arXiv:2005.13962v1 [cs.CL] 28 May 2020

A Corpus for Large-Scale Phonetic Typology

Elizabeth Salesky^c Eleanor Chodroff^γ Tiago Pimentel³ Matthew Wiesner^c Ryan Cotterell^{3,ð} Alan W Black⁴ Jason Eisner^c

^cJohns Hopkins University ³University of York

John's Hopkins University University of for

³University of Cambridge ^dETH Zürich ⁴Carnegie Mellon University esalesky@jhu.edu eleanor.chodroff@york.ac.uk

Abstract

A major hurdle in data-driven research on typology is having sufficient data in many languages to draw meaningful conclusions. We present VoxClamantis v1.0, the first largescale corpus for phonetic typology, with aligned segments and estimated phonemelevel labels in 690 readings spanning 635 languages, along with acoustic-phonetic measures of vowels and sibilants. Access to such data can greatly facilitate investigation of phonetic typology at a large scale and across many languages. However, it is nontrivial and computationally intensive to obtain such alignments for hundreds of languages, many of which have few to no resources presently available. We describe the methodology to create our corpus, discuss caveats with current methods and their impact on the utility of this data, and illustrate possible research directions through a series of case studies on the 48 highest-quality readings. Our corpus and scripts are publicly available for non-commercial use at https:// voxclamantisproject.github.io.

1 Introduction

Understanding the range and limits of crosslinguistic variation is fundamental to the scientific study of language. In speech and particularly phonetic typology, this involves exploring potentially universal tendencies that shape sound systems and govern phonetic structure. Such investigation requires access to large amounts of cross-linguistic data. Previous cross-linguistic phonetic studies have been limited to a small number of languages with available data (Disner, 1983; Cho and Ladefoged, 1999), or have relied on previously reported measures from many studies (Whalen and Levitt, 1995; Becker-Kristal, 2010; Gordon and Roettger, 2017; Chodroff et al., 2019).



Figure 1: The 635 languages of our corpus geo-located with mean Mel Cepstral Distortion (MCD) scores.

Existing multilingual speech corpora have similar restrictions, with data too limited for many tasks (Engstrand and Cunningham-Andersson, 1988; Ladefoged and Maddieson, 2007) or approximately 20 to 30 recorded languages (Ardila et al., 2020; Harper, 2011; Schultz, 2002).

The recently developed CMU Wilderness corpus (Black, 2019) constitutes an exception to this rule with over 600 languages. This makes it the largest and most typologically diverse speech corpus to date. In addition to its coverage, the CMU Wilderness corpus is unique in two additional aspects: cleanly recorded, read speech exists for all languages in the corpus, and the same content (modulo translation) exists across all languages.

However, this massively multilingual speech corpus is challenging to work with directly. Copyright, computational restrictions, and sheer size limit its accessibility. Due to copyright restrictions, the audio cannot be directly downloaded with the sentence and phoneme alignments. A researcher would need to download original audio MP3 and text through links to bible.is, then segment these with speech-to-text sentence alignments distributed in Black (2019).¹ For phonetic research, subsequently identifying examples of specific phonetic segments in the audio is also a near-essential

¹The stability of the links and recording IDs is also questionable. Since the release of Black (2019), many of the links have already changed, along with a few of the IDs. We have begun identifying these discrepancies, and plan to flag these in a future release.

step for extracting relevant acoustic-phonetic measurements. Carrying out this derivative step has allowed us to release a stable-access collection of token-level acoustic-phonetic measures to enable further research.

Obtaining such measurements requires several processing steps: estimating pronunciations, aligning them to the text, evaluating alignment quality, and finally, extracting phonetic measures. This work is further complicated by the fact that, for a sizable number of these languages, no linguistic resources currently exist (e.g., language-specific pronunciation lexicons). We adapt speech processing methods based on Black (2019) to accomplish these tasks, though not without noise: in §3.4, we identify three significant caveats when attempting to use our extended corpus for large-scale phonetic studies.

We release a comprehensive set of standoff markup of over 400 million labeled segments of continuous speech.² For each segment, we provide an estimated phoneme-level label from the X-SAMPA alphabet, the preceding and following labels, and the start position and duration in the audio. Vowels are supplemented with formant measurements, and sibilants with standard measures of spectral shape.

We present a series of targeted case studies illustrating the utility of our corpus for large-scale phonetic typology. These studies are motivated by potentially universal principles posited to govern phonetic variation: **phonetic dispersion** and **phonetic uniformity**. Our studies both replicate known results in the phonetics literature and also present novel findings. Importantly, these studies investigate current methodology as well as questions of interest to phonetic typology at a large scale.

2 Original Speech

The CMU Wilderness corpus (Black, 2019) consists of recorded readings of the New Testament of the Bible in many languages and dialects. Following the New Testament structure, these data are broken into 27 books, each with a variable number of chapters between 1 and 25. Bible chapters contain standardized verses (approximately sentence-level segments); however, the speech is originally split only by chapter. Each chapter



Figure 2: The extraction process for the measurements released in VoxClamantis v1.0.

has an average of 13 minutes of speech for a total of \approx 20 hours of speech and text per language. These recordings are clean, read speech with a sampling rate of 16 kHz. In most languages, they are non-dramatic readings with a single speaker; in some, they are dramatic multi-speaker readings with additive music.³ The release from Black (2019) includes several resources for processing the corpus: scripts to download the original source data from bible.is, 'lexicons' created using grapheme-to-phoneme (G2P) conversion, and scripts to apply their generated sentence alignments, which facilitates downstream language processing tasks, including phoneme alignment.

3 The VoxClamantis v1.0 Corpus

Our VoxClamantis v1.0 corpus is derived from 690 audio readings of the New Testament of the Bible⁴ in 635 languages.⁵ We mark estimated speech seg-

²For some languages, we provide multiple versions of the markup based on different methods of predicting the pronunciation and generating time alignments (§3.1).

³Information about the recordings available can be found at https://www.faithcomesbyhearing.com/mission/recordings ⁴Nine of the readings from Black (2019) could not be aligned.

⁵We specify number of distinct languages by the number of

distinct ISO 639-3 codes, which may not distinguish dialects.

ments labeled with phonemic labels, and phonetic measures for the tokens that are vowels or sibilants. The extraction process is diagrammed in Figure 2. In the sections below, we detail our procedures for extracting labeled audio segments and their phonetic measures, in both high- and low-resource languages. We then outline important caveats to keep in mind when using this corpus.

3.1 Extracting Phoneme Alignments

We use a multi-pronged forced alignment strategy to balance broad language coverage ($\S3.1.1$) with utilization of existing high-quality resources (\$3.1.2). We assess the quality of our approaches in \$3.1.3. We release the stand-off markup for our final alignments as both text files and Praat TextGrids (Boersma and Weenink, 2019).⁶

Using scripts and estimated boundaries from Black (2019), we first download and convert the audio MP3s to waveforms, and cut the audio and text into 'sentences' (hereafter called 'utterances' as they are not necessarily sentences). This step creates shorter-length speech samples to facilitate forced alignment; utterance boundaries do not change through our processing.

To extract labeled segments, we first require pronunciations for each utterance. A pronunciation is predicted from the text alone using some graphemeto-phoneme (G2P) method. Each word's predicted pronunciation is a sequence of categorical labels, which are 'phoneme-level' in the sense that they are usually intended to distinguish the words of the language. We then align this predicted sequence of 'phonemes' to the corresponding audio.

3.1.1 All Languages

Most of our languages have neither existing pronunciation lexicons nor G2P resources. To provide coverage for all languages, we generate pronunciations using the simple 'universal' G2P system Unitran (Qian et al., 2010, as extended by Black, 2019), which deterministically expands each grapheme to a fixed sequence of phones in the Extended Speech Assessment Methods Phonetic Alphabet (X-SAMPA) (Wells, 1995/2000). This naive process is error-prone for languages with opaque orthographies, as we show in §3.1.3 below and discuss further in §3.4 (Caveat **B**). Even so, it provides a starting point for exploring low-resource languages: after some manual inspection, a linguist may be able to correct the labels in a given language by a combination of manual and automatic methods.

For each reading, to align the pronunciation strings to the audio, we fit a generative acoustic model designed for this purpose: specifically, eHMM (Prahallad et al., 2006) as implemented in Festvox (Anumanchipalli et al., 2011) to run full Baum–Welch from a flat start for 15 to 30 iterations until the mean mel cepstral distortion score (see §3.1.3) converges. Baum-Welch does not change the predicted phoneme labels, but obtains a language-specific, reading-specific, contextual (triphone) acoustic model for each phoneme type in the language. We then use Viterbi alignment to identify an audio segment for each phoneme token.

3.1.2 High-Resource Languages

A subset of the languages in our corpus are supported by existing pronunciation resources. Two such resources are Epitran (Mortensen et al., 2018), a G2P tool based on language-specific rules, available in both IPA and X-SAMPA, and WikiPron (Lee et al., 2020), a collection of crowd-sourced pronunciations scraped from Wiktionary. These are mapped from IPA to X-SAMPA for label consistency across our corpus. Epitran covers 29 of our languages (39 readings), while WikiPron's 'phonemic' annotations⁷ provide partial coverage of 13 additional languages (18 readings). We use Epitran for languages with regular orthographies where it provides high-quality support, and WikiPron for other languages covered by WikiPron annotations. While Unitran and Epitran provide a single pronunciation for a word from the orthography, WikiPron may include multiple pronunciations. In such cases, Viterbi alignment (see below) chooses the pronunciation of each token that best fits the audio.

For most languages covered by WikiPron, most of our corpus words are out-of-vocabulary, as they do not yet have user-submitted pronunciations on Wiktionary. We train G2P models on WikiPron annotations to provide pronunciations for these words. Specifically, we use the WFST-based tool Phonetisaurus (Novak et al., 2016). Model hyperparameters are tuned on 3 WikiPron languages from SIGMORPHON 2020 (Gorman et al., 2020) (see Appendix C for details). In general, for languages that are not easily supported by Epitran-style G2P *rules*, training a G2P *model* on sufficiently many

⁶Corresponding audio will need to be downloaded from source and split by utterance using scripts from Black (2019).

⁷WikiPron annotations are available at both the phonemic and phonetic level, with a greater number of phonemic annotations, which we use here.

ISO 639-3	tpi	ron	azj	msa	ceb	tur	tgl	spa	ilo	rus	hau	ind	tgk	jav	kaz
# Types	1398	9746	18490	7612	8531	21545	9124	11779	15063	16523	4938	5814	12502	10690	20502
Unitran PER	18.4	21.3	26.9	30.1	30.1	31.2	34.4	34.4	35.0	37.4	37.6	38.8	39.8	49.9	46.8
# Tokens	291k	169k	125k	157k	190k	125k	185k	168k	169k	130k	201k	170k	159k	177k	142k
Weighted PER	20.1	21.3	26.1	31.1	35.9	28.5	40.1	32.6	32.7	36.8	36.7	40.5	38.8	54.1	47.7
ISO 639-3	swe	kmr	som	tir	pol	hae	vie	tha	lao	ben	tel	hin	mar	tam	
# Types	8610	8127	14375	22188	18681	15935	2757	23338	31334	8075	23477	7722	17839	31642	
Unitran PER	46.9	54.3	54.6	57.8	67.1	67.3	73.8	80.3	89.1	90.0	90.3	95.7	97.8	100.5	
# Tokens	165k	176k	156k	121k	141k	164k	211k	26k	36k	173k	124k	191k	159k	139k	
Weighted PER	49.5	53.9	56.0	57.4	66.8	64.8	80.6	80.4	89.4	86.2	88.3	91.3	97.8	102.1	

Table 1: Phoneme Error Rate (PER) for Unitran treating Epitran as ground-truth. 'Types' and 'Tokens' numbers reflect the number of unique word types and word tokens in each reading. We report PER calculated using word types for calibration with other work, as well as frequency-weighted PER reflecting occurrences in our corpus.

high-quality annotations may be more accurate.

We align the speech with the high-quality labels using a multilingual ASR model (see Wiesner et al., 2019). The model is trained in Kaldi (Povey et al., 2011) on 300 hours of data from the IARPA BABEL corpora (21 languages), a subset of Wall Street Journal (English), the Hub4 Spanish Broadcast news (Spanish), and a subset of the Voxforge corpus (Russian and French). These languages use a shared X-SAMPA phoneme label set which has high coverage of the labels of our corpus.

Our use of a pretrained multilingual model here contrasts with §3.1.1, where we had to train reading-specific acoustic models to deal with the fact that the same Unitran phoneme label may refer to quite different phonemes in different languages (see §3.4). We did not fine-tune our multilingual model to each language, as the cross-lingual ASR performance in previous work (Wiesner et al., 2019) suggests that this model is sufficient for producing phoneme-level alignments.

3.1.3 Quality Measures

Automatically generated phoneme-level labels and alignments inherently have some amount of noise, and this is particularly true for low-resource languages. The noise level is difficult to assess without gold-labeled corpora for either modeling or assessment. However, for the high-resource languages, we can evaluate Unitran against Epitran and WikiPron, pretending that the latter are ground truth. For example, Table 1 shows Unitran's phoneme error rates relative to Epitran. Appendix B gives several more detailed analyses with examples of individual phonemes.

Unitran pronunciations may have acceptable phoneme error rates for languages with transparent orthographies and one-to-one grapheme-tophoneme mappings. Alas, without these conditions they prove to be highly inaccurate.

That said, evaluating Unitran labels against Epitran or WikiPron may be unfair to Unitran, since some discrepancies are arguably not errors but mere differences in annotation granularity. For example, the 'phonemic' annotations in WikiPron are sometimes surprisingly fine-grained: WikiPron frequently uses /t/ in Cebuano where Unitran only uses /t/, though these refer to the same phoneme. These tokens are scored as incorrect. Moreover, there can be simple systematic errors: Unitran always maps grapheme $\langle a \rangle$ to label /a/, but in Tagalog, all such tokens should be /a/. Such errors can often be fixed by remapping the Unitran labels, which in these cases would reduce PER from 30.1 to 6.8 (Cebuano) and from 34.4 to 7.8 (Tagalog). Such rules are not always this straightforward and should be created on a language-specific basis; we encourage rules created for languages outside of current Epitran support to be contributed back to the Epitran project.

For those languages where we train a G2P system on WikiPron, we compute the PER of the G2P system on held-out WikiPron entries treated as ground truth. The results (Appendix C) range from excellent to mediocre.

We care less about the pronunciations themselves than about the segments that we extract by aligning these pronunciations to the audio. For high-resource languages, we can again compare the segments extracted by Unitran to the higher-quality ones extracted with better pronunciations. For each Unitran token, we evaluate its label and temporal boundaries against the high-quality token that is closest in the audio, as measured by the temporal distance between their midpoints (Appendix B).

Finally, the segmentation of speech and text into corresponding utterances is not perfect. We use the utterance alignments generated by Black (2019), in which the text and audio versions of a putative

utterance may have only partial overlap. Indeed, Black (2019) sometimes failed to align the Unitran pronunciation to the audio at all, and discarded these utterances. For each remaining utterance, he assessed the match quality using Mel Cepstral Distortion (MCD)-which is commonly used to evaluate synthesized spoken utterances (Kominek et al., 2008)-between the original audio and a resynthesized version of the audio based on the aligned pronunciation. Each segment's audio was resynthesized given the segment's phoneme label and the preceding and following phonemes, in a way that preserves its duration, using CLUSTER-GEN (Black, 2006) with the same reading-specific eHMM model that we used for alignment. We distribute Black's per-utterance MCD scores with our corpus, and show the average score for each language in Appendix E. In some readings, the MCD scores are consistently poor.

3.2 Phonetic measures

Using the phoneme-level alignments described in §3.1, we automatically extract several standard acoustic-phonetic measures of vowels and sibilant fricatives that correlate with aspects of their articulation and abstract representation.

3.2.1 Vowel measures

Standard phonetic measurements of vowels include the formant frequencies and duration information. Formants are concentrations of acoustic energy at frequencies reflecting resonance points in the vocal tract during vowel production (Ladefoged and Johnson, 2014). The lowest two formants, F1 and F2, are considered diagnostic of vowel category identity and approximate tongue body height (F1) and backness (F2) during vowel production (Figure 3). F3 correlates with finer-grained aspects of vowel production such as rhoticity (/r/-coloring), lip rounding, and nasality (House and Stevens, 1956; Lindblom and Sundberg, 1971; Ladefoged et al., 1978), and F4 with high front vowel distinctions and speaker voice quality (Eek and Meister, 1994). Vowel duration can also signal vowel quality, and denotes lexical differences in many languages.

We extracted formant and duration information from each vowel using Praat (Boersma and Weenink, 2019). The first four formants (F1–F4) were measured at each quartile and decile of the vowel. Formant estimation was performed with the Burg algorithm in Praat with pre-emphasis from 50 Hz, a time window of 25 ms, a time



Figure 3: Vowel Chart

step of 6.25 ms, a maximum of five formants permitted, and a formant ceiling of 5000 Hz, which is the recommended value for a male vocal tract (Boersma and Weenink, 2019). Note that the speakers in this corpus are predominantly male.

3.2.2 Sibilant measures

Standard phonetic measurements of sibilant fricatives such as /s/, /z/, /ʃ/, and /ʒ/ include measures of spectral shape, and also segment duration. Measures of spectral shape frequently distinguish sibilant place of articulation: higher concentrations of energy generally reflect more anterior constriction locations (e.g., /s z/ are produced closer to the teeth than /ʃ ʒ/). Segment duration can also signal contrasts in voicing status (Jongman et al., 2000).

Our release contains the segment duration, spectral peak, the spectral moments of the frequency distribution (center of gravity: COG, variance, skewness, and kurtosis), as well as two measures of the mid-frequency peak determined by sibilant quality. These are the mid-frequency peak between 3000 and 7000 Hz for alveolar sibilants, and between 2000 and 6000 Hz for post-alveolar sibilants (Koenig et al., 2013; Shadle et al., 2016). The spectral information was obtained via multitaper spectral analysis (Rahim and Burr, 2017), with a time-bandwidth parameter (nw) of 4 and 8 tapers (k) over the middle 50% of the fricative (Blacklock, 2004). Measurements were made using the methods described in Forrest et al. (1988) for spectral moments and Koenig et al. (2013) for spectral peak varieties.

3.3 Computation times

Generating phoneme-level alignments and extracting subsequent phonetic measures takes significant time, computational resources, and domain knowledge. Our release enables the community to use this data directly without these prerequisites. Table 2 shows that the time to extract our resources,

	Computation Time					
Resource	Per Language	Total Time				
Utterance Alignments	30m	14d 13h				
Phoneme Alignments	3d 3h 37m	6y 12d 16h				
Vowel Measures	45m	21d 20h				
Sibilant Measures	20m	9d 17h				
	3d 5h 0m	6y 58d 19h				

Table 2: Computation time to generate the full corpus.

once methods have been developed, was more than 6 CPU years, primarily for training eHMM models.

3.4 General caveats

We caution that our labeling and alignment of the corpus contains errors. In particular, it is difficult to responsibly draw firm linguistic conclusions from the Unitran-based segments (§3.1.1). In §5 we suggest future work to address these issues.

- A Quality of Utterance Pairs: For some utterances, the speech does not correspond completely to the text, due to incorrect cosegmentation. In our phonetic studies, we threshold using reading-level MCD as a heuristic for overall alignment quality, and further threshold remaining readings using utterance-level MCD. We recommend others do so as well.
- **B** Phoneme Label Consistency and Accuracy: Phoneme-level labels are predicted from text without the aid of audio using G2P methods. This may lead to systematic errors. In particular, Unitran relies on a 'universal' table that maps grapheme $\langle s \rangle$ (for example) to phoneme /s/in every context and every language. This is problematic for languages that use $\langle s \rangle$ in some or all contexts to refer to other phonemes such as f/ or s/, or use digraphs that contain < s>, such as $\langle sh \rangle$ for /[/. Thus, the predicted label /s/ may not consistently refer to the same phoneme within a language, nor to phonetically similar phonemes across languages. Even WikiPron annotations are user-submitted and may not be internally consistent (e.g., some words use $/d_{3}/d_{3}$ or /t/ while others use $/d_{z}/$ or /t/), nor comparable across languages.

'Phoneme' inventories for Unitran and WikiPron have been implicitly chosen by whoever designed the language's orthography or its WikiPron pages; while this may reflect a reasonable folk phonology, it may not correspond to the inventory of underlying or surface phonemes that any linguist would be likely to posit.

- C Label and Alignment Assessment: While alignment quality for languages with Epitran and WikiPron can be assessed and calibrated beyond this corpus, it cannot for those languages with only Unitran alignments; the error rate on languages without resources to evaluate PER is unknown to us. The Unitran alignments should be treated as a first-pass alignment which may still be useful for a researcher who is willing to perform quality control and correction of the alignments using automatic or manual procedures. Our automatically-generated alignment offers an initial label and placement of the boundaries that would hopefully facilitate downstream analysis.
- **D** Corpus Representation: It is difficult to draw conclusions about 'average behavior' across languages. Some language families are better represented in the corpus than others, with more languages, more Bible readings per language, more hours of speech per reading, or more examples of a given phoneme of interest.⁸ Additionally, the recordings by language are largely single-speaker (and predominantly male). This means that we can often draw conclusions only about a particular speaker's idiolect, rather than the population of speakers of the language. Metadata giving the exact number of different speakers per recording do not exist.

4 Phonetic Case Studies

We present two case studies to illustrate the utility of our resource for exploration of crosslinguistic typology. Phoneticians have posited several typological principles that may structure phonetic systems. Though previous research has provided some indication as to the direction and magnitude of expected effects, many instances of the principles have not yet been explored at scale. Our case studies investigate how well they account for cross-linguistic variation and systematicity for our phonetic measures from vowels and sibilants. Below we present the data filtering methods for our case studies, followed by an introduction to and evaluation of phonetic dispersion and uniformity.

4.1 Data filtering

For quality, we use only the tokens extracted using high-resource pronunciations (Epitran and WikiPron) and only in languages with mean

⁸See our corpus website for exact numbers of utterances and our phonetic measures per each language.

MCD lower than 8.0.9 Furthermore, we only use those utterances with MCD lower than 6.0. The vowel analyses focus on F1 and F2 in ERB taken at the vowel midpoint (Zwicker and Terhardt, 1980; Glasberg and Moore, 1990).¹⁰ The sibilant analyses focus on mid-frequency peak of /s/ and /z/, also in ERB. Vowel tokens with F1 or F2 measures beyond two standard deviations from the labeland reading-specific mean were excluded, as were tokens for which Praat failed to find a measurable F1 or F2, or whose duration exceeded 300 ms. Sibilant tokens with mid-frequency peak or duration measures beyond two standard deviations from the label- and reading-specific mean were also excluded. When comparing realizations of two labels such as /i/-/u/ or /s/-/z/, we excluded readings that did not contain at least 50 tokens of each label. We show data representation with different filtering methods in Appendix D.

After filtering, the vowel analyses included 48 readings covering 38 languages and 11 language families. The distribution of language families was 21 Indo-European, 11 Austronesian, 3 Creole/Pidgin, 3 Turkic, 2 Afro-Asiatic, 2 Tai-Kadai, 2 Uto-Aztecan, 1 Austro-Asiatic, 1 Dravidian, 1 Hmong-Mien, and 1 Uralic. Approximately 8.2 million vowel tokens remained, with a minimum of $\approx 31,000$ vowel tokens per reading. The sibilant analysis included 22 readings covering 18 languages and 6 language families. The distribution of language families was 10 Indo-European, 6 Austronesian, 3 Turkic, 1 Afro-Asiatic, 1 Austro-Asiatic, and 1 Creole/Pidgin. The decrease in total number of readings relative to the vowel analysis primarily reflects the infrequency of /z/ cross-linguistically. Approximately 385,000 /s/ and 83,000 /z/ tokens remained, with a minimum of \approx 5,200 tokens per reading.

4.2 Phonetic dispersion

Phonetic dispersion refers to the principle that contrasting speech sounds should be distinct from one another in phonetic space (Martinet, 1955; Jakobson, 1968; Flemming, 1995, 2004). Most studies investigating this principle have focused on its validity within vowel systems, as we do here. While languages tend to have seemingly well-dispersed vowel inventories such as {/i/, /a/, /u/} (Joos, 1948; Stevens and Keyser, 2010), the actual phonetic realization of each vowel can vary substantially (Lindau and Wood, 1977; Disner, 1983). One prediction of dispersion is that the number of vowel categories in a language should be inversely related to the degree of per-category acoustic variation (Lindblom, 1986). Subsequent findings have cast doubt on this (Livijn, 2000; Recasens and Espinosa, 2009; Vaux and Samuels, 2015), but these studies have been limited by the number and diversity of languages investigated.

To investigate this, we measured the correlation between the number of vowel categories in a language and the degree of per-category variation, as measured by the *joint* entropy of (F1, F2) conditioned on the vowel category. We model $p(F1, F2 \mid V)$ using a bivariate Gaussian for each vowel type v. We can then compute the joint conditional entropy under this model as $H(F1, F2 \mid V) = \sum_{v} p(v) H(F1, F2 \mid V = v) =$ $\sum_{v} p(v) \frac{1}{2} \ln \det(2\pi e \Sigma_v)$, where Σ_v is the covariance matrix for the model of vowel v.

Vowel inventory sizes per reading ranged from 4 to 20 vowels, with a median of 8. Both Spearman and Pearson correlations between entropy estimate and vowel inventory size across analyzed languages were small and not significant (Spearman $\rho = 0.11$, p = 0.44; Pearson r = 0.11, p = 0.46), corroborating previous accounts of the relationship described in Livijn (2000) and Vaux and Samuels (2015) with a larger number of languages—a larger vowel inventory does not necessarily imply more precision in vowel category production.¹¹

4.3 Phonetic uniformity

Previous work suggests that F1 is fairly uniform with respect to phonological height. Within a single language, the mean F1s of /e/ and /o/—which share a height—have been found to be correlated across speakers (*Yorkshire English*: Watt, 2000; *French*: Ménard et al., 2008; *Brazilian Portuguese*: Oushiro, 2019; *Dutch, English, French, Japanese*, *Portuguese, Spanish*: Schwartz and Ménard, 2019). Though it is physically possible for these vowels

⁹In the high-MCD languages, even the low-MCD utterances seem to be untrustworthy.

¹⁰The Equivalent Rectangular Bandwidth (ERB) scale is a psychoacoustic scale that better approximates human perception, which may serve as auditory feedback for the phonetic realization (Fletcher, 1923; Nearey, 1977; Zwicker and Terhardt, 1980; Glasberg and Moore, 1990). The precise equation comes from Glasberg and Moore (1990, Eq. 4).

¹¹Since differential entropy is sensitive to parameterization, we also measured this correlation using formants in hertz, instead of in ERB, as ERB is on a logarithmic scale. This change did not the influence the pattern of results (Spearman $\rho = 0.12, p = 0.41$; Pearson r = 0.13, p = 0.39).



(a) F1 of /i/-/u/ in ERB

(b) Mid-frequency peak of /s/-/z/ in ERB

Figure 4: Correlations of mean F1 (ERB) between *ii* and *u* and of mean mid-frequency peak (ERB) between *is* and *iz*. The paired segments share a relevant phonological feature specification that is approximated by the acoustic-phonetic measurement: vowel height by F1 and sibilant place by mid-frequency peak. Each reading is represented by an ellipsoid, centered on the paired means and shaped by $\frac{1}{10}$ of their respective standard deviations. The solid line reflects the best-fit linear regression line with standard error in gray shading; the dashed line shows the line of equality. Marginal histograms show the range of variation in the segment-specific means.

to differ in F1 realization, the correlations indicate a strong tendency for languages and individual speakers to yoke these two representations together.

Systematicity in the realization of sibilant place of articulation has also been observed across speakers of American English and Czech (Chodroff, 2017). Phonetic correlates of sibilant place strongly covary between /s/ and /z/, which share a [+anterior] place of articulation and are produced the alveolar ridge, and between /ʃ/ and /ʒ/, which share a [-anterior] place of articulation and are produced behind the alveolar ridge.

A principle of uniformity may account for these above findings. Uniformity here refers to a principle in which a distinctive phonological feature should have a consistent phonetic realization, within a language or speaker, across different segments with that feature (Keating, 2003; Chodroff et al., 2019). Similar principles posited in the literature include Maximal Use of Available Controls, in which a control refers to an integrated perceptual and motor phonetic target (Ménard et al., 2008), as well as a principle of gestural economy (Maddieson, 1995). Phonetic realization refers to the mapping from the abstract distinctive feature to an abstract phonetic target. We approximate this phonetic target via an acoustic-phonetic measurement, but we emphasize that the acoustic measurement is not necessarily a direct reflection of an underlying phonetic target (which could be an articulatory gesture, auditory goal, or perceptuo-motor representation of the sound). We make the simplifying assumption that the acoustic-phonetic formants (F1, F2) directly correspond to phonetic targets linked to the vowel features of height and backness.

More precisely, uniformity of a phonetic measure with respect to a phonological feature means that any two segments sharing that feature will tend to have approximately equal measurements in a given language, even when that value varies across languages. We can observe whether this is true by plotting the measures of the two segments against each other by language (e.g., Figure 4).

Vowels. As shown in Figure 4 and Table 3, the strongest correlations in mean F1 frequently reflected uniformity of height (e.g., high vowels /i/–/u/: r = 0.79, p < 0.001, mid vowels /e/–/o/: r = 0.62, p < 0.01).¹² Nevertheless, some vowel pairs that differed in height were also moderately correlated in mean F1 (e.g., /o/–/a/: r = 0.66, p < 0.001). Correlations of mean F1 were overall moderate in strength, regardless of the vowels' phonological specifications.

Correlations of mean F2 were also strongest among vowels with a uniform backness specification (e.g., back vowels /u/–/o/: r = 0.69, p < 0.001; front vowels /i/–/ɛ/: r = 0.69, p < 0.05; Table 4). The correlation between front tense vowels /i/ and /e/ was significant and in the ex-

¹²*p*-values are corrected for multiple comparisons using the Benjamini-Hochberg correction and a false discovery rate of 0.25 (Benjamini and Hochberg, 1995).

pected direction, but also slightly weaker than the homologous back vowel pair (r = 0.41, p < 0.05). Vowels differing in backness frequently had negative correlations, which could reflect influences of category crowding or language-/speaker-specific differences in peripheralization. We leave further exploration of those relationships to future study.

The moderate to strong F1 correlations among vowels with a shared height specification are consistent with expectations based on previous studies, and also with predictions of uniformity. Similarly, we find an expected correlation of F2 means for vowels with a shared height specification. The correlations of vowel pairs that were predicted to have significant correlations, but did not, tended to have small sample sizes (< 14 readings).

Nevertheless, the correlations are not perfect; nor are the patterns. For instance, the back vowel correlations of F2 are stronger than the front vowel correlations. While speculative, the apparent peripheralization of /i/ (as revealed in the negative F2 correlations) could have weakened the expected uniformity relation of /i/ with other front vowels. Future research should take into account additional influences of the vowel inventory composition, as well as articulatory or auditory factors for a more complete understanding of the structural forces in the phonetic realization of vowels.

Sibilants. The mean mid-frequency peak values for /s/ and /z/ each varied substantially across readings, and were also strongly correlated with one another (r = 0.87, p < 0.001; Figure 4).¹³ This finding suggests a further influence of uniformity on the realization of place for /s/ and /z/, and the magnitude is comparable to previous correlations observed across American English and Czech speakers, in which r was ≈ 0.90 (Chodroff, 2017).

5 Directions for Future Work

We hope our corpus may serve as a touchstone for further improvements in phonetic typology research and methodology. Here we suggest potential steps forward for known areas (§3.4) where this corpus could be improved:

A Sentence alignments were generated using Unitran, and could be improved with higherquality G2P and verse-level text segmentation to standardize utterances across languages.

- **B** Consistent and comparable phoneme labels are the ultimate goal. Concurrent work on universal phone recognition (Li et al., 2020) addresses this issue through a universal phone inventory constrained by language-specific PHOIBLE inventories (Moran and McCloy, 2019). However, free-decoding phones from speech alone is challenging. One exciting possibility is to use the orthography and audio jointly to guide semi-supervised learning of per-language pronunciation lexicons (Lu et al., 2013; Zhang et al., 2017).
- C Reliable quality assessment for current methods remains an outstanding research question for many languages. For covered languages, using a universal label set to map additional high quality lexicons (e.g., hand-annotated lexicons) to the same label space as ours would enable direct label and alignment assessment through precision, recall, and PER.
- **D** Curating additional resources beyond this corpus would improve coverage and balance, such as contributing additional Epitran modules. Additional readings exist for many languages on the original bible.is site and elsewhere. Annotations with speaker information are not available, but improved unsupervised speaker clustering may also support better analysis.

6 Conclusion

VoxClamantis v1.0 is the first large-scale corpus for phonetic typology, with extracted phonetic features for 635 typologically diverse languages. We present two case studies illustrating both the research potential and limitations of this corpus for investigation of phonetic typology at a large scale. We discuss several caveats for the use of this corpus and areas for substantial improvement. Nonetheless, we hope that directly releasing our alignments and token-level features enables greater research accessibility in this area. We hope this corpus will motivate and enable further developments in both phonetic typology and methodology for working with cross-linguistic speech corpora.

Acknowledgments

The authors gratefully acknowledge Colin Wilson for his guidance and discussion on the topic, Florian Metze for resources, and Carlos Aguirre for helpful feedback.

 $^{^{\}rm 13}{\rm The}$ magnitude of this correlation did not change when using hertz (r=0.86, p<0.001).

References

- Gopala Krishna Anumanchipalli, Kishore Prahallad, and Alan W. Black. 2011. Festvox: Tools for creation and analyses of large speech corpora. In Workshop on Very Large Scale Phonetics Research, UPenn, Philadelphia.
- Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. 2020. Common Voice: A massivelymultilingual speech corpus. In Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020).
- Roy Becker-Kristal. 2010. Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus. Ph.D. thesis, University of California, Los Angeles.
- Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.
- Alan W. Black. 2006. CLUSTERGEN: A statistical parametric synthesizer using trajectory modeling. In *Proceedings of INTERSPEECH*.
- Alan W. Black. 2019. CMU Wilderness Multilingual Speech Dataset. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5971–5975, Brighton, UK. IEEE.
- Oliver Blacklock. 2004. Characteristics of Variation in Production of Normal and Disordered Fricatives, Using Reduced-Variance Spectral Methods. Ph.D. thesis, University of Southampton.
- Paul Boersma and David Weenink. 2019. Praat: Doing phonetics by computer [computer program]. version 6.0.45.
- Taehong Cho and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27(2):207–229.
- Eleanor Chodroff. 2017. *Structured Variation in Obstruent Production and Perception*. Ph.D. thesis, Johns Hopkins University.
- Eleanor Chodroff, Alessandra Golden, and Colin Wilson. 2019. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. *The Journal of the Acoustical Society of America*, 145(1):EL109–EL115.
- Sandra Ferrari Disner. 1983. Vowel Quality: The Relation between Universal and Language-specific Factors. Ph.D. thesis, UCLA.

- Gary F. Simons Eberhard, David M. and Charles D. Fennig, editors. 2020. *Ethnologue: Languages of the world*, 23 edition. SIL international. Online version: http://www.ethnologue.com.
- Arvo Eek and Einar Meister. 1994. Acoustics and perception of Estonian vowel types. *Phonetic Experimental Research*, XVIII:146–158.
- Olle Engstrand and Una Cunningham-Andersson. 1988. Iris - a data base for cross-linguistic phonetic research.
- Edward S. Flemming. 1995. Auditory Representations in Phonology. Ph.D. thesis, UCLA.
- Edward S. Flemming. 2004. Contrast and perceptual distinctiveness. In Bruce Hayes, R. Kirchner, and Donca Steriade, editors, *The Phonetic Bases* of *Phonological Markedness*, 1968, pages 232–276. University Press, Cambridge, MA.
- Harvey Fletcher. 1923. Physical measurements of audition and their bearing on the theory of hearing. *Journal of the Franklin Institute*, 196(3):289–326.
- Karen Forrest, Gary Weismer, Paul Milenkovic, and Ronald N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, 84(1):115–123.
- Brian R. Glasberg and Brian C.J. Moore. 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138.
- Matthew Gordon and Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, 3(1).
- Kyle Gorman, Lucas F.E. Ashby, Aaron Goyzueta, Arya D. McCarthy, Shijie Wu, and Daniel You. 2020. The SIGMORPHON 2020 shared task on multilingual grapheme-to-phoneme conversion. In *Proceedings of the SIGMORPHON Workshop*.
- Mary Harper. 2011. The IARPA Babel multilingual speech database. Accessed: 2020-05-01.
- Arthur S. House and Kenneth N. Stevens. 1956. Analog studies of the nasalization of vowels. *The Journal of Speech and Hearing Disorders*, 21(2):218–232.
- Roman Jakobson. 1968. *Child Language, Aphasia and Phonological Universals*. Mouton Publishers.
- Allard Jongman, Ratree Wayland, and Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263.
- Martin Joos. 1948. Acoustic phonetics. *Language*, 24(2):5–136.

- Patricia A. Keating. 2003. Phonetic and other influences on voicing contrasts. In Proceedings of the 15th International Congress of Phonetic Sciences, pages 20–23, Barcelona, Spain.
- Laura Koenig, Christine H. Shadle, Jonathan L. Preston, and Christine R. Mooshammer. 2013. Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, 56(4):1175–1189.
- John Kominek, Tanja Schultz, and Alan W. Black. 2008. Synthesizer voice quality of new languages calibrated with mean mel cepstral distortion. In *Spoken Languages Technologies for Under-Resourced Languages*.
- Peter Ladefoged, Richard Harshman, Louis Goldstein, and Lloyd Rice. 1978. Generating vocal tract shapes from formant frequencies. *The Journal of the Acoustical Society of America*, 64(4):1027–1035.
- Peter Ladefoged and Keith Johnson. 2014. A Course in *Phonetics*. Nelson Education.
- Peter Ladefoged and Ian Maddieson. 2007. The UCLA phonetics lab archive.
- Jackson L. Lee, Lucas F.E. Ashby, M. Elizabeth Garza, Yeonju Lee-Sikka, Sean Miller, Alan Wong, Arya D. McCarthy, and Kyle Gorman. 2020. Massively multilingual pronunciation mining with WikiPron. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC* 2020). European Language Resources Association (ELRA). Resources downloadable from https:// github.com/kylebgorman/wikipron.
- Xinjian Li, Siddharth Dalmia, Juncheng Li, Matthew Lee, Patrick Littell, Jiali Yao, Antonios Anastasopoulos, David R Mortensen, Graham Neubig, Alan W. Black, et al. 2020. Universal phone recognition with a multilingual allophone system. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 8249–8253. IEEE.
- Mona Lindau and Patricia Wood. 1977. Acoustic vowel spaces. UCLA Working Papers in Phonetics, 38:41–48.
- Björn Lindblom. 1986. Phonetic universals in vowel systems. In John J. Ohala and Jeri Jaeger, editors, *Experimental Phonology*, pages 13–44. Academic Press, Orlando.
- Björn Lindblom and Johan Sundberg. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, 50(4B):1166–1179.
- Peder Livijn. 2000. Acoustic distribution of vowels in differently sized inventories-hot spots or adaptive dispersion. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PER-ILUS)*, 11.

- Liang Lu, Arnab Ghoshal, and Steve Renals. 2013. Acoustic data-driven pronunciation lexicon for large vocabulary speech recognition. In 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, pages 374–379. IEEE.
- Ian Maddieson. 1995. Gestural economy. In Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm, Sweden.
- André Martinet. 1955. Économie Des Changements Phonétiques: Traité de Phonologie Diachronique, volume 10. Bibliotheca Romanica.
- Lucie Ménard, Jean-Luc Schwartz, and Jérôme Aubin. 2008. Invariance and variability in the production of the height feature in French vowels. *Speech Communication*, 50:14–28.
- Steven Moran and Daniel McCloy, editors. 2019. *PHOIBLE 2.0.* Max Planck Institute for the Science of Human History, Jena.
- David R. Mortensen, Siddharth Dalmia, and Patrick Littell. 2018. Epitran: Precision G2P for many languages. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Paris, France. European Language Resources Association (ELRA).
- Terrance M. Nearey. 1977. Phonetic Feature Systems for Vowels. Ph.D. thesis, University of Alberta. Reprinted 1978 by Indiana University Linguistics Club.
- Josef Robert Novak, Nobuaki Minematsu, and Keikichi Hirose. 2016. Phonetisaurus: Exploring graphemeto-phoneme conversion with joint n-gram models in the WFST framework. *Natural Language Engineering*, 22(6):907–938.
- Livia Oushiro. 2019. Linguistic uniformity in the speech of Brazilian internal migrants in a dialect contact situation. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, pages 686–690, Melbourne, Australia. Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society. IEEE Catalog No.: CFP11SRW-USB.
- Kishore Prahallad, Alan W. Black, and Ravishankhar Mosur. 2006. Sub-phonetic modeling for capturing pronunciation variations for conversational speech synthesis. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 1. IEEE.

- Ting Qian, Kristy Hollingshead, Su-youn Yoon, Kyoung-young Kim, and Richard Sproat. 2010. A Python toolkit for universal transliteration. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA).
- Karim Rahim and Wesley S. Burr. 2017. multitaper: Multitaper spectral analysis. *R package version 1.0-14*.
- Daniel Recasens and Aina Espinosa. 2009. Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication*, 51(3):240– 258.
- Tanja Schultz. 2002. GlobalPhone: A multilingual speech and text database developed at Karlsruhe University. In *Seventh International Conference on Spoken Language Processing*, pages 345–348, Denver, CO.
- Jean-Luc Schwartz and Lucie Ménard. 2019. Structured idiosyncrasies in vowel systems. OSF Preprints.
- Christine H. Shadle, Wei-rong Chen, and D. H. Whalen. 2016. Stability of the main resonance frequency of fricatives despite changes in the first spectral moment. *The Journal of the Acoustical Society of America*, 140(4):3219–3220.
- Kenneth N. Stevens and Samuel J. Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics*, 38(1):10–19.
- Andreas Stolcke. 2002. SRILM an extensible language modeling toolkit. In Seventh International Conference on Spoken Language Processing, pages 901–904.
- Bert Vaux and Bridget Samuels. 2015. Explaining vowel systems: Dispersion theory vs natural selection. *Linguistic Review*, 32(3):573–599.
- Dominic J. L. Watt. 2000. Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, 12(1):69–101.
- John C. Wells. 1995/2000. Computer-coding the IPA: A proposed extension of SAMPA.
- D.H. Whalen and Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23:349–366.
- Matthew Wiesner, Oliver Adams, David Yarowsky, Jan Trmal, and Sanjeev Khudanpur. 2019. Zero-shot pronunciation lexicons for cross-language acoustic model transfer. In *Proceedings of IEEE Association* for Automatic Speech Recognition and Understanding (ASRU).

- Xiaohui Zhang, Vimal Manohar, Daniel Povey, and Sanjeev Khudanpur. 2017. Acoustic data-driven lexicon learning based on a greedy pronunciation selection framework. *arXiv preprint arXiv:1706.03747*.
- Eberhard Zwicker and Ernst Terhardt. 1980. Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America*, 68(5):1523–1525.

A Pairwise Correlations between Vowel Formant Measures (§4 Case Studies)

Table 3 and Table 4 respectively show Pearson correlations of mean F1 and mean F2 in ERB between vowels that appear in at least 10 readings. As formalized in the present analysis, phonetic uniformity predicts strong correlations of mean F1 among vowels with a shared height specification, and strong correlations of mean F2 among vowels with a shared backness specification. The respective "Height" and "Backness" columns in Table 3 and Table 4 indicate whether the vowels in each pair match in their respective specifications. *p*-values are corrected for multiple comparisons using the Benjamini-Hochberg correction and a false discovery rate of 0.25 (Benjamini and Hochberg, 1995). Significance is assessed at $\alpha = 0.05$ following the correction for multiple comparisons; rows that appear in gray have correlations that are not significant according to this threshold.

V1	V2	Height	# Readings	r	p
/i/	/i:/	\checkmark	12	0.81	0.006
/eː/	/oː/	\checkmark	10	0.81	0.015
/i/	/u/	\checkmark	40	0.79	0.000
/ɛ/	/ɔ/	\checkmark	11	0.68	0.053
/o/	/a/		37	0.66	0.000
/i:/	/oː/		11	0.65	0.070
/i:/	/uː/	\checkmark	12	0.64	0.061
/e/	/o/	\checkmark	35	0.62	0.001
/e/	/u/		36	0.59	0.001
/e/	/a/		34	0.58	0.002
/u/	/ə/		12	0.58	0.105
/i:/	/eː/		11	0.58	0.118
/i/	/e/		38	0.54	0.002
/ɛ/	/a/		12	0.54	0.127
/u/	/o/		38	0.49	0.007
/ɛ/	/u/		14	0.49	0.135
/i/	/o/		39	0.46	0.011
/e/	/ɛ/	\checkmark	12	0.46	0.204
/u/	/a/		37	0.42	0.027
/i:/	/e/		11	0.42	0.288
/u/	/uː/	\checkmark	10	0.41	0.334
/i:/	/u/	\checkmark	11	0.33	0.430
/i:/	/a/		11	0.28	0.496
/i/	/a/		39	0.27	0.173
/i/	/ɛ/		14	0.24	0.496
/i:/	/o/		13	0.19	0.624
/i/	/ə/		13	0.10	0.785
/u/	/ɔ/		12	0.09	0.785
/ɛ/	/o/	\checkmark	13	-0.09	0.785
/e/	/ɔ/	\checkmark	10	-0.12	0.785
/uː/	/o/		10	-0.12	0.785
/i/	/ɔ/		11	-0.42	0.288
/o/	/ə/	\checkmark	11	-0.51	0.173
/ə/	/a/		11	-0.90	0.001

Table 3: Pearson correlations (r) of **mean F1** in ERB between vowel categories.

V1	V2	Backness	# Readings	r	p
/e/	/ɛ/	\checkmark	12	0.77	0.019
/u/	/uː/	, ,	10	0.77	0.037
/i/	/i:/	\checkmark	12	0.70	0.038
/u/	/o/	\checkmark	38	0.69	0.000
/i/	/ɛ/	\checkmark	14	0.69	0.031
/uː/	/o/	\checkmark	10	0.62	0.130
/u/	/ə/		12	0.60	0.107
/u/	/ɔ/	\checkmark	12	0.52	0.168
/i/	/e/	\checkmark	38	0.41	0.038
/ɛ/	/a/		12	0.32	0.519
/o/	/a/		37	0.30	0.159
/eː/	/oː/		10	0.27	0.666
/e/	/a/		34	0.24	0.339
/o/	/ə/		11	0.21	0.724
/ə/	/a/	\checkmark	11	0.16	0.830
/i:/	/e/	\checkmark	11	0.11	0.911
/i/	/a/	\checkmark	39	0.06	0.911
/i:/	/eː/	\checkmark	11	0.06	0.965
/e/	/o/		35	0.01	0.965
/u/	/a/		37	0.00	0.985
/ɛ/	/ɔ/		11	-0.03	0.965
/i:/	/a/		11	-0.04	0.965
/ɛ/	/o/		13	-0.04	0.965
/e/	/u/		36	-0.12	0.666
/ɛ/	/u/		14	-0.22	0.666
/i/	/ə/		13	-0.23	0.666
/iː/	/oː/		11	-0.42	0.345
/i/	/o/		39	-0.48	0.017
/i:/	/o/		13	-0.52	0.149
/i/	/u/		40	-0.55	0.003
/i/	/ɔ/		11	-0.63	0.107
/e/	/၁/		10	-0.65	0.107
/iː/	/u/		11	-0.80	0.019
/iː/	/uː/		12	-0.83	0.009

Table 4: Pearson correlations (r) of **mean F2** in ERB between vowel categories.

B Distributions of Unitran Segment Accuracy (§3.1.3 Quality Measures)

Here we evaluate the quality of the Unitran dataset in more detail. The goal is to explore the variation in the quality of the labeled Unitran segments across different languages and phoneme labels. This evaluation includes only readings in high-resource languages, where we have not only the aligned Unitran pronunciations but also aligned high-resource pronunciations (Epitran or WikiPron) against which to evaluate them. The per-token statistics used to calculate these plots are included in the corpus release to enable closer investigation of individual phonemes than is possible here.

B.1 Unitran Pronunciation Accuracy

First, in Figures 5 and 6, we consider whether Unitran's utterance pronunciations are accurate without looking at the audio. For each utterance, we compute the unweighted Levenshtein alignment between the Unitran pronunciation of the utterance and the high-resource pronunciation. For each reading, we then score the percentage of Unitran 'phoneme' tokens that were aligned to high-resource 'phoneme' tokens with exactly the same label.¹⁴ We can see in Figure 6 that many labels are highly accurate in many readings while being highly inaccurate in many others. Some labels are noisy in some readings.¹⁵



Figure 5: Unitran pronunciation accuracy per language, evaluated by Levenshtein alignment to WikiPron pronunciations (hatched bars) or Epitran pronunciations (plain bars). Where a language has multiple readings, error bars show the min and max across those readings.



Figure 6: Unitran pronunciation accuracy per language, for selected phonemes. Accuracy is evaluated by Levenshtein alignment as in Figure 5. Each curve is a kernel density plot with integral 1. For the /z/ curve, the integral between 80% and 100% (for example) is the estimated probability that in a high-resource language drawn uniformly at random, the fraction of Unitran /z/ segments that align to high-resource /z/ segments falls in that range. The 'all' curve is the same, but now the uniform draw is from all pairs of (high-resource language, Unitran phoneme used in that language).

¹⁴By contrast, PER in Table 1 aligns at the word level rather than the utterance level, uses the number of symmetric alignment errors (insertions + deletions + substitutions) rather than the number of correct Unitran phonemes, and normalizes by the length of the high-resource 'reference' pronunciation rather than by the length of the Unitran pronunciation.

¹⁵Note that as §3.1.3 points out, it may be unfair to require exact match of labels, since annotation schemes vary.)

B.2 Unitran Segment Label Accuracy

In Figures 7 and 8, we ask the same question again, but making use of the audio data. The match for each Unitran segment is now found not by Levenshtein alignment, but more usefully by choosing the high-resource segment with the closest midpoint. For each reading, we again score the percentage of Unitran 'phoneme' tokens whose aligned high-resource 'phoneme' tokens have exactly the same label. Notice that phonemes that typically had high accuracy in Figure 6, such as /p/ and /b/, now have far more variable accuracy in Figure 8, suggesting difficulty in aligning the Unitran pronunciations to the correct parts of the audio.



Figure 7: Unitran pronunciation accuracy per language, as in Figure 5 but with audio midpoint alignment in place of Levenshtein alignment.



Figure 8: Unitran pronunciation accuracy per language, for selected phonemes, as in Figure 6 but with audio midpoint alignment in place of Levenshtein alignment.

B.3 Unitran Segment Boundary Accuracy

Finally, in Figures 9 and 10, we measure whether Unitran segments with the "correct" label also have the "correct" time boundaries, where "correctness" is evaluated against the corresponding segments obtained using Epitran or WikiPron+G2P.



Figure 9: **Mean error per language in the temporal boundaries of Unitran segments.** Each Unitran segment is evaluated against the WikiPron segment (hatched bars) or Epitran segment (plain bars) with the closest midpoint, as if the latter were truth. The error of a segment is the absolute offset of the left boundary plus the absolute offset of the right boundary. Only segments where the Unitran label matches the Epitran/WikiPron label are included in the average. Where a language has multiple readings, error bars show the min and max across those readings.



Figure 10: Mean error per language in the temporal boundaries of Unitran segments, for selected phonemes. Each curve is a kernel density plot with integral 1. For the /z/ curve, the integral between 50ms and 100ms (for example) is the estimated probability that in a high-resource language drawn uniformly at random, the Unitran /z/ segments whose corresponding Epitran or WikiPron segments are also labeled with /z/ have mean boundary error in that range. Small bumps toward the right correspond to individual languages where the mean error of /z/ is unusually high. The 'all' curve is the same, but now the uniform draw is from all pairs of (high-resource language, Unitran phoneme used in that language). The boundary error of a segment is evaluated as in Figure 9.

C WikiPron Grapheme-to-Phoneme (G2P) Accuracy (§3.1.3 Quality Measures)

For each language where we used WikiPron, Table 5 shows the phoneme error rate (PER) of Phonetisaurus G2P models trained on WikiPron entries, as evaluated on held-out WikiPron entries. This is an estimate of how accurate our G2P-predicted pronunciations are on out-of-vocabulary words, insofar as those are distributed similarly to the in-vocabulary words. (It is possible, however, that out-of-vocabulary words such as Biblical names are systematically easier or harder for the G2P system to pronounce, depending on how they were transliterated.)

The same G2P configuration was used for all languages, with the hyperparameter settings shown in Table 6. (seq1_max and seq2_max describe how many tokens in the grapheme and phoneme sequences can align to each other.). These settings were tuned on SIGMORPHON 2020 Task 1 French, Hungarian, and Korean data (Gorman et al., 2020), using 20 random 80/20 splits.

	lat	nhx	srn	mah	por-po	mfe	mww	por-bz	eng	khm	mlg	ori	ban	urd
Train size 41741	34181	126	157	813	9633	203	227	10077	54300	3016	114	211	172	704
$\mathbf{PER} \qquad \boxed{0.8} \\ +0.02$	2.4 ± 0.04	4.1		9.6		10.7		11.4 ± 0.16	14.5			16.1	19.5	

Table 5: WikiPron G2P Phone Error Rate (PER) calculated treating WikiPron annotations as ground-truth. We perform 20 trials with random 80/20 splits per language, and report PER averaged across trials with 95% confidence intervals for each language.

Phonetisaurus Alignment Hyperparameters	seq1_max	seq2_max 3	seq1_del True	seq2_del True	grow True	max EM iterations 11	
Graphone Language Model	n-gram order	LM type	discounting	gt2min	gt3min	gt4min	gt5min
Hyperparameters	5	max-ent	Kneser-Ney	2	2	3	4

Table 6: Table of final G2P hyperparameter settings. Alignment parameters not listed here for phonetisaurus-align use the default values. The language model was trained using SRILM (Stolcke, 2002) ngram-count using default values except for those listed above.

D Retention Statistics (§4.1 Data Filtering)

Table 7 shows what percentage of tokens would be retained after various methods are applied to filter out questionable tokens from the readings used in \$4.1. In particular, the rightmost column shows the filtering that was actually used in \$4.1. We compute statistics for each reading separately; in each column we report the minimum, median, mean, and maximum statistics over the readings. The top half of the table considers vowel tokens (for the vowels in Appendix A); the bottom half considers sibilant tokens (/s/ and /z/).

On the left side of the table, we consider three filtering techniques for Unitran alignments. **Midpoint** retains only the segments whose labels are "correct" according to the midpoint-matching methods of Appendix B. **MCD** retains only those utterances with MCD < 6. **Outlier** removes tokens that are outliers according to the criteria described in §4.1. Finally, **AGG.** is the aggregate retention rate retention rate after all three methods are applied in order.

On the right side of the table, we consider the same filtering techniques for the high-resource alignments that we actually use, with the exception of **Midpoint**, as here we have no higher-quality annotation to match against.

			High-Resource Alignments							
		# Tokens	Midpoint	MCD	Outlier	AGG.	# Tokens	MCD	Outlier	AGG.
	Min	50,132	2%	42%	83%	1%	61,727	42%	84%	37%
ls	Median	21,5162	23%	88%	90%	16%	232,059	88%	90%	79%
Vowei	Mean	23,9563	25%	81%	89%	20%	223,815	81%	90%	73%
Vo	Max	662,813	65%	100%	93%	60%	468,864	100%	93%	93%
	# Readings	49	46	48	49	45	49	48	49	48
	Min	7,198	10%	42%	89%	13%	7184	44%	91%	43%
ıts	Median	28,690	70%	87%	97%	59%	27569	87%	97%	85%
ilaı	Mean	30,025	63%	80%	95%	56%	27083	81%	96%	79%
Sibilants	Max	63,573	89%	100%	98%	79%	45,290	100%	99%	96%
	# Readings	36	26	35	36	19	25	22	25	22

Table 7: Summary of quality measure retention statistics for **vowels** and **sibilants** over unique readings with reading-level MCD < 8 for Unitran and high-resource alignments.

E All VoxClamantis v1.0 Languages

All 635 languages from 690 readings are presented here with their language family, ISO 639-3 code, and mean utterance alignment quality in Mel Cepstral Distortion (MCD) from Black (2019). Languages for which we release Epitran and/or WikiPron alignments in addition to Unitran alignments are marked with ^e and ^w respectively. MCD ranges from purple (*low*), blue–green (*mid*), to yellow (*high*). Lower MCD typically corresponds to better audio-text utterance alignments and higher quality speech synthesis, but judgments regarding distinctions between languages may be subjective. ISO 639-3 is not intended to provide identifiers for *dialects* or other *sub-language* variations, which may be present here where there are multiple readings for one ISO 639-3 code. We report the most up-to-date language names from the ISO 639-3 schema (Eberhard and Fennig, 2020). Language names and codes in many schema could be pejorative and outdated, but where language codes cannot be easily updated, language names can and often are.

NIGER-CONGO: 159		Koonzime ozm	8.0	Rigwe iri	7.3	Gorontalo gor	6.2
Abidji abi	6.3	Kouya kyf	8.2	Rundi run	8.3	Hanunoo hnn	6.0
Adele ade	6.9	Kukele kez	7.8	Saamia lsm	6.8	Hiligaynon hil	6.7
Adioukrou adi	7.4	Kunda kdn	6.4	Sango sag	6.7	Iban iba	6.5
Akan aka	7.8	Kuo xuo	6.7	Sekpele lip	6.6	^e Iloko ilo	6.5
Akebu keu	7.0	Kusaal kus	7.0	Selee snw	6.5	^e Indonesian ind	7.2
Akoose bss	7.2	Kutep kub	6.9	Sena seh	6.6	^e Indonesian ind	6.8
Anufo cko	6.9	Kutu kdc	5.7	Shambala ksb	6.4	^e Indonesian ind	6.4
Avatime avn	6.3	Kutu kuc	7.4	Sissala sld	7.6	Itawit itv	6.6
Bafut bfd	7.3	Kuwaataay cwt	7.5	Siwu akp	6.3	^e Javanese jav	7.3
	7.0		7.9		6.9		8.5
Bandial bqj	7.3	Lama (Togo) las	7.3	Soga xog	7.7	Kadazan Dusun dtp	6.2
Bekwarra bkv	7.5	Lelemi lef	7.0	South Fali fal		Kagayanen cgc	6.2 5.9
Bete-Bendi btt	7.6	Lobi lob		Southern Birifor biv	7.6 7.6	Kalagan kqe	
Biali beh		Lokaa yaz	6.6	Southern Bobo Madaré bwq		Kankanaey kne	5.7
Bimoba bim	7.0	Lukpa dop	8.0	Southern Dagaare dga	6.5	Keley-I Kallahan ify	6.2
Bokobaru bus	6.9	Lyélé lee	8.1	Southern Nuni nnw	7.6	Khehek tlx	9.1
Bomu bmq	7.0	Machame jmc	6.8	Southwest Gbaya gso	7.6	Kilivila kij	6.2
Buamu box	8.1	Mada (Nigeria) mda	6.6	Supyire Senoufo spp	8.3	Kinaray-A krj	6.3
Buli (Ghana) bwu	7.3	Makaa mep	6.9	Talinga-Bwisi tlj	6.5	Kisar kje	6.3
Bum bmv	6.4	Makhuwa vmw	6.8	Tampulma tpm	7.1	Koronadal Blaan bpr	6.4
Cameroon Mambila mcu	7.6	Malawi Lomwe lon	5.8	Tharaka thk	7.8	Lampung Api ljp	6.4
Central-Eastern Niger fuq	7.1	Malba Birifor bfo	6.5	Tikar tik	7.8	Lauje law	6.4
Cerma cme	8.5	Mamara Senoufo myk	8.0	Timne tem	7.2	Ledo Kaili lew	7.0
Cerma cme	6.1	Mampruli maw	7.6	Toura (Côte d'Ivoire) neb	6.8	Luang lex	6.1
Chopi cce	6.3	Mankanya knf	6.6	Tsonga tso	5.2	Lundayeh Ind	6.5
Chumburung neu	7.3	Masaaba myx	6.1	Tumulung Sisaala sil	8.0	Ma'anyan mhy	6.4
Delo ntr	8.0	Meta' mgo	6.4	Tuwuli bov	6.2	Madurese mad	7.4
Denya anv	6.7	Miyobe soy	7.2	Tyap kcg	7.5	Mag-antsi Ayta sgb	6.4
Ditammari tbz	7.7	Moba mfq	8.1	Vengo bav	6.7	Makasar mak	6.4
Djimini Senoufo dyi	7.1	Moba mfq	7.2	Vunjo vun	6.5	^w Malagasy mlg	8.8
Duruma dug	6.7	Mochi old	6.9	West-Central Limba lia	7.3	^w Malagasy mlg	7.3
Eastern Karaboro xrb	8.1	N	7.2		7.0	W Malagana and	6.3
Eastern Raraboro Arb	0.1	Mossi mos	1.2	Yocoboué Dida gud	7.0	wanagasy mig	0.5
Ekajuk eka	7.5	Mossi mos	7.5	AUSTRONESIAN: 106	7.0	^W Malagasy mlg Malay (macrolanguage) msa	6.3
				0	6.5	Malay (macrolanguage) msa	
Ekajuk eka	7.5	Mossi mos	7.5	AUSTRONESIAN: 106			6.3
Ekajuk eka Ewe ewe Ewe ewe	7.5 6.3	Mossi mos Mumuye mzm	7.5 7.7	AUSTRONESIAN: 106 Achinese ace	6.5	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa	6.3 6.0 6.3
Ekajuk eka Ewe ewe	7.5 6.3 6.7	Mossi mos Mumuye mzm Mundani mnf	7.5 7.7 6.8	AUSTRONESIAN: 106 Achinese ace Agutaynen agn	6.5 6.1	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqj	6.3 6.0 6.3
Ekajuk eka Ewe ewe Ewe ewe Farefare gur	7.5 6.3 6.7 8.1	Mossi mos Mumuye mzm Mundani mnf Mwan moa	7.5 7.7 6.8 7.8	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj	6.5 6.1 5.9	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqj Manado Malay xmm	6.3 6.0 6.3 5.2
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur	7.5 6.3 6.7 8.1	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw	7.5 7.7 6.8 7.8	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp	6.5 6.1 5.9 6.3	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqj Manado Malay xmm Mapos Buang bzh	6.3 6.0 6.3 5.2 5.8
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Farefare gur	7.5 6.3 6.7 8.1 8.3 8.7	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh	7.5 7.7 6.8 7.8 6.5 8.4	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk	6.5 6.1 5.9 6.3 5.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manasa mqj Manado Malay xmm Mapos Buang bzh Maranao mrw	 6.3 6.0 6.3 5.2 5.8 6.0
Ekajuk eka Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd	7.5 6.3 6.7 8.1 8.3 8.7 7.7	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr	7.5 7.7 6.8 7.8 6.5 8.4 6.8	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa	6.5 6.1 5.9 6.3 5.4 5.9	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manasa mqj Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁹ Marshallese mah	6.3 6.0 6.3 5.2 5.8 6.0 7.9
Ekajuk eka Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq	6.5 6.1 5.9 6.3 5.4 5.9 6.7	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manado Malay xmm Manado Malay xmm Mapos Buang bzh Maranao mrw ^W Marshallese mah Matigsalug Manobo mbt	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0
Ekajuk eka Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqj Manado Malay xmm Mapos Buang bzh Maranao mrw ^W Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0
Ekajuk eka Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Nateni ntm	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa nqj Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy	 6.3 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.6
Ekajuk eka Ewe ewe Farefare gur Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0	Mossi mos Mumuye mzm Mundani mnf Mwani wmw Mündü muh Nafaanra nfr Nande nhb Nateni ntm Navdm nmz Ndogo ndz	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min	 6.3 6.0 6.3 5.8 6.0 7.9 6.4 6.0 6.6 6.3
Ekajuk eka Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Nawdm nmz Ndogo ndz Ngangam gng	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁹ Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpy	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.6 6.3 6.3
Ekajuk eka Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündi muh Nafaanra nfr Nande nnb Nateni ntm Nawdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balanese ban	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.6 6.3 6.3 6.3
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mindi muh Nafaanra nfr Nande nnb Nateni ntm Navdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.9 6.7	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balinese ban Bambam ptu Batad Ifugao ifb	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 5.8	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa nqj Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Itugo itu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.4 6.0 6.6 6.3 6.3 6.7 6.7
Ekajuk eka Ewe ewe Earefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1	Mossi mos Mumuye mzm Mundani mnf Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Nateni ntm Nateni ntm Natogo ndz Ngangam gng Nigeria Mambla mzk Nilamba nim Ninzo nin	 7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balanese ban Balinese ban	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 5.8 6.2	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁹ Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb	6.3 6.0 5.2 5.8 6.0 7.9 6.4 6.0 6.3 6.3 6.3 6.3 6.7 6.7 6.7
Ekajuk eka Ewe ewe Earefare gur Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifè ife	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Navdam nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko	 7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balinese ban Balinese ban Batad Ifugao ifb Batak Dairi btd	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 5.8 6.2 6.4 5.8	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mww Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij	 6.3 6.0 6.3 5.8 6.0 7.9 6.4 6.0 6.3 6.3 6.3 6.7 6.7 7.3
Ekajuk eka Ewe ewe Earefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifè ife Iybie North-Okpela-Ar atg	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8 7.7	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Navden mmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.9 6.7 5.9 6.8 7.2	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batak Dairi btd Batak Dairi btd Batak Karo btx	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 5.8 6.2 6.1 6.4 5.8 6.2 6.1 6.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manado Malay xmm Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mww Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia	 6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.3 6.7 6.7 6.7 7.3 6.8
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifê ife Ivbie North-Okpela-Ar at g Izere izr	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8 7.7 6.8	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündi muh Nafaanra nfr Nande nnb Nateni ntm Nawdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.9 6.7 5.9 6.8 7.2	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arogi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batak Lari btx Batak Simalungun bts	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.1 6.2 6.1 6.2 6.1 6.2 6.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁹ Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.4 6.3 6.4 6.3 6.4 6.3 6.4 6.3 6.4 6.3 6.7 6.7 6.8 5.8
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Farefare gur Gitongar egur Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gavere gwr Hanga hag Haya hay Ifê ife Iybie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo	7.5 6.3 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8 7.7 6.8 7.7 6.8 7.1	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nub Nateni ntm Navdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nicham bud Nyabwa nwb	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 8.0 6.9 8.0 6.9 8.0 6.9 8.0 6.7 5.9 8.8 7.2 7.3 8.8	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balinese ban Bambam ptu Batad Ifugao ifb Batak Dairi btd Batak Karo btx Batak Simalungun bts Besoa bep	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 6.2 6.4 6.4 6.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn	6.3 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.4 6.3 6.4 6.3 6.4 6.3 6.4 6.5 6.7 6.7 6.8 5.8 6.3
Ekajuk eka Ewe ewe Farefare gur Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifè ife Ivbie North-Okpela-Ar atg Izere izr Jala-Fonyi dyo Jola-Kasa csk	7.5 6.3 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8 7.7 6.8 7.1 7.5	Mossi mos Mumuye mzm Mundani mnf Mwan iwnw Mündü muh Nafaanra nfr Nande mho Nateni ntm Nawdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Ntcham bud	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 8.0 6.9 8.0 6.9 8.0 6.7 5.9 6.8 7.2 7.3 8.8 7.7	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak bla Balantak bla Balantak bla Balantak ban Bambam ptu Batad Ifugao ifb Batak Karo bts Batak Simalungun bts Besoa bep Brooke's Point Palawa plw	6.5 6.1 5.9 6.3 5.4 5.7 6.2 6.1 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpy Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.6 6.3 6.3 6.7 6.7 7.3 6.8 5.8 6.3 6.3 6.3
Ekajuk eka Ewe ewe Earefare gur Farefare gur Farefare gur Farefare gur Farefare gur Farefare gur Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifè ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 7.7 6.8 7.3 6.1 7.2 7.1 7.8 7.7 6.8 7.7 6.8 7.1 7.5 7.9	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Navden mmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nteham bud Nyabwa nwb	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arap-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batad Ifugao ifb Batak Jairi btd Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn	6.5 6.1 5.9 6.3 5.4 5.7 6.2 6.1 6.2 6.1 6.2 6.4 6.2 6.4 6.2 6.4 6.4 6.4 6.4 6.4 6.4	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau Pamona pmf Pampanga pam	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.3 6.3 6.7 7.3 6.8 5.8 6.3 6.7 6.7 7.3 6.8
Ekajuk eka Ewe ewe Earefare gur Farefare gur Fan fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifê ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kaas csk Jukun Takum jbu Kabiyè kbp	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 7.7 6.8 7.3 6.1 7.2 7.1 7.8 7.7 6.8 7.7 6.8 7.1 7.5 7.9 7.4	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Navden nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Niteham bud Nyakyusa-Ngonde nyy Nyankole nyn	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7 8.0	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batak Dairi btd Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn [©] Cebuano ceb	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 5.8 6.2 6.1 6.2 6.4 6.4 6.2 6.4 6.2 6.4 6.2	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwv Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.3 6.4 6.3 6.7 6.7 6.7 6.7 6.8 5.8 6.3 6.7 6.4 6.3 6.7 6.3 6.7 6.3 6.3 6.3 6.4 6.3
Ekajuk eka Ewe ewe Eare fare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifé ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu Kabiyé kbp Kagulu kki	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 7.0 8.0 7.1 7.2 7.1 7.2 7.1 7.2 7.1 7.8 7.1 7.8 7.1 7.5 7.9 7.4 6.6	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwan i wmw Miindii muh Nafaanra nfr Nande nnb Nateni ntm Nawdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Ntcham bud Nyakyusa-Ngonde nyy Nyakube nyu	7.5 7.7 6.8 7.8 6.5 8.4 6.5 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7 8.0 6.7	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Batak Dairi btd Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn [©] Cebuano ceb	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 5.8 6.2 6.1 6.4 5.8 6.2 6.1 6.4 6.2 6.4 6.4 6.4 6.2 6.4 6.4 6.4 6.2 6.4 6.4 6.4 6.5 6.5	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw [*] Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nj Nias nia Obo Manobo obo Owa stn Palauan pau Pamona pmf Pampanga pam	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.3 6.7 6.7 6.7 6.8 5.8 6.3 6.8 5.8 6.3 6.7 6.3 6.5
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gawene gwr Hanga hag Haya hay Ifê ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu Kabiyè kbp Kagulu kki	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 7.0 8.0 7.1 7.2 7.1 7.2 7.1 7.8 7.1 7.8 7.1 7.8 7.1 7.5 7.9 7.4 6.6 7.9	Mossi mos Mumuye mzm Mundani mnf Mwan iwaw Mündü muh Nafaanra nfr Nande nub Nateni ntm Nateni ntm Nateni ntm Nateni ntm Nateni ntm Nateni ntm Nateni ntm Nateni ntm Nateni ntm Nateni ntm Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nigetia Mud Nyabwa nwb Nyakyusa-Ngonde nyy Nyankole nyn Nyankole nyn Nyakule nyn	7.5 7.7 6.8 7.8 6.5 8.4 6.5 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7 8.0 6.7 5.9	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak biz Balantak biz Balantak biz Batad Ifugao ifb Batak Jarir btd Batak Karo btx Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn [©] Cebuano ceb Central Bikol bel Central Bikol bel	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.5 6.4 6.5 6.6	Malay (macrolanguage) msa ⁶ Malay (macrolanguage) msa Mamasa nqj Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁰ Marshallese mah Matigsalug Manobo mbt Mayoyao Itugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau Pampanga pam Pangasinan pag Paranan prf	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.3 6.7 6.7 6.3 6.8 6.8 6.3 6.7 6.3 6.6 6.3 6.3 6.7 6.3
Ekajuk eka Ewe ewe Earefare gur Farefare gur Farefare gur Fon fon Gikyode acd Gityama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifè ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu Kabiyè kbp Kagulu kki Kako kkj	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 7.0 8.0 7.1 7.2 7.1 7.2 7.1 7.8 7.1 7.8 7.1 7.8 7.1 7.5 7.9 7.4 6.6 7.9	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Mündü muh Nafaanra nfr Nande nib Nateni ntm Navdin nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nigetia mbud Nyabwa nwb Nyakyusa-Ngonde nyy Nyaturu rim Nyote nuj Nyote nuj	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7 8.8 7.7 6.7 8.0 6.7 5.9 7.1	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batad Ifugao ifb Batad Ifugao ifb Batad Jifugao ifb Batad Karo btx Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn Cachtral Bikol bel Central Bikol bel Central Malay pse Central Malay pse	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 5.4 6.1 6.4 6.2 6.4 6.4 6.2 6.8 6.9 6.5 6.6 6.0 6.0	Malay (macrolanguage) msa ⁶ Malay (macrolanguage) msa Mamasa mqj Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁰ Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwv Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau Pampanga pam Pangasinan pag Paranan prf Rejang rej	6.3 6.3 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.3 6.3 6.3 6.4 6.6 6.3 6.7 6.7 6.3 6.8 5.8 6.3 6.7 6.3 6.6 6.3 6.6 6.3 6.6 6.7 6.3 6.6 6.6 6.7 6.3
Ekajuk eka Ewe ewe Earefare gur Farefare gur Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifê ifê Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu Kabiyê kbp Kagulu kki Kako kkj Kasem xsm	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.5 7.9 7.4 6.6 7.9 7.4 6.6 7.9 7.7 8.9	Mossi mos Mumuye mzm Mundani mnf Mwan awa Mwani wmw Mündü muh Nafaanra nfr Nande nnb Nateni ntm Navden nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nteham bud Nyabwa nwb Nyakyusa-Ngonde nyy Nyankole nyn Nyaturu rim Nyote nuj Nyote nuj Nyoro nyo Nzima nzi	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.7 5.9 7.7 6.8 7.2 7.3 8.8 7.7 6.7 8.0 6.7 5.9 7.1 7.2	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batad Ifugao ifb Batad Arao btx Batad Karo btx Batak Simalungun bts Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jvn [©] Cebuano ceb Central Malay pse Central Malay pse Central Malay pse	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 6.1 6.4 5.8 6.2 6.1 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.5 6.5 6.5 6.5 6.5	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Mamasa mqj Manado Malay xmm Mapos Buang bzh Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwv Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau Pampanga pam Pangasinan pag Paranan prf Rejang rej Roviana rug	6.3 6.3 6.3 5.2 5.8 6.0 7.9 6.4 6.0 6.3 6.7 6.7 6.8 5.8 6.3 6.7 6.3 6.7 6.3 6.7 6.3 6.7 6.3 6.7 6.3 5.8 6.3 5.7
Ekajuk eka Ewe ewe Earefare gur Farefare gur Fan fon Gikyode acd Giryama nyf Götonga toh Gogo gog Gökana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifê ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kaas csk Jukun Takum jbu Kabiyè kbp Kagulu kki Kako kkj Kasem xsm	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8 7.7 6.8 7.1 7.5 7.9 7.4 6.6 7.9 7.4 8.9 7.4	Mossi mos Mumuye mzm Mundani mnf Mwan moa Mwani wmw Miindii muh Nafaanra nfr Nande nnb Nateni ntm Nawdm nmz Ndogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nicham bud Nyakyusa-Ngonde nyy Nyakkusa-Ngonde nyy Nyakkusa-Ngonde nyy Nyakkusa-Ngonde nyy Nyakusa-Ngonde nyy Nyakusa-Ngonde nyy Nyaturu rim Nyole nuj Nyoto nyo Nyiam azi Obolo ann	7.5 7.7 6.8 7.8 6.5 8.4 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7 8.0 6.7 5.9 7.1 7.2 8.5	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Batak Dairi btd Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn [©] Cebuano ceb Central Bikol bel Central Bikol bel Central Malay pse Central Mang cmo Central Sama sml Da'a Kaili kzf	6.5 6.1 5.9 6.3 5.4 5.9 6.7 6.2 5.6 6.1 6.4 5.8 6.2 6.1 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.3 6.5 6.5	Malay (macrolanguage) msa ^e Malay (macrolanguage) msa Manado Malay xmm Manado Malay xmm Mapos Buang bzh Maranao mrw ^w Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau Pamona pmf Pampanga pam Pangasinan pag Paranan pf Rejang rej Roviana rug Sambal xsb	6.3 6.3 5.2 5.8 6.0 6.4 6.3 6.3 6.3 6.3 6.7 6.7 6.7 6.7 6.8 5.8 6.3 6.7 6.3 6.6 6.7 6.3 6.5 5.8 6.3 6.7 6.3 6.5 7 6.0 6.3 6.5 7 9 6.4 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gavere gwr Hanga hag Haya hay Ifè ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu Kabiyè kbp Kasgulu kki Kako kkj Kasem xsm Kasem xsm Kasem xsm	7.5 6.3 6.7 8.1 8.3 8.7 7.7 6.8 6.8 6.8 7.0 7.3 6.8 7.3 6.1 7.2 7.1 7.2 7.1 7.5 7.7 6.8 7.1 7.5 7.9 4 6.6 7.7 8.9 7.4 6.8	Mossi mos Mumuye mzm Mundani mnf Mwan awa Mwani wmw Mündü muh Nafaanra nfr Nande nub Nateni ntm Nateni ntm Ninzo nin Nkonya nko Northern Dagara dgi Nicham bud Nyabwa nwb Nyakyusa-Ngonde nyy Nyankole nyn Nyatwole nyi Nyankole nyi nyankole nyi Nyankole nyi nyankole nyi nyankole nyi nyankole nyi nyankole nyi nyankole nyi nyankole nyankole nyi nyankole nyi nyankol	7.5 7.7 6.8 7.8 6.5 8.4 6.5 8.4 6.8 7.2 7.4 8.0 6.9 6.7 5.9 6.8 7.2 7.3 8.8 7.7 6.7 8.0 6.7 5.9 6.8 7.2 7.3 8.8 7.7 8.0 6.7 8.0 7.2 7.3 8.8 7.2 7.4 8.3 8.8 7.2 7.4 8.3 8.5 8.8 7.2 7.4 8.3 8.5 8.8 7.2 8.5 8.8 7.2 8.5 8.8 7.2 8.5 8.8 7.2 8.5 8.8 7.2 8.5 8.6 8.0 8.0 8.0 8.0 8.0 8.0 8.0 8.0 8.0 8.0	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alij Alune alp Ambai amk Araganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi alia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak bar Batad Ifugao ifb Batad Ifugao ifb Batak Karo btx Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jvn Caritabean Javanese jvn Caritabean Javanese jvn Caritabean Javanese jvn Caritabean Javanese jvn Carital Sama sml Da'a Kaili kzf Duri mvp Fataleka far	6.5 6.1 5.9 6.3 5.4 5.4 6.7 5.6 5.4 6.1 6.2 5.6 5.4 6.1 6.2 6.1 6.2 6.1 6.2 6.3 6.2 5.6 6.2 5.6 6.2 5.6 6.2 5.6 6.2 5.6 6.2 5.6 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 7 7 7 7 7 7 7 7	Malay (macrolanguage) msa ⁶ Malay (macrolanguage) msa Mamasa nqj Manado Malay xmm Mapos Buang bzh Maranao mrw ¹⁰ Marshallese mah Matigsalug Manobo mbt Mayoyao Itugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nij Nias nia Obo Manobo obo Owa stn Palauan pau Pampanga pam Pangasinan pag Paranan prf Rejang rej Roviana rug Sambal xsb Sambal xsb	6.3 6.3 6.3 5.2 5.8 6.0 6.4 6.3 6.3 6.3 6.4 6.3 6.7 6.7 7.3 6.8 6.8 6.3 6.7 6.3 6.8 5.8 6.3 6.7 6.8 5.8 6.3 6.7 6.8 5.8 6.3 6.7 6.8 5.8 6.3 6.7 6.8 5.8 6.3 6.6 6.7 6.0 6.2 5.7 6.0 6.0
Ekajuk eka Ewe ewe Ewe ewe Farefare gur Farefare gur Fon fon Gikyode acd Giryama nyf Gitonga toh Gogo gog Gokana gkn Gourmanchéma gux Gwere gwr Hanga hag Haya hay Ifê ife Ivbie North-Okpela-Ar atg Izere izr Jola-Fonyi dyo Jola-Kasa csk Jukun Takum jbu Kabiyè kbp Kagulu kki Kako kkj Kasem xsm Kasem xsm	7.5 6.3 6.7 8.1 8.3 8.7 6.8 7.0 8.0 7.3 6.1 7.2 7.1 7.8 7.1 7.5 7.1 7.5 7.9 7.4 6.8 6.3 7.4 6.8 6.3	Mossi mos Mumuye mzm Mundani mnf Mwan dia Mwani wmw Mündü muh Nafaanra nfr Nande mb Naten intb Nawdin mmz Nidogo ndz Ngangam gng Nigeria Mambila mzk Nilamba nim Ninzo nin Nkonya nko Noone nhu Northern Dagara dgi Nteham bud Nyabwa nwb Nyakyusa-Ngonde nyy Nyankole nyn Nyakuru rim Nyole nuj Nyoro nyo Nzima njo	7.5 7.7 6.8 7.8 6.5 8.4 6.8 7.2 7.4 8.3 6.9 8.0 6.9 6.9 6.9 6.9 6.9 6.9 6.9 6.9 6.8 7.2 7.3 8.8 7.7 6.7 8.0 6.7 5.9 8.0 6.7 5.9 8.0 6.7 7.2 8.0 6.5 8.3 7.2 7.2 7.3 8.3 7.2 7.2 8.3 7.2 7.2 8.3 7.2 7.2 8.3 7.2 7.2 8.3 7.2 7.2 8.3 7.2 7.2 8.3 8.3 7.2 7.2 7.3 8.3 8.3 7.2 7.2 7.3 8.3 7.2 7.2 7.3 8.3 7.2 7.2 7.4 8.3 8.3 7.2 7.2 7.4 8.3 7.2 7.2 7.4 8.3 8.3 7.2 7.2 7.4 8.3 8.3 7.2 7.2 7.4 8.3 8.3 7.2 7.2 7.4 8.3 8.3 7.2 7.2 7.4 8.3 8.3 7.2 7.2 7.3 8.3 8.3 7.2 7.2 8.3 8.3 7.2 7.3 8.3 8.3 7.2 7.3 8.3 8.3 7.2 7.3 8.3 7.2 7.3 8.3 7.2 7.3 8.3 7.2 7.3 8.3 7.2 7.5 8.3 7.2 7.5 8.3 7.2 7.5 8.3 7.2 7.5 8.3 7.7 7.5 8.3 7.7 7.5 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 8.3 7.7 7.7 8.3 7.7 7.7 8.3 7.7 7.7 8.3 7.7 7.7 8.3 7.7 7.7 7.7 7.7 7.7 7.7 7.7 7.7 7.7 7	AUSTRONESIAN: 106 Achinese ace Agutaynen agn Alangan alj Alune alp Ambai amk Amganad Ifugao ifa Aralle-Tabulahan atq Arop-Lokep apr Arosi aia Bada (Indonesia) bhz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak blz Balantak biz Batad Ifugao ifb Batat Jarib bta Batak Karo btx Batak Simalungun bts Besoa bep Brooke's Point Palawa plw Caribbean Javanese jyn [©] Cebuano ceeb Central Bikol bel Central Bikol bel Central Bikol bel Central Manog emo Central Sima sml Da'a Kali kzf Duri myp	6.5 6.1 5.9 6.3 5.4 5.9 6.2 5.6 5.4 6.1 6.2 5.4 6.1 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.4 6.2 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.6 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5 6.5	Malay (macrolanguage) msa ⁶ Malay (macrolanguage) msa Mamasa mqi Manado Malay xmm Mapos Buang bzh Maranao mrw ⁹ Marshallese mah Matigsalug Manobo mbt Mayoyao Ifugao ifu Mentawai mwy Minangkabau min Misima-Panaeati mpx Mongondow mog Muna mnb Napu npy Ngaju nji Nias nia Obo Manobo obo Owa stn Palauan pau Pampanga pam Pangasinan pag Paranan prf Rejang rej Roviana rug Sambal xsb	6.3 6.0 6.3 5.2 5.8 6.0 7.9 6.4 6.3 6.3 6.7 6.7 7.3 6.8 5.8 6.3 6.7 6.7 6.3 6.3 6.5 6.3 6.5 6.3 6.5 6.3 6.3 6.3 6.3 6.3 6.3 6.3 6.3 6.3 6.3

Pular fuf

7.0

Sasak sas

Gilbertese gil

Kono (Sierra Leone) kno

Yue Chinese yue 72 Zyphe Chin zyp **QUECHUAN: 22** Avacucho Ouechua quy Cajamarca Ouechua gyc Cañar Highland Quichu qxr 5.6 Cusco Quechua quz 6.8 Huallaga Huánuco Quec qub 6.2 Huamalíes-Dos de Mayo gyh 6.6 Huaylas Ancash Quechu qwh Huaylla Wanca Quechua qvw Inga inb 6.8 6.9 Lambayeque Quechua quf Margos-Yarowilca-Laur qvm Napo Lowland Quechua qvo 6.4 North Bolivian Quechu qul 57 North Junín Quechua qvn 5.9 Northern Conchucos An qxn 6.1 Northern Pastaza Quic qvz Panao Huánuco Quechua qxh San Martín Quechua qvs South Bolivian Quechu guh 6.5 7.0 South Bolivian Quechu guh Southern Pastaza Quec qup 6.1 Tena Lowland Quichua quw **EASTERN SUDANIC: 19** Acoli ach 6.5 Adhola adh Alur alz 5.2 Bari bfa 6.9 Datooga tcc Kakwa keo Karamojong kdj 6.5 Kumam kdi 6.2 6.7 Kupsabiny kpz Lango (Uganda) laj Luwo lwo 6.7 Mabaan mfz Markweeta enb Murle mur Nuer nus Sabaot spy Shilluk shk Southwestern Dinka dik Teso teo TURKIC: 18 Bashkir bak Chuvash chv Crimean Tatar crh 5.4 5.3 Gagauz gag Gagauz gag 5.6 Kara-Kalpak kaa 6.7 Karachay-Balkar krc ⁷ Kazakh kaz 6.8 5.4 Khakas kih 6.5 Kumyk kum 5.4 Nogai nog 6.8 ^e North Azerbaijani azj Southern Altai alt Tatar tat ^e Turkish tur Turkish tur 6.2 Tuvinian tyv 6.2 Uighur uig

5.6

6.6

6.2

6.1

6.8

6.6

6.8

6.1

7.0

6.0

6.5

6.6

5.6

6.8

6.7

6.0

5.9

6.3

6.1

6.0

7.0

6.8

6.1

6.7

6.3

6.3

6.2

6.0

6.7

5.8

6.1

5.5

6.6

^e Romanian ron 5.8 Russian rus Sinte Romani rmo Spanish spa 6.6 ^e Spanish spa ^e Spanish spa ^e Spanish spa Spanish spa Swedish swe ² Swedish swe ^e Tajik tgk " Urdu urd 5.5 Vlax Romani rmy 6.8 **OTO-MANGUEAN: 27** 7.0 Atatláhuca Mixtec mib Ayutla Mixtec miy Central Mazahua maz Chicahuaxtla Triqui trs 6.3 **Diuxi-Tilantongo Mixt xtd** 6.7 Jalapa De Díaz Mazate maj Jamiltepec Mixtec mxt Lalana Chinantec cnl Lealao Chinantec cle Magdalena Peñasco Mix xtm Mezquital Otomi ote Nopala Chatino cya Ozumacín Chinantec chz 6.8 Peñoles Mixtec mil 6.0 Pinotepa Nacional Mix mio 6.8 San Jerónimo Tecóatl maa 6.0 San Jerónimo Tecóatl maa San Juan Atzingo Popo poe 6.2 San Marcos Tlacoyalco pls San Pedro Amuzgos Amu azg 6.8 Santa María Zacatepec mza Sochiapam Chinantec cso Southern Puebla Mixte mit **Tepetotutla Chinantec cnt** Tezoatlán Mixtec mxb Usila Chinantec cuc Yosondúa Mixtec mpn SINO-TIBETAN: 24 Achang acn 6.6 Akeu aeu 5.9 Akha ahk 6.9 Bawm Chin bgr Eastern Tamang taj 6.9 Falam Chin cfm Hakka Chinese hak 6.5 Kachin kac 7.3 Khumi Chin cnk 5.8 Kulung (Nepal) kle 6.4 Lahu lhu 7.2 Lashi lsi 6.5 Lolopo yel 6.8 Mandarin Chinese cmn Maru mhx 7.0 Min Nan Chinese nan Mro-Khimi Chin cmr Newari new Pwo Northern Karen pww Sherpa xsr Sunwar suz Tedim Chin ctd

Huastec hus Ixil ixl Ixil ixl Ixil ixl K'iche' que K'iche' que K'iche' quc K'iche' que K'iche' auc K'iche' que Kaqchikel cak Kaqchikel cak Kaqchikel cak Kauchikel cak Kaqchikel cak Kaqchikel cak Kekchí kek Kekchí kek Mam man Mam mam Mam man Mam mam Mopán Maya mop Popti' jac Popti' jac Poqomchi' poh Pogomchi' poh O'aniob'al kib Tektiteko ttc Tz'utujil tzj Tzeltal tzh Tzeltal tzh Tzotzil tzo Tzotzil tzo Western Kanjobal knj Yucateco yua INDO-EUROPEAN: 40 Albanian sqi Awadhi awa ^e Bengali ben ^e Bengali ben e Bengali ben Caribbean Hindustani hns Chhattisgarhi hne Dari prs ^wEnglish eng Fiji Hindi hif French fra French fra ^e Hindi hin Iranian Persian pes [•] Latin lat Magahi mag Maithili mai Malvi mup ^e Marathi mar ^e Northern Kurdish kmr ^V Oriya (macrolanguage) ori Ossetian oss ^e Polish pol Portuguese por ⁹ Portuguese por Portuguese por Portuguese por

Sudest tgo Sundanese sun Tagalog tgl Tangoa tgp 6.1 Termanu twu Tombonuo txa Toraja-Sa'dan sda 6.3 Tuwali Ifugao ifk Uma ppk Western Bukidnon Mano mbb 6.6 Western Tawbuid twb AFRO-ASIATIC: 45 Bana bew Daasanach dsh Daba dbq 7.0 Dangaléat daa Dawro dwr ^e Eastern Oromo hae Egyptian Arabic arz Gamo gmv 7.2 Gen gej Gofa gof Gofa gof Gude gde Hamer-Banna amf 6.5 Hausa hau Hdi xed Iragw irk Kabyle kab Kafa kbr Kambaata ktb Kamwe hig Kera ker 7.3 Kimré kqp 6.7 Konso kxc 6.6 Koorete kqy Lele (Chad) lln Male (Ethiopia) mdy 6.6 Marba mpg Mbuko mqb Merey meq Mesopotamian Arabic acm Mofu-Gudur mif Muyang muy 6.6 Mwaghavul sur North Mofu mfk 7.0 Parkwa pbi 6.9 Pévé lme Sebat Bet Gurage sgw ² Somali som Standard Arabic arb Sudanese Arabic apd Tachelhit shi 5.0 Tamasheq taq 7.1 ^e Tigrinya tir Tumak tmc 6.6 Wandala mfi MAYAN: 42 Achi acr Aguacateco agu Chol ctu 7.0Chortí caa Chuj cac Chuj cac

Cortal Hustes Nutu 40Sign Light 40Sign Light 40Sign Light 40Chipplan cxLaster Hustes Nutu 40Kiman (Paga New Konikaer 2)TAK-KANAF4Chipplan cxGareres Nutu 40Norricent Nith 40Norricent Nith 40Gareres Nutu 40Gareres Nutu 40Norricent Nith 40Norricent Nith 40DOGON 1Highan Puchk Nahan 20Sunt Tatera mrgTa Dam MitDocons Daga addIsthum-Mearyapan Nah haSunt Tatera mrgTUCANAN: 11Fast BIRDS Haba 1Norricent Data Nahan 20Gareres Naha AddGaranamayMathema 1Norricent Puchk Nahan 2)GaranamayGaranamayMathema 1Norricent Puchk Nahan 2)GaranamayGaranamayGaranamayStern Nerg Nahad NahaGaranamayGaranamayGaranamayGaranamayStern Nerg Nahad NahaGaranamayGaranamayGaranamayGaranamayCREDLES XNP DECINSI 1Stern Nerg NahadSama andGaranamayGaranamayBelke Krid Egght MGaranamayGaranamayGaranamayGaranamayCREDLES XNP DECINSI 1Stern Nerg NahadGaranamayGaranamayBelke Krid Egght MGaranamayGaranamayGaranamayJamadar Creek Egglt MGaranamayGaranamayGaranamayJamadar Creek Egglt MGaranamayGaranamayGaranamayNerfer BridgingGaranamayGaranamayGaranamayNerfer BridgingGaranamayGaranamayGaranamayJamadar Creek EggltGaranamayGaranamayGaranamay		_		
Extern Hussiers Natural 65 Kuman (Papea New Coni, us. 20 TALEADAL4 CORAN: 1 Guerren Namer Cons and 10 50 Lowe Grand Valley Da.dit 10 60 Colina 10 51 Tot Colina 10 51 Tot Colina 10 51 Colina 10 52 Colina 10 53 Colina 10 53 <td>UTO-AZTECAN: 15</td> <td>Huli hui 7.7</td> <td>Sharanahua med 6.0</td> <td>CHIQUITO: 1</td>	UTO-AZTECAN: 15	Huli hui 7.7	Sharanahua med 6.0	CHIQUITO: 1
El Nayar Coract So Kyaka 1/2 72 Guerreron Nahardi uga Maria (Maria Maria Mari				Chiquitano cax 7.3
Guerrers Nahuad up 50 Lower Grand Valley Da.dit 60 Narthern That and 51 Colline constant of the DOGON:1 "Indums-Mecyapan Shah ha 60 South Tarres due 67 "Torton Acn.4 Torton Acn.4 Torton Acn.4 EAST BIRDS IEAD:1 May and bis 60 South Tarres due 60 Torton Acn.4 EAST BIRDS IEAD:1 May and bis 60 South Tarres due 60 Ital and acn.4 60 EAST BIRDS IEAD:1 South acter act Acn.4 60 Fortant Acn.4 Fortant Acn.4 EAST BIRDS IEAD:1 Nardiad fast South acter act Acn.4 60 Fortant Acn.4 Fortant Acn.4 Fortant Acn.4 EAST BIRDS IEAD:1 Nardiad fast South acter act Acn.4 61 Fortant Acn.4 Fortant Acn.4 EAST BIRDS IEAD:1 Torton Acn.4 EAST BIRDS IEAD:1 Torton Acn.4 EAST BIRDS IEAD:1 Eastern Marken Conte Acn.4 Fortant Acn.4 EAST BIRDS IEAD:1 Torton Acn.4 EAST BIRDS IEAD:1 Torton Acn.4 EAST BIRDS IEAD:1				COFÁN: 1
Highand Fuchis Namu az, 61 Lover Grand Valley Da.ult 71 Pishnaw-Necepapon Nah ab, 62 South Tairon was, 53 Torro So Degon dt. Withow Oxace Namu 65 South Tairon was, 53 Torro No Congen dt. Northern Davise Nation 46 Dama 64 54 Northern Davise Nation 46 Dama 64 54 Northern Paubin Nahun 45 Dama 64 54 Start Arcess Cara cd, 64 Koragang ce 64 South startin Paubin Nahun 45 Barano 64 54 South startin Paubin Nahun 46 Pausagang ce 64 South startin Paubin Nahun 46 Pausagang ce 64 South startin Paubin Nahun 46 Pausagang ce 64 South startin Paubin Nahun 46 Pausagang ce 74 South startin Paubin Nahun 46 74 Galillo Carlia cer 72 Galillo Carlia cer 74 Galillo Carlia cer 72 Galillo Carlia cer 74 Galillo Carlia cer 72 Balander Creele Egglis rd 74 Tarcan 0m< 72				Cofán con 6.8
* Intrans-Mecayapan Nahah 6.1 Value 6.2 * Frank 5.3 TOTON CNN. 4 Northern Savan Matah 6.3 South Torreson Car. 6 7 TOTON CNN. 4 EAST BIRDS BLED: 1 Northern Savan Matah 6.3 South Torreson Car. 6.4 6 South Torreson Car. 6.4 5 Mayama Matah 6 EAST BUCK NILLE: 1 Nasiation and torreson Car. 6.4 5 Nasiation and torreson Car. 6.4 6 Carl BUCK NILLE: 1 Torreson Car. 7.6 Call AUS NILLE: 1 Carl BUCK NILLE: 1 Torreson Car. 7.6 Call AUS NILLE: 1 Torreson Car. 7.6 Call AUS NILLE: 1 Torreson Car. 7.6 Gal AUS NILLE: 1 San Matec Del Mar Hua hun. 1 Belance Torele Edgil and 5 Torreson Car. 7.6 Gal AUS NILLE: 1 San Matec Del Mar Hua hun. 1 <td>0</td> <td></td> <td></td> <td>DOGON: 1</td>	0			DOGON: 1
Ishmus-Meregrapan Nahur (2) South Tairon (2) TOTONACAN: 4 FABRUPS HEAD: 1 Northern Davies Nahur (2) TITCANDA: 11 Fighhand Toimac (2) Fighhand Toimac (2) Fighhand Toimac (2) Fighhand Toimac (2) Fighhand (2)	-			Toro So Dogon dts 8.2
Mayo of y Norther Duola Natura of Southers Parola Natura Norther Oxac Natura M Southers Parola Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura Natura <td></td> <td></td> <td></td> <td>EAST BIRD'S HEAD: 1</td>				EAST BIRD'S HEAD: 1
Northern Pasters Nature 10 Sorthern Pasters Negra Nature 10 Sinta Teress Cora col. 53 Sorthestern Negra Nature 10 Sinta Teress Cora col. 54 Sorthestern Negra Nature 10 Sorthestern Pasters Negra Nature 10 Sorthestern Huasters Nature 10 Sorthestern Huaste	• •			
Norther Packa Natura of Samt Teres Cars cot Sierra Negen Naturd ma Gamano gy Gamano gy Samt Serra Negen Naturd ma Gamano gy Gamano gy Gamano gy Gamano gy Samt Serra Negen Naturd ma Gamano gy Gamano gy	• •			
Santa Terese Carros, 53 Santa Terese Carros, 53 Surta Terese Carros, 54 Surta Aces, 54 Western Haasteea Nahard and 54 Western Haasteea Nahard and 54 Western Haasteea Nahard and 54 Western Haasteea Nahard and 54 CREOLES AND PIDGINS: 14 Beitze Krift English by 164 Beitze Krift Krift English by 164 Beitze Krift			· ·	
Sierra Negra Naturati nu of Skifild-ALDUT; Sierra Negra Naturati	•		• • • •	
Southasher I retent av lup Secoya sey Galabit Carbon Galabit Carbon <th< td=""><td>Sierra Negra Nahuatl nsu</td><td></td><td></td><td></td></th<>	Sierra Negra Nahuatl nsu			
Western Husstein Nahu word 01 Scoops op 40 03 Galbi Carlb ard 72 GLAICURCAN: 1 Taraettin-Ahmaettin-Tahi Sionas on 50 Partaman physics 10 Toha inde Belize Kriel English hig 05 Tucano too 70 Culla cull 12 11 "Hunong Daw maw Belize Kriel English hig 05 Tucano too 70 Guabbe ogi 14 20 "Hunong Daw maw Belize Kriel English hig 05 Tucano too 70 Guabbe ogi 14 20 "Hunong Daw maw Belize Kriel English hig 05 Tucano too 70 Guabbe ogi 26 Barabas 58 KHOE-KWADI: 1 Jamaican Creek Englism 6 05 Eastern Bolivina Guarg at 26 Marai Huitoto hua 64 KOMAN: 1 Jamaican Creek Fardi Sing 10 Guarjajára gub 64 Marai Huitoto hua 64 KOMAN: 1 Workyen nife 70 Guarjajára gub 64 Marai Huitoto hua 64 KOMAN: 1 Uduk 40 Siat Lucian Creek Fardi Sing Guarjajára gub 64 Marai Huitoto hua 64 KOMAN: 1 Siat Lucian Creek Fardi Sing Gardina Guarg gub 64 Marai Huitoto hua 64 KOMAN: 1 Siat Lucian Creek Fardi Sing Gardina Guarg gub 64 Guafjafar gub 64 Marai Huitoto hua 63 Marai Huitoto hua 64 Stranactan Tope yan 74 Western Bolivina Guarg gub 72	0			Central Siberian Yupi ess 5.8
Zazattian-Ahmeutlian-Thili 65 Strain with 50 Patamona phe 65 Tota unb CREOLES AND PRICINS: 14 Strain with 67 GUAHBAN: 15 GUAHBAN: 16 HMONG-MIEN: 1 Beizer Kriol English tori 68 Turcano tori 72 Guayabere gua 64 HHONG ANIEN: 1 Battere Monor Crede Engli 73 Aché gua 64 Bara Marco Mar Han Invo Hattian hat 70 TUPLNN 8 HUITOTOAN: 3 Sam Marco Marc March Marco 65 Sam Marco March March Marco Sam Marco March Marco Sam Marco March Marco 64 MKHOF-KWADI: 1 Jannaican Crede Engli 63 Eastern Biolivan Guarg gui 62 MIKE-ZOQUE: 3 Eduk undu KKOKDF Nigerian Pidgin peru 64 Paragauyan Guarani ggi 62 MIKE-ZOQUE: 3 Eduk undu More March Statu Lucian Crede Fari 55 Urbubic Kaapari yr 62 March March 63 March March Statu Lucian Crede Fari 54 Verstern Bolivan Guarg yr 74 URALC: 3 March March 10	Western Huasteca Nahu nhw		Galibi Carib car 7.3	GUAICURUAN: 1
CREOLES AND PIDCINS: 14 Striano str. 72 GUAHBAN: 3 HINONG-MIEX: 1 Beitze Kriel Ragish bzj. 63 Tucano tuo 76 Cashiba gab 76 "Hinong Daw mwy Eastern Maroon Crool cijk 78 Tucano tuo 76 Guayabero gao 76 San Mateo Del Mar Hua huy Islande Creole Engli and 63 Eastern Bolivan Gaze guo 64 Marci Huiton huu 64 KHOE-KWADI: 1 Jamaelac Creole Engli and 63 Eastern Bolivan Gaze guo 64 Marci Huiton huu 64 KORDOFANIAN: 1 Migerian Pidgin pern 60 Guarani guo 66 Marci Huiton huu 65 KORDOFANIAN: 1 Saint Lucian Creole Faci Weitern Bolivan Gaze guo 72 Queztatteree Mitexen 70 KORDOFANIAN: 1 Saint Lucian Creole Faci 64 Paraguayan Gazarani guo 60 Gatifan tabi 57 Marchi Huiton Huu huv Saint Lucian Creole Faci 70 RAWAKAN: 7 Vientoikano tun 70 KORDOFANIA: 1 Hitighalan Popolece pai 60 Marchi uun 70 Marchi uun	Zacatlán-Ahuacatlán-T nhi		Patamona pbc 6.3	Toba tob 6.2
Belize Keiol English bz/ Bishama bit 65 7 Tucano tuo 76 70 Cuiba cui 75 70 Cuiba cui 75 70 Manong Daw mawa Bishama bit Eastern Maroon Creol City 13 Tuyuea tuo 76 Guabibo gub 76 San Mateo Del Mar Hua huv Haitan hat 70 TUPLAN: 8 HUTOTOAN: 3 KHOE-KWADI: 1 Jamadean Creole Englis 63 Aché guq 64 Minica Huitoo huv 64 KOMAN: 1 Morisyen mit 70 Guapaizyen gud 64 Minica Huitoo huv 64 KOMAN: 1 Morisyen mit 70 Guarayu gur 68 Minica Huitoo huv 64 KOMAN: 1 Sinamacan sm 64 Paraguayan Guaranigo 66 Huiton huv 64 KORDOFANIAN: 1 Optertan Hugin pen 65 Garfuna cub 74 Western Bioliva Garar guo 76 MACRO-GEI; 1 Sraman Tongo yrn 76 Ashainka cub 78 78 MacRo-GEI; 1 MacRo-GEI; 1 CENTRAL SUDANIC: 13 Garfuna cub 74 Bacher Kun kuv 63 Tabaru hyv 57 MACRO-GEI; 1 Aringa hu 65 Garfuna cub 76 Ashainka cub 72 Minica Huitoo hito 73 CENTRAL SUDANIC: 13 <td< td=""><td>CREOLES AND PIDGINS: 14</td><td>Siriano sri 7.3</td><td></td><td>HMONG-MIEN: 1</td></td<>	CREOLES AND PIDGINS: 14	Siriano sri 7.3		HMONG-MIEN: 1
Bislanak is67 2Turgue tur76 2Gauakibo gal60 2HUAVEAN: 1 San Mateo Del Mar Hua hurHaitan hart70TUPLAN: 8 TUPLAN: 8HUITOTON: 3 3 Bandatco Cele Engli Jan64Aché gan64Bora ban58Janacaca Creole Engli Jan64Eastern Bolivian Garar gan64Murai Huitoo huu64KOHANN: 1Janacaca Creole Engli Jan64Gauayabi kyz62Murai Huitoo huu64KOHANN: 1Morisyen mit To Grayayan Gaeran Gyan66MKEZ-ZOUER66KORDOFANIAN: 1Uduk dulNgerian Pidgin pem69Kayabi kyz62Coatita Mixe areo60KORDOFANIAN: 1Sint Lacian Creole Farf53Urubi-Kaapor urb72Urubi-Kaapor urb72Uduk dulSint Lacian Creole Farf54RAVWAKAN: 7Finishin 52Aruanu may"Tok Pisin tpi62Ashainika cai78Komi-Zyriaa kyv67MACRO-GE: 1CENTRAL SUDANC: 13Garifuna Garag gan63Garifuna Garag gan63Galarabi gan71MACRO-GE: 1Gor gar65Parecis pab64Garabi gan72MMACRO-GE: 1MACRO-GE: 1Gor gar74Border Knan kyn63Garabi gan73MMACRO-GE: 1Jar Modo bey65Parecis pab64Garabi gan73MMACRO-RE: 1Mixel Modo by74Border Knan kyn63Garabi gan74MMACRO-RE: 1Mora gan gan74Border K	Belize Kriol English bzj	6.8 Tucano tuo 7.7	Cuiba cui 7.3	
Eastern Maroon Creeke dijk 78 TUPLAN: 8 HUTTOTON: 3 San Mateo Del Mar Hua huv Islander Creeke Engli ier 63 Aché geq 61 Bora boa San Mateo Del Mar Hua huv Jamaican Creeke Engli ier 63 Eastern Bolivian Guar grag 64 Muri Huitoi hun 64 KOMAN: 1 Wortsyna mfe 70 Guaraya gyr 68 Muri Huitoi hun 64 KOMAN: 1 Wortsyna mfe 70 Guaraya gyr 68 Muri Huitoi hun 64 KORDOFANAN: 1 Santan Creeke Fact 58 Urubi-Kaapo urb 62 Quetzing Kisc pm 61 Udka dud Sarana Creeke Fact 58 Urubi-Kaapo urb 72 Finishi fin 50 MACROOFANAN: 1 Sarana Creeke Fact 54 Urubi-Kaapo urb 72 Western Bolivian Guar grav 72 MACROO-GE: 1 CENTRAL SUDANIC: 13 Garifina cab 71 Udmurt udm 50 Kayapó fu Gorg gr 65 Regiond biy 70 Nomatiguega and 63 Tabaru biy 50 MACROO-GE: 1 Mad dim hi 74 Border Kuna kvn 65 ArMARAN: 2 MACROO-GE: 1 Eastern Bolivian Guar grav 72 Mistiko miq Guarg gvt 78 Tereco ter	• •	6.7 Tucano tuo 7.6	Guahibo guh 7.6	
Hairian hat 22 TUPIAN: 8 HUTOTOAN: 3 KHOE-KWADI: 1 Jamaican Creole Engli G 3 Eastern Bolivian Guar gui 82 Minica Huitoto hu 64 KOMAN: 1 Morisyen mfe 70 Guarayu 97 68 Minica Huitoto hu 64 KOMAN: 1 Minica Huitoto hu 64	Eastern Maroon Creole djk	7.8 Tuyuca tue 7.6	Guayabero guo 7.6	
Namer Jannicar Creek Englin Fore Creek Englin Moriey and For Sigrian Pidgin pen Sain Lacian Creek Fact Sain Lacian C	Haitian hat	7.0 TUPIAN: 8	HUITOTOAN: 3	
Jamincali Creor Engly JamEastern Borivan Guargu gyrGauargu gyrGauargu gyrGauargu gyrMurcal Hautoo huuGauKOMAN: 1"Morisyen mfc70Guargu gyr68Murcal Hautoo huu64KOMAN: 1Nigerian Pidgi peed61Paraguayan Guarani gug66Gautafia Mixe mco50KORDOFANIAN: 1Pijin pis61Paraguayan Guarani gug66Gautafia Mixe mco50More moriSaint Lucian Creole F act58Urubó-Kaapor urb72Querzaltopee Mixe park61Siant Lucian Creole F act58Urubó-Kaapor urb72URALIC: 3Toruanu anySranan Congo srn76ARWAKAN: 7"Finnish fin52Aruanu any"G Tok Pisin (pi62Ashininka cni78Komi-Zyrian kpi55Kayapó txuAruoga yu60Machigueng meb63Galel gbi52Eanset caxAruokaya yu60Nomatsiguenga not63Tabaru thy57MATACOAN: 1Gor ger65Central Aymara ayr73MISUMALPAN: 1Gula yu73Treeno ter63AYMARAN: 2Mistio mingJur Modo bes67CHIBCHAN: 7Central Aymara ayr73MISUMALPAN: 1Ma'di mhi74Central Tunebo tur67Gaurary angr74Mistio mingJur Modo bes67CHIBCHAN: 7Central Aymara ayr73MISUMALPAN: 1Mard al dri	Islander Creole Engli icr	6.3 Aché guq 6.1	Bora boa 5.8	KHOE-KWADI: 1
** Morisyen mfe7.0Garayu gyr Kayabi kyz 6.26.3MIXE-ZOQUE: 3Uduk uduNigerian Pidgin pem6.4Kayabi kyz Paraguayan Guarang gug 6.26.2Coatfain Mixe mod Poplotea pol 8.047.0KORDOFANIAN: 1Saint Lacian Creole F act * Saramacen arm * Tak Pisin pi 6.35.4Urubú-Kaapor ub Vatta KAWAKAN: 77.2Finnish fin 5.27.2LOWER SPIK-KAMU: 1Sranan Tongo ri * Tak Pisin pi * Tak Pisin pi 6.36.5Garifuna cab 7.47.2Finnish fin 5.27.2Aruanu may* Tak Pisin pi * 7.06.5Garifuna cab * 7.47.8Komi-Zyrian kyp * 6.76.7MACRO-CE: 1CENTRAL SUDANIC: 13 * Garigun ch * 6.5Garifuna cab * 7.47.4MACRO-CE: 1MACRO-CE: 1Avikaya avu * 6.6Garifuna cab * 7.46.6Garlea gbi 7.2MACRO-CE: 1Madigueng anot * 6.3Tereno ter * 6.5AYMARAN: 7*MATACOAN: 1* Gargur ch * 6.4Tereno ter * 6.5AYMARAN: 7*MISUMALPAN: 1* Kenga kyq * 7.4Border Kuna kyn * 6.86.8Central Aymara ayr * 7.27.2MISUMALPAN: 1* Kenga kyq * 7.4Border Kuna kyn * 6.76.8Prizz PhiMiski omig * 6.7Maza mera * 7.2Miski omig * 7.2* Maru Minhi 7.1Central Tunbe tir * 6.6Garabiang gum 6.4Puinave gum * 7.2PhiZaAN: 1* Maru Modo bas * 6.8Ter	Jamaican Creole Engli jam	6.3 Eastern Bolivian Guar gui 8.2	Minica Huitoto hto 7.0	Southern Samo sbd 8.3
Nigerian Pidgin pen (1)60Kayabi kyz Praguyan Guarani gug (6)62Contrain Mixemeo (1)73KORDOFAMLAN: 1Saint Lacian Creole Fad Saramaccan srm74Urubić-Kangor urb V72URALIC: 3UWER SEPIK-RAMU: 1Sranan Tongo srn74Western Bolivian Guarg guy Tak Pisin pi (2)72URALIC: 3UWER SEPIK-RAMU: 1"Tak Pisin pi (4)64Ashainika cni Ignaciono ga78Komano SiAKayapó Kuu Udmurt dudi50Kayapó Kuu MACRO-GE: 1CENTRAL SUDANIC: 13Garifuna cab Ignaciono ga71Udmurt dudi50Kayapó Kuu MacNologeAring Jue Gor gar65Garifuna cab Ignaciono ga66Galela gbi Tabaru by 57MACRO-GE: 1Mathio Androger Guay gyt73Tereno ter65AYMARAN: 2Macana Mathion qJur Modo bes Gor gar Gar ga67Central Aymara ayr72Mistion qMady gyt73Bordor Kuna kvn68Central Aymara ayr73Mistion qMady gyt74Bordor Kuna kvn68Central Aymara ayr74Mistion qMady gyt75San Bus kvn akan aka Konga kyg74Bordor Kuna kvn68Pácz pbbMator md Gor gal kogCograf kogGambiano gum64PUINAYE: 1Puinave puiMator md Mator md74Central Aymara ayr74Mistion qMabra bay75San Bus kon akan akan aka Cograf kog66Central Aymara ayr73	Krio kri	7.0 Guajajára gub 6.4	Murui Huitoto huu 6.4	KOMAN: 1
Pijip js Saint Lucian Creole F act Saint Lucian Creole F act Saranaccan swaCall Urubé-Kaapor urb 72 Western Bolivan Guar gan 72 URALLC: 3Highland Popolaca poi Queztatepee Nice park Queztatepee Nice park Noman Togo srn 62ROMON AUXI: 1 UWER SEPIK-RAMU: 1 LOWER SEPIK-RAMU: 1 UWER SEPIK-RAMU: 1 OWER SEME 0 OWER SEME 0 <b< td=""><td>W Morisyen mfe</td><td>7.0 Guarayu gyr 6.8</td><td>MIXE-ZOQUE: 3</td><td>Uduk udu 6.8</td></b<>	W Morisyen mfe	7.0 Guarayu gyr 6.8	MIXE-ZOQUE: 3	Uduk udu 6.8
Pijin pis Saint Lucian Creote F act Saramaccan sem Saramaccan sem (74)Paraguayan Cuarani gu Urubú-Kaapou the 2260 Quetzitepec Mixe pxm61 COMER SEPIK-RAMU: 1Saramaccan sem Saramaccan sem (74)Western Bolivian Guar guv Western Bolivian Guar guv (74)72 URALIC: 3IOWER SEPIK-RAMU: 1 LOWER SEPIK-RAMU: 1Sarama Congo sem (76)62ARAWAKAN: 7* Finnish fin Saramaccan sem (76)Araamu nsy MACRO-CE: 1CENTRAL SUDANIC: 13Garifuna cai (74)71 Udmurt udm (75)Komi-Zyrian kpv (76)76CENTRAL SUDANIC: 13Garifuna cai (76)71 MASCOLN: 1MASCOLN: 1 MASCOLN: 1Avokaya ava Bedjond biy (76)70Nomatsiguenga not (76)63 Catel ag bi (76)72 MATCOAN: 1Gorager (76)65Parecis pab (76)Tobelo the (76)66MatzOAN: 2Jur Modo bec (77)70CHIBCHAN: 7 Central Aymara ayr (72)23 Miskitomiq (74)Miskitomiq (74)Macinga gene (74)67CHIBCHAN: 7 Central Aymara ayr (74)24 Miskitomiq (74)Miskitomiq (74)Madi mini (74)71Central Yunara ayr (74)23 Miskitomiq (74)Miskitomiq (74)Mardi mini (74)72Central Yunara ayr (74)73 Miskitomiq (74)Miskitomiq (74)Mardi mini (75)74Central Yunara ayr (74)74 Miskitomiq (74)Miskitomiq (74)Mardi mini (75)74Central Yunara ayr (74)74Miski	Nigerian Pidgin pcm	6.9 Kayabí kyz 6.2	Coatlán Mixe mco 7.9	KORDOFANIAN: 1
Saint Lucian Creok F acf53Urubú-Kaapor urb72Quetzaltepee Mise pam61LOWER SEPIK-RAMU:1**Saramacea srm74Western Bolivian Guar gnv72URALIC: 3Aruamu msySrana Tongo srn76ARAWAKAN: 7* Finnish fin52Aruamu msy**Tok Pisin tpi62Asháninka cni78Komi-Zyrian kpv67MACRO-CE: 1CENTRAL SUDANIC: 13Garifuna cni78West PAPUAN: 3MASCOIAN: 1Avokaya avu65Ignaciano ign63Galda gbi72Enxet cnxBedjond biy70Nomatsiguenga not63Galda gbi72Enxet cnxGorg gr65Parecis pab61Tobelo th66Maca mcaGulay gvi78Tereno ter65AYMARAN: 2MISUMALPAN: 1Jur Modo bex67CHIBCHAN: 7Central Aymara ayr73MISUMALPAN: 1Kenga kyq74Border Kuna kvn68Central Aymara ayr74Miskito miqLugbara lgg68Cabécar cjp62BARBACOAN: 2PAÉZAN: 1Ma'di mhi71Central Tunebu th67Galambiano gun64PUINAVE: 1Moru mgd65Ngibere gwi64CHOCO: 2Poinave pulNgambaysba75Kanuada kan66Halb Mongolian klk74SUKA: 1Moru mgd75Kanuada kan66Halb Mongolian klk74SUKA: 1Maru may mgb75Kuru klkr70MAKHACO: 2	Pijin pis	6.1 Paraguayan Guaraní gug 6.6	Highland Popoluca poi 8.0	
Saramacan sm Vestern Bolivian Cuar gow Vestern Bolivian Bolivian Cuar gow Vestern Bolivian	Saint Lucian Creole F acf	5.8 Urubú-Kaapor urb 7.2	Quetzaltepec Mixe pxm 6.1	
Sranan long stri100 (10)ARAWAKAN: 7 (10)Finnins lin100 (10)100 (10)CENTRAL SUDANIC: 13Garifuna cab7.1Udmurt udm59Kayapó txuCENTRAL SUDANIC: 13Ignaciano ign60WEST PAPLAN: 3MASCOIAN: 1Avokaya va60Machigueng mcb60Galela gbi72Enxet enxMacting luc6.5Ignaciano ign60Galela gbi72Enxet enxBedjond bjy70Nomatsigueng mcb6.3Tabaru tby57MATACOAN: 1Gor gri6.5Parecis pab6.1Tobelo th6.6Maca mcaJur Modo bex6.7CHIBCHAN: 7Central Aymara ayr7.3MISUMALPAN: 1Kenga kyq7.4Border Kuna kyn6.8Central Aymara ayr7.3Miskito miqLugbara ign6.8Cabécar cjp6.2BARBACOAN: 2PÁEZAN: 1Moru ugd6.5Ngäbere gyn6.4Chortenr Enberáre mp7.3SULKA: 1Moru ugd6.5San Blas Kuna cuk6.1Epena sig7.3SULKA: 1Moru ugd6.6Teribe tfr6.6Northern Enberáre mp7.3SULKA: 1Maxmande lem6.0Teribe tfr6.6Northern Enberáre mp7.3SULKA: 1Maxmand kini7.7Kurukh kru7.0Kalanyk xal6.9Purepecha tszMoru ugd6.7Gambarabam7.5MONGOLIC: 2Sulka: sunSulka: sunMoru ugd6.7Kuruk	[₩] Saramaccan srm	7.4 Western Bolivian Guar gnw 7.2	URALIC: 3	
CENTRAL SUDANC: 13Carifuna can tanga lucKayapó txuCENTRAL SUDANC: 13Garifuna can tanga lucCarifuna can tanga lucKayapó txuAnoma can anoma can Avokaya avu65Kayapó txuMachiguenga net6.6Garifuna can tanga lucKayapó txuGor gr6.5Rance Call Parceis pabCentral Aymara ayr7.2MISUMALPAN: 1Machiguenga net6.6Gentral Aymara ayr7.2MISUMALPAN: 1Machiguenga net6.6Central Aymara ayr	Sranan Tongo srn	7.6 ARAWAKAN: 7	^w Finnish fin 5.2	
Aringa luc6.5Ignaciano ign6.9WEST PAPUAN: 3MASCOIAN: 1Avokaya avu6.9Machiguenga meb6.6Galela gbi7.2Enset ensBedjond bjv7.0Nomatsiguenga not6.3Tabaru thy5.7MATACOAN: 1Gor gr6.5Parecis pab6.1Tobelo thi6.6Maca macaGulay gvl7.8Tereno ter6.5AYMARAN: 2MISUMALPAN: 1Jur Modo bec6.7CHIBCHAN: 7Central Aymara ayr7.3MISUMALPAN: 1Kenga kvg7.4Border Kuna kvn6.8Central Aymara ayr7.3MISUMALPAN: 1Ma'di mhi7.1Central Tunebo tuf6.7Awa-Cuaniguer kvi6.8Pézz pbbMbay myb8.4Coogui kog6.7Guambiano gun6.4PUINAVE: 1Manbara bam7.5San Blas Kuna cuk6.1Epena sja7.3SULLKa: 1Nomaande lem7.5Karanda kan6.6Halt Mongolian khk7.4Sulka suaBanbara bam7.5Karanda kan6.6Halt Mongolian khk7.4TARASCAN: 1Boko (Benin) bc7.0Malaylam mal7.0NAKH-DAGHESTAN: 2TICUNA: 1Dyula dyu7.4CHSUGNASTIC: 4TACANAN: 2TOL: 1Mandinka muk6.1Juang jun7.0Taceana tna6.0Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 1TOR-ORYA: 1Boko (Benin)7.7É tatem Bru bru6.3E set gji aces6.5 <td>^e Tok Pisin tpi</td> <td>6.2 Asháninka eni 7.8</td> <td>Komi-Zyrian kpv 6.7</td> <td>MACRO-GE: 1</td>	^e Tok Pisin tpi	6.2 Asháninka eni 7.8	Komi-Zyrian kpv 6.7	MACRO-GE: 1
Avokaya avu Bedjond bjy60Galela ghi72Enxet enxBedjond bjy70Nomatsiguenga not Barea is pab6.1Tabaru by5.7MATACOAN: 1Gor gq6.5Paree is pab6.1Tobelo th6.6Maca meaGulay gv17.8Tereno ter6.5AYMARAN: 2MISUMALPAN: 1Jur Modo bes6.7CHIBCHAN: 7Central Aymara ayr7.2MisUMALPAN: 1Lugbara lgg6.8Cabécar cjp6.2BARBACOAN: 2PÁEZAN: 1Ma'il mit71Central Tunebo tif6.7Ganabiano gun6.4PúEZAN: 1Moru mgd6.5San Bia Kuna cuk6.7Ganabiano gun6.4PUINAVE: 1Moru mgd6.5San Bia Kuna cuk6.1Epena sja7.3SULKA: 1Nomaande lem6.6Teribe tfr6.6Northern Emberá emp7.3SULKA: 1MANDE: 13DRAVIDIAN: 5MONGOLIC: 2PurepechatszBabara ban7.5Kannada kan6.6Halh Mongoliankhk7.4TARASCAN: 1Busa bqp6.7Cintral tam 6.6Avarie ava6.0Ticuna teaDyula dyu7.1Chechen fr7.1TicUNA: 1Madika kan6.6Central Ojibwa ojc6.5Ticuna teaMANDE: 13Gurange si7.3Tacana tan6.6Totheore siMaxi kan kan6.5Central Aymara ayr7.1TiCUNA: 1Boko (Benin) bq7.0Malayalam mal8.7NAKH-DAGHESTAN:	CENTRAL SUDANIC: 13		Udmurt udm 5.9	Kayapó txu 7.7
Bedjond bjy7.0Nomatsiguenga not6.3Tabaru by5.7MATACOAN: 1Gor ggr6.5Parceis pab6.1Tobelo th6.6MATACOAN: 1Gulay gvl7.8Tereno ter6.5AYMARAN: 2MISUMALPAN: 1Jur Modo bes6.7CHIBCHAN: 7Central Aymara ayr7.3MISUMALPAN: 1Kenga kyq7.4Border Kuna kyn6.8Central Aymara ayr7.2Miskito miqLugbara igg6.8Cabécar cjp6.2BARBACOAN: 2PÁEZAN: 1Ma'di mhi7.1Central Innebo tuř6.7Guambiano gun6.4PUINAVE: 1Mora mgd6.5Ngäbere gym6.4CHOCO: 2Puinave puiNgambay sba7.5San Blas Kuna cuk6.6Hath Mongolian khk7.4SULKA: 1Nomaande len6.6Teribe tifr6.6Hath Mongolian khk7.4TARASCAN: 1Bambara bam7.5Kannada kan6.6Hath Mongolian khk7.4TARASCAN: 1Boko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Boko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Loko lok7.1Crugu tell4AUSTRO-ASIATIC: 4TACANAN: 2Tol.1Mandink amk6.8Khmer khw7ACCACA-VANA: 2Tol.2Mandink amk6.4Kuranko khk7Chechen fer7.1Tol.1Dyula dyu7.1Eastern Bru bru6.3E	Aringa luc			MASCOIAN: 1
Gor gqr Gulay gvl6.5Parecis pah Freeno ter6.1Tobelo th 6.66.0MATACOAN: 1Gulay gvl Jur Modo bex (Kenga kvq Lugbara lgg Ma'di mbi M'a'di mbi M'a'di mbi7.1CHIBCHAN: 7Central Aymara ayr (Cabécar cip 6.27.2MISUMALPAN: 1Ma'di mbi M'a'di mbi M'a'di mbi7.1Central Tunebo tuf (Cabécar cip 6.76.7Cambiana ayr (Cambiana ayr 6.77.2Miskito miq (A (Cabécar cip 6.77.2Miskito miq (Cabécar cip 6.4Ma'di mbi M'a'di mbi M'a'di mbi Morumg 6.57.5San Blas Kuna cuk (Cabécar cip 6.76.7Awa-Cuaiquer kwi (Cabécar cip 6.86.6Pézz pbbMorumg Maymbay sha A.57.5San Blas Kuna cuk (Cabécar cip 6.46.1Epena sja (Cabécar cip 7.37.3SULKA: 1MANDE: 13DRAVIDIAN: 5MOROGLIC: 2Sulka suaSULKA: 1Bambara bam Daup (A7.7Kurukh kru (A7.0Kalmyk xal (A6.9Purepecha tszBoko (Benin) bqc Dyula dyu Loko tok (A7.1AUSTRO-ASIATIC: 4TACANAN: 2TICUNA: 1Mandika muk Sus us Mandika muk (AS6.8Central Ojibwa oje6.1URACOAN: 1Mandika muk Sus us Mandika muk A6.8Central Aymara ayr (A7.3TICUNA: 1Mandika muk A6.8Khmer khm A8.7ALCIC: 1Orya uryMandika muk A6.8Khmer khm (A7.0Tacana tna (A6.6TOR-ORYA	Avokaya avu		0	Enxet enx 6.0
Gulay gyl7.8Tereno ter6.5AYMARAN: 2Maca meaJur Modo bex6.7CHIBCHAN: 7Central Aymara ayr7.3MISUMALPAN: 1Kenga kyq7.4Border Kuna kyn6.8Central Aymara ayr7.2Miskito miqLugbara lge6.8Cabécar cjp6.2BARBACOAN: 2PÁEZAN: 1Ma'di mhi7.1Central Tunebo tf6.7Awa-Cuaiquer kwi6.8Páez pbbMbay myb8.4Cogui kog6.7Guambiano gum6.4PUINAVE: 1Moru mgd6.5Ngäbere gyn6.4CHOCO: 2Puinave puiNgambay sba7.5San Blas Kuna cuk6.1Northern Emberá emp7.3SULKA: 1Nomaande lem6.6Terleb tf6.6Northern Emberá emp7.3SULKA: 1Bambara bam7.5Kannada kan6.6Halh Mongolian khk7.4TARASCAN: 1Bissa bib7.7Kurunk kru7.0Kalmyk xal6.9Purepecha ts.2Boko (Beni) bog7.0Malayalam8.7NAKH-DACHESTAN: 2TICUNA: 1Dyula dyu7.1C'Tamil tam6.6Avaric ava6.0Ticuna tcaDyula dyu7.1Central Junebo K6.1Chechen che7.1TOL: 1Dyula dyu7.1C'Tamil tam6.6Avaric ava6.0Ticuna tcaDyula dyu7.1C'Tamil tam6.6Chechen che7.1TOL: 1Dyula dyu7.1C'Tamil tam6.6Chechen che	• •			MATACOAN: 1
Gulay gvl78Tereno ter65AYMARAN: 2MISUMALPAN: 1Jur Modo bex6.7CHIBCHAN: 7Central Aymara ayr7.2MISUMALPAN: 1Kenga kyq7.4Border Kuna kyn6.8Central Aymara ayr7.2Miskito miqLugbara lgg6.8Cabécar cjp6.2BARBACOAN: 2PÁEZAN: 1Ma'di mii7.1Central Tunebo tif6.7Awa-Cuaiquer kyi6.8Páez pbbMbay myb8.4Cogui kog6.7Guambiano gum6.4PúINAVE: 1Moru mgd6.5Ngäbere gym6.4CHOCO: 2Puinave puiNgambay sba7.5San Blas Kuna cuk6.1Epena sja7.3SULKA: 1Nomaande lem6.6Teribe ffr6.6Northern Emberá emp7.3SulkA: 1MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka sua1Bambara bam7.5Kurukh kru7.0Kalmyk xal6.9Purepecha tszBoko (Benin) buc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Dyula dyu7.1C'Telugu tel9.1Chechen che7.1Tol.: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol.: 1Tol.: 1Mandinka muk6.8Khmer khn8.7ALGIC: 1Orya uryMende (Sierra Loone) me7.7 ⁶ Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bio8.1JIVAROAN: 4ARUAUAN: 1Urarina ura <td></td> <td>· · · · · · · · · · · · · · · · · · ·</td> <td></td> <td>Maca mea 6.1</td>		· · · · · · · · · · · · · · · · · · ·		Maca mea 6.1
Jur Modo besOrdChink HAN 7Central Aymara ayr7.2Miskito miqKenga kyq7.4Border Kuna kvn6.8Central Aymara ayr7.2Miskito miqLugbara lgg6.8Cabécar cip6.2BARBACOAN: 2PÁEZAN: 1Ma'di mhi7.1Central Tunebo tuf6.7Awa-Cuaiquer kwi6.8Páez pbbMbay myb8.4Cogui kog6.7Guambiano gun6.4PUINAVE: 1Moru mgd6.5Ngäbere gym6.4CHOCO: 2Puinave pui8Nomaande lem6.6Teribe tir6.6Northern Emberá emp7.3SULKA: 1MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka sua9Bambara bam7.5Kannada kan6.6Halh Mongolian khk7.4TARASCAN: 1Bissa bib7.7Kurukh kn7.0Kalamyk val6.9Purepecha tszBoko (Benin) hqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol. 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol. 1Mandinka mnk6.8Khmer khm8.7ALCIC: 1Orya uryMende (Sierra Leone) men7.4Central Ojibwa ojc6.1URARINA: 1Northern Bobo Maaré bob8.5Aguaruna agr6.4CACUA-NUKAK: 1Urarina uraNorthern Bobo Maaré bob8.5Aguaruna agr6.4CACUA-NUKAK: 1Urarina uraMandinka mak f				
Kenga KyiFor Central Tunebo tifGold Central Ary (marked)Central Central Centra				
Ma'di mit Ma'di mit Mbay myb7.1Central Tunebo tri 6.76.7Awa-Cuaiquer kwi Guambiano gum 6.4Pácz pbbMbay myb8.4Cogui kog 6.56.7Guambiano gum Guambiano gum 6.4PUINAVE: 1Moru mgd Ngambay sba 7.55San Blas Kuna cuk 6.16.1Epena sja 7.37.3SULKA: 1Nomaande lem 6.66.6Teribe tir 6.66.6Halth Mongolian khk 7.37.4MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka sua 9Bambara bam Bissa bib 7.7Kurukh kru 7.07.0Kalmyk xal 6.96.9Boko (Benin) bqc 7.07.0Malayalam mal 6.78.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bqp 7.16.7-Malayalam mal 6.78.7NAKH-DAGHESTAN: 2Tol L: 1Busa bqp 7.17.1Tamitan 6.66.6Avaric ava 6.06.07.1Buka dyu dyu 7.17.1Tol L: 1Byula dyu 7.17.1Tol L: 1Tol L: 1Byula dyu 7.17.1Juang jun 7.07.0Tacana tan 6.66.5TOR-ORYA: 1Chechen coney men 7.77.7Vietnamese vie 6.66.6Central Ojibwa ojc 6.10.1URARINA: 1Mandinka mnk 8.88.2Achuar-Shiwiar acu 6.26.2Paumarí pad 6.45.4CACUA-NUKAK: 1Northern Bobo Madaré bbo 8.18.1JIVAROAN: 4ARAUAN: 1				
Mbay myb Moru mgd8.4Cogui kog Si6.7Guambiano gum Guambiano gum 6.46.4PUINAVE: 1Moru mgd6.5Ngäbere gym San Blas Kuna cuk 6.66.1Epena sja Teribe tfr7.3Puinave pui SULKA: 1Nomaande lem MANDE: 130Teribe tfr6.6Northern Emberá emp MOROGLIC: 27.3SULKA: 1Bambara bam Bissa bib 7.77.7Karnada kan Kannada kan 6.66.6Halh Mongolian khk Kalmyk xal 6.97.4TARASCAN: 1Bissa bib Dyula dyu 7.17.0Malayalam mal Cogui tell8.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bap Dyula dyu Mandinka mnk 6.87.4Chechen che Col7.1TOL: 1Dyula dyu Mandinka mnk 6.87.4AUSTRO-ASIATIC: 4TACANAN: 2Tol: 1Mandinka mnk 6.87.1Castern Bru bru Col6.3Ese Ejja ese Col6.5TOR-ORYA: 1Mandinka mnk 6.87.7Vietnamese vie Col6.6Central Ojibwa ojc Cacu-ANN: 1URARINA: 1Urarina uraMandinka mnk 6.83.2Achuar-Shiwiar acu Aguarun agr 6.46.4Basque cus Cacu-ANK: 16.4VANOMAM: 1Northern Bobo Madaré bbo Alaré bbo Alaminka mnk 6.8JIVAROAN: 4ARAUAN: 1URARINA: 1Northern Bobo Madaré bbo Alaré bbo Alaguarun agr Alaguarun ag		JI		
Mark 19965Nögber gyn Ngäber gyn 6664CHOC0: 2Puinave pui Puinave puiNgambay sba7.5San Blas Kuna cuk6.1Epena sja7.3SULKA: 1Nomaande lem6.6Teribe ffr6.6Northern Emberá emp7.3SULKA: 1MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka suaSulka suaBambara bam7.5Kannada kan6.6Halh Mongolian khk7.4TARASCAN: 1Bassa bib7.7Kurukh kru7.0Kalayk xal6.9Purepecha tszBoko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bqp6.7e* Tamil tam6.6Avaric ava6.0Ticuna tcaDyula dyu7.1e* Telugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol jicKuranko knk9.2Eastern Bru bru6.3Ese Ejja ese6.5Loko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mandinka muk6.8Khmer khn8.7ALGIC: 1Orya ury1Mende (Sierra Leone) men7.7e* Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina uraSusu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URUC-CHIPAYA: 1Shuar jiv6.9CACUA-NUKAK: 1Chipaya cap <td></td> <td></td> <td></td> <td>Páez pbb 7.2</td>				Páez pbb 7.2
Ngambay sha7.5 6.6San Blas Kuna cuk6.1Epena sja7.3Puinave puiNomaande lem6.6Teribe tfr6.6Northern Emberá emp7.3SULKA: 1MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka sua7.4Bambara bam7.5Kannada kan6.6Halh Mongolian khk7.4TARASCAN: 1Bissa bib7.7Kurukh kru7.0Kalmyk xal6.9Purepecha tsz8Boko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bqp6.7CTelugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol.: 1Tol.: 1Loko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mende (Sierra Leone) men7.7C Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JUXAROAN: 4ARUUAN: 1Urarina uraVarina uraSus usu8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1Amasongaxango kao8.5Aguaruna agr6.4BASQUE: 1Chipaya capTRANS-NEW GUINEA: 1:5Shuar jiv6.9CACUA-NUKAK: 1YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuAwa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsu	• •			PUINAVE: 1
Nomaande lem MANDE: 136.6Teribe ffr Teribe ffr6.6Northern Emberá emp MONGOLIC: 27.3SULKA: 1MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka suaSulka sua	0	0		Puinave pui 5.2
MANDE: 13DRAVIDIAN: 5MONGOLIC: 2Sulka suaBambara bam7.5Kannada kan6.6Halh Mongolian khk7.4TARASCAN: 1Bissa bib7.7Kurukh kru7.0Kalmyk xal6.9Purepecha tsz8Boko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bap6.7 ℓ Tamil tam6.6Avaric ava6.0Ticuna tcaDyula dyu7.1 ℓ Telugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol.: 1Loko lok7.1Juang jun7.0Tacana tha6.6Kuranko knk9.2Eastern Bru bru6.3Ese Ejja ese6.5Monde (Sierra Leone) men7.7 ℓ Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina uraSus uss8.2Achuar-Shiwar acu6.2Paumarí pad5.4URU-CHIPAYA: 1Anano sus8.5Aguaruna agr6.4BASQUE: 1Chipaya capChipaya cap7.6YANOMAM: 1Awa (Papua New Guine a wb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuAwa (Papua New Guine a wb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuAwa (Papua New Guine a wb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuAwa (Papua New Guine a wb7.8PANOAN: 4Cacua cbv7.6 <t< td=""><td>• •</td><td></td><td>1 0</td><td>SULKA: 1</td></t<>	• •		1 0	SULKA: 1
Bambara bam7.5Kannada kan6.6Halh Mongolian khk7.4TARASCAN: 1Bissa bib7.7Kurukh kru7.0Kalmyk xal6.9Purepecha tsz7Boko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bqp6.7 ℓ Tamil tam6.6Avaric ava6.0Ticuna tcaDyula dyu7.1 ℓ Telugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol.: 1Kuranko knk9.2Eastern Bru bru6.3Eastern Bru bru6.3Eastern Bru bruMandinka mnk6.8Khmer khm8.7ALGIC: 1Orya uryMende (Sierra Leone) men7.7 ℓ Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JUVAROAN: 4ARAUAN: 1Urarina uraSasquage as6.4Basque eus6.2YANOMAM: 1Anano boj6.2Shua ji bi6.9CACUA-NUKAK: 1Cacua cbv7.6Sanumá xsuAwa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1				Sulka sua 6.8
Difficient7.7Kinnik kru7.0Kinnik kru7.0Kinnik kru7.0Purepecha tsz7.1Boko (Benin) bqc7.0Malayalam mal8.7NAKH-DAGHESTAN: 2TICUNA: 1Busa bqp6.7° Tamil tam6.6Avaric ava6.0Ticuna tcaDyula dyu7.1° Telugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol. 1Loko lok7.1Juang jun7.0Tacana tna6.6Kuranko knk9.2Eastern Bru bru6.3Ese Ejja ese6.5Mandinka mnk6.8Khmer khm8.7ALGIC: 1Orya uryMende (Sierra Leone) men7.7° Vietnamese vie6.6Central Ojibwa ojc6.1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina ura0Susu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1TRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá susEse mcq6.6Cashinahua des6.4CAHUAPANAN: 1ZAMUCOAN: 1				TARASCAN: 1
Boko (Benin) beq7.0Malayalam mal8.7NAKH-DAGHESTAN.: 2TICUNA: 1Busa bap6.7 $^{\circ}$ Tamil tam6.6Avaric ava6.0Ticuna tca1Dyula dyu7.1 $^{\circ}$ Telugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol.: 1Loko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mandinka mk6.8Khmer khm8.7ALGIC: 1Orya ury1Mende (Sierra Leone) me7.7 $^{\circ}$ Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina ura0Susu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1TRANS-NEW GUINEA: 12Huambisa hub6.0Basque cus6.2YANOMAM: 1YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1			-	
Bus bop Bus bop Dyula dyu6.7C C Tamil tam C Telugu telC Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point Point <td></td> <td></td> <td></td> <td>· · · ·</td>				· · · ·
Dyula dyu7.1C Telugu tel9.1Chechen che7.1TOL: 1Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2Tol.: 1Kuranko knk9.2Eastern Bru bru6.3Ese Ejja ese6.5TOR-ORYA: 1Loko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mandinka mnk6.8Khmer khm8.7ALGIC: 1Orya ury8Mende (Sierra Leone) me7.7C Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina ura0Susu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1Kaasongaxango kao8.5Aguaruna agu6.4BASQUE: 1Chipaya cap0TRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1				
Dyula dyu8.4AUSTRO-ASIATIC: 4TACANAN: 2101.: 1Mandinka mak9.2Eastern Bru bru6.3Ese Ejja ese6.5Tol.icLoko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mandinka mak6.8Khmer khm8.7ALGIC: 1Orya ury5.4Mende (Sierra Leone) men7.7C Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina ura0Susu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1TRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2YANOMAM: 1Anjam boj6.2Shuar jiv6.9CACUA-NUKAK: 1YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1				
Kuranko knk9.2Eastern Bru bru6.3Ese Ejja ese6.5Tot jieLoko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mandinka mnk6.8Khmer khm8.7ALGIC: 1Orya uryMende (Sierra Leone) men7.7C Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina uraUrarina uraSus usus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1TRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2Chipaya capCACUA-NUKAK: 1Chipaya capYANOMAM: 1Anjam boj6.2Shuar jiv6.9CACUA-NUKAK: 1Cacua cbv7.6Sanumá xsuAwa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1	• •			TOL: 1
Loko lok7.1Juang jun7.0Tacana tna6.6TOR-ORYA: 1Mandinka mnk6.8Khmer khm8.7ALGIC: 1Orya ury7Mende (Sierra Leone) men7.7C Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina uraSus usv8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1Xaasongaxango kao8.5Aguaruna agr6.4BASQUE: 1Chipaya capChipaya capTRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2YANOMAM: 1Anjam boj6.2Shuar jiv6.9CACUA-NUKAK: 1YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1	• •			Tol jic 7.2
Mandinka mmk6.8Khmer khm8.7ALGIC: 1Orya ury9Mende (Sierra Leone) men7.7C Vietnamese vie6.6Central Ojibwa ojc6.1URARINA: 1Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina uraSusu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1Xaasongaxango kao8.5Aguaruna agr6.4BASQUE: 1Chipaya capTRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2YANOMAM: 1Anjam boj6.2Shuar jiv6.9CACUA-NUKAK: 1YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1			**	TOR-ORYA: 1
Manunka min Manu				Orya ury 5.7
Northern Bobo Madaré bbo8.1JIVAROAN: 4ARAUAN: 1Urarina uraSusu sus8.2Achuar-Shiwiar acu6.2Paumarí pad5.4URU-CHIPAYA: 1Xaasongaxango kao8.5Aguaruna agr6.4BASQUE: 1Chipaya capTRANS-NEW GUINEA: 12Huambisa hub6.0Basque eus6.2YANOMAM: 1Anjam boj6.2Shuar jiv6.9CACUA-NUKAK: 1YANOMAM: 1Awa (Papua New Guinea awb7.8PANOAN: 4Cacua cbv7.6Sanumá xsuEse mcq6.6Cashinahua cbs6.4CAHUAPANAN: 1ZAMUCOAN: 1				
Susu sus 8.2 Achuar-Shiwiar acu 6.2 Paumarí pad 5.4 URU-CHIPAYA: 1 Xaasongaxango kao 8.5 Aguaruna agr 6.4 BASQUE: 1 Chipaya cap Chipaya cap TRANS-NEW GUINEA: 12 Huambisa hub 6.0 Basque cus 6.2 YANOMAM: 1 Anjam boj 6.2 Shuar jiv 6.9 CACUA-NUKAK: 1 YANOMAM: 1 Awa (Papua New Guinea awb 7.8 PANOAN: 4 Cacua cbv 7.6 Sanumá xsu Ese mcq 6.6 Cashinahua cbs 6.4 CAHUAPANAN: 1 ZAMUCOAN: 1	· · · ·			
Xaasongaxango kao 8.5 Aguaruna agr Huambisa hub 6.4 BASQUE: 1 Othe Chin Arman TRANS-NEW GUINEA: 12 Huambisa hub 6.0 Basque cus 6.2 Chipaya cap Chipaya cap Anjam boj Awa (Papua New Guinea awb 6.2 Shuar jiv 6.9 CACUA-NUKAK: 1 YANOMAM: 1 Ese mcq 6.6 Cashinahua cbs 6.4 CAHUAPANAN: 1 ZAMUCOAN: 1				
TRANS-NEW GUINEA: 12 Huambisa hub 6.0 Basque eus 6.2 Chipaya cap Anjam boj 6.2 Shuar jiv 6.9 CACUA-NUKAK: 1 YANOMAM: 1 Awa (Papua New Guinea awb 7.8 PANOAN: 4 Cacua cbv 7.6 Sanumá xsu Ese mcq 6.6 Cashinahua cbs 6.4 CAHUAPANAN: 1 ZAMUCOAN: 1				
Anjam boj Awa (Papua New Guinea awb 6.2 Shuar jiv 6.9 CACUA-NUKAK: 1 YANOMAM: 1 Awa (Papua New Guinea awb 7.8 PANOAN: 4 Cacua cbv 7.6 Sanumá xsu 3 Ese mcq 6.6 Cashinahua cbs 6.4 CAHUAPANAN: 1 ZAMUCOAN: 1				
Awa (Papua New Guinea awb 7.8 PANOAN: 4 Cacua cbv 7.6 Sanumá xsu Ese mcq 6.6 Cashinahua cbs 6.4 CAHUAPANAN: 1 ZAMUCOAN: 1				
Ese mcq 6.6 Cashinahua cbs 6.4 CAHUAPANAN: 1 ZAMUCOAN: 1	• •			Sanumá xsu 6.6
	· •			ZAMUCOAN: 1
Gwahatike dah 7.0 Panoan Katukína knt Chayahuita cbt 8.0 Chamacoco ceg				Chamacoco ceg 6.5