**1) Sindhi (Laari-aware) phone inventory**

Practical, field-tested set balancing linguistic fidelity with ASR/aligner needs. Add/remove once you observe speaker-specific realizations.

**1.1 Consonants (IPA → Kaldi-style symbol)**

**Plosives (voiceless / aspirated / voiced / breathy-voiced):**

* Bilabial: p → P, pʰ → PH, b → B, bʱ → BH
* Dental: t̪ → T\_D, t̪ʰ → TH\_D, d̪ → D\_D, d̪ʱ → DH\_D
* Retroflex: ʈ → T\_R, ʈʰ → TH\_R, ɖ → D\_R, ɖʱ → DH\_R
* Palatal(=alveolo-palatal stop/affricate base): c → C, cʰ → CH, ɟ → J, ɟʱ → JH
* Velar: k → K, kʰ → KH, ɡ → G, ɡʱ → GH

**Implosives (characteristic to Sindhi):**

* ɓ → IM\_B, ɗ → IM\_D, ʄ → IM\_J, ɠ → IM\_G

**Affricates:**

* t͡ʃ → CH\_F, t͡ʃʰ → CHH\_F, d͡ʒ → JH\_F, d͡ʒʱ → JHH\_F

**Fricatives:**

* f → F, s → S, z → Z, ʃ → SH, x → KHX, ɣ → GX, h → H

**Nasals:**

* m → M, n̪ → N\_D, ɳ → N\_R, ɲ → N\_P, ŋ → NG

**Liquids / Rhotics / Laterals:**

* r/ɾ (tap) → R, ɽ (retroflex flap) → RD, l → L, ɭ → L\_R

**Approximants / Glides:**

* w → W, j → Y

**Laari-salient realization flags (for annotation, not separate phones unless you decide to):**

* RD→R (retroflex flap realized as alveolar tap)
* **Aspiration weakening**: PH→P, TH\_\*→T\_\*, KH→K in fast speech
* **Cluster simplification**: dr/tr → d/t (mark at lexicon-variant level)

**1.2 Vowels (short/long; add nasalized counterparts with “~” if you model them)**

| **IPA** | **Sym** | **IPA** | **Sym** |
| --- | --- | --- | --- |
| i | I | iː | II |
| e | E | eː | EE |
| ə | AX | a | A |
| aː | AA | æ (contextual) | AE |
| o | O | oː | OO |
| u | U | uː | UU |

**Notes:**

* Many corpora collapse e~eː and o~oː depending on evidence; keep both for now.
* Add nasalized (I~, A~, etc.) only if consistently present and you can annotate them.

**1.3 phones.txt (starter)**

<eps>

SIL

SPN

P PH B BH T\_D TH\_D D\_D DH\_D T\_R TH\_R D\_R DH\_R C CH J JH K KH G GH

IM\_B IM\_D IM\_J IM\_G

CH\_F CHH\_F JH\_F JHH\_F

F S Z SH KHX GX H

M N\_D N\_R N\_P NG

R RD L L\_R W Y

I II E EE AX A AA AE O OO U UU

**2) Minimal lexicon skeleton**

Provide **canonical** and **Laari-variant** entries where relevant. Keep everything **space-separated Kaldi style**.

**2.1 lexicon.txt (canonical)**

<eps> SIL

ؙ, SPN

سنڌ S AX N\_D D\_D

سنڌو S AX N\_D D\_D U

پاڻي P AA N\_R II

حق H AX Q

زبان Z AX B AA N\_D

ماڻهو M AA N\_R H O

زمين Z AX M II N\_D

ثقافت S AX Q AA F AX T\_D

انسان AX N\_D S AA N\_D

دريا D\_D AX R Y AA

بااختيار B AA AX KH T\_D I AA R

*(Use U+06A9/06AF etc. per your script; the above is illustrative.)*

**2.2 lexicon\_laari.txt (observed variants)**

سنڌ S AX N\_D D\_D

سنڌو S AX N\_D D\_D U

پاڻي P AA N\_R II

حق H AX Q

زبان Z AX B AA N\_D

ماهو M AA H O ; ماهو ~ ماڻهو (cluster simplification / vowel reduction)

دريا D\_D AX R Y AA

درياھ D\_D AX R Y AA H ; final /h/ in some tokens

اختيار AX KH T\_D I AA R

باختيار B AA KH T\_D I AA R ; aspiration weakening bʱ→b; normalize as needed

Keep a **notes/lexicon\_decisions.md** logging why each variant exists (evidence: timestamped examples).

**3) ToBI-LR (Sindhi/Laari) quick reference**

**Tiers you’ll label:**

* words, phones, breaks (0,1,2,3,4), tones (pitch accents & boundaries), disc (discourse pragmatics)

**3.1 Break indices**

* **0** clitic/strong sandhi (no audible break)
* **1** word boundary (default)
* **2** minor phrase (phonological phrase)
* **3** intermediate phrase (ip; often pre-final hold)
* **4** intonation phrase (IP; clear cadence + pause)

**3.2 Pitch accents (pick a compact, reliable set)**

* **H\***: prominent high target on stressed syllable
* **L\***: prominent low (emphatic gravity)
* **L+H\***: rising accent (common on contrast/new info)
* **H+L\***: falling complex (firm assertions/slogans)

**3.3 Phrase tones & boundary tones**

* **ip tones**: L-, H-
* **IP tones**: L-L%, L-H%, H-L%, H-H%

**3.4 Discourse tags (in disc tier)**

CALL-RESP, SLOGAN, APPL-COVERED (applause masked), CHANT-UNDER, CODE:URDU/EN, LAARI-INDEX (salient dialect token)

**3.5 Ten common label patterns (how to use)**

1. **Triadic list** (water–land–rights): H\* ... H\* ... L+H\* L-H% (penultimate rise, final continuation)
2. **Slogan close**: H+L\* H-L% (decisive fall) + break 4
3. **Appeal to identity**: L+H\* ... L-L% (warm but final)
4. **Contrast pair** (“they say… we answer…”): first clause H\* L-H%, second H+L\* H-L%
5. **Questioning challenge** (rhetorical): L+H\* L-H%
6. **Definition cadence**: H\* H- L-L% (mid phrase hold then full stop)
7. **Enumerative parallelism**: each item H\* with break 2/3, final L+H\* L-L%
8. **Moral coda**: L\* ... L-L% (grave terminal)
9. **Call-and-response cue**: pre-pause L+H\* H- + break 3, then reset
10. **Climactic rise** (crowd swell): L+H\* H-H% (only if truly non-final continuation)

**Reliability tip:** Label **fewer, clearer accents** rather than over-label; consistency over granularity.

**4) Dataset templates**

**4.1 metadata/segments.csv (header + sample rows)**

segment\_id,source\_id,t\_start\_s,t\_end\_s,sr\_hz,channel,snr\_db,lufs,diar\_spk,denoise\_profile,notes

RBP\_JSH\_1979\_0001,JSH\_1979\_raw,12.540,19.880,48000,mono,14.2,-23.1,spk0,wpe\_v1+mmse\_v1,opening definition (clean)

RBP\_TV\_1986\_0007,PTV\_1986\_raw,305.120,318.900,48000,mono,21.7,-23.0,spk0,wpe\_v1+mmse\_v1,studio; code-switch EN

RBP\_RALLY\_1994\_0210,SUJ\_1994\_raw,1260.000,1271.300,48000,mono,9.8,-23.5,spk0,wpe\_v1+rnnoise\_v2,applause tail (tag APPL-COVERED)

**4.2 Praat TextGrid (minimal textual form)**

File type = "ooTextFile"

Object class = "TextGrid"

xmin = 0

xmax = 7.34

tiers? <exists>

size = 4

item []:

item [1]:

class = "IntervalTier"

name = "words"

xmin = 0

xmax = 7.34

intervals: size = 3

intervals [1]: xmin = 0; xmax = 2.10; text = "پاڻي"

intervals [2]: xmin = 2.10; xmax = 5.60; text = "زمين"

intervals [3]: xmin = 5.60; xmax = 7.34; text = "حق"

item [2]:

class = "IntervalTier"

name = "phones"

xmin = 0

xmax = 7.34

intervals: size = 6

intervals [1]: xmin=0; xmax=0.55; text="P"

intervals [2]: xmin=0.55;xmax=1.10; text="AA"

intervals [3]: xmin=1.10;xmax=2.10; text="N\_R I I"

intervals [4]: xmin=2.10;xmax=3.40; text="Z AX"

intervals [5]: xmin=3.40;xmax=5.60; text="M II N\_D"

intervals [6]: xmin=5.60;xmax=7.34; text="H AX Q"

item [3]:

class = "IntervalTier"

name = "breaks"

xmin = 0

xmax = 7.34

intervals: size = 4

intervals [1]: xmin=0; xmax=2.10; text="1"

intervals [2]: xmin=2.10;xmax=5.60; text="3"

intervals [3]: xmin=5.60;xmax=7.34; text="4"

intervals [4]: xmin=7.34;xmax=7.34; text=""

item [4]:

class = "TextTier"

name = "tones"

xmin = 0

xmax = 7.34

points: size = 4

points [1]: number=0.30; mark="L+H\*"

points [2]: number=2.05; mark="H-"

points [3]: number=5.55; mark="H+L\*"

points [4]: number=7.30; mark="H-L%"

*(Add a fifth tier disc if you want discourse tags as point events.)*

**5) Feature-extraction configs / commands**

**5.1 Pitch (F0) with REAPER (robust, open source)**

* **Command (per file):**  
  reaper -i input.wav -f f0.txt -p pitchmarks.txt -a -u 50 -l 500 -s 0.01
  + -u/-l = F0 bounds (adjust to speaker; start 50–500 Hz)
  + -s 0.01 = 10 ms hop
* **Post:** Convert f0.txt to CSV with time\_s, f0\_hz, voiced\_flag.

**5.2 openSMILE (industry standard)**

**eGeMAPS (paralinguistics low-dim):**

SMILExtract -C config/gemaps/eGeMAPSv02.conf \

-I segments/RBP\_JSH\_1979\_0001.wav \

-O features/egemaps/RBP\_JSH\_1979\_0001.csv

**ComParE (large feature set):**

SMILExtract -C config/ComParE\_2016.conf \

-I segments/RBP\_JSH\_1979\_0001.wav \

-O features/compare/RBP\_JSH\_1979\_0001.csv

*(Lock versions; record exact config paths.)*

**5.3 Torchaudio (Log-Mel & MFCC; Python sketch)**

import torch, torchaudio as ta

wav, sr = ta.load("segments/RBP\_JSH\_1979\_0001.wav") # mono, float32 [-1,1]

target\_sr = 16000

if sr != target\_sr:

wav = ta.functional.resample(wav, sr, target\_sr)

sr = target\_sr

# Log-mel (80 bins)

mel = ta.transforms.MelSpectrogram(

sample\_rate=sr, n\_fft=1024, hop\_length=160, win\_length=400, # 25 ms win / 10 ms hop

f\_min=20.0, f\_max=7600.0, n\_mels=80, power=2.0

)(wav)

logmel = torch.log(mel + 1e-6) # [1, 80, T]

# MFCC (40 coeffs) + deltas

mfcc = ta.transforms.MFCC(

sample\_rate=sr, n\_mfcc=40,

melkwargs={"n\_fft":1024, "hop\_length":160, "win\_length":400,

"f\_min":20.0, "f\_max":7600.0, "n\_mels":80}

)(wav) # [1, 40, T]

delta = ta.functional.compute\_deltas(mfcc)

delta2 = ta.functional.compute\_deltas(delta)

torch.save({"logmel": logmel, "mfcc": mfcc, "d1": delta, "d2": delta2},

"features/llf/RBP\_JSH\_1979\_0001.pt")

**5.4 Praat/Parselmouth (formants, CPP, jitter/shimmer)**

* **Formants:** LPC order start at 5, max formant 5 kHz (male) / 5.5 kHz (adjust after pilot), window 25 ms, pre-emphasis 50 Hz.
* **CPP (Cepstral Peak Prominence):** 60 dB range, 10–500 Hz cepstral domain.

Keep a features/docs/measurements.md describing parameters and justifications.

**6) Pilot “gold” annotations (recommendation)**

* Select **30 minutes** across contexts; hand-label **phones/words/breaks/tones** with two annotators.
* Compute **Cohen’s κ** on breaks and tones tiers; revise the guide until κ ≥ 0.70.

**7) Checklist to start tomorrow**

* Create folders from Section 4.
* Drop 10–20 cleaned segments in segments/.
* Populate phones.txt, lexicon.txt, lexicon\_laari.txt.
* Export 3–5 TextGrids using the template above.
* Run REAPER + openSMILE + torchaudio on those files.
* Write docs/orthography\_guidelines.md (diacritics & punctuation policy).
* Begin notes/lexicon\_decisions.md with your first 50 variants.