Consonant Codes in 94 Austroasiatic Languages

Twelve balanced branches: the family that separates word length from sample size

Alejandro Toledo Martínez · ORCID 0009-0000-1277-9697 · Integrative Corpus · 2026

Abstract

We present a catalogue of 2,135 consonant codes grouping 14,851 lexical forms from 94 Austroasiatic languages, built with no etymology, no proto-forms and no cognacy judgements.

It is the first family in the series whose widest-reaching codes are not a single consonant. `m·t` for *eye* appears in 64 languages and eleven of the twelve branches; `m·h` for *nose*, in 58 and ten. They are Austroasiatic *mat* and *muh*, the two comparanda this family's comparison has been citing for a century, and the procedure did not look for them: they emerged from requiring a skeleton to repeat for the same concept across three languages.

And this family falsifies a finding of the series. The Sino-Tibetan paper registered, and confirmed out of sample, that the grouping null explodes below 2.8 consonants per word. Austroasiatic has 2.75 consonants and a null of 14.6%; Austronesian has 2.78 and a null of 39.0%. Same word length, twenty-four points apart. What separates them is the languages: 94 against 582.

Across the nine families measured, r(arity, null) = −0.84 and r(log languages, null) = +0.70, and the two variables are largely independent of one another. The null is a function of word length AND sample size, and the two had been conflated because until now every short-word family was also a large one.

We also document a bug found before the catalogue was built — the fourth time a non-consonant entered the skeleton through a source convention, and the first caught in time.

1. What is sought, and why this family

The procedure is described once for the whole series in "The catalogue of consonantal codes" and is not repeated. The essentials: a code is a consonant skeleton appearing for the same concept in three or more distinct languages, and it is not claimed that its members are cognate, that they descend from a proto-form, or that the code means anything.

This family was in the plan from the start and was skipped twice. It ran third — "twelve balanced branches and tone in 30% of the forms: the exact midpoint of the axis" — and was where the tone question was to be decided "in a family where the rest of the structure is favourable". Tone ended up being decided on the fly in Sino-Tibetan, which is precisely what the plan wanted to avoid.

It arrives ninth for a different reason, and a better one. With eight families done, the criterion for choosing the next stopped being "balanced branches" and became balanced branches × long words: in short-word families the branch crossing goes to the unary codes, where the catalogue cannot distinguish a genuine monoconsonantal etymon from a coincidence of availability. Austroasiatic is the only pending family that had both.

2. The population

94 languages · 49,783 forms · 1,570 concepts · 12 branches.

Branch	Languages	Forms
Bahnaric	23	6,291
Aslian	15	5,694
Vietic	13	7,221
Khasi-Palaung	13	13,469
Mundaic	8	753
Khmuic	6	5,424
Katuic · Pearic	4 + 4	5,564
Mangic	3	3,454
Nicobaric · Khmeric · Monic	2+2+2	1,913

100% of the forms obtain a skeleton, and no language is excluded. The largest branch is 24% — the best balance in the series after Pama-Nyungan's 18% — and nine of the twelve branches have three or more languages.

A single macroarea: the anomaly list of §6.1-bis does not exist here, and that is said before the results.

2.1 No language is excluded, and the three proposals are rejected

Two pairs of content duplicates, both below the rule. `KNB (a Pearic variety)` and `Surin Khmer` coincide on 65% of 936 concepts; `Kensiw Kedah` and `Kintaq`, on 61% of 145. The published rule requires 80% and their glottocodes differ. They are neighbours within the same branch, which is what one expects to happen.

`Korku` sits 2,845 km from the centre, 4.9 times the median. It is Mundaic, spoken in Madhya Pradesh, and the family stretches from India to Vietnam. It is the fifth time this check has asked for too much, and the corrected rule applies: distance alone does not propose, it accompanies.

Fifteen languages mark tone in digits where the rest do not, nearly all Vietic. That is a property of the source and not of the language, and the skeleton ignores tone in any case.

2.2 A bug found BEFORE the catalogue was built

Four Pearic languages — Chong, Samre, Kasong, KNB — plus Wa and seven others write register tone as a letter after a period: `pi?.A`, `klaŋ.B`, `krih.C`. It is the Diffloth–Ferlus convention, and it is tone: not a segment.

`A` is a vowel and dropped out harmlessly. **`B`**, **`C`** and **`D`** are consonants and entered the skeleton:

```
Wa  klaŋ.A_ć ɛ.B          →  k·l·n·s·b       ← that final "b" was tone B
Wa  pau?.B_ʔaik.A_taŋ.A   →  p·h·b·h·k·t·n   ← the "b" in third position was tone B
```

12,974 forms in 12 languages carry the mark; 5,866 had the letter as a segment — in the rest the source had already converted it to a superscript — and 3,235 ended up with an invented consonant inside the skeleton.

It is the fourth time a non-consonant has entered the skeleton through a source convention, after Danish stød, Swedish `ɵ` read as `θ`, and prenasalised digraphs. The previous three were found *after* publishing the affected family. This one was found by phase 0 before the catalogue was built, which is exactly what that phase exists for, and the first time the order came out right.

The fix is conservative, like the other three: it aligns the words of the spelling with the segment groups and removes a group's last segment only when its word ends in `.X` and that segment is exactly that letter in lower case. A genuine final `b` is not lost. Checked over all 5,866 forms: zero discrepancies between the stored skeleton and the recomputed one.

3. What the instrument costs here

Mean arity: 2.75 consonants per form. Collision: 56.2%. More than half the forms share a skeleton with another concept inside their own language.

Both figures place it on the expensive side of the axis, between Austronesian (2.78) and Trans-New-Guinea (2.41). What §5 measures is that the story does not end there.

Tone does not enter the skeleton, and thirty languages mark it one way or another. It is declared as a dimension the skeleton does not capture, like the vowels.

Five concepts are duplicated by case in the source list — `HAIR/hair`, `KNOW/know`, `SMELL/smell`, `THINK/think`, `HIDE/hide` — so the same skeleton can count twice. It is the list collision the Trans-New-Guinea paper measures over the forms, here in its crudest form: the same gloss written two ways.

4. The catalogue

2,135 codes over 1,570 concepts, grouping 14,851 of 49,783 forms (29.8%). Of these, 1,695 are binary or ternary and 309 are unary: 14.5%.

Arity	Codes	%
1	309	14.5%
2	1,243	58.2%
3	452	21.2%
4	114	5.3%
5+	17	0.8%

14.5% unary is the second lowest in the series, after Pama-Nyungan's 3.9% and far from Sino-Tibetan's 35.9%. And 58.2% binary is the highest: this family concentrates its catalogue at exactly the arity the instrument handles best.

Precision: 6.63% against 0.144% chance = 46× [45–46× across seeds].

And on grouping without the three-language threshold: 41.7% observed against 14.6% null — an advantage of 27.1 points, 2.9× chance, $\kappa = 0.317$.

4.1 The widest-reaching codes are binary, and cross eleven branches

The fifteen of widest reach, unfiltered:

Code	Concept	Languages	Arity	Branches	Enrich.
m·t	*eye*	64	2	11	147×
m·h	*nose*	58	2	10	100×
k	*fish*	44	1	8	42×
t	*hand*	44	1	11	49×
s·m	*bird*	42	2	8	93×
s·k	*hair* (HAIR)	40	2	8	82×
c·n	*foot*	37	2	8	61×
s·n	*foot*	37	2	9	33×
k·p	*bite*	36	2	5	83×
s	*dog*	36	1	8	32×
p·r	*fly*	35	2	9	143×
m	*thou*	32	1	8	36×
t·n	*weave*	32	2	8	78×
m·j	*one*	31	2	7	153×
t·h	*breast*	30	2	7	45×

Eleven of the fifteen are binary, and so are the first two. It is the first time in the series: in Sino-Tibetan the top twelve were unary, in Trans-New-Guinea fourteen of fifteen, and in Austronesian and Pama-Nyungan the ceiling was also taken by single-consonant codes.

m·t|eye is Austroasiatic *mat* and **m·h|nose** is *muh*, the two textbook comparanda. Their distributions:

Code	Concept	Languages	Branches	The five largest
m·t	eye	64	11	Bahnaric 21 · Aslian 15 · Vietic 13 · Khmuic 4 · Pearic 3
m·h	nose	58	10	Bahnaric 23 · Aslian 14 · Khmuic 5 · Khasi-Palaung 4 · Katuic 4

They are not concentrated in one branch with a tail: they cover the family. The procedure consulted no reconstruction.

5. The finding: the null is not a function of arity alone

The Sino-Tibetan paper published, over seven families and confirmed out of sample on the eighth, that the grouping null explodes below three consonants per word. The prediction was registered before computing and it held: arity 2.22, null 42.0%.

Austroasiatic breaks it.

Family	Arity	Languages	Null	κ
Sino-Tibetan	2.22	267	42.0%	0.340
Trans-New-Guinea	2.41	196	37.4%	0.232
Austroasiatic	2.75	94	14.6%	0.317
Austronesian	2.78	582	39.0%	0.380
Turkic	2.85	32	13.7%	0.498
Afro-Asiatic	3.16	111	4.7%	0.262
Uralic	3.27	27	5.0%	0.301
Indo-European	3.35	207	11.8%	0.364
Pama-Nyungan	3.35	154	11.3%	0.310

Read the third and fourth rows together. Austroasiatic and Austronesian have practically the same word length — 2.75 and 2.78 — and their nulls differ by twenty-four points. What separates them is the number of languages: 94 against 582.

And it is not an isolated case. Across the nine:

	r
arity ↔ null	−0.84
log languages ↔ null	+0.70
arity ↔ log languages	−0.34

Both variables predict the null and are largely independent of one another. The correlation between them is weak, so they are not measuring the same thing twice.

5.1 Why, and why it was hidden

The mechanism is no mystery. The null asks how often any two languages share a skeleton for the same concept under shuffling. With 582 languages there are 169,071 pairs where that can happen; with 94, 4,371. More languages means more chances to coincide by accident, exactly as more consonants means fewer.

What kept it hidden is an accident of the working order: the three short-word families the series had done were also the large ones — Sino-Tibetan 267, Trans-New-Guinea 196, Austronesian 582 — while the long-word ones included the two smallest — Turkic 32, Uralic 27. Arity and size came correlated in the sample without being correlated in the world.

Austroasiatic is the first short-word, small-sample family, and that is why it separates the two.

And it is not the number of concepts. That is the obvious objection — a corpus with more concepts spreads the forms over more buckets and lowers the coincidence rate under shuffling — and it does not hold: r(log concepts, null) = −0.09 across the nine. The deciding case is the same one: Austroasiatic and Austronesian have 1,570 and 1,823 concepts, a 16% difference, and nulls of 14.6% and 39.0%.

It is worth noting because it was not predictable: the chance rate for *precision* does depend strongly on the number of concepts — the Indo-European paper publishes this — and one would expect the grouping null to behave the same way. They are two nulls of two different questions, and they answer to different variables.

5.2 What survives of the earlier finding

Almost all of it, with an added condition. Arity remains the stronger predictor (−0.84 against +0.70), and the step below 2.8 remains real *at equal size*. What does not hold is reading it as a function of a single variable, and in particular:

A raw grouping figure cannot be compared across families without its null beside it, and the null cannot be estimated from arity alone. The Trans-New-Guinea paper stated the first half of that rule; this is the second.

And κ follows neither. Austroasiatic gives 0.317 with a low null; Sino-Tibetan, 0.340 with a very high one. Cost and performance are different things, which is what §5 of the Sino-Tibetan paper had already separated.

6. What the family gives

6.1 The branch crossing, and for once without the unary codes taking it

641 codes cross three or more branches, and 233 cross four. Decomposed by arity, which is the reading the Sino-Tibetan paper forced into publication:

Arity	Codes	Cross ≥3 branches	Rate
1 (unary)	309	143	46.3%
2	1,243	413	33.2%
3	452	65	14.4%
4+	131	20	15.3%

Only 22% of those that cross are unary, against 58% in Sino-Tibetan and 64% in Trans-New-Guinea. The defensible figure is 478 — the binary and ternary ones — and it is the second best in the series after Indo-European's 1,237, with less than half its languages.

The rate still falls with arity, as in every family: a short code crosses more easily. But here the fall is from 46% to 33%, not from 28% to 3%, because there are real branches to cross.

6.2 Substitution pairs: 631 disjoint candidates

Of 754 pairs of codes for the same concept differing in exactly one position, 631 are disjoint in languages and only those can be proposed as an isogloss. The 16% that overlap is the second cleanest proportion in the series, after Pama-Nyungan's 5%.

The two codes for 'foot' — `c·n` in 37 languages and eight branches, `s·n` in 37 and nine — are the best example of what this section produces: same concept, same second consonant, and the first one splits the family.

6.3 Four emergent classes, and it is the third family of nine to produce any

Pair	Times (discover)	Ratio	Times (check)	Ratio
b ↔ ɓ	183	45.7×	91	21.1×
c ↔ s	3,345	19.5×	3,115	16.8×
p ↔ ɓ	106	9.1×	172	13.8×
s ↔ z	42	4.8×	87	9.2×
d ↔ t	931	5.3×	1,549	8.1×
b ↔ p	657	5.4×	828	6.3×
c ↔ j	299	2.8×	647	5.7×

The closure at ≥2× gives four clean classes:

```
CE1  { c  d  h  j  s  t  z }      CE2  { b  p  ɓ }
CE3  { l  r  x }                  CE4  { g  k }
```

Fifteen segments in four classes, and all four are recognisable: coronals, labials, liquids and velars. Only Austronesian and Pama-Nyungan had produced clean classes; the other six families of the series publish none, because the closure puts nearly their whole inventory into one.

b ↔ **ɓ** at 21× and **p** ↔ **ɓ** at 13.8× are the bilabial implosive against its two pulmonic counterparts — a real alternation of this linguistic area, which the procedure finds without knowing what an implosive is. They are not regular correspondences: environment and direction are missing. They are candidates, with the place to look.

6.4 Geography, with an inversion that must be declared

Codes shared	Pairs	Mean distance	vs chance
1	1,485	1,243 km	1.07
2	486	1,475 km	1.26
3–5	1,135	1,135 km	0.97
6–10	790	887 km	0.76
11–20	317	581 km	0.50
21+	416	288 km	0.25

It falls to 0.25 of chance, the second best in the series after Trans-New-Guinea's 0.09. But the second bucket runs backwards — 1.26, further than chance — and that is published rather than smoothed away: with only 486 pairs it is the smallest bucket in the table, and sharing exactly two codes is barely distinguishable from sharing one. Monotonicity resumes from the third and is clean thereafter.

7. Limits

A single macroarea: §6.1-bis does not exist.

Tone does not enter the skeleton, and thirty languages mark it. In those where it is contrastive, the skeleton sees less than the language distinguishes.

14.5% of the catalogue is a single consonant, and the catalogue cannot distinguish the genuine monoconsonantal etymon from coincidence. It is the second lowest proportion in the series and it is still one in seven: the unary codes are counted and not read.

Bahnaric is 24% of the languages. It is the second best balance in the series and it is still a branch that weighs more than the others — and it weighs especially at the top of §4.1, where it leads the first seven codes.

Five concepts are duplicated by case in the source list, so five skeletons may count twice.

Automatic segmentation is not used, as in the eight before. Here it costs $\Delta\kappa = -0.005$, the second smallest drop in the series — and it is worth saying that this happens in a family famous for its infixation, where one might expect an affix discoverer to have more to gain. It does not gain it.

8. Conclusion

The series plan placed this family third and skipped it twice. It arrives ninth, and it arrives better than expected for a reason the plan itself had to correct beforehand: what makes a family productive is not that it has balanced branches, but that it has them AND has long enough words. Austroasiatic is the only pending family that had both, and it is the first in the series whose reach ceiling is not taken by unary codes: `m·t` for *eye* across eleven branches of twelve, `m·h` for *nose* across ten, with no reconstruction consulted.

And the methodological result comes from breaking something. The grouping null had been read as a property of word length, with a registered and confirmed prediction sustaining it. This family splits it

in two: with the same arity as Austronesian and a sixth of its languages, its null is twenty-four points lower. Word length and sample size are two axes, not one, and they had been conflated because in the eight preceding families they travelled together by accident of the working order.

Neither would have been visible without doing the ninth. And the harder one to admit is the second: the finding this series registered in advance, confirmed out of sample and published as its most solid regularity, was half a regularity.

References

Diffloth, G. (2005). The contribution of linguistic palaeontology to the homeland of Austroasiatic. In *The Peopling of East Asia*. RoutledgeCurzon. — [register tone written as a letter, §2.2]

Ferlus, M. (1979). Formation des registres et mutations consonantiques dans les langues môn-khmer. *Mon-Khmer Studies* 8.

Hammarström, H., Forkel, R., Haspelmath, M. and Bank, S. *Glottolog*. Max Planck Institute for Evolutionary Anthropology.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification* 2(1). — [the chance-correction this paper uses as κ]

Jäger, G. (2018). Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Data* 5.

List, J.-M. (2014). *Sequence Comparison in Historical Linguistics*. Düsseldorf University Press.

List, J.-M., Forkel, R., Greenhill, S. J. *et al.* (2022). Lexibank, a public repository of standardized wordlists. *Scientific Data* 9.

List, J.-M., Cysouw, M. and Forkel, R. (2016). Concepticon: a resource for the linking of concept lists. *LREC*.

Shorto, H. (2006). *A Mon-Khmer Comparative Dictionary*. Pacific Linguistics. — [the comparanda of §4.1, which this work did not consult]

Sidwell, P. and Blench, R. (2011). The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. In *Dynamics of Human Diversity*. Pacific Linguistics.

Toledo Martínez, A. (2026). *The catalogue of consonantal codes: a comparative method without genealogy, cognacy or reconstruction*. — [the method of the series, described once]

Toledo Martínez, A. (2026). *Consonant Codes in 267 Sino-Tibetan Languages*. — [the arity-step finding this paper splits in two]

Toledo Martínez, A. (2026). *Integrative Corpus: a multi-layer lexical database of 225 macrosystems*. — [the data] · https://doi.org/10.13140/RG.2.2.31244.68486

Data and reproducibility

Catalogue	`familias/austroasiatic/output/catalogo-austroasiatic.tsv` — 2,135 rows
Per-form detail	`…-lenguas.tsv` — 13,092 rows
Hygiene and decisions (§2.1)	`pipeline/fases/fase0_higiene.py` → `runs/fase0-decisiones.md`

Tone-letter bug (§2.2)	`corpus_integrativo/ingest/recompute_tono_letra.py` · backup `skeleton_bak_tonoletra_20260909`
Normalisation table (§3)	`pipeline/fases/tabla_normalizacion.py` → `runs/normalizacion.md`
Catalogue (§4)	`pipeline/fases/catalogo_familia.py`
Grouping and nulls (§4, §5)	`pipeline/fases/agrupamiento_familia.py` → `runs/agrupamiento.md`
The nine-family table (§5)	`pipeline/tools/tabla_serie.py` — derived, not written by hand
Automatic segmentation (§7)	`pipeline/fases/fase2_despiece.py` → `runs/fase2-despiece.md`
Branch crossing (§6.1)	`pipeline/fases/fase6_por_rama.py` → `runs/fase6-por-rama.md`
Sibling codes (§6.2)	`pipeline/fases/codigos_hermanos.py` → `runs/codigos-hermanos.md`
Emergent classes (§6.3)	`pipeline/fases/fase6_clases_emergentes.py` → `runs/fase6-clases.md`
Geography (§6.4)	`pipeline/fases/codigos_vs_parentesco.py` → `runs/codigos-parentesco.md`

Every figure comes from those scripts over the same population declared in §2, with the same exclusions — which here are none. None was written by hand.

Appendix A · One hundred codes, with commentary

The catalogue holds 2,135 codes and the complete data are in the family's `output/` directory. What follows is a readable sample: the hundred binary and ternary codes of widest reach, each with up to sixteen languages and their words.

Three things about the presentation. Languages are spread along the alphabetical list, not taken from its head. Long names are trimmed; the forms are not altered. And the thematic grouping is a reading order declared by hand, not a semantic classification.

This appendix does carry an emergent-code column, for the first time since Pama-Nyungan. The phase-6 closure gives four clean classes here — `{c d h j s t z}`, `{b p ɓ}`, `{l r x}`, `{g k}` — and this family is the third of nine to produce any. In the other six the closure puts nearly the whole inventory into a single class and the appendix is published without the column.

The column says what it says and nothing more: two codes with the same digit differ in segments that phase 6 found interchangeable *in this family*. They are not regular correspondences — environment and direction are missing — and a class's value is relative to the family that produced it.

What this appendix selects, measured

Sorting by reach is not neutral. Measured with `pipeline/tools/audita_anexo.py`:

	catalogue	this appendix
binary codes	58.2%	84%
ternary codes	21.2%	16%
Bahnaric as dominant branch	21.3%	58%
Khasi-Palaung as dominant branch	24.4%	11%
Pearic as dominant branch	16.6%	0%
crossing four branches or more	10.9%	60%
mean enrichment	200×	131×

It is the least deformed appendix in the series on the arity side: 84/16 against the catalogue's 58/21, where Sino-Tibetan gave 97/3 against 47/13. The reason is the family's arity — at 2.75 consonants on average there are plenty of ternaries reaching ten languages or more.

But the branch bias is the strongest of the three, and it runs both ways. Bahnaric nearly triples its weight; and the branch contributing most codes to the catalogue, Khasi-Palaung, ends below half of its own. That had not happened in the other two audits, where the large branch simply inflated. The explanation is reach: Khasi-Palaung has thirteen languages with long lists — 13,469 forms, the most of any branch — so it produces many codes, but codes covering few languages. Bahnaric has twenty-three languages and produces codes that travel.

And Pearic, the catalogue's third branch at 16.6%, has not a single code here.

The largest bias is the one that suits the reader, which is why it must be said first: codes crossing four or more branches are 10.9% of the catalogue and 60% of this appendix. It is by far the highest proportion in the series. The appendix shows the genealogically interesting because reach and branch crossing travel together in a family with twelve balanced branches — not because they were chosen.

Audit: thirty-one of a hundred depend on the table

Thirty-one codes have at least one position where the normalisation merges two segments that three or more languages write differently. That sits between Afro-Asiatic's 9 of 91 and Sino-Tibetan's 68 of

100, and nearly all of them fall into two classes:

1 · Nasal place. t·n|weave is ɲ in 20 languages and n in 8; h·n|fire, ɲ in 15 and n in 5; b·n|full, ŋ in 14 and ɲ in 4. The table merges n, ŋ and ɲ into n, and in a family with three contrastive nasals that shows up in one code out of ten.

2 · The glottal. n·h|this is ʔ in 7 and h in 5; l·h|leaf, the same; t·h|earth, ʔ in 17 and h in 4. The table writes both as h, and in Austroasiatic the final glottal stop is a segment in its own right.

And there is a third case, which is the interesting one. s·n|foot gathers dʑ in 19 languages and ts in 6; s·m|bird, s in 22 and ts in 15; c·n|foot, c in 23 and ɟ in 12. These are voicing and affrication alternations in the palatal series, which is exactly where phase 6 found its strongest class — c ↔ s at 16.8×. The table merges them *before* looking and phase 6 rediscovers them *afterwards*: the result coincides and the grounding does not, and that distinction is the whole paper.

What this appendix prints that the code does not see

Of the 2,257 forms the hundred codes cite, 374 (17%) carry at least one mark the skeleton ignores, almost all tone. It sits between the Trans-New-Guinea appendix's 0% and Sino-Tibetan's 75%, and it is the figure to keep in view to know how far to trust what is on the page: in one form out of six, the reader sees something the procedure did not use.

Enrichment, once again, does not rank by antiquity

The ten highest-enrichment codes of this appendix:

Code	Concept	Languages	Enrich.	Branches
p·l·m	leech	15	1417×	1
p·r·t	banana	19	1303×	1
p·l·j	fruit	17	427×	1
d·m	ripe or cooked	14	379×	1
h·w	go up (ascend)	14	301×	1
m·k	headgear	14	277×	5
r·t	buy	12	253×	1
h·r	chicken	14	245×	3
t·r·n	path or road	18	243×	2
s·n·m	year	12	238×	4

Seven of the ten sit inside a single branch, and it is the same pattern the Trans-New-Guinea appendix gave in pure form: a local object or creature, with a name that is long and rare inside its branch's inventory, hits the maximum. Enrichment measures distinctiveness, not depth.

And at the other end sit the codes that cross most branches: h·n|I at 18× across six, k·t|bite at 21× across six, s·n|bone at 14× across five. The appendix's best genealogical candidates are at the very bottom by enrichment, because they are made of frequent consonants. The two columns do not measure the same thing and should not be read together.

And one thing this appendix does better than the two before it

"Other" takes only 22 of the 100. In the Sino-Tibetan appendix it was 55 and in Trans-New-Guinea 47.

That is no credit of ours: the Austroasiatic Lexibank lists are genuinely basic vocabulary, whereas the Sino-Tibetan ones carried medicine, cotton, steel and ploughs. It shows in the sections — seventeen

body-part codes, sixteen verbs, twelve of nature — and it shows in the loan problem, which weighs far less here.

Reverse index: codes recurring across concepts

Nineteen codes cover fifty of the hundred entries, and it is the cleanest index of the three audited appendices: in Sino-Tibetan it was twenty-three covering seventy-nine, with `s·n` taking sixteen concepts on its own. Here the largest takes five.

The table below shows the six recurring three times or more, and one of them must be read slowly, because it is not collision:

`s·k` appears with 'hair' (40 languages), 'hair' again (26), 'feather' (23) and 'feather or fur or hair' (13). All four are the same field, and two of them are the same gloss written in different case — `HAIR` and `hair`, which the corpus carries as distinct concepts. It is not that the skeleton fails to discriminate: it is that the concept list asks four times for the same thing. §3 of the paper counts five such cases in this family.

The other five are collision, and of the manageable kind. `h·n` covers I, fire, smell, goose and eat with enrichments from 11× to 58×; `k·t`, frog, be born, bite and scrape. Unrelated concepts under a short skeleton: the code groups by availability as much as by history, and the code alone does not say which.

Code	Concepts it appears with, and their enrichment
`h·n`	i (20, 18×) · fire (20, 20×) · smell (14, 22×) · goose (13, 58×) · eat (12, 11×)
`s·k`	HAIR (40, 82×) · hair (26, 75×) · feather (23, 51×) · feather or fur or hair (13, 62×)
`n·m`	cry (25, 89×) · blood (16, 30×) · piss (15, 66×) · sweet (14, 51×)
`k·t`	frog (20, 69×) · be born (20, 83×) · bite (14, 21×) · scrape (13, 92×)
`k·n`	child (19, 42×) · boy or son (17, 68×) · muroid (mouse or rat) (14, 27×) · wide (12, 30×)
`c·h`	sick (13, 92×) · louse (12, 38×) · sour (11, 76×)

Numerals

Four codes, and they are unit numerals, not decade compounds as in Sino-Tibetan. `m·j|one` in 31 languages and seven branches, at 153× enrichment, is the best of the section: it is Austroasiatic *muəy*, and it crosses more than half the family's branches.

`c·t|ten` is the audit case: `c` in 7 languages and `ɟ` in 5, a palatal voicing alternation the table merges and phase 6 rediscovers as a class.

`m·j|one` · binary · 31 languages · enrichment 153× · emergent `m·1`

```
Alak mo:j     Brao mu:j    Chong moj.D          Cua muj
Jeh mu̯:j      K.Surin môj  Kasseng mu:j         Khmu mo:j
Mah Meri muy  Mlabri mɔj   Nyaheun muj          Oy muj
Rengao mo̯j    Sapuan mu:j  Semaq Beri Brwe muj  Stieng muoj
```

`b·r|two` · binary · 26 languages · enrichment 146× · emergent `2·3`

```
Brao ba:r     Ceq Wong ber         Chrau bar    Halang ba:r
Kasseng ba:r  Khmer Siem Reap bir  Khmu ba:r    Lao–Katu bar
Mlabri bɛr    MnongRlm ba:r        Nge bar      Nyakur bar.1
Pakoh bar     Santali bar          Sapuan ba:r  Stieng ba:r
```

p·n|four · binary · 22 languages · enrichment 78× · emergent 2·n

Alak poːn	Brao puan	Ceq Wong pɑn	Chrau puon
Halang puan	Jeh puan	Kasseng puən	Mel poan
MnongRlm puon	Nyaheun puan	Palaung phon	Ramam pɔn̥
Rengao puːn	Rumai phon	Sedang pu̱n	Sre puan

c·t|TEN · binary · 12 languages · enrichment 152× · emergent 1·1

Alak c’it	BahnarPleiku ɟit	Brao cit	Cheng cit
Halang ɟəːt	JruLaven cet	Mel c.i̱ːt	Nyaheun cit
Ramam cuot	Sapuan ɟit	Sedang cat	Sre ɟət

Body parts

Seventeen codes, the most populated section, and it holds the paper’s ceiling: m·t|eye in 64 languages and eleven branches, m·h|nose in 58 and ten. *mat* and *muh*.

Worth reading beside it is the ’foot’ pair that §6.2 uses as its best isogloss candidate: **c·n** in 37 languages and eight branches against **s·n** in 37 and nine. Same concept, same second consonant, and the first one splits the family. With the emergent column in view one also sees that c and s belong to the same class, so the pair is exactly the palatal alternation phase 6 found on its own.

m·t|eye · binary · 64 languages · enrichment 147× · emergent m·1

Ahlao mat	Batek Deq Koh mɛt	Cheng mat	Cua mat
Jeh mat	Kensiw Kedah met	Ksinmul mat	Mah Meri mɛt
Mang Ch mat.51	Mlabri mat	Nge mat	Oy mat
Rengao mat	Semai Kampar mat	Sre mat	Tenen Paborn mɛt

m·h|nose · binary · 58 languages · enrichment 100× · emergent m·1

Alak muh	Brao muh	Chrau muh	Halang mu̱h
Kasseng moh	Ksinmul moh	M.Bunong muh	Mel moh
Mon muh	Nyaheun muh	Parengi muʔ	Prai muh
Samre muh.B	Semaq Beri Brwe muh	Sora mʔu	Taoi moh.A

s·k|HAIR · binary · 40 languages · enrichment 82× · emergent 1·4

Alak sɔk	Bugan sak.T.55	Didra sak	Jah Hut sɔk
JruLaven sok	Kasseng sok	Kintaq sɔk	Lanoh Kertei sɔok
Mah Meri suk	MnongRlm sɔk	Nyaheun sɔk	Oy sɔk
Samre suk.A	Semai Kampar sɒːk	Stieng chɔk	Taoi sɔk.A

c·n|foot · binary · 37 languages · enrichment 61× · emergent 1·n

Ahlao ciːŋ¹	Batek Deq Koh can	Ceq Wong cɑn	Chrau ɟəŋ
Jeh ɟ.o̤ːŋ	Kensiw Kedah can	Khmer Siem Reap cəːŋ	Maleng Bro ciŋ꜀
Mang Ch cuaŋ.⁴	Mel c.ə̤ːŋ	Nyaheun ɟiŋ	Palaung ɟɯŋ
Phong ciːŋ.²¹	Sapuan ɟiŋ	Sre ɟəŋ	Tampuan ci̤ŋ

s·n|foot · binary · 37 languages · enrichment 33× · emergent 1·n

Arem tɕìŋ	Chrau ʒəŋ	Didra ʒeŋ.L	Jah Hut ʒɔŋ
K.Surin cə̂ŋ	Khmer Siem Reap ʒaəŋ	M.Bunong ʒəŋ	Mlabri ʒəŋ
Mon ʒoŋ	Nyakur chuŋ.2	Pangkham ʈeŋ	Rumai ʑân.51
Sedang ceaŋ.L	Sre ʒəŋ	Tampuan cuŋ	Wa ćauŋ.B

t·h|BREAST · binary · 30 languages · enrichment 45× · emergent 1·1

Alak tɔh	BahnarPleiku tɔh	Bulang tɨh.⁵¹	Cheng tɔh
Cua toh	Halang tuh	JruLaven toh	M.Bunong tɔh
MnongRlm tɔh	Mon tah	Nyaheun tɔh	Oy tɔh
Ramam tɔh	Sedang tuh.T	Semelai tuh	Tampuan toh

s·k|hair · binary · 26 languages · enrichment 75× · emergent 1·4

Ahlao sɔk¹	Alak sɔk	Ceq Wong sɔk	Cheng sɔk
Halang sɔk	Jeh suk	JruLaven sok	Kasseng sok
Mel sɔk	MnongRlm sɔk	Nyaheun sɔk	Phong suk
Rengao tsə̤k	Sapuan sɔk	Sedang sak	Tampuan sɔk

l·m|liver · binary · 19 languages · enrichment 74× · emergent 3·m

Ahlao lɔɔm.C	Arem lùo̱m	Chong lo:m.C	Chut lɔɔm.2
Iduh lû:m	Kasong lɔ:m.C	Lao-Katu lɔm	Maleng Bro lɔɔmꟾ
Malieng lɔm.33	Malieng lɔmꟾꟾ	May lə̤m²	Nge lɔm
Pakoh lɔm	Phong lom	Samre luəm.B	Taoi lɔm.C

k·l·m|liver · ternary · 19 languages · enrichment 136× · emergent 4·3·m

Alak klə:m	BahnarPleiku klə:m	Brao klə:m	Cheng klə:m
Chrau khləm	Halang k.l.ə̤:.m	JruLaven klʌ:m	Kasseng klɔ:m
M.Bunong kləm	Mel klə:m	Nyaheun klʌ:m	Rengao k.l.ə̤:.m
Sapuan klə:m	Sedang kliam	Sre klə:m	Stieng klə:m

k·j|head · binary · 17 languages · enrichment 50× · emergent 4·1

Batek Deq Koh kuy	Car kuj	Ceq Wong kuy	Jah Hut kɔj
Jahai Banun kuj	Kensiw Kedah kuj	Kintaq kuy	Lanoh Kertei kuy
Mah Meri koy	Menriq Lah kuy	Nancawri kúj	Semai Kampar ku:y
Semaq Beri Brwe koj	Semelai khoj	Semnam Bal ku:y	Temiar Kelantan ku:y

n·m|blood · binary · 16 languages · enrichment 30×

Bulang nam.⁵¹	Chrau nʰa:m	Juang iŋam	Kharia eŋam
Kme-2 na:m.³¹	Mel ŋʰa:m	Meng Dan nɨm.45	Palaung n̥a:m
Paliu nˡam.⁵¹	Pangkham n̥a:m	Pear n̪a:m	Prai ńiam
Rumai nham.412	Taoi ŋam.C	Wa nham.A	Xiang Zhai Tang n̥a:m

s·n|bone · binary · 16 languages · enrichment 14× · emergent 1·n

Arem ɕiə̤ŋ	Chut sàŋ	Kharia ʒaŋ	Khasi sh'ieŋ
Malieng saŋ.55	Malieng saŋꟾꟾ	Mang Ch čaŋ.31	May ɕaŋ²
Mundari ʒaŋ	Mương [Hoa Binh] sɨaŋ¹	Ngu□ n C□ Liêm sɨəŋ¹	Paliu si.⁵⁵_aŋ.⁵¹
Phong siaŋ.²⁴	Santali ʒaŋ	Sora əʒáŋ	Vietnamese swəŋ

m·h·m|blood · ternary · 14 languages · enrichment 220× · emergent m·1·m

Car ma-ham	Ceq Wong mhə̃m	Chong məha:m.A	Didra məhiam
Halang məha:m	Kensiw Kedah mhə̃m	Kintaq mhɨ̃m	M.Bunong m'ham
Mah Meri maham	Rengao məha:m	Sedang məheam	Semaq Beri Brwe maham
Semelai maham	Tenen Paborn mahɨm		

k·l·k|INTESTINES · ternary · 13 languages · enrichment 144× · emergent 4·3·4

Alak kla:k	BahnarPleiku kla:k	Brao kla:k	Cua kla:k
Halang kla:k	Jeh kla:k	JruLaven kla:k	Kasseng kla:k
Nyaheun kla:k	Ramam kla:k	Rengao kla:k	Sapuan kla:k
Tampuan kla:k			

s·m|blood · binary · 13 languages · enrichment 28× · emergent 1·m

Chut ăs□ m	K.Surin chiəm	Kha Poong asam³³ʔ	Khasi snam
Khmer Siem Reap ʒham	Lyngngam snam	Maleng Bro asəmꟾ	Malieng asam.51ʔ
Malieng asamꟾꟾ	Mon chim	Nyakur chim.1	Phong sa:m.⁴⁴
Toum sa:m			

m·m|blood · binary · 11 languages · enrichment 60×

Ceq Wong mhə̃m	Iduh mîm	Khmu ma:m	Ksinmul miəm
M.Bunong mham	Mel mʰa:m	MnongRlm mham	Parengi miam
Phong mim	Sre mham	Stieng mham	

s·h·n|bone · ternary · 11 languages · enrichment 151× · emergent 1·1·n

Bulang sa.³¹_ʔaŋ.³³	Jah Hut ʒəʔaŋ	Jahai Banun ʒəʔeŋ	K.Surin cʔʌŋ
Khasi ʃʔeŋ	Khmer Siem Reap chəʔɨŋ	Mlabri ca ʔɛŋ	Phong sʔiŋ
Semaq Beri Brwe ʒaʔaŋ	Semelai ʒəʔaŋ	Wa səʔaŋ.B	

Kinship and person

Four codes. k·n appears with 'child' (19 languages) and 'boy or son' (17) — two glosses of one field, the list again — and also with 'mouse' and 'wide', which are not.

h·n|I · binary · 20 languages · enrichment 18× · emergent 1·n

BahnarPleiku ʔiŋ	Bulang ʔoŋ.³³	Ceq Wong ʔiŋ	Chong ʔiń.A
K.Surin ʔaŋ	Kasong ʔiń.A	Ksinmul ʔaŋ	Menriq Lah ʔiŋ
Phong ʔeŋ	Prai ʔəń	Samre ʔiń.A	Semai Kampar ʔen
Semelai ʔəŋ	Semnam Bal ʔĩ:ŋ	Sre ʔaŋ	Tampuan ʔaŋ

k·n|CHILD · binary · 19 languages · enrichment 42× · emergent 4·n

Ahlao kɔ:n¹	Arem kɒn	Batek Deq Koh kɛn	Chut kɔn
Jahai Banun kɛn	Khasi kʰun	Khmu kɔ:n	Lanoh Kertei kɔn
Lyngngam kʰon	May kɔn¹	Menriq Lah kɛn	Palaung kuən
Phong kɔ:n	Rumai kɔ:n	Semnam Bal kuo:n	Toum kɔ:n

m·h|THOU · binary · 15 languages · enrichment 27× · emergent m·1

Ahlao miʔ.B	Batek Deq Koh mɔh	Car mɛh	Ceq Wong mɨ̃ʔ
Jahai Banun miʔ	Kintaq muh	Kme-2 meʔ.⁵³	Ksinmul mih
Lanoh Kertei mih	Menriq Lah mɔh	Prai mah	Semnam Bal mih
Tenen Paborn mɔh	Wa maiʔ.A	Xu meʔ.55	

n·h|THIS · binary · 12 languages · enrichment 32× · emergent n·1

Car ŋih	Cua nɛʔ	JruLaven neʔ	Khmer Siem Reap nih
Khmu niʔ	Lanoh Kertei nɔoh	Lao-Katu neʔ	Mon ṇaʔ
Mundari ni=iʔ	Pakoh nnəh.L	Semaq Beri Brwe naʔ	Semnam Bal nɔh

Nature and sky

Twelve codes. d·k|water in 30 languages is Austroasiatic *ɗaak*; h·n|fire in 20 and p·l·n|sky in 15 complete the expected. All three depend on the table to some degree — d/ɗ in the first, ŋ/n in the second, ŋ/ɲ in the third — and all three are real codes regardless.

d·k|water · binary · 30 languages · enrichment 99× · emergent 1·4

Ahlao dak.C	Arem ɗɛa̤k	Brao da:k	Chut da:k.³
Halang da:k	JruLaven da:k	Kha Poong da:k.³³	M.Bunong dak
Malieng dak.33	Malieng dakˌˌ	Mel ɗ.a̤:.k	Mương [Hoa Binh] da:k⁷
Nyakur dak.1	Phong da:k.²¹	Sapuan da:k	Stieng da:k

t·h|earth · binary · 21 languages · enrichment 29× · emergent 1·1

BahnarPleiku tɛh	Batek Deq Koh teʔ	Ceq Wong tɛʔ	Cheng tɛʔ
Jah Hut teʔ	Jahai Banun teʔ	Kensiw Kedah teʔ	Lanoh Kertei tɛʔ
Mah Meri teʔ	Mel tɛh	MnongRlm tɛh	Nyakur tiʔ.1
Semai Kampar teʔ	Tampuan tɛh	Temiar Kelantan tɛʔ	Tenen Paborn tieʔ

h·n|fire · binary · 20 languages · enrichment 20× · emergent 1·n

Alak ʔu:ɲ	BahnarPleiku ʔuɲ	Brao ʔuɲ	Cheng ʔuɲ
Halang ʔun	Jeh ʔo̤n	JruLaven ʔuɲ	Kasseng ʔu:ɲ
Mel ʔoŋ	MnongRlm ʔɨɲ	Nyaheun ʔuɲ	Oy ʔuɲ
Rengao ʔṳn	Sapuan ʔuɲ	Sedang ʔɔn	Stieng ʔoŋ

h·s|fire · binary · 20 languages · enrichment 117× · emergent 1·1

Batek Deq Koh ʔɔs	Ceq Wong ʔɔs	Jah Hut ʔɔs	Jahai Banun ʔɔs
Kintaq ʔɔs	Lanoh Kertei ʔos	Lao-Katu ʔos	Mah Meri ʔus
Nge ʔujh	Pakoh ʔujh	Phong ʔos	Semai Kampar ʔɒ:s
Semelai ʔus	Semnam Bal ʔo:s	Sre ʔous	Temiar Kelantan ʔo:s

m·n|night · binary · 20 languages · enrichment 33×

Alak maŋ	BahnarPleiku maŋ	Brao maŋ	Cheng maŋ
Halang maŋ	Jeh maŋ	JruLaven maŋ	Kasseng maŋ
Mel maŋ	MnongRlm maŋ	Nyaheun maŋ	Oy maŋ
Sapuan maŋ	Sedang maŋ	Sre maŋ	Stieng maŋ

t·m|stone · binary · 16 languages · enrichment 39× · emergent 1·m

Alak tamɔː	BahnarPleiku təmɔː	Brao təmɑː	Chrau təmoː
Halang təmou	Jeh təmou.T	JruLaven təmɔː	K.Surin thmâ
Kasseng təmɔ	Khmer Siem Reap thmɔː	Nge tamɔ	Oy təmɔ
Samre thamo.C	Sapuan tmɔː	Semelai təmu	Stieng təmau

p·l·n|sky · ternary · 15 languages · enrichment 124× · emergent 2·3·n

BahnarPleiku plɛŋ	Chong phlɨŋ.A	Halang pliːŋ	Jeh pliːŋ
Kasong phliń.A	Kasseng pliːŋ	Mang Ch pliːń.[6]	Palaung pleŋ
Pear pleŋ	Pusing plɛŋ	Ramam plɛ̊ŋ	Rengao plɨ̈ŋ
Rumai plân.[412]	Sedang plɛŋ	Tampuan plaiŋ	

t·r·n|ROAD · ternary · 14 languages · enrichment 138× · emergent 1·3·n

Chrau trɔŋ	Didra truoŋ	Halang truoŋ	Jeh truaŋ.L
Kasseng truaŋ	M.Bunong trɔn	MnongRlm troŋ	Nyaheun truəŋ
Oy truəŋ	Rengao troŋ.T	Sedang troaŋ.T	Semaq Beri Brwe təruŋ
Semelai təroŋ	Stieng trɔŋ		

k·j|moon · binary · 14 languages · enrichment 42× · emergent 4·1

BahnarPleiku kʰɛːj	Brao kʰaj	Cheng kʰaj	Chrau khaj
Halang khaj	Iduh kâj	Jeh kʰɛːj	JruLaven khaj
M.Bunong khaj	Mel kʰəj	Ramam kʰʌj	Rengao kʰɛj
Stieng khaj	Tampuan kʰaj		

b·t·h|stone · ternary · 13 languages · enrichment 195× · emergent 2·1·1

Batek Deq Koh batuʔ	Jahai Banun batuʔ	Kensiw Kedah batuʔ	Kintaq batuʔ
Lanoh Kertei batuʔ	Mah Meri batuʔ	Menriq Lah batuʔ	Semai Kampar batuh
Semaq Beri Brwe batuʔ	Semelai batuʔ	Semnam Bal batuʔ	Temiar Kelantan batuʔ
Tenen Paborn batuʔ			

s·m·n|star · ternary · 12 languages · enrichment 197× · emergent 1·m·n

Bulang sa.[31]_məiŋ.[51]	Chrau səmaŋ	Iduh samɛ̂ŋ	Mel səmɑŋ
Mon saman	Palaung sameŋ	Pangkham samaŋ	Pear s`man
Phong smeŋ	Sre səmaŋ	Stieng səmaŋ	Xiang Zhai Tang samɛŋ

s·n·m|year · ternary · 12 languages · enrichment 238× · emergent 1·n·m

BahnarPleiku sənam	Cua sanɨm	Kha Poong sənam.[33]	Meng Dan sənam
Palaung sanɔm	Pangkham sanam	Pear s`nam	Rumai sanam
Sre sənam	Stieng sənam	Tampuan sanam	Xiang Zhai Tang sanəm

Animals

Nine codes, and two of the appendix's three highest enrichments are here: **p·l·m|leech** at 1417× and h·r|chicken at 245×. The first is single-branch; the second crosses three.

The leech is the perfect example of the header's warning: a creature with a specific name inside one branch tops a column the reader interprets as depth.

s·m|bird · binary · 42 languages · enrichment 93× · emergent 1·m

Alak cem	BahnarPleiku sɛːm	Chrau sum	Jah Hut cɛm
Kasong chim.B	Khmu siːm	Lyngngam sim	Mel sim
Nge cem	Nyaheun cem	Pangkham sim	Phong som
Rengao cim.L	Sedang cem.L	Sre seːm	Vietnamese chim

c·m|bird · binary · 28 languages · enrichment 158× · emergent 1·m

Ahlao ciːm¹	Alak cɛːm	Cheng ceːm	Chut icim.1
Iduh cîəm	Jah Hut cɛm	JruLaven ceːm	Kasseng ceːm
Mương [Hoa Binh] chim	Nyaheun ceːm	Ramam cem̥	Samre chiːm.C
Sedang cem	Semelai cim	Stieng com̥	Vietnamese chim

s·k|feather · binary · 23 languages · enrichment 51× · emergent 1·4

Bulang suk.³¹	Ceq Wong sɔk	Chong suk.A	Halang sok
Jah Hut sɔk	JruLaven sok	Kasong suk.A	Kensiw Kedah sok
Kme-2 suk.⁵³	Ksinmul sok	Nyaheun sɔk	Oy sɔk
Paliu suk.³³	Phong suk.⁴⁵	Sedang sak.L	Semai Kampar sɔk

k·h|fish · binary · 17 languages · enrichment 30× · emergent 4·1

Ahlao kaʔ.B	Arem akɛa̰ʔ	Ceq Wong kiəʔ	Jah Hut kaʔ
Kha Poong aka.¹³ʔ	Khmu kaʔ	Kintaq kaʔ	Kme-2 a_khaʔ.⁵³
Lanoh Kertei kaʔ	Mlabri kaʔ	Nyakur kaʔ.1	Paliu kha.ʔ⁵⁵
Semai Kampar kaʔ	Semaq Beri Brwe kah	Semnam Bal kaʔ	Temiar Kelantan kaʔ

s·j|louse · binary · 15 languages · enrichment 116× · emergent 1·1

Cua saj	JruLaven cej	K.Surin caj	Kasseng caj
Khmer Siem Reap caj	Ksinmul cej	Malieng cəj.51ʔ	Mon caj
Nancawri séj	Nyakur caj.1	Oy caj	Sedang caj.L
Sre saj	Tampuan səj	Vietnamese tɕəj	

h·r|CHICKEN · binary · 14 languages · enrichment 245× · emergent 1·3

Alak ʔɛːr	Cheng ʔiar	Chrau ʔier	Halang ʔiər
Iduh ʔiər	JruLaven ʔiar	Kasseng ʔiər	MnongRlm ʔier
Pear ʔiar	Ramam ʔaiɻ	Rengao ʔiːr	Sapuan ʔiar
Sre ʔiar	Stieng ʔier		

s·t|BEE · binary · 13 languages · enrichment 85× · emergent 1·1

Alak suːt	BahnarPleiku suːt	Cheng sut	Chrau sɨt
Halang sut	JruLaven suːt	Kasseng suːt	Nyaheun sut
Rengao tsʉːt	Sapuan sut	Sedang sɔt	Sre sut
Stieng sot			

h·n|GOOSE · binary · 13 languages · enrichment 58× · emergent 1·n

Ahlao han.C	Bulang han.³³	Chong haːn.A	Kasong haːn.E
Kme-2 han.¹³	Nyakur haːŋ	Pear han	Phong haːn.⁴²
Prai haːn	Rumai han.⁵⁵	Samre haːn.A	Wa han.A
Xu han.³⁵			

c·h|louse · binary · 12 languages · enrichment 38× · emergent 1·1

Batek Deq Koh ceʔ	Ceq Wong cɛʔ	Jah Hut cɛʔ	Jahai Banun ciʔ
Kensiw Kedah ceʔ	Kintaq ceʔ	Lanoh Kertei cɛʔ	Menriq Lah ceʔ
Semai Kampar cɛːʔ	Semaq Beri Brwe cih	Semnam Bal ceʔ	Temiar Kelantan cɛːʔ

Plants and food

Nine codes, and **p·r·t|banana** at 1303× is the appendix's second highest enrichment, from a single branch. Together with p·l·j|fruit (427×) and d·m|ripe or cooked (379×) they form a block of agricultural vocabulary that is highly distinctive and highly local — which is another way of saying that enrichment does not distinguish the ancient from the shared-by-proximity.

r·h|root · binary · 23 languages · enrichment 114× · emergent 3·1

Alak rɛːh	BahnarPleiku rəh	Car rɛh	Didra reh
Halang riəh	K.Surin rɨh	Kasong reːh.A	M.Bunong rɛh
Mel əɾiɛh	MnongRlm rɛh	Nyaheun riah	Nyakur rih
Pusing riah	Rengao rih	Samre riːh.C	Semaq Beri Brwe rɛh

s·c|meat · binary · 18 languages · enrichment 197× · emergent 1·1

Batek Deq Koh sɛc	Ceq Wong sac	Chong chuc.C	Jah Hut sɛc
Jahai Banun sɛc	Kensiw Kedah sec	Khmer Siem Reap sac	Kintaq sec
Menriq Lah sɛc	Pakoh sec.T	Prai sec	Semai Kampar sɛc
Semelai səc	Semnam Bal sĕc	Taoi səc.A	Temiar Kelantan sec

p·l·j|fruit · ternary · 17 languages · enrichment 427× · emergent 2·3·1

Alak palaj	BahnarPleiku plɛːj	Brao plaj	Cheng plaj
Chrau plaj	Cua plaj	Halang plaj	Jeh plaj
JruLaven plaj	Kasseng plaj	Mel plaːj	Ramam plaj
Sapuan plaj	Sedang plaj	Sre plaj	Stieng plaj

h·l|leaf · binary · 17 languages · enrichment 72× · emergent 1·3

BahnarPleiku hlaː	Chut ʔula.3	Cua hlaː	Didra hla
Kasseng ʔəlaː	Lao-Katu ʔala	May ʔula³	Nge hla
Nyaheun hla	Nyakur hlâː	Oy hla	Palaung hla
Ramam hlaː	Rengao hlaː	Sedang hla	Tampuan hlaː

b·h|salt · binary · 15 languages · enrichment 62× · emergent 2·1

Alak bɔh	BahnarPleiku bɔh	Brao bɑh	Cheng bɔh
Cua bʌh	Halang bɔh	Jeh bɔh	JruLaven bɔh
Kasseng bɔh	MnongRlm bɔh	Nyaheun bɔh	Rengao bɔh
Sapuan bɔh	Sre bɔh	Stieng bɔh	

r·s|root · binary · 14 languages · enrichment 143× · emergent 3·1

Jeh riajh.T	JruLaven rias	Kasseng rias	Khmer Siem Reap rɨs
Khmu riəs	Lao-Katu rias	Nge riəjh	Oy riajh
Pakoh rɨajh.T	Phong riəs	Semaq Beri Brwe rɛs	Semelai rɛs
Sre rias	Taoi riajh.C		

h·n|eat · binary · 12 languages · enrichment 11× · emergent 1·n

Ahlao ʔân	Arem ʔăn	Bulang ʔeŋ.³³	Chut ʔan.1
Kha Poong ʔan.³³	Maleng Bro ʔanˈ	Malieng ʔan.33	Malieng ʔăn
May ʔăn¹	Phong ʔan	Toum ʔan	Vietnamese ăn

l·h|leaf · binary · 12 languages · enrichment 40× · emergent 3·1

Arem ulɛa̯ʔ	Chong laʔ.A	Khmu ʰlaʔ	Kme-2 laʔ.⁵³
M.Bunong l'ha	Mang Ch la.⁵¹_hɔ.³⁵	Meng Dan l̥ɛː hi	Palaung l̥aː heːj
Parengi olaʔ	Rumai l̥aː hɔːj	Wa lhaʔ.A	Xu laʔ.55

s·s|meat · binary · 12 languages · enrichment 207× · emergent 1·1

Alak sac	Jah Hut sɛc	Jahai Banun sɛc	K.Surin sac
Kasseng ʒec	Kensiw Kedah sec	Khmer Siem Reap sac	Lao-Katu sec
Nge sac	Semai Kampar sɛc	Semelai səc	Temiar Kelantan seʒ

Verbs of action and state

Sixteen codes, the second most populated section, and it is a large difference from the two preceding families: Sino-Tibetan contributed a single verb and Trans-New-Guinea five. Austroasiatic verbs have enough consonants to form binary codes with reach.

k·t covers ’be born’, ’bite’ and ’scrape’ with enrichments from 21× to 92×, and h·w|go up reaches 301× within a single branch.

k·p|bite · binary · 36 languages · enrichment 83× · emergent 4·2

Alak kap	Batek Deq Koh kap	Car kap	Cheng kap
Didra kap	Jah Hut kap	Jeh kap	Kasseng kap
Kme-2 kap.⁵⁵	M.Bunong kap	Menriq Lah kap	Nyaheun kap
Ramam kap	Sapuan kap	Semnam Bal kap	Stieng kap

p·r|fly_V · binary · 35 languages · enrichment 143× · emergent 2·3

Alak par	Brao par	Cheng par	Chut parh.1
JruLaven par	Kha Poong pɤɽˈ	Maleng Bro pərˈ	Mlabri par
Mundari apir	Nyakur phar.1	Pakoh par	Phong par
Pusing pɨr	Rengao par	Sre par	Tampuan par

n·m|cry · binary · 25 languages · enrichment 89×

Alak ɲiːm	BahnarPleiku ɲəm	Cheng ɲɨːm	Chrau ɲiːm
Halang ɲi̤əm	Iduh ńîəm	JruLaven ɲɨəm	Kme-2 ńaːm.³¹
Mel ɲim	Mương [Hoa Binh] nhạm	Ngu□ n C□ Liêm ɲam⁴	Phong ńaːm.⁴²
Prai ńem	Ramam ɲim	Sapuan ɲɨəm	Stieng ɲiːm

n·m|PISS · binary · 15 languages · enrichment 66×

Brao noːm	Bulang nɨm.⁵¹	Chrau noːm	Kme-2 num.⁵⁵
Mlabri nɔm	MnongRlm noːm	Pear n̥am	Pusing nom
Rengao nɔm	Sapuan noːm	Sedang nu̱m	Sre nom
Stieng noːm	Wa nɨm.B	Xu num.⁵⁵	

k·l·h|fall · ternary · 15 languages · enrichment 166× · emergent 4·3·1

BahnarPleiku klih	Brao kliːh	Cua kleh	Halang klih
JruLaven kleh	Kasseng kleh	Lanoh Kertei klɨʔ	Nyaheun klieh
Prai khlih	Ramam klauh	Sapuan klieh	Sedang klɛh
Semnam Bal kləʔ	Stieng kleh	Temiar Kelantan klɨʔ	

k·h|vomit · binary · 15 languages · enrichment 42× · emergent 4·1

Batek Deq Koh kəʔ	Ceq Wong kəʔ	Jahai Banun kɨʔ	Kensiw Kedah kəʔ
Kintaq kəʔ	Lanoh Kertei kɔoʔ	Mah Meri kuʔ	Menriq Lah kəʔ
Paliu ku.ʔ⁵⁵	Semai Kampar kəːʔ	Semaq Beri Brwe kuʔ	Semelai kʰuʔ
Semnam Bal koʔ	Temiar Kelantan koːʔ	Tenen Paborn kɨʔ	

k·t|bite · binary · 14 languages · enrichment 21× · emergent 4·1

Ahlao kə̰ːt	Alak kiat	Bulang kiat.³³	Cua kat
Kasong khat.A	Kha Poong kaːt.³³	Malieng kat.43	Malieng katꟾꟾ
Mon kit	Nyakur kɨt.1	Phong kat.²¹	Samre khat.A
Toum kaːt	Wa kiat.B		

c·n|sew · binary · 14 languages · enrichment 48× · emergent 1·n

Alak c’iːŋ	Brao ɟiŋ	Chrau ɟiŋ	JruLaven ɟiːŋ
Kasseng ɟiːŋ	Mel ɟʊŋ	MnongRlm ɟɨŋ	Nyaheun ɟiŋ
Nyakur chîːŋ	Palaung ɟiŋ	Pangkham cin	Sre ɟiŋ
Stieng ɟeŋ	Tampuan cɨ̤ŋ		

h·n|smell · binary · 14 languages · enrichment 22× · emergent 1·n

Batek Deq Koh ʔɔ̃ŋ	Ceq Wong ʔɔ̃ŋ	Jah Hut ʔɔ̃ŋ	Jahai Banun ʔɔ̃ŋ
Kensiw Kedah ʔɔ̃ŋ	Kintaq ʔɔ̃ŋ	Lanoh Kertei ʔũŋ	Mah Meri hɔn
Menriq Lah ʔɔ̃ŋ	Phong huːń.²⁴	Semaq Beri Brwe ʔun	Semnam Bal ʔɔ̃ːŋ
Temiar Kelantan ʔuːŋ	Tenen Paborn ʔɔ̃ŋ		

h·m|give · binary · 13 languages · enrichment 27× · emergent 1·m

Alak ʔam	BahnarPleiku ʔam	Brao ʔam	Cua ʔam
Didra ʔam	Halang ʔam	JruLaven ʔam	Nyaheun ʔəm
Oy ʔam	Ramam ʔam	Rengao ʔam	Sapuan ʔam
Sedang ʔə̤m			

h·k|give · binary · 13 languages · enrichment 30× · emergent 1·4

Batek Deq Koh ʔɛk	Ceq Wong ʔak	Jah Hut ʔɔk	Jahai Banun ʔɛk	
Kensiw Kedah ʔek	Kintaq ʔɛk	Lanoh Kertei ʔɔok	Menriq Lah ʔɛk	
Samre ʔuək.C	Semai Kampar ʔɒːk	Semnam Bal ʔoːk	Temiar Kelantan ʔoːk	
Tenen Paborn ʔɛk				

p·l|fly_V · binary · 12 languages · enrichment 45× · emergent 2·3

Arem păl	Chut pəl	Cua pʌl	Kha Poong pɤ̆l²³
Ksinmul pal	Malieng pɤ̆l	May pɤ̆l¹	Meng Dan pil.55
Mương [Hoa Binh] pal¹	Phong pil	Sapuan pal	Toum pʌl

c·p|WALK · binary · 11 languages · enrichment 70× · emergent 1·2

Batek Deq Koh cip	Ceq Wong cɨp	Jah Hut cip	Jahai Banun cɨp
Kensiw Kedah cɨp	Kintaq cɨp	Lanoh Kertei cip	Menriq Lah cɨp
Semai Kampar ci:p	Semnam Bal ci:p	Temiar Kelantan ci:p	

n·p|count · binary · 11 languages · enrichment 231× · emergent n·2

Ahlao năp	Chong nap.C	Iduh na²ɔp	Kasong nap.E
Khmu nap	Mlabri nap	Nyakur nap	Phong nap.⁴⁴
Prai nap	Pusing nap	Samre nap.B	

j·m|cry · binary · 11 languages · enrichment 149× · emergent 1·m

Bulang jam.⁵¹	Chong ja:m.C	Chut jạm	Kasong ja:m.C
Khmu ja:m	Nyakur jâ:m	Rumai jam.⁴¹²	Samre ja:m.B
Toum ja:m	Wa jiam.B	Xu jam.³¹	

s·t|die · binary · 11 languages · enrichment 32× · emergent 1·1

Alak si:t	Arem tɕi:t	Chrau cɨt	Cua se:t
Kasseng cit	Lao-Katu cet	Mang Ch sit.55	Pakoh cet.T
Sre chət	Stieng chet	Vietnamese ch□ t	

Objects and culture

A single code carries this label, t·r·n|path or road, in 18 languages and two branches at 243×. Material culture weighs little in this appendix, and that is a property of the source lists: they are basic vocabulary.

t·r·n|PATH OR ROAD · ternary · 18 languages · enrichment 243× · emergent 1·3·n

BahnarPleiku trɔ:ŋ	Brao tru:ŋ	Cheng tru:ŋ	Chrau trɔ:ŋ
Halang trɔaŋ	Jeh trṵaŋ	JruLaven truoŋ	Kasseng truəŋ
MnongRlm tro:ŋ	Nyaheun truoŋ	Rengao trɔ:ŋ	Sapuan tru:ŋ
Sedang trə̯əŋ	Semaq Beri Brwe tʁoŋ	Semelai troŋ	Stieng trɔ:ŋ

Qualities

Six codes, with middling enrichments and short reach. It is the section where the instrument has least to say in every family of the series, and here too.

b·n|full · binary · 19 languages · enrichment 62× · emergent 2·n

Ahlao bə̰̀n	Brao biŋ	Chrau beɲ	Halang bi:ŋ
Jeh biŋ	JruLaven bi:ŋ	Khmer Siem Reap bəɲ	Lao-Katu biŋ
M.Bunong beŋ	Mel b.ẹ:.ɲ	MnongRlm bi:ŋ	Oy biŋ
Pakoh biɲ	Rengao biŋ	Sapuan biŋ	Sre be:ŋ

t·n|bitter · binary · 16 languages · enrichment 31× · emergent 1·n

Ahlao taŋ⁴	BahnarPleiku taŋ	Chrau taŋ	Chut ət□ □ ŋ
Kha Poong tăŋ³³ʔ	Maleng Bro táŋˊ	Malieng taŋ³	May tăŋ³
Mel taŋ	MnongRlm taŋ	Mương [Hoa Binh] taŋ³	Ngu□ n C□ Liêm tăŋ³
Phong taŋ	Stieng taŋ	Tampuan taŋ	Toum taŋ

n·m|sweet · binary · 14 languages · enrichment 51×

Alak ŋaːm	BahnarPleiku ŋaːm	Brao ŋaːm	Chong ŋaːm.D
Cua ŋɨːm	Kasong ŋaːm.D	Kasseng ŋaːm	Pear ŋam
Ramam ŋaːm	Rengao ŋaːm	Rumai ŋam.⁴¹²	Samre ŋaːm.C
Sapuan ŋaːm	Sedang ŋɛ̯əm		

b·k|white · binary · 13 languages · enrichment 93× · emergent 2·4

Brao bɑːk	Cheng bɔːk	Kasseng bɔk	Lao–Katu bɔk
Meng Dan buk.35	MnongRlm buok	Nge bɔk	Nyaheun bɔːk
Oy bɔk	Pusing buk	Sapuan bɔːk	Stieng bɔːk
Tampuan baːk			

k·r·h|old · ternary · 12 languages · enrichment 190× · emergent 4·3·1

Alak kraʔ	BahnarPleiku kraʔ	Brao kraʔ	Cheng kraʔ
Cua karaʔ	Halang kraʔ	JruLaven kraʔ	Kasseng kraʔ
Nyaheun kraʔ	Rengao kraʔ	Sapuan kraʔ	Tampuan kraʔ

k·n|wide · binary · 12 languages · enrichment 30× · emergent 4·n

Ahlao kw̯aŋ	Chong kwaːŋ.A	Iduh kwâːŋ	Kasong kwaːŋ.E
Kha Poong kwaŋ³³ʔ	Maleng Bro kwáːŋˊ	Mlabri kwaŋ	Nyakur kwâːŋ
Phong kwaːŋ.⁴⁴	Prai khwaːŋ	Pusing kuaŋ	Samre kwaːŋ.B

Other

Twenty-two codes, less than a quarter of the appendix. In Sino-Tibetan it was fifty-five and in Trans-New-Guinea forty-seven.

It is not a residual category, but neither is it the dumping ground it was in the other two: these are concepts the thematic list — written for Swadesh-type basic vocabulary — does not contemplate by their label, not by their nature. `m·k|headgear` at 277× across five branches and `s·n·m|year` at 238× across four are the two best, and both are wide-reaching codes that cross branches: exactly the opposite of the ”Other” of the two preceding families, where the ceiling was a seventeenth-century loan and a list of single-branch objects.

t·n|WEAVE · binary · 32 languages · enrichment 78× · emergent 1·n

Alak taːɲ	Brao taːɲ	Ceq Wong teŋ	Chong thaːɲ́.A
Halang taːn	JruLaven taːɲ	Kasseng taːɲ	Mang Ch taːn.⁶
MnongRlm taːɲ	Nyakur tâːñ	Phong taːɲ́.²⁴	Ramam taŋ̊
Samre thaːɲ́.A	Sedang tɛ̯n	Stieng taːɲ	Wa taiŋ.A

p·n|SHOOT · binary · 29 languages · enrichment 58× · emergent 2·n

Alak paŋ	Brao paŋ	Cheng paŋ	Halang pɛŋ
Jeh pɛŋ	Kasseng paŋ	Khmu píŋ	Mel paŋ
Nyaheun paŋ	Pear pɔn	Prai pheń	Pusing päŋ
Rengao pɛŋ	Sedang pe̤ŋ	Stieng paŋ	Wa puiŋ.A

k·t|FROG · binary · 20 languages · enrichment 69× · emergent 4·1

Alak kiːt	BahnarPleiku kit	Brao kiːt	Cheng kiːt
Halang kiːt	Jeh kɨat	Kasseng kiːt	Kha Poong kɔːt.³³
MnongRlm kɨt	Nyaheun kiet	Paliu a.³³_khɛt.³³	Ramam kʌt
Sedang kɛt	Sre kit	Stieng kɨt	Tampuan kət

k·t|BE BORN · binary · 20 languages · enrichment 83× · emergent 4·1

Alak kəːt	Brao kəːt	Bulang kət.³³	Chong kəːt.A
Iduh kə̂ːt	JruLaven kəːt	Kasong kəːt.E	Kme-2 kət.¹³
Nyaheun kəːt	Nyakur kə̂ːt	Paliu kət.⁵⁵	Pear kə́t
Pusing kɨt	Rumai kɨt.⁵⁵	Samre kəːt.A	Stieng kəːt

p·r·t|BANANA · ternary · 19 languages · enrichment 1303× · emergent 2·3·1

Alak priːt	BahnarPleiku priːt	Brao prɨːt	Cheng prɨːt
Chrau priːt	Cua pareːt	Jeh prɨat	JruLaven prɨət
Kasseng priət	Mel prit	MnongRlm priːt	Rengao pret
Sapuan prɨat	Sedang priət	Sre priːt	Stieng priːt

k·n|BOY OR SON · binary · 17 languages · enrichment 68× · emergent 4·n

Alak kɔːn	BahnarPleiku kɔːn	Brao kuan	Chrau kɔːn
Halang koan	Jeh kɔːn	JruLaven kuan	Mel kɔːn
MnongRlm kuon	Nyaheun kuan	Ramam kɔːn	Rengao kɔːn
Sapuan kuan	Sedang kɔn	Sre kon	Stieng kɔːn

p·l·m|LEECH · ternary · 15 languages · enrichment 1417× · emergent 2·3·m

Alak pləːm	BahnarPleiku pləːm	Brao pləːm	Chrau pləːm
Halang p.l.e̤ː.m	Jeh p.l.e̤ː.m	JruLaven plʌːm	Kasseng plɤːm
MnongRlm pləːm	Nyaheun pleːm	Rengao p.l.e̤ː.m	Sedang pliəm
Sre plɔm	Stieng pləːm	Tampuan plaəm	

m·k|HEADGEAR · binary · 14 languages · enrichment 277× · emergent m·4

Ahlao muok	Bulang mhɔk.⁵¹	Chong muak.A	Kasong muək.D
Kha Poong muok.³³	Kme-2 mok.¹³	Mang Ch mɔk.⁵⁵	Mlabri muak
Paliu m¹uək.⁵¹	Phong muak.⁴²	Prai muək	Rumai mhok.⁵⁵
Samre muək.A	Wa mhok.A		

k·n|MUROID (MOUSE OR RAT) · binary · 14 languages · enrichment 27× · emergent 4·n

BahnarPleiku kənɛː	Brao kəneː	Cheng kneː	Chong khɔːn.B
Chrau kənɛː	Cua kaneː	Halang kənɛː	JruLaven knɛː
Kasong khɔːn.B	Mel kənaɛ	Phong kə̆ne.⁴⁴	Rengao kəniː
Samre khuən.C	Wa kiaŋ.B		

b·k|tie · binary · 14 languages · enrichment 141× · emergent 2·4

Brao bɨk	Cheng bɨk	JruLaven bɨk	Kasseng bɤk
Lanoh Kertei bɨk	Mah Meri bək	Nyaheun bɨk	Rumai bak.⁵⁵
Sapuan bɨk	Semai Kampar bək	Semaq Beri Brwe bək	Semelai bək
Semnam Bal bəːk	Temiar Kelantan bək		

h·w|GO UP (ASCEND) · binary · 14 languages · enrichment 301× · emergent 1·w

BahnarPleiku haːw	Brao haːw	Chrau haːw	Halang haːw
Jeh haːw	JruLaven haw	Mel haːw	MnongRlm haːw
Nyaheun haːw	Ramam haːw	Rengao haːw	Sre haːw
Stieng haːw	Tampuan haːw		

d·m|RIPE OR COOKED · binary · 14 languages · enrichment 379× · emergent 1·m

BahnarPleiku ɗuːm	Brao dum	Cheng dum	Chrau dum
Halang duːm	Jeh dṵm	JruLaven duːm	Kasseng doːm
Mel ɗʊm	MnongRlm dɨːm	Nyaheun dum	Sapuan dɨm
Sre dum	Stieng dum		

h·h|WE (INCLUSIVE) · binary · 14 languages · enrichment 93× · emergent 1·1

Batek Deq Koh heʔ	Ceq Wong hɛʔ	Jahai Banun heʔ	Kensiw Kedah heʔ
Kintaq heʔ	Kme-2 ʔɛʔ.⁵³	Lanoh Kertei ʔɛʔ	Mang Ch ʔa.³¹_hi.³¹
Menriq Lah hiʔ	Semai Kampar heːʔ	Semaq Beri Brwe hɛh	Semnam Bal ʔɛːʔ
Temiar Kelantan ʔɛːʔ	Wa ʔeʔ.A		

s·k|FEATHER OR FUR OR HAIR · binary · 13 languages · enrichment 62× · emergent 1·4

BahnarPleiku sɔk	Cheng sɔk	Chut ŭsúk	Halang sɔk
Jahai Banun sɔk	Kasseng sok	Kha Poong usuk˧˥	Maleng Bro uso̯k
Malieng uʂŏk˧˥	Mel sɔk	Ramam tsak	Rengao tsə̤k
Sedang sak			

c·h|SICK · binary · 13 languages · enrichment 92× · emergent 1·1

Alak c’iʔ	BahnarPleiku ɟiʔ	Brao ɟiʔ	Cheng ɟiʔ
Cua ɟiʔ	Halang ɟɨʔ	Jeh ɟɨʔ	JruLaven ɟiʔ
Nyaheun ɟiʔ	Pusing cŏʔ	Ramam cuʔ	Rengao ɟɨʔ
Sapuan ɟiʔ			

k·t|SCRAPE · binary · 13 languages · enrichment 92× · emergent 4·1

Bulang khut.⁵¹	Chong khuːt.A	Kasong khuːt.B	Kme-2 khut.¹³
Nyakur khuːt	Paliu khɨt.⁵¹	Pear kat	Phong khuːt.²¹
Prai khuːt	Pusing kɔt	Rumai khiet.⁵⁵	Wa kɔt.A
Xu kuat.³³			

r·t|BUY · binary · 12 languages · enrichment 253× · emergent 3·1

Alak roːt	BahnarPleiku rɔt	Brao ruat	Cua rɔːt
Halang ruat	Jeh ruat	JruLaven ruat	Mel ruət
MnongRlm ruot	Nyaheun ruat	Rengao r.o̤ːt	Sre roat

l·k|CHOOSE · binary · 12 languages · enrichment 180× · emergent 3·4

Bulang lhək.³³	Chong lɨak.C	Kasong lɨək.D	Kha Poong alɨ:k.³³
Kme-2 lək.¹³	Nyakur lə:k	Phong liak.²¹	Prai lɨək
Pusing lɨak	Samre lɨək.C	Wa laɨk.B	Xu lək.³³

m·m|IRON · binary · 12 languages · enrichment 165×

Alak ma:m	BahnarPleiku ma:m	Brao ma:m	Cheng ma:m
Halang ma:m	Jeh ma:m	JruLaven mə:m	Kasseng ma:m
Nyaheun ma:m	Rengao ma:m	Sapuan ma:m	Sedang mɛ̰əm

t·c|SELL · binary · 12 languages · enrichment 194× · emergent 1·1

BahnarPleiku tɛc	Chrau tac	Cua tɛc	JruLaven tec
Kasseng tac	Mel tac	MnongRlm tac	Nyaheun tec
Prai thec	Sapuan tɛc	Sre tac	Stieng tac

k·c·t|kill · ternary · 12 languages · enrichment 169× · emergent 4·1·1

Ahlao kacə̰:t	Brao kəce:t	Cheng kce:t	Chut kacit
JruLaven kce:t	Lao-Katu kəcet	Malieng kacet.24	Malieng kacét
Nyakur kacét	Oy kəcet	Pakoh kəcet.T	Sapuan kacɛt

c·h|SOUR · binary · 11 languages · enrichment 76× · emergent 1·1

Ahlao cuʔ.B	Cheng ɟuʔ	Chong choʔ.C	Cua ɟuaʔ
Halang ɟɯəʔ	Iduh cɔ²ɛʔ	Jeh ɟɯaʔ	Kasseng ɟɔʔ
Kha Poong cɔ.³³ʔ	Rengao ɟo̤ʔ	Sapuan ɟuoʔ	