9953488372093020	0.9976689976689980	0.0
bib	Bissa	590,000	1000	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06
bzi	Bisu	10,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
blr	Blang	15,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
bdv	Bodo Parja	156,354	794	1.0	0.9987405541561710	0.9993698802772530	0.0	1.0	0.9987405541561710	0.9993698802772530	0.0
bvw	Boga	26,000	292	0.9965635738831620	0.9931506849315070	0.9948542024013720	1.09892162820624E-06	0.9965397923875430	0.9863013698630140	0.9913941480206540	1.09892162820624E-06
bor	Borôro	1,400	1000	1.0	0.992	0.9959839357429720	0.0	1.0	0.989	0.9944695827048770	0.0
sbl	Botolan Sambal	32,867	1000	0.9969788519637460	0.99	0.9934771700953340	3.29933188529323E-06	0.9969788519637460	0.99	0.9934771700953340	3.29933188529323E-06
bgg	Bugun	1,700	791	0.9987357774968390	0.9987357774968390	0.9987357774968390	1.09952456557784E-06	0.9987357774968390	0.9987357774968390	0.9987357774968390	1.09952456557784E-06
bwu	Buli (Ghana)	168,000	1000	0.9989858012170390	0.985	0.9919436052366570	1.09977729509774E-06	0.9989858012170390	0.985	0.9919436052366570	1.09977729509774E-06
bds	Burunge	28,000	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
kyb	Butbut Kalinga	12,000	1000	0.9979879275653920	0.992	0.9949849548645940	2.19955459019549E-06	0.9979838709677420	0.99	0.9939759036144580	2.19955459019549E-06
bwc	Bwile	32,000	169	0.9640718562874250	0.9526627218934910	0.9583333333333330	6.59263865967261E-06	0.9636363636363640	0.9408284023668640	0.9520958083832340	6.59263865967261E-06
tbk	Calamian Tagbanwa	10,000	1000	0.9959595959595960	0.986	0.9909547738693470	4.39910918039097E-06	0.9969666329625890	0.986	0.9914529914529920	3.29933188529323E-06
mcu	Cameroon Mambila	130,000	1000	0.9838709677419360	0.976	0.9799196787148590	1.75964367215639E-05	0.9858442871587460	0.975	0.9803921568627450	1.53968821313684E-05
ram	Canela	2,500	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.994	0.9969909729187560	0.0
cam	Cemuhî	3,300	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.997	0.99849774661993	0.0
cmo	Central Mnong	30,000	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	1.0	0.998	0.998998998998999	0.0
sml	Central Sama	80,000	1000	0.994994994994995	0.994	0.9944972486243120	5.49888647548871E-06	0.9959919839679360	0.994	0.994994994994995	4.39910918039097E-06
syb	Central Subanen	140,000	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998	0.998	0.998	2.19955459019549E-06
hne	Chhattisgarhi	16.2 million	1000	0.9949596774193550	0.987	0.9909638554216870	5.49888647548871E-06	0.9949545913218970	0.986	0.9904570567553990	5.49888647548871E-06
dao	Chin, Daai	37,000	792	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06	0.9987389659520810	1.0	0.9993690851735020	1.09952577453344E-06
cje	Chru	19,314	870	0.9988412514484360	0.9908045977011490	0.994806693594922	1.09962008126192E-06	0.9988399071925750	0.9896551724137930	0.9942263279445730	1.09962008126192E-06

Table 15: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
asg	Cishingini	100,000	1000	0.9979529170931420	0.975	0.9863429438543250	2.19955459019549E-06	0.9979508196721310	0.974	0.98582995951417	2.19955459019549E-06
mta	Cotabato Manobo	30,000	1000	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06	0.9989939637826960	0.993	0.995987963891675	1.09977729509774E-06
crl	Cree, Northern East	6,870	27	1.0	0.9259259259259260	0.9615384615384620	0.0	1.0	0.9259259259259260	0.9615384615384620	0.0
cul	Culina	4,000	1000	1.0	0.996	0.9979959919839680	0.0	1.0	0.996	0.9979959919839680	0.0
cug	Cung	2,000	1000	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06	0.9989909182643800	0.99	0.9944751381215470	1.09977729509774E-06
dnj	Dan	1.58 million	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
dnv	Danu	100,000	793	0.9987405541561710	1.0	0.9993698802772530	1.0995269834917E-06	0.9987389659520810	0.9987389659520810	0.9987389659520810	1.0995269834917E-06
ntr	Delo	10,900	1000	0.997	0.997	0.997	3.29933188529323E-06	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06
des	Desano	2,460	1000	1.0	0.987	0.9934574735782590	0.0	1.0	0.986	0.9929506545820750	0.0
dso	Desiya	50,000	795	0.9962406015037590	1.0	0.9981167608286250	3.29858820424858E-06	0.9987437185929650	1.0	0.9993714644877440	1.09952940141619E-06
dhn	Dhanki	163,000	1000	0.9959266802443990	0.978	0.9868819374369320	4.39910918039097E-06	0.9969293756397130	0.974	0.9853313100657560	3.29933188529323E-06
div	Maldivian	356,000	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mbd	Dibabawon Manobo	10,000	1000	0.9989816700611000	0.981	0.9899091826437940	1.09977729509774E-06	0.9989785495403470	0.978	0.988377968671046	1.09977729509774E-06
dig	Digo	88,000	1000	0.9939455095862770	0.985	0.9894525364138620	6.59866377058646E-06	0.9959473150962510	0.983	0.9894313034725720	4.39910918039097E-06
djj	Djeebbana	100	6	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
dyi	Djimini Senoufo	95,500	1000	0.997002997002997	0.998	0.9975012493753120	3.29933188529323E-06	0.997002997002997	0.998	0.9975012493753120	3.29933188529323E-06
sce	Dongxiang	250,000	1000	1.0	0.982	0.9909182643794150	0.0	1.0	0.977	0.9883662114314620	0.0
bru	Eastern Bru	69,000	1000	0.995995995995996	0.995	0.9954977488744370	4.39910918039097E-06	0.995995995995996	0.995	0.9954977488744370	4.39910918039097E-06
xrb	Eastern Karaboro	35,000	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.994	0.9969909729187560	0.0
eka	Ekajuk	30,000	1000	0.9968879668049790	0.961	0.9786150712830960	3.29933188529323E-06	0.9968717413972890	0.956	0.9760081674323630	3.29933188529323E-06
ekk	Estonian, Standard	953,000	419	0.9881235154394300	0.9928400954653940	0.9904761904761910	5.4953750923223E-06	0.9880952380952380	0.9904534606205250	0.9892729439809300	5.4953750923223E-06
utr	Etulo	10,000	1000	1.0	0.986	0.9929506545820750	0.0	1.0	0.984	0.9919354838709680	0.0
fin	Finnish	4,700,000	1000	0.9979508196721310	0.974	0.98582995951417	2.19955459019549E-06	0.9989733059548260	0.973	0.9858156028368790	1.09977729509774E-06
fag	Finongan	1,300	162	1.0	0.9876543209876540	0.9937888198757760	0.0	1.0	0.9876543209876540	0.9937888198757760	0.0
frd	Fordata	50,000	1000	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06	1.0	0.991	0.9954796584630840	0.0
flr	Fuliiru	300,000	1000	0.9940179461615160	0.997	0.9955067398901650	6.59866377058646E-06	0.995004995004995	0.996	0.9955022488755620	5.49888647548871E-06
gdg	Ga'dang	6,000	1000	0.9890547263681590	0.994	0.9915211970074810	1.20975502460752E-05	0.991991991991992	0.991	0.991495747873937	8.79821836078194E-06
gbl	Gamit	400,000	922	1.0	0.9945770065075920	0.9972811310494830	0.0	1.0	0.9945770065075920	0.9972811310494830	0.0
gaq	Gata'	3,060	795	0.9924906132665830	0.9974842767295600	0.9949811794228360	6.59717640849716E-06	0.9924906132665830	0.9974842767295600	0.9949811794228360	6.59717640849716E-06
gyl	Gayil	55,700	912	1.0	0.9901315789473690	0.9950413223140500	0.0	1.0	0.9890350877192980	0.9944873208379270	0.0
ghn	Ghanongga	2,510	1000	0.9909638554216870	0.987	0.9889779559118240	9.89799565587969E-06	0.9909547738693470	0.986	0.9884711779448620	9.89799565587969E-06
gri	Ghari	12,100	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
acd	Gikyode	10,400	1000	0.996003996003996	0.997	0.9965017491254370	4.39910918039097E-06	0.997	0.997	0.997	3.29933188529323E-06
nyf	Giryama	623,000	1000	0.9949849548645940	0.992	0.9934902353530300	5.49888647548871E-06	0.9949748743718590	0.99	0.9924812030075190	5.49888647548871E-06
glw	Glavda	28,500	453	0.9955947136563880	0.9977924944812360	0.9966923925027560	2.19823218167949E-06	0.9955947136563880	0.9977924944812360	0.9966923925027560	2.19823218167949E-06
gqr	Gor	87,000	1000	0.9989868287740630	0.986	0.992450931051837	1.09977729509774E-06	0.9989847715736040	0.984	0.9914357682619650	1.09977729509774E-06
goj	Gowlan	20,200	1000	0.9878296146044630	0.974	0.9808660624370590	1.31973275411729E-05	0.9897959183673470	0.97	0.9797979797979800	1.09977729509774E-05
gok	Gowli	35,000	1000	0.998995983935743	0.995	0.9969939879759520	1.09977729509774E-06	0.998995983935743	0.995	0.9969939879759520	1.09977729509774E-06
gde	Gude	68,000	1000	1.0	0.987	0.9934574735782590	0.0	1.0	0.986	0.9929506545820750	0.0
gdf	Guduf-Gava	55,900	312	0.9967320261437910	0.9775641025641030	0.9870550161812300	1.09894578131199E-06	0.9967320261437910	0.9775641025641030	0.9870550161812300	1.09894578131199E-06
rub	Gungu	49,000	1000	0.9979818365287590	0.989	0.99347061778001	2.19955459019549E-06	0.997979797979798	0.988	0.9929648241206030	2.19955459019549E-06
guf	Gupapuyngu	450	2	1.0	0.5	0.6666666666666670	0.0	1.0	0.5	0.6666666666666670	0.0
gge	Guragone	20	1	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
gue	Gurinji	540	31	1.0	0.967741935483871	0.9836065573770490	0.0	1.0	0.967741935483871	0.9836065573770490	0.0
gvr	Gurung	325,622	1000	1.0	0.992	0.9959839357429720	0.0	1.0	0.991	0.9954796584630840	0.0
gwr	Gwere	409,000	1000	0.993006993006993	0.994	0.993503248375812	7.6984410656842E-06	0.9959919839679360	0.994	0.994994994994995	4.39910918039097E-06
hlb	Halbi	500,000	1000	0.9807497467071940	0.968	0.9743331655762460	2.08957686068571E-05	0.9847094801223240	0.966	0.9752650176678450	1.64966594264661E-05
bgc	Haryanvi	13m	790	0.98625	0.9987341772151900	0.9924528301886790	1.20947569228739E-05	0.9874843554443050	0.9987341772151900	0.9930774071743240	1.0995233566249E-05
mey	Hassaniyya	2.7m	91	1.0	0.978021978021978	0.9888888888888890	0.0	1.0	0.978021978021978	0.9888888888888890	0.0
hav	Havu	506,000	351	0.997093023255814	0.9772079772079770	0.9870503597122300	1.09899288292209E-06	0.9970760233918130	0.9715099715099720	0.9841269841269840	1.09899288292209E-06
hoy	Holiya	500	795	0.9949937421777220	1.0	0.9974905897114180	4.39811760566478E-06	0.9962406015037590	1.0	0.9981167608286250	3.29858820424858E-06
hre	Hre	113,000	1000	0.9969879518072290	0.993	0.9949899799599200	3.29933188529323E-06	0.9969849246231160	0.992	0.9944862155388470	3.29933188529323E-06
hru	Hruso	4,000	135	1.0	0.9851851851851850	0.9925373134328360	0.0	1.0	0.9851851851851850	0.9925373134328360	0.0
var	Huarijio	2,840	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.997	0.99849774661993	0.0
dud	Hun-Saare	73,000	1000	0.996996996996997	0.996	0.9964982491245620	3.29933188529323E-06	0.996996996996997	0.996	0.9964982491245620	3.29933188529323E-06
hwo	Hwana	32,000	1000	0.9979736575481260	0.985	0.9914443885254150	2.19955459019549E-06	0.9989847715736040	0.984	0.9914357682619650	1.09977729509774E-06
ibl	Ibaloi	111,000	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.995	0.9974937343358400	0.0
ivb	Ibatan	1,350	1000	0.9826707441386340	0.964	0.9732458354366480	1.86962140166616E-05	0.9856262833675560	0.96	0.972644376899696	1.53968821313684E-05
isl	Icelandic	230,000	989	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ifa	Ifugao, Amganad	271,000	1000	0.9736070381231670	0.996	0.9846762234305490	2.96939869676391E-05	0.9764474975466140	0.995	0.9856364536899460	2.63946550823458E-05
ifu	Ifugao, Mayoyao	30,000	1000	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06
ifk	Ifugao, Tuwali	30,000	1000	0.9870259481037920	0.989	0.988011988011988	1.42971048362707E-05	0.9899699097291880	0.987	0.9884827240861290	1.09977729509774E-05
ibo	Igbo	18m	1000	0.9950199203187250	0.999	0.9970059880239520	5.49888647548871E-06	0.996011964107677	0.999	0.9975037443834250	4.39910918039097E-06
ikz	Ikizu	132,000	1000	0.996969696969697	0.987	0.9919597989949750	3.29933188529323E-06	0.9969666329625890	0.986	0.9914529914529920	3.29933188529323E-06
iqw	Ikwo	260,000	1000	0.9673267326732670	0.977	0.9721393034825870	3.62926507382255E-05	0.9672619047619050	0.975	0.9711155378486060	3.62926507382255E-05
isd	Isnag	30,000	1000	0.9959798994974870	0.991	0.993483709273183	4.39910918039097E-06	0.9959798994974870	0.991	0.993483709273183	4.39910918039097E-06
ivv	Ivatan	35,000	997	0.989010989010989	0.9929789368104310	0.9909909909909910	1.20975103323736E-05	0.9909819639278560	0.9919759277833500	0.9914786967418550	9.89796299921476E-06
izz	Izii	540,000	1000	0.9787878787878790	0.969	0.9738693467336680	2.30953231970526E-05	0.9787878787878790	0.969	0.9738693467336680	2.30953231970526E-05
jaa	Jamamadí	480	1000	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06	1.0	0.994	0.9969909729187560	0.0
jib	Jibu	30,000	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
jum	Jumjum	25,000	357	0.9971671388101980	0.9859943977591040	0.9915492957746480	1.09900012968202E-06	0.9971671388101980	0.9859943977591040	0.9915492957746480	1.09900012968202E-06
juy	Juray	801,000	794	0.9962358845671270	1.0	0.9981143934632310	3.29858457735786E-06	0.9974874371859300	1.0	0.9987421383647800	2.19905638490524E-06
kkj	Kako	100,000	1000	0.9930278884462150	0.997	0.9950099800399200	7.6984410656842E-06	0.9940119760479040	0.996	0.995004995004995	6.59866377058646E-06
kqe	Kalagan	21,400	1000	0.9989658738366080	0.966	0.9822064056939500	1.09977729509774E-06	0.9989658738366080	0.966	0.9822064056939500	1.09977729509774E-06
ksc	Kalinga, Southern	13,000	1000	0.994994994994995	0.994	0.9944972486243120	5.49888647548871E-06	0.9959839357429720	0.992	0.9939879759519040	4.39910918039097E-06
ify	Kallahan, Keley-i	8,000	1000	0.9979919678714860	0.994	0.9959919839679360	2.19955459019549E-06	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06
tbx	Kapin	2,350	756	1.0	0.9920634920634920	0.9960159362549800	0.0	1.0	0.9907407407407410	0.9953488372093020	0.0
krx	Karon	9,000	864	0.9988358556461000	0.9930555555555560	0.9959373186302960	1.09961282632385E-06	0.9988358556461000	0.9930555555555560	0.9959373186302960	1.09961282632385E-06
xsm	Kasem	130,000	1000	0.999	0.999	0.999	1.09977729509774E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
xtc	Katcha-Kadugli-Miri	75,000	769	0.9986945169712790	0.9947984395318600	0.996742671009772	1.09949796922725E-06	1.0	0.9947984395318600	0.9973924380704040	0.0
kbo	Keliko	10,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.997	0.99849774661993	0.0
kca	Khanty	13,600	373	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kij	Kilivila	20,000	1000	0.9979859013091640	0.991	0.9944806823883590	2.19955459019549E-06	0.9979838709677420	0.99	0.9939759036144580	2.19955459019549E-06
kqp	Kimré	20,000	950	0.9989473684210530	0.9989473684210530	0.9989473684210530	1.0997168229181E-06	0.9989473684210530	0.9989473684210530	0.9989473684210530	1.0997168229181E-06
krj	Kinaray-a	378,000	985	0.9989506820566630	0.966497461928934	0.9824561403508770	1.09975915274555E-06	0.9989462592202320	0.9624365482233500	0.9803516028955530	1.09975915274555E-06
gia	Kitja	410	24	1.0	0.9166666666666670	0.9565217391304350	0.0	1.0	0.9166666666666670	0.9565217391304350	0.0
cdz	Koda	1,300	503	1.0	0.9781312127236580	0.9889447236180910	0.0	1.0	0.9781312127236580	0.9889447236180910	0.0
nit	Kolami, Southeastern	10,000	1000	1.0	0.99	0.9949748743718590	0.0	1.0	0.988	0.993963782696177	0.0
kxw	Konai	600	1000	1.0	0.989	0.9944695827048770	0.0	1.0	0.989	0.9944695827048770	0.0
kfc	Konda-Dora	20,000	1000	0.9969909729187560	0.994	0.9954932398597900	3.29933188529323E-06	0.9989939637826960	0.993	0.995987963891675	1.09977729509774E-06
kma	Konni	3,800	840	0.9988066825775660	0.9964285714285710	0.9976162097735400	1.09958380752885E-06	0.9988038277511960	0.9940476190476190	0.996420047732697	1.09958380752885E-06
kor	Korean	42m	1000	1.0	0.991	0.9954796584630840	0.0	1.0	0.99	0.9949748743718590	0.0
coe	Koreguaje	2,210	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.995	0.9974937343358400	0.0
kff	Koya	300,000	1000	1.0	0.98	0.98989898989899	0.0	1.0	0.979	0.989388580090955	0.0
kux	Kukatja	580	59	1.0	0.8813559322033900	0.936936936936937	0.0	1.0	0.864406779661017	0.9272727272727270	0.0
wlg	Kunbarlang	75	3	0.75	1.0	0.8571428571428570	1.09857273430359E-06	0.75	1.0	0.8571428571428570	1.09857273430359E-06
xuo	Kuo	11,800	1000	0.9950149551345960	0.998	0.9965052421367950	5.49888647548871E-06	0.996003996003996	0.997	0.9965017491254370	4.39910918039097E-06
key	Kupia	6,600	1000	0.9950099800399200	0.997	0.9960039960039960	5.49888647548871E-06	0.996003996003996	0.997	0.9965017491254370	4.39910918039097E-06
krh	Kurama	40,300	101	1.0	0.9702970297029700	0.9849246231155780	0.0	1.0	0.9702970297029700	0.9849246231155780	0.0
khj	Kuturmi	10,500	162	1.0	0.9876543209876540	0.9937888198757760	0.0	1.0	0.9814814814814820	0.9906542056074770	0.0
kxv	Kuvi	350,000	786	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
kwo	Kwomtari	600	451	0.9977528089887640	0.9844789356984480	0.9910714285714290	1.0991136747327E-06	1.0	0.9844789356984480	0.9921787709497210	0.0
lbj	Ladakhi	150,000	49	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
lhi	Lahu Shi	117,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
nrz	Lala	3,000	1000	1.0	0.99	0.9949748743718590	0.0	1.0	0.988	0.993963782696177	0.0
lro	Laro	40,000	473	0.9978768577494690	0.9936575052854120	0.9957627118644070	1.0991402524945E-06	1.0	0.9915433403805500	0.9957537154989390	0.0
llu	Lau	16,900	1000	0.9969758064516130	0.989	0.9929718875502010	3.29933188529323E-06	0.9989888776541960	0.988	0.9934640522875820	1.09977729509774E-06
lbv	Lavatbura-Lamusong	1,310	130	1.0	0.9692307692307690	0.9843750000000000	0.0	1.0	0.9692307692307690	0.9843750000000000	0.0
lcp	Lawa, Western	8,000	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
lik	Lika	60,000	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.994	0.9969909729187560	0.0

Table 16: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
loq	Lobala	60,000	1000	0.9969818913480890	0.991	0.9939819458375130	3.29933188529323E-06	0.9969788519637460	0.99	0.9934771700953340	3.29933188529323E-06
lbm	Lodhi	25,000	795	0.9718482252141980	0.9987421383647800	0.9851116625310180	2.52891762325725E-05	0.9754299754299750	0.9987421383647800	0.9869484151646990	2.19905880283239E-05
log	Logo	210,000	1000	1.0	1.0	1.0	0.0	1.0	0.998	0.998998998998999	0.0
ztp	Loxicha Zapotec	75,000	1000	0.9989517819706500	0.953	0.9754350051177070	1.09977729509774E-06	1.0	0.945	0.9717223650385610	0.0
knb	Lubuagan Kalinga	14,000	1000	1.0	0.67	0.8023952095808380	0.0	1.0	0.669	0.8016776512881970	0.0
lmd	Lumun	45,000	833	1.0	0.9903961584633850	0.9951749095295540	0.0	1.0	0.9891956782713090	0.9945684972842490	0.0
lga	Lungga	2,770	1000	0.9888888888888890	0.979	0.9839195979899500	1.20975502460752E-05	0.9908814589665650	0.978	0.984398590840463	9.89799565587969E-06
lwo	Luwo	80,000	1000	0.9969849246231160	0.992	0.9944862155388470	3.29933188529323E-06	0.9979879275653920	0.992	0.9949849548645940	2.19955459019549E-06
lyn	Luyana	222,000	1000	1.0	0.972	0.9858012170385400	0.0	1.0	0.97	0.9847715736040610	0.0
ayz	Mai Brat	20,000	1000	1.0	0.985	0.9924433249370280	0.0	1.0	0.983	0.9914271306101870	0.0
mai	Maithili	31.9m	967	0.9989669421487600	1.0	0.999483204134367	1.09973738271301E-06	0.9989669421487600	1.0	0.999483204134367	1.09973738271301E-06
pmy	Malay, Papuan	500,000	1000	0.9989615784008310	0.962	0.9801324503311260	1.09977729509774E-06	0.9989572471324300	0.958	0.9780500255232260	1.09977729509774E-06
mgq	Malila	65,000	1000	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06
mmn	Mamanwa	5,150	1000	0.9979818365287590	0.989	0.99347061778001	2.19955459019549E-06	0.9979818365287590	0.989	0.99347061778001	2.19955459019549E-06
kby	Manga Kanuri	280,000	1000	1.0	0.992	0.9959839357429720	0.0	1.0	0.99	0.9949748743718590	0.0
mge	Mango	52,200	996	1.0	0.9909638554216870	0.9954614220877460	0.0	1.0	0.9909638554216870	0.9954614220877460	0.0
mbi	Manobo, Ilianen	14,600	1000	0.997	0.997	0.997	3.29933188529323E-06	0.996996996996997	0.996	0.9964982491245620	3.29933188529323E-06
mfm	Marghi South	166,000	872	0.997709049255441	0.9988532110091740	0.9982808022922640	2.19924499919178E-06	0.9988518943742830	0.9977064220183490	0.9982788296041310	1.09962249959589E-06
mfh	Matal	18,000	1000	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06	1.0	0.994	0.9969909729187560	0.0
mgv	Matengo	150,000	1000	0.9938650306748470	0.972	0.9828109201213350	6.59866377058646E-06	0.9958974358974360	0.971	0.9832911392405060	4.39910918039097E-06
mfy	Mayo	40,000	1000	1.0	0.991	0.9954796584630840	0.0	1.0	0.99	0.9949748743718590	0.0
mdm	Mayogo	100,000	1000	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06
mfo	Mbe	14,300	1000	1.0	0.987	0.9934574735782590	0.0	1.0	0.986	0.9929506545820750	0.0
mbu	Mbula-Bwazza	40,600	1000	1.0	0.989	0.9944695827048770	0.0	1.0	0.988	0.993963782696177	0.0
mfx	Melo	20,200	322	0.9487179487179490	0.9192546583850930	0.9337539432176660	1.75833257322081E-05	0.9516129032258070	0.9161490683229810	0.9335443037974680	1.64843678739451E-05
mxm	Meramera	2,000	1000	0.9969818913480890	0.991	0.9939819458375130	3.29933188529323E-06	0.9979757085020240	0.986	0.9919517102615700	2.19955459019549E-06
mxv	Metlatónoc Mixtec	46,600	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.997	0.99849774661993	0.0
mif	Mofu-Gudur	60,000	1000	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06	0.998995983935743	0.995	0.9969939879759520	1.09977729509774E-06
mzq	Mori Atas	16,100	190	0.9578947368421050	0.9578947368421050	0.9578947368421050	8.79038771103798E-06	0.9574468085106380	0.9473684210526320	0.9523809523809520	8.79038771103798E-06
mpa	Mpoto	80,000	1000	0.9959473150962510	0.983	0.9894313034725720	4.39910918039097E-06	0.9979654120040690	0.981	0.9894099848714070	2.19955459019549E-06
moz	Mukulu	12,000	1000	0.998992950654582	0.992	0.9954841946813850	1.09977729509774E-06	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06
unx	Munda	519,000	792	0.9949748743718590	1.0	0.9974811083123430	4.39810309813378E-06	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06
mur	Murle	60,000	1000	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06
sur	Mwaghavul	295,000	667	1.0	0.9850074962518740	0.9924471299093660	0.0	1.0	0.9850074962518740	0.9924471299093660	0.0
moa	Mwan	17,000	1000	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06
muh	Mündü	23,000	1000	0.9979939819458380	0.995	0.9964947421131700	2.19955459019549E-06	0.9979939819458380	0.995	0.9964947421131700	2.19955459019549E-06
nag	Naga Pidgin	30,000	795	0.9987405541561710	0.9974842767295600	0.9981120201384520	1.09952940141619E-06	0.9987405541561710	0.9974842767295600	0.9981120201384520	1.09952940141619E-06
kfw	Naga, Kharam	1,400	795	0.9962358845671270	0.9987421383647800	0.9974874371859300	3.29858820424858E-06	0.9962358845671270	0.9987421383647800	0.9974874371859300	3.29858820424858E-06
tvt	Naga, Tutsa	25,000	791	0.9962168978562420	0.9987357774968390	0.9974747474747470	3.29857369673353E-06	0.9962168978562420	0.9987357774968390	0.9974747474747470	3.29857369673353E-06
nck	Nakara	75	2	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
nal	Nalik	5,140	792	0.9962264150943400	1.0	0.998109640831758	3.29857732360033E-06	0.9962264150943400	1.0	0.998109640831758	3.29857732360033E-06
naw	Nawuri	14,000	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.996	0.9979959919839680	0.0
ncr	Ncane	15,500	1000	0.9979919678714860	0.994	0.9959919839679360	2.19955459019549E-06	0.9989949748743720	0.994	0.9964912280701750	1.09977729509774E-06
ndh	Ndali	150,000	1000	0.993006993006993	0.994	0.993503248375812	7.6984410656842E-06	0.9949899799599200	0.993	0.993993993993994	5.49888647548871E-06
dne	Ndendeule	100,000	1000	0.9989785495403470	0.978	0.988377968671046	1.09977729509774E-06	1.0	0.978	0.9888776541961580	0.0
ndz	Ndogo	48,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
njd	Ndonde Hamba	10,000	144	0.9470198675496690	0.9930555555555560	0.9694915254237290	8.78994342572663E-06	0.9466666666666670	0.9861111111111110	0.9659863945578230	8.78994342572663E-06
ndv	Ndut	38,600	1000	0.9890764647467730	0.996	0.992526158445441	1.20975502460752E-05	0.9900596421471170	0.996	0.9930209371884350	1.09977729509774E-05
niy	Ngiti	100,000	1000	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
xnj	Ngoni	170,000	795	0.9925	0.9987421383647800	0.9956112852664580	6.59717640849716E-06	0.9924906132665830	0.9974842767295600	0.9949811794228360	6.59717640849716E-06
pcm	Nigerian Pidgin	120m	1000	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06
num	Niuafo'ou	690	785	0.9987244897959180	0.9974522292993630	0.9980879541108990	1.09951731190008E-06	1.0	0.9961783439490450	0.998085513720485	0.0
nkn	Nkangala	22,300	886	1.0	0.9785553047404060	0.9891614375356530	0.0	1.0	0.9774266365688490	0.9885844748858450	0.0
snf	Noon	32,900	1000	0.9979919678714860	0.994	0.9959919839679360	2.19955459019549E-06	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06
stb	Northern Subanen	10,000	1000	0.998992950654582	0.992	0.9954841946813850	1.09977729509774E-06	0.9989909182643800	0.99	0.9944751381215470	1.09977729509774E-06
nob	Norwegian	4.64m	772	0.9922779922779920	0.9987046632124350	0.9954809554551320	6.5970095755594E-06	0.9935567010309280	0.9987046632124350	0.9961240310077520	5.49750797963283E-06
ncf	Notsi	1,840	1000	0.9969262295081970	0.973	0.9848178137651820	3.29933188529323E-06	0.9979338842975210	0.966	0.9817073170731710	2.19955459019549E-06
nxl	Nuaulu, South	1,500	1000	1.0	0.976	0.9878542510121460	0.0	1.0	0.972	0.9858012170385400	0.0
nuq	Nukumanu	700	1000	0.9979253112033200	0.962	0.9796334012219960	2.19955459019549E-06	0.9979144942648590	0.957	0.9770290964777950	2.19955459019549E-06
nkr	Nukuoro	860	1000	0.997997997997998	0.997	0.9974987493746870	2.19955459019549E-06	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06
nnw	Nuni, Southern	168,000	1000	1.0	0.998	0.998998998998999	0.0	1.0	0.998	0.998998998998999	0.0
nwb	Nyabwa	42,700	1000	1.0	0.989	0.9944695827048770	0.0	1.0	0.989	0.9944695827048770	0.0
nyy	Nyakyusa-Ngonde	805,000	1000	0.999	0.999	0.999	1.09977729509774E-06	1.0	0.999	0.9994997498749370	0.0
nuj	Nyole	341,000	1000	0.9861111111111110	0.994	0.9900398406374500	1.53968821313684E-05	0.9890547263681590	0.994	0.9915211970074810	1.20975502460752E-05
gaz	Oromo, West Central	8.92m	1000	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
ury	Orya	1,600	1000	0.99000999000999	0.991	0.9905047476261870	1.09977729509774E-05	0.991	0.991	0.991	9.89799565587969E-06
oyd	Oyda	16,600	272	0.9325842696629210	0.9154411764705880	0.9239332096474950	1.9780154570919E-05	0.9431818181818180	0.9154411764705880	0.9291044776119400	1.64834621424325E-05
pac	Pacoh (Cado)	16,000	1000	1.0	0.994	0.9969909729187560	0.0	1.0	0.994	0.9969909729187560	0.0
pdo	Padoe	6,000	172	0.9763313609467460	0.9593023255813950	0.967741935483871	4.39510692745766E-06	0.9761904761904760	0.9534883720930230	0.9647058823529410	4.39510692745766E-06
bfz	Pahari, Mahasu	1m	792	0.9912390488110140	1.0	0.9956002514142050	7.69668042173411E-06	0.9924717691342540	0.9987373737373740	0.9955947136563880	6.59715464720066E-06
slm	Pangutaran Sama	35,200	1000	0.992964824120603	0.988	0.9904761904761910	7.6984410656842E-06	0.9929577464788730	0.987	0.9899699097291880	7.6984410656842E-06
peg	Pengo	350,000	783	0.9936548223350250	1.0	0.9968173138128580	5.49757447014377E-06	0.9949174078780180	1.0	0.9974522292993630	4.39805957611502E-06
tpp	Pisaflores Tepehua	4,000	1000	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06	0.998997995991984	0.997	0.997997997997998	1.09977729509774E-06
ktj	Plapo Krumen	100	1000	0.9910358565737050	0.995	0.9930139720558880	9.89799565587969E-06	0.9910358565737050	0.995	0.9930139720558880	9.89799565587969E-06
plj	Polci	22,000	929	0.9956663055254600	0.9892357373519910	0.99244060475162	4.3987657063428E-06	0.9956616052060740	0.9881593110871910	0.9918962722852510	4.3987657063428E-06
pwr	Powari	214,000	955	0.9852320675105490	0.9780104712041890	0.9816079873883340	1.53961201777152E-05	0.9862433862433860	0.9759162303664920	0.9810526315789470	1.42963973078784E-05
pub	Purum	500	102	0.9801980198019800	0.9705882352941180	0.9753694581280790	2.19738445328525E-06	1.0	0.9705882352941180	0.9850746268656720	0.0
pbb	Páez	77,400	1000	1.0	0.994	0.9969909729187560	0.0	1.0	0.994	0.9969909729187560	0.0
qva	Quechua, Ambo-Pasco	90,000	1000	0.9987980769230770	0.831	0.9072052401746720	1.09977729509774E-06	0.998780487804878	0.819	0.9000000000000000	1.09977729509774E-06
rki	Rakhine	730,000	793	0.9962264150943400	0.9987389659520810	0.9974811083123430	3.29858095047511E-06	0.9962264150943400	0.9987389659520810	0.9974811083123430	3.29858095047511E-06
rei	Reli	19,000	557	0.9981751824817520	0.9820466786355480	0.9900452488687780	1.09924174304565E-06	0.9981718464351010	0.9802513464991020	0.9891304347826090	1.09924174304565E-06
rmb	Rembarunga	69	2	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mnv	Rennell-Bellona	4,390	482	0.995850622406639	0.995850622406639	0.995850622406639	2.19830225117142E-06	0.995850622406639	0.995850622406639	0.995850622406639	2.19830225117142E-06
rhg	Rohingya	1m	786	1.0	0.9720101781170480	0.9858064516129030	0.0	1.0	0.9681933842239190	0.9838396897220430	0.0
rmn	Romani, Balkan	30,000	228	1.0	0.9956140350877190	0.9978021978021980	0.0	1.0	0.9956140350877190	0.9978021978021980	0.0
rol	Romblomanon	200,000	1000	0.9909365558912390	0.984	0.9874560963371800	9.89799565587969E-06	0.9919354838709680	0.984	0.9879518072289160	8.79821836078194E-06
sav	Saafi-Saafi	114,000	1000	0.9989785495403470	0.978	0.988377968671046	1.09977729509774E-06	0.9989775051124740	0.977	0.9878665318503540	1.09977729509774E-06
sch	Sakachep	25,000	790	0.9937027707808560	0.9987341772151900	0.9962121212121210	5.49761678312452E-06	0.9937027707808560	0.9987341772151900	0.9962121212121210	5.49761678312452E-06
xsb	Sambal	103,000	1000	0.9959839357429720	0.992	0.9939879759519040	4.39910918039097E-06	0.9969849246231160	0.992	0.9944862155388470	3.29933188529323E-06
cok	Santa Teresa Cora	7,000	1000	1.0	0.997	0.99849774661993	0.0	1.0	0.997	0.99849774661993	0.0
asj	Sari	7,600	1000	0.9979818365287590	0.989	0.99347061778001	2.19955459019549E-06	0.997979797979798	0.988	0.9929648241206030	2.19955459019549E-06
stw	Satawalese	460	531	0.9980952380952380	0.9868173258003770	0.9924242424242420	1.09921032730087E-06	0.9980952380952380	0.9868173258003770	0.9924242424242420	1.09921032730087E-06
lip	Sekpele	23,400	1000	0.9979879275653920	0.992	0.9949849548645940	2.19955459019549E-06	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06
slu	Selaru	8,000	1000	0.9898167006109980	0.972	0.9808274470232090	1.09977729509774E-05	0.9907881269191400	0.968	0.9792615073343450	9.89799565587969E-06
snw	Selee	11,300	1000	0.996	0.996	0.996	4.39910918039097E-06	0.996	0.996	0.996	4.39910918039097E-06
set	Sentani	30,000	1000	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06	0.9989969909729190	0.996	0.9974962443665500	1.09977729509774E-06
xsr	Sherpa	50,000	1000	0.997	0.997	0.997	3.29933188529323E-06	0.9979959919839680	0.996	0.9969969969969970	2.19955459019549E-06
shr	Shi	654,000	1000	1.0	0.982	0.9909182643794150	0.0	1.0	0.981	0.9904088844018170	0.0
sle	Sholaga	24,000	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.994	0.9969909729187560	0.0
pll	Shwe Palaung	148,000	68	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
sil	Sisaala, Tumulung	105,000	1000	0.998989898989899	0.989	0.9939698492462310	1.09977729509774E-06	0.9989888776541960	0.988	0.9934640522875820	1.09977729509774E-06
akp	Siwu	27,000	1000	0.998992950654582	0.992	0.9954841946813850	1.09977729509774E-06	0.998991935483871	0.991	0.9949799196787150	1.09977729509774E-06
slk	Slovak	4,61m	1000	1.0	0.963	0.9811512990320940	0.0	1.0	0.957	0.9780275932549820	0.0
sby	Soli	54,400	794	0.9924906132665830	0.9987405541561710	0.9956057752667920	6.59716915471571E-06	0.993734335839599	0.9987405541561710	0.9962311557788940	5.49764096226309E-06
sor	Somrai	7,410	695	1.0	0.9942446043165470	0.9971139971139970	0.0	1.0	0.9928057553956840	0.996389891696751	0.0
wib	Southern Toussian	19,500	1000	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
syl	Sylheti	7m	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
tap	Taabwa	250,000	754	0.9986357435197820	0.9708222811671090	0.98453261600538	1.09947983608955E-06	0.9986282578875170	0.9655172413793100	0.9817936614969660	1.09947983608955E-06
klg	Tagakolu	83,000	1000	1.0	0.99	0.9949748743718590	0.0	1.0	0.99	0.9949748743718590	0.0
tgj	Tagin	59,000	792	0.9949748743718590	1.0	0.9974811083123430	4.39810309813378E-06	0.9974811083123430	1.0	0.9987389659520810	2.19905154906689E-06

Table 17: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.

Code (ISO-639-3)	Name	Speakers	Support ($\theta = 0.3$)	Precision ($\theta = 0.3$)	Recall ($\theta = 0.3$)	F1 ($\theta = 0.3$)	FPR ($\theta = 0.3$)	Precision ($\theta = 0.5$)	Recall ($\theta = 0.5$)	F1 ($\theta = 0.5$)	FPR ($\theta = 0.5$)
thv	Tahaggart Tamahaq	25,000	395	1.0	0.979746835443038	0.989769820971867	0.0	1.0	0.979746835443038	0.989769820971867	0.0
blt	Tai Dam	699,000	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tlj	Talinga-Bwisi	68,500	1000	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06	0.998001998001998	0.999	0.9985007496251870	2.19955459019549E-06
ttq	Tamajaq, Tawallammat	450,000	913	1.0	0.9956188389923330	0.9978046103183320	0.0	1.0	0.9934282584885000	0.9967032967032970	0.0
taq	Tamasheq	250,000	702	1.0	0.9971509971509970	0.9985734664764620	0.0	1.0	0.9971509971509970	0.9985734664764620	0.0
tpu	Tampuan	31,100	84	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
rif	Tarifit	1.5m	1000	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
yer	Tarok	300,000	1000	1.0	0.992	0.9959839357429720	0.0	1.0	0.99	0.9949748743718590	0.0
tsg	Tausug	900,000	1000	0.9959919839679360	0.994	0.994994994994995	4.39910918039097E-06	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06
tvn	Tavoyan	400,000	822	1.0	0.9975669099756690	0.9987819732034110	0.0	1.0	0.9975669099756690	0.9987819732034110	0.0
tyx	Teke-Tyee	14,400	1000	1.0	0.977	0.9883662114314620	0.0	1.0	0.976	0.9878542510121460	0.0
kdh	Tem	204,000	1000	1.0	0.995	0.9974937343358400	0.0	1.0	0.995	0.9974937343358400	0.0
ted	Tepo Krumen	28,300	1000	0.9969788519637460	0.99	0.9934771700953340	3.29933188529323E-06	0.9969758064516130	0.989	0.9929718875502010	3.29933188529323E-06
nod	Thai, Northern	6m	50	1.0	0.94	0.9690721649484540	0.0	1.0	0.94	0.9690721649484540	0.0
thr	Tharu, Rana	336,000	1000	0.9867751780264500	0.97	0.978315683308119	1.42971048362707E-05	0.9907786885245900	0.967	0.9787449392712550	9.89799565587969E-06
bod	Central Tibetan	1.07m	947	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
tcn	Tichurong	2,420	269	0.9924812030075190	0.9814126394052050	0.9869158878504670	2.19778770689424E-06	0.9962264150943400	0.9814126394052050	0.9887640449438200	1.09889385344712E-06
tkr	Tsakhur	13,000	618	0.9983844911147010	1.0	0.9991915925626520	1.09931545626538E-06	0.9983818770226540	0.9983818770226540	0.9983818770226540	1.09931545626538E-06
tsn	Tswana	3.41m	758	0.9986807387862800	0.9986807387862800	0.9986807387862800	1.09948467153445E-06	0.9986807387862800	0.9986807387862800	0.9986807387862800	1.09948467153445E-06
bov	Tuwuli	11,400	1000	0.994	0.994	0.994	6.59866377058646E-06	0.9959839357429720	0.992	0.9939879759519040	4.39910918039097E-06
udi	Udi	4,200	1000	0.9989837398373980	0.983	0.9909274193548390	1.09977729509774E-06	0.9989806320081550	0.98	0.989399293286219	1.09977729509774E-06
vaa	Vaagri Booli	9,300	1000	0.9979939819458380	0.995	0.9964947421131700	2.19955459019549E-06	0.9979939819458380	0.995	0.9964947421131700	2.19955459019549E-06
vgr	Vaghri	10,000	1000	0.9979899497487440	0.993	0.9954887218045110	2.19955459019549E-06	0.9979838709677420	0.99	0.9939759036144580	2.19955459019549E-06
wbq	Waddar	1.93m	510	0.9863013698630140	0.9882352941176470	0.9872673849167480	7.69429468049441E-06	0.9921259842519690	0.9882352941176470	0.9901768172888020	4.39673981742538E-06
fad	Wagi	3,380	224	1.0	0.9419642857142860	0.9701149425287360	0.0	1.0	0.9419642857142860	0.9701149425287360	0.0
wja	Waja	60,000	252	1.0	0.9880952380952380	0.9940119760479040	0.0	1.0	0.9880952380952380	0.9940119760479040	0.0
cou	Wamey	18,400	1000	1.0	0.99	0.9949748743718590	0.0	1.0	0.99	0.9949748743718590	0.0
mfi	Wandala	23,500	1000	0.998989898989899	0.989	0.9939698492462310	1.09977729509774E-06	1.0	0.988	0.993963782696177	0.0
hrw	Warwar Feni	4,200	1000	1.0	0.971	0.9852866565195330	0.0	1.0	0.968	0.983739837398374	0.0
guc	Wayuu	170,000	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
kyu	Western Kayah	100,000	1000	1.0	0.999	0.9994997498749370	0.0	1.0	0.999	0.9994997498749370	0.0
suc	Western Subanon	75,000	1000	0.996	0.996	0.996	4.39910918039097E-06	0.9979919678714860	0.994	0.9959919839679360	2.19955459019549E-06
wlo	Wolio	65,000	1000	0.9989764585465710	0.976	0.9873545776428930	1.09977729509774E-06	0.9989754098360660	0.975	0.9868421052631580	1.09977729509774E-06
wob	Wè Northern	156,000	1000	1.0	0.993	0.9964877069744110	0.0	1.0	0.993	0.9964877069744110	0.0
nce	Yale	600	350	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
ymp	Yamap	1,580	502	0.9979959919839680	0.9920318725099600	0.9950049950049950	1.09917528878083E-06	0.9979959919839680	0.9920318725099600	0.9950049950049950	1.09917528878083E-06
yam	Yamba	40,800	1000	1.0	0.993	0.9964877069744110	0.0	1.0	0.993	0.9964877069744110	0.0
yns	Yansi	100,000	1000	1.0	0.99	0.9949748743718590	0.0	1.0	0.99	0.9949748743718590	0.0
zae	Yareni Zapotec	2,900	1000	1.0	0.993	0.9964877069744110	0.0	1.0	0.992	0.9959839357429720	0.0
jnj	Yemsa	81,600	147	0.9931972789115650	0.9931972789115650	0.9931972789115650	1.09874654993583E-06	1.0	0.9931972789115650	0.9965870307167240	0.0
yij	Yindjibarndi	330	5	1.0	1.0	1.0	0.0	1.0	0.8	0.888888888888889	0.0
zak	Zanaki	100,000	1000	0.9919597989949750	0.987	0.9894736842105260	8.79821836078194E-06	0.9919436052366570	0.985	0.9884596086302060	8.79821836078194E-06
zin	Zinza	138,000	1000	1.0	0.984	0.9919354838709680	0.0	1.0	0.983	0.9914271306101870	0.0
gel	ut-Ma'in	36,000	869	0.9976798143851510	0.9896432681242810	0.9936452917388790	2.19923774419786E-06	0.9976798143851510	0.9896432681242810	0.9936452917388790	2.19923774419786E-06
bwo	Borna	19,900	794	0.9987421383647800	1.0	0.9993706733794840	1.09952819245262E-06	1.0	1.0	1.0	0.0
ena	Apal	980	990	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
mni	Meitei	1.37m	790	0.9987341772151900	0.9987341772151900	0.9987341772151900	1.0995233566249E-06	0.9987341772151900	0.9987341772151900	0.9987341772151900	1.0995233566249E-06
nde	Ndebele	1.55m	1000	0.998	0.998	0.998	2.19955459019549E-06	0.998998998998999	0.998	0.9984992496248120	1.09977729509774E-06
nlx	Nahali	15,000	794	0.9924812030075190	0.9974811083123430	0.9949748743718590	6.59716915471571E-06	0.9949748743718590	0.9974811083123430	0.9962264150943400	4.39811276981047E-06
sna	Shona	10.7m	1000	0.9940357852882700	1.0	0.9970089730807580	6.59866377058646E-06	0.9970059880239520	0.999	0.998001998001998	3.29933188529323E-06
tdx	Malagasy, Tandroy-Mahafaly	1m	1000	0.998003992015968	1.0	0.9990009990009990	2.19955459019549E-06	0.999000999000999	1.0	0.9995002498750630	1.09977729509774E-06
uro	Ura	1,900	789	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
yao	Yao	1m	794	0.9900124843945070	0.9987405541561710	0.9943573667711600	8.79622553962095E-06	0.9912390488110140	0.9974811083123430	0.9943502824858760	7.69669734716833E-06
yom	Kiyombe	1.47m	1000	0.9920634920634920	1.0	0.9960159362549800	8.79821836078194E-06	0.9930417495029820	0.999	0.9960119641076770	7.6984410656842E-06

Table 18: LID Performance on the Full-LID Set with Thresholds: 0.3 and 0.5.