# Consonant Codes in 196 Trans-New Guinea Languages

The floor of the method, and what can be seen from there

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

## Abstract

We present a catalogue of 3,929 consonant codes grouping 33,781 lexical forms from 196 Nuclear Trans-New Guinea languages. A code is a consonant skeleton shared by words for the same concept across distinct languages, built with no etymology, no proto-forms and no cognacy judgements.

This is the family where the instrument performs worst in the whole series, and that is the reason for doing it. Its precision is 13 times above chance — against 216× for Turkic and 104× for Indo-European — and its grouping barely 1.4 times: 51.9% observed against 37.4% for the null. Corrected for chance, $\kappa = 0.232$.

The cause is measured and it is a single one: the short word. At 2.41 consonants per form the skeleton space saturates, and shuffling concepts within each language already groups more than a third of the forms. 35% of the resulting codes are a single consonant, and fourteen of the fifteen widest-reaching ones are too: n groups a hundred languages for *eat*, and as many again for *thou*, *I*, *mother* and *sleep*.

With this family the series reaches seven, and a regularity appears that six did not show: the null explodes below three consonants per word. The two families with arity under three — this one at 2.41 and Austronesian at 2.79 — have nulls of 37% and 44%; the five above 3.03 sit between 4.7% and 15%. Eight times the difference on either side of a narrow step.

And not everything collapses. The geographic signal is the best in the series: languages sharing twenty-one codes or more sit at 9% of the distance expected by chance, on a monotone curve without the reversal that spoils the Afro-Asiatic one. The emergent classes come out clean — an entirely labial and fricative series — and the 1,383 disjoint sibling-code pairs more than triple the Afro-Asiatic count.

## 1. What is sought, and why this family

The procedure — what a code is, in what order it is built, what nulls it uses and what beating them authorises — is described once for the whole series in "The catalogue of consonantal codes" and is not repeated here. 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 chosen for its low performance, not in spite of it. The corpus profile gave it 13 times chance, the lowest figure among the sixty evaluable families with sufficient material. An instrument is understood as well by seeing where it stops working as by seeing where it shines, and a described floor is worth more than an assumed one.

It also brings something Afro-Asiatic lacked: ten branches, six of them with seven or more languages, and the largest — Madang — at 51% rather than Chadic's 72%. The genealogical structure missing there is present here.

## 2. The population

196 languages · 94,306 forms · 895 concepts · 10 branches.

| Branch | Languages | Forms |
|---|---|---|
| Madang | 100 | 65,861 |
| Asmat-Awyu-Ok | 22 | 8,305 |
| Finisterre-Huon | 20 | 6,755 |
| Kainantu-Goroka | 20 | 5,491 |
| Greater Binanderean | 11 | 4,003 |
| Enga-Kewa-Huli | 10 | 1,997 |
| Dani | 7 | 899 |
| Mek | 4 | 537 |
| Chimbu-Wahgi · Paniai Lakes | 1 + 1 | 458 |

100% of the forms obtain a skeleton. A single source, without the 7% loss IDS cost in Afro-Asiatic or the 21% of Turkic Cyrillic. It is the cleanest population in the series.

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 two proposals revealed defects in the check

Phase 0 proposed two exclusions. Both are rejected.

**`Danaru`**, 634 forms. It was proposed for exclusion ”for marking tone in digits”. It has two forms with a superscript, and both are the same word: `sæg¹ro–`. That is 0.32%, and the report rounded it to ”0%”, so the proposal contradicted itself on its own line.

**`East Kewa`** against **`West Kewa`**, 73% identical forms. The mechanical rule this series publishes is *same glottocode, or more than 80%*. The glottocodes differ (`east2516` / `west2599`) and 73% is below the threshold. They are two distinct dialects and both stay in.

Both checks were corrected: the tone one now requires 5% before proposing, and the duplicate one only proposes above the threshold the series publishes. It is the third phase-0 check that proposed more than its evidence authorises — the first was the geographic one in Afro-Asiatic, which wanted to delete Semitic entirely — and the pattern deserves a name: a check that proposes a drastic action on thin evidence is worse than not having it, because its proposal is accepted by inertia.

### 2.2 A segmentation bug, found while building the normalisation table

Prenasalised stops written as a plain digraph — `nd`, `mb`, `ŋg` — were not split into two positions, while those written with a superscript — `ⁿd`, `ᵐb`, `ᵑg` — were. And three languages use both spellings at once: Enga has 35 digraph forms and 57 superscript ones, Merina Malagasy 80 and 57, Guaraní 15 and 107.

The result was that the same phoneme received two treatments within the same language: `monda–` gave `m·n·d` and `k.e.nd.e` gave `k·n`, losing the stop entirely. Corrected with an explicit table of the fifteen attested digraphs — without generalising to unseen sequences — over 478 forms of the corpus. Isolated over 20,000 forms, two change and none without a digraph.

## 3. What the instrument costs here, and it is what defines the paper

Mean arity: 2.41 consonants per form. The lowest in the series, below Austronesian (2.79) and far below Indo-European and Pama-Nyungan (3.35).

Collision: 57.2%. More than half the forms share a skeleton with another concept within their own language.

The two figures are the same thing said twice, and everything else follows from them.

### 3.1 The third kind of collision, and with it the series has them all

Turkic showed two. Collision of the language: *kül* 'ash' and *kül-* 'laugh' are already homophones in Old Turkic, so not even keeping the vowel would separate them. And collision manufactured by our own table, when the normalisation writes velar ŋ as n.

Here a third appears: collision of the concept list. Ten languages give a single word for the six ways of cutting the questionnaire asks about separately — CHOP, CUT, CUT (WITH AXE), CUT (WITH KNIFE), CUT (WITH SCISSORS), CUT DOWN: `san–` in Watiwa covers eight concepts, `o–` in Kare seven.

It is neither the language's nor ours: it belongs to whoever designed the questionnaire. And it behaves differently from the other two, because it varies across families according to how much each one colexifies.

### 3.2 And it must be measured before it is discounted, because it is not what it seems

Two opposite readings give the same symptom. (a) The list asks six nearly identical things and any language would answer with one word: an artefact of elicitation. (b) These languages genuinely colexify, and then it is a typological property to be reported, not discounted.

They are separated by asking the same questions in other families. Over the 89 concept clusters sharing a gloss head:

| Family | Clusters with $\geq$2 concepts | With a repeated skeleton |
|---|---|---|
| Nuclear Trans New Guinea | 3,055 | 41.8% |
| Indo-European | 4,004 | 23.8% |
| Afro-Asiatic | 1,808 | 9.9% |
| Pama-Nyungan | 1,756 | 9.6% |
| Austronesian | 7,036 | 8.4% |

If it were the list, all would give the same. They do not: this family collapses four or five times more. Hypothesis (a) alone does not explain the data.

But the control inverts the easy reading. Its words are short, so any two forms coincide more by chance. Comparing within cluster against outside, in the same language:

| Family | Within | Outside | Ratio |
|---|---|---|---|
| NTNG | 38.8% | 0.8% | 47× |
| Austronesian | 29.4% | 0.5% | 58× |
| Pama-Nyungan | 18.4% | 0.2% | 80× |
| Indo-European | 15.8% | 0.3% | 51× |
| Afro-Asiatic | 14.4% | 0.2% | 83× |

In absolute terms it colexifies more than any; relative to its own baseline, less than any. Its collision outside the cluster is already four times Afro-Asiatic's. The property is real and the driver is word length, not unusual semantics.

And it weighs little: of the observed collisions, only 9.9% occur entirely within a cluster. Discounting it would take collision from 29.1% to 26.3% — 2.8 points — and would hide a real property behind a supposed artefact. It is declared and not discounted.

## 4. The catalogue

3,929 codes over 895 concepts, grouping 33,781 of 92,608 forms (36.5%). Of these, 2,464 are binary or ternary and 1,375 are unary: 35.0%.

| Arity | Codes | % |
|---|---|---|
| 1 | 1,375 | 35.0% |
| 2 | 1,880 | 47.8% |
| 3 | 584 | 14.9% |
| 4 | 83 | 2.1% |
| 5+ | 7 | 0.2% |

Precision: 3.79% against 0.294% chance = 13× [13–13× across seeds]. Median enrichment of the binaries and ternaries: 16×, against 100× for Afro-Asiatic and 155× for Uralic.

And on grouping without the three-language threshold: 51.9% observed against 37.4% null — an advantage of 14.5 points, 1.4× chance, $\kappa = 0.232$.

### 4.1 The widest-reaching codes are a single consonant

The fifteen of widest reach, unfiltered:

| Code | Concept | Languages | Arity |
|---|---|---|---|
| n | *eat* | 100 | 1 |
| n | *your* (sg.) | 87 | 1 |
| n | *thee* | 75 | 1 |
| n | *thou* | 65 | 1 |
| n | *I* | 65 | 1 |
| n | *mother* | 60 | 1 |
| m | *taro* | 59 | 1 |
| m | *die* | 53 | 1 |
| m | *louse* | 51 | 1 |
| j | *my* | 51 | 1 |
| **m·m** | *father* | 48 | 2 |
| n | *sleep* | 46 | 1 |
| m | *breast* | 43 | 1 |
| m | *eye* | 41 | 1 |
| n | *his or her* | 38 | 1 |

Fourteen of the fifteen are unary, and n alone appears eight times for unrelated concepts: eat, your, thee, thou, I, mother, sleep, his-or-her.

This does not mean they are false. The review of the Afro-Asiatic paper showed a unary code can be a perfectly real etymon — Chadic l|cow is the lateral ɬ of *ɬa*. It means something harder to handle: the

catalogue cannot distinguish a genuine monoconsonantal etymon from a coincidence of availability, because both give it the same evidence, and in this family that indistinction takes the top of the list.

The group's only binary, **`m·m|father`** in 48 languages, is the labial reduplication pair that appears in every family of the series, and about which the Austronesian paper already warned: its shape is explained by child-rearing before history.

## 5. The series finding: the null explodes below three consonants

With seven families measured, a regularity appears that six did not show. Ordered by arity:

| Family | Arity | Groups | Null | Advantage | κ |
|---|---|---|---|---|---|
| Nuclear Trans New Guinea | 2.41 | 51.9% | 37.4% | 14.5 | 0.232 |
| Austronesian | 2.79 | 64.8% | 44.0% | 20.8 | 0.371 |
| Turkic | 3.03 | 56.9% | 15.0% | 41.9 | 0.493 |
| Afro-Asiatic | 3.16 | 29.7% | 4.7% | 25.0 | 0.262 |
| Uralic | 3.26 | 33.9% | 5.9% | 28.0 | 0.298 |
| Indo-European | 3.35 | 44.3% | 12.9% | 31.4 | 0.361 |
| Pama-Nyungan | 3.35 | 40.2% | 14.0% | 26.2 | 0.305 |

The two families below three consonants per word have nulls of 37% and 44%; the five above 3.03 sit between 4.7% and 15%. Eight times the difference on either side of a step a quarter of a consonant wide.

Until this work Austronesian was the series' only short-word case, and its high null could be read as a property of its own. With two cases it stops being an anecdote. With seven families it is not a law — and is not presented as one — but it is a regularity the series must declare.

The practical consequence is what matters. This family groups 51.9% of its forms, more than Afro-Asiatic (29.7%) and more than Uralic (33.9%). Read alone, that figure looks like a good result. Against its null it is worth 1.4 times chance, against Afro-Asiatic's 6.3. The raw grouping figure is not comparable across families without its null beside it, and this work is the clearest demonstration of why.

### 5.1 Automatic segmentation, and why it falls harder here

Automatic affix discovery — phase 2 of the protocol, which this series' catalogue does not use — raises grouping from 51.9% to 67.9%, sixteen points. And it raises the null by twenty-six, from 37.4% to 63.7%: after segmentation, two of every three forms group by chance.

| Step | In a group | Null | Advantage | κ | Codes of 4+ consonants |
|---|---|---|---|---|---|
| Whole word | 51.9% | 37.4% | 14.5 | 0.232 | 14.8% |
| + discovered affixes | 67.9% | 63.7% | 4.2 | 0.116 | 2.4% |

$\Delta\kappa = -0.118$, the second largest drop in the series after Austronesian ($-0.151$).

And here something must be said that the series itself later disproved. The first version of this section claimed that the order of the drops follows the same axis as the null — the shorter the word, the more segmentation destroys. Sino-Tibetan, measured afterwards, falsifies it: at arity 2.22, the lowest of the eight families, its drop is $-0.020$, the third smallest.

| Family | Arity | Δκ |
|---|---|---|
| Afro-Asiatic | 3.16 | −0.004 |
| Turkic | 2.85 | −0.009 |
| Sino-Tibetan | 2.22 | −0.020 |
| Uralic | 3.27 | −0.035 |
| Indo-European | 3.35 | −0.094 |
| Pama-Nyungan | 3.35 | −0.102 |
| Nuclear Trans New Guinea | 2.41 | −0.118 |
| Austronesian | 2.78 | −0.151 |

The two shortest-word families sit at opposite ends of the table, and the two longest-word ones in the middle. The association does not exist. What does stand is the §5 finding, which is about the catalogue's null and not about segmentation: there Sino-Tibetan confirms — arity 2.22, null 41.5%.

It is left written because the error is instructive: with seven families a pattern was visible that the eighth breaks, and the only reason it was visible is that the seven fitted in one table.

## 6. What the family does give, and gives better than any

### 6.1 Geography, with no U-turn

| Codes shared | Pairs | Mean distance | vs chance |
|---|---|---|---|
| 1 | 3,856 | 297 km | 0.93 |
| 2 | 1,432 | 201 km | 0.63 |
| 3–5 | 1,295 | 91 km | 0.29 |
| 6–10 | 503 | 42 km | 0.13 |
| 11–20 | 313 | 31 km | 0.10 |
| 21+ | 212 | 30 km | 0.09 |

Monotone to the end. Languages sharing twenty-one codes or more sit at 9% of the distance expected by chance, and the curve does not reverse in any band.

It is worth saying why, because it also explains someone else's failure. In Afro-Asiatic this same section rose again from eleven shared codes upward, and the cause was the uneven spread of the branches: Chadic compact, Semitic from Morocco to Oman. Here there is no dispersed branch — the whole family has a median of 95 km from its centre — and the curve falls clean. Compactness and geographic signal are the same property seen twice, and an extensive family cannot give this section however well its method works.

### 6.2 Substitution pairs: 1,383 disjoint candidates

Of 1,493 pairs of codes for the same concept differing in exactly one position, 1,383 are disjoint in languages — no language is in both — and only those can be proposed as an isogloss. They more than triple Afro-Asiatic's 450, and the reason is structure: ten branches against four.

And they must be read with a warning. Four of the seven widest-reaching pairs are TOBACCO: `k·b`, `k·h` and `k·p` against `k·s`, which reaches 33 languages. Tobacco arrived in New Guinea in the seventeenth century through trade. It is a modern loan the catalogue groups exactly as it groups any inheritance, like the cassava in Afro-Asiatic, and it heads the list. From 274 forms in 131 languages.

Those that are not tobacco: `g·n` `m·n` in *louse* (10 and 26 languages, crossing Madang, Finisterre-Huon and Mek), `n·n` `s·n` in *we*, `m·l` `m·r` in *vine*, `n·n` `t·n` in the oblique of *thou*.

### 6.3 Emergent classes: the labial and fricative series

The 196 languages are split in two halves, one to discover and one to check. The pairs surviving in both:

| Pair | Times (discover) | Ratio | Times (check) | Ratio |
|---|---|---|---|---|
| f ↔ ɸ | 80 | 9.2× | 91 | 10.3× |
| j ↔ z | 580 | 7.7× | 714 | 9.3× |
| v ↔ w | 745 | 6.2× | 958 | 7.8× |
| b ↔ f | 522 | 2.4× | 1,315 | 5.9× |
| v ↔ β | 69 | 5.5× | 60 | 4.7× |
| f ↔ w | 337 | 2.1× | 676 | 4.1× |
| b ↔ ɸ | 94 | 2.2× | 171 | 4.0× |

Six of the seven are labial, and the set reads as a single series: stop, fricative, approximant and their bilabial against labiodental variants. It is the most coherent result in the series — Afro-Asiatic gave a mix of voicing and manner spread across several places of articulation — and they are not regular correspondences: environment and direction are missing. They are candidates, with the place to look.

## 7. Limits

A single macroarea. The anomaly list of §6.1-bis does not exist, and not through a fault of the method but because the family occupies a continuous area.

35% of the catalogue is a single consonant, and the catalogue cannot distinguish the genuine mono-consonantal etymon from a coincidence of availability. The unary codes are counted and not read: the paper works with the 2,464 binaries and ternaries, 62.7%.

The labial-velars **kp** and **gb** merge into **k** and **g**. They are characteristic of this linguistic area, and the table strips their labial component. The first version of this paper left that "unmeasured" as the family's first pending task; the adversarial review measured it, and it costs little: 95 forms in 11 languages, touching 16 of the catalogue's 3,929 codes — 0.4% —, and in 15 of those 16 the merge is load-bearing, meaning the code gathers forms with `kp` alongside forms with plain `k`. Appendix A has a single affected code, `k·n|left`, with `kp` in 4 languages and `k` in 9.

The warning stands and the figure bounds it: the table erases a real contrast of this area, and it does so in less than half a percentage point of the catalogue. Declaring a limit is not the same as leaving it without a number.

Madang is 51% of the languages. Better balanced than Chadic's 72%, but still a dominant branch, and every figure for the family is weighted towards it.

Tobacco heads the substitution pairs and is a seventeenth-century loan. There is no way, within this method, to separate what arrived from what was inherited.

## 8. Conclusion

This family was done for its low performance, and it has delivered precisely that: 13 times chance, $\kappa = 0.232$, a null of 37.4% and 35% of the catalogue in single-consonant codes. It is the floor of the series, and it is now described rather than assumed.

What that floor teaches holds for the other six. The null explodes below three consonants per word, and with two families on that side of the step it can no longer be attributed to one of them. The

consequence is a reading rule for the whole series: this family groups 51.9% of its forms, more than Afro-Asiatic and more than Uralic, and is worth far less than either. The raw grouping figure says nothing without its null beside it.

And the floor is not a hole. The geographic signal is the best of the seven — 9% of the expected distance, with no reversal — the emergent classes come out as a coherent labial series, and the 1,383 disjoint pairs triple those of the previous family. The same compactness that sinks precision, because it compresses the skeleton space, is what makes geography legible.

An instrument is not understood by its best case. Austronesian looked like the hard case disguised as an easy one; this is the hard case undisguised, and it is the one that says where the limit lies and why.

## References


Brown, C. H., Holman, E. W., Wichmann, S. and Velupillai, V. (2008). Automated classification of the world's languages. *Sprachtypologie und Universalienforschung* 61.

Dolgopolsky, A. B. (1964). Gipoteza drevnejšego rodstva jazykovyx semej Severnoj Evrazii s verojatnostnoj točki zrenija. *Voprosy Jazykoznanija* 2.

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 $\kappa$]

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

List, J.-M. (2012). SCA: phonetic alignment based on sound classes. In *New Directions in Logic, Language and Computation*. Springer.

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*.

Pawley, A. and Hammarström, H. (2018). The Trans New Guinea family. In B. Palmer (ed.), *The Languages and Linguistics of the New Guinea Area*. De Gruyter Mouton. — [the reference classification]

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). *Integrative Corpus: a multi-layer lexical database of 225 macrosystems, with row-level provenance and its verification apparatus*. — [the data] · https://doi.org/10.13140/RG.2.2.31244.68486

Turchin, P., Peiros, I. and Gell-Mann, M. (2010). Analyzing genetic connections between languages by matching consonant classes. *Journal of Language Relationship* 3.


## Data and reproducibility

| | |
|---|---|
| Catalogue | `familias/nuclear-trans-new-guinea/output/catalogo-nuclear-trans-new-guinea.tsv` — 3,929 rows |
| Per-form detail | `…-lenguas.tsv` — 21,817 rows |
| Hygiene (§2.1) | `pipeline/fases/fase0_higiene.py` → `runs/fase0-higiene.md` |
| Data decisions (§2.1, §3) | `runs/fase0-decisiones.md` |
| Prenasalised stops (§2.2) | `corpus_integrativo/ingest/recompute_prenasal_digrafo.py` |
| Normalisation table (§3) | `pipeline/fases/tabla_normalizacion.py` → `runs/normalizacion.md` |
| Loading, population, warnings | `cod/familia.py` |
| Catalogue (§4) | `pipeline/fases/catalogo_familia.py` |
| Grouping and nulls (§4, §5) | `pipeline/fases/agrupamiento_familia.py` → `runs/agrupamiento.md` |
| $\kappa$ (§5) | `cod/metricas.py::kappa` |
| Automatic segmentation (§5.1) | `pipeline/fases/fase2_despiece.py` → `runs/fase2-despiece.md` |
| Geography (§6.1) | `pipeline/fases/codigos_vs_parentesco.py` → `runs/codigos-parentesco.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` |
| Substitutions by branch | `pipeline/fases/fase6_por_rama.py` → `runs/fase6-por-rama.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 3,929 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 carries no emergent-code column. The classes of §6.3 come out as a labial series — f↔ɸ, v↔w, b↔f, v↔β — and not as a closed partition of the inventory: the closure at ≥2× puts sixteen segments into a single class. §6.3 publishes them as candidates, which is what they are, and `ci/clases.py` declares an empty dictionary for this family.

## What this appendix selects, measured

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

| | catalogue | this appendix |
|---|---|---|
| binary codes | 47.8% | 93% |
| ternary codes | 14.9% | 7% |
| Madang as dominant branch | 66.3% | 85% |
| Asmat-Awyu-Ok as dominant branch | 10.3% | 0% |
| crossing four branches or more | 1.4% | 5% |
| mean enrichment | 69× | 43× |

Madang goes from 66% to 85%, and the catalogue's second branch, Asmat-Awyu-Ok, has not a single code here. §7 of the paper warns that "every figure of the family is weighted towards Madang"; in this appendix that weighting reaches five codes out of six.

And enrichment drops by a third. Reach and enrichment are bought with one another: a code reaching many languages is usually made of frequent consonants, which is exactly what enrichment penalises.

## What this appendix shows and the paper did not see: sixteen pronoun codes

Sixteen of the hundred codes are pronouns, and their distribution by skeleton is this:

| Skeleton | Codes | Concepts |
|---|---|---|
| n·n | 8 | they, their, oblique of *you*, we, oblique of *thee*, you (2), your (pl.) |
| n·g | 5 | he or she, oblique of *thee*, oblique of *you*, thou, you |
| s·n | 2 | we, our |
| t·n | 1 | oblique of *you* |

This deserves reading with both readings on the table, because both are defensible.

(a) Ross's Trans-New-Guinea classification is built precisely on pronoun paradigms, and the *na / ni / nu* series is its central material. That sixteen of this family's hundred widest-reaching codes are pronouns, and that thirteen of them begin with n, is the same observation reached another way — and this procedure consulted no tree.

(b) n is the most frequent consonant in this family's inventory: 34,037 positions, more than any other. n·n is, by construction, the most available skeleton that exists here. A short pronoun made of the commonest consonant is exactly what availability produces with no need of history.

The catalogue does not separate (a) from (b), and neither does this appendix. What can be said is the narrow thing: the enrichments of those sixteen run from 8× to 30×, the lower half of the list. A pronoun is not a good genealogical candidate *for this instrument*, even if it is one for the comparative method.

And there is a third, more uncomfortable warning. Nine of the sixteen depend on the table writing velar ŋ as n: n·n|they is n in 18 languages and ŋ in 8; n·n|their is 15 and 9; n·n|oblique of you is ŋ in 11 and n in 9. The pronoun evidence of this appendix rests on a merge in the normalisation.

## The table audit, and why it comes out cheap here

Twenty-three of the hundred codes have at least one position where the normalisation merges two segments that three or more languages write differently. In the Sino-Tibetan appendix it was sixty-eight. The difference is one of family: there s gathers eleven sibilant and affricate symbols; here the table is short and twenty-one of the twenty-three cases are the same merge, **n**/**ŋ**.

The other two: g·r|skin, with r in 16 languages and ř in 4, and k·n|left, with k in 9 and **kp** in 4. That is the appendix's only code touched by the labial-velar merge, the limit §7 of the paper declares and the adversarial review bounded at 0.4% of the catalogue.

## And something the Sino-Tibetan appendix cannot say

Not one of the 1,545 forms this appendix cites carries a single mark the skeleton ignores. Zero percent. In the Sino-Tibetan appendix they are 75%, almost all tone.

That is not a virtue of this family nor a defect of the other: it is that here what is seen written is what the code uses, and there it is not. When two forms in this appendix look alike, they look alike in the very thing the procedure joined them by. It is the only place in the series where that can be said without qualification.

## High enrichment is local, and here it shows in pure form

The ten highest-enrichment codes of this appendix:

| Code | Concept | Languages | Enrich. | Branches |
|---|---|---|---|---|
| g·n·m | possum | 13 | 309× | 1 |
| w·g·j | crocodile | 11 | 278× | 1 |
| g·l·m | drum | 16 | 262× | 1 |
| g·r·m | drum | 13 | 190× | 1 |
| s·p·r | axe | 12 | 170× | 1 |
| b·r·r | axe | 11 | 168× | 1 |
| f·n | fishing spear | 12 | 156× | 1 |
| g·g·r | grass-skirt | 10 | 155× | 1 |
| b·m | g-string | 23 | 127× | 1 |
| b·b | same-sex older sibling | 15 | 122× | 1 |

Ten of ten sit inside a single branch, and nine are material culture. Possum, crocodile, drum, axe, fishing spear, grass-skirt, g-string: objects and animals of one place, with names that are long and rare inside that branch's inventory.

That is what enrichment measures: distinctiveness, not depth. A local object with a specific name hits the maximum. The Sino-Tibetan paper found this by way of the loan — a Persian groundnut at 1086× — and here it comes by way of the endemic. It is the same phenomenon: enrichment rewards what resembles nothing else, and has no opinion about since when.

### And four codes are counted twice

`s·m` appears with 'pot' and with 'cooking pot', twelve languages each, and the twelve are the same. `n·g` appears with 'female' and 'female person', seventeen and seventeen, the same seventeen. `k·k` with 'carry' and 'carry on back', thirteen of fourteen shared. `m·n` with 'hair' and 'hair (body)', all ten shared.

Eight of this appendix's hundred entries are four codes written twice, because the questionnaire asks twice for the same thing. It is the list collision §3.2 of the paper measures over the forms, seen here in the catalogue of codes. §3.2 concludes it weighs little and is declared rather than discounted; this appendix shows how much it weighs when read top to bottom.

## Reverse index: codes recurring across concepts

Eighteen codes cover seventy-four of the hundred entries of the appendix; the table below shows the ten that recur three times or more. Three quarters of the blocks that follow repeat a skeleton already seen, and that is what §3's 61% collision means put into a list.

**`m·n`** takes twelve concepts: louse, breast, be, snake, stone, tongue, axe, hair (body), boy, mother, house and hair. No semantic reading sustains that, and none is attempted: `m` and `n` are the family's two most frequent consonants, and the enrichments here run from 5× to 12×, the lower half of the whole appendix.

**`n·n`** takes another twelve, and eight are pronouns: they, their, oblique of *you*, we, oblique of *thee*, you, your (pl.), thou. The other four — mother, sister, thumb, eat — are not. Those are the two extremes with the same row format: a code that appears to gather a paradigm and a code that gathers whatever fits into two frequent consonants. And the same code does both at once.

| Code | Concepts it appears with, and their enrichment |
|---|---|
| `m·n` | louse (26, 12×) · breast (23, 11×) · be (19, 16×) · snake (17, 8×) · stone (16, 8×) · tongue (14, 6×) · axe (13, 7×) · hair (body) (12, 5×) · boy (11, 6×) · mother (10, 5×) · house (10, 5×) · hair (10, 8×) |
| `n·n` | they (26, 23×) · their (23, 28×) · mother (22, 13×) · you (oblique case of you) (19, 14×) · we (17, 12×) · thee (oblique case of you) (13, 10×) · you (11, 11×) · your (plural) (11, 13×) · sister (10, 19×) · thumb (10, 8×) · eat (10, 6×) · thou (10, 8×) |
| `n·g` | female person (17, 25×) · female (17, 28×) · he or she (17, 29×) · sleep (16, 15×) · thee (oblique case of you) (15, 17×) · you (oblique case of you) (14, 15×) · thou (14, 16×) · you (11, 16×) |
| `n·m` | name (23, 18×) · breast (18, 15×) · tree (15, 13×) · bird (12, 10×) · louse (12, 10×) · mother (10, 8×) |
| `m·m` | father (48, 58×) · taro (15, 21×) · mother (12, 14×) · dog (10, 12×) |
| `k·n` | sleep (14, 12×) · left (14, 16×) · moon (11, 8×) · fill (10, 14×) |
| `k·m` | die (25, 28×) · bow (14, 23×) · liver (11, 14×) |
| `m·t` | sharpen (something) (16, 40×) · snake (11, 16×) · nose (10, 14×) |
| `g·n` | skin (12, 21×) · how many (12, 38×) · louse (10, 17×) |
| `s·m` | pot (12, 78×) · cooking pot (12, 89×) · vulva (12, 38×) |

## Body parts

Twelve codes, and six of the twelve come in pairs for the same concept: `m·n` and `n·m` for ’breast’, `g·r` and `g·n` for ’skin’, `g·m` and `k·m` for ’liver’. These are not list duplicates — they are distinct codes for the same concept — and that is why they go to §6.2 as candidate isoglosses.

The enrichments are low: `m·n|tongue` 6×, `m·n|hair` 8×, `m·n|breast` 11×. `d·h|ear`, at 65×, is the exception and belongs to a single branch.

**`m·n|BREAST`** · binary · 23 languages · enrichment 11×

| | | | |
|---|---|---|---|
| Alekano amina | Amaimon amun | Anam amaŋ | Atemble aman |
| Bepour mɛna | Faita aməna | Gende ami–ne | Kesawai amne |
| Kuman amu–na | Kâte moŋ– | Mape moŋ– | Moere mɛna |
| Pal mɛŋ | Sausi amene | Tauya amena | Tokano amina |

**`g·m|liver`** · binary · 22 languages · enrichment 43×

| | | | |
|---|---|---|---|
| Amele gema– | Bagupi geme– | Gal gami– | Gumalu gemʌ– |
| Hunjara–Kaina Ke gomo | Korak –gom | Murupi gʌm– | Nake gam– |
| Orokaiva gomo | Panim gemʌ– | Rempi gʌma– | Saruga gam– |
| Sinsauru gʌma | Siroi gamu | Utu gemu– | Wamas game– |

**`g·r|skin`** · binary · 20 languages · enrichment 26×

| | | | |
|---|---|---|---|
| Anjam gaɽa | Bagupi garʌ– | Baimak gara– | Bongu gare |
| Lemio geɽ | Matepi gare– | Mawan gari– | Mosimo gara– |
| Nake gara– | Nobonob gaɽa– | Rapting gara– | Rempi gʌɽa |
| Saruga gara– | Siroi garo | Utu gara | Wagi gara |

**`n·m|BREAST`** · binary · 18 languages · enrichment 15×

| | | | |
|---|---|---|---|
| Dedua namu– | Gadsup naami | Kambaira nami | Komba nam– |
| Kube namu– | Mesem nam– | Mungkip nom | Nabak nam– |
| Nek num | Nuk nom | Ono ŋamu– | Selepet nam– |
| Tairora naama | Tairora nama | Timbe nam– | Tobo namo– |

**`m·n|tongue`** · binary · 14 languages · enrichment 6×

| | | | |
|---|---|---|---|
| Amowe muna | Anjam meŋ | Asmat emene | Bongu muin |
| Girawa mino– | Kein minɔ– | Male moin | Migum muni |
| Saep men– | Sam men | Siroi mane | Sumau mino– |
| Uya mino– | Yabong meŋu– | | |

**`d·h|ear`** · binary · 12 languages · enrichment 65×

| | | | |
|---|---|---|---|
| Amele dahi– | Bagupi dahu– | Gal dahu– | Isebe dahi– |
| Kein dahu– | Matepi dahu– | Mawan dahu– | Munit dahi– |
| Nake dahu– | Panim dahi– | Saruga dah– | Silopi dahu– |

**`g·t|head`** · binary · 12 languages · enrichment 31×

```
Alekano góto  Anjam gate   Bongu gate      Ewage–Notu giti
Isebe gati–   Male gati    Orokaiva _giti  Panim gati–
Suena giti    Yegha 'giti  Yekora giti_    Zia gitau
```

**g·n|skin** · binary · 12 languages · enrichment 21×

```
Amowe gɑn–a  Bilakura guan  Gumalu ganau–    Isebe gʌna–
Panim gana–  Parawen goʌna  Ukuriguma guʌːn  Usan goʌn
Wagi gone–   Waskia goaŋ    Yaben guʌːnu     Yarawata guana
```

**k·m|liver** · binary · 11 languages · enrichment 14×

```
Aeka komo     Asas kame–    Bau kema–   Hunjara–Kaina Ke komo
Mauwake kema  Mawak kema    Moere kema  Sausi kamo
Sihan kema–   Silopi kemu–  Sumau kʌma
```

**t·n|belly** · binary · 10 languages · enrichment 18×

```
Doghoro 'tini  Ewage–Notu tini  Gaina 'tini  Hunjara–Kaina Ke tini
Korak –taŋu    Orokaiva tini    Suena tini   Yegha 'tini
Yekora tini    Zia tini
```

**m·t|nose** · binary · 10 languages · enrichment 14×

```
Amaimon meta  Amele mete–   Amowe mutu  Bau mete–
Faita mota    Gumalu mete–  Maia –met   Nakai 'mitɔ
Ogea mutu–    Uya mutu–
```

**m·n|HAIR** · binary · 10 languages · enrichment 8×

```
Apali mʉnʉ    Korak –munɛ     Mala –imun    Miani –umun
Mum mʉnʉ      Parawen muna    Sileibi mɨnə  Utarmbung mɨnɨ–
Waskia moːnɛ  Yarawata munia
```

## Kinship and person

Seventeen codes, the appendix's most populated section, and eight of them are pronouns. This appendix's header discusses them together because they do not read separately.

The two heading the section are the ones to read most carefully. **m·m|father** in 48 languages is the widest-reaching code of the whole family, and §4.1 of the paper already warns: a reduplicated labial nasal for 'father' is explained by child-rearing before history, and appears in unrelated families. **m·m|mother**, twelve languages, is the same case. And there are three further codes for 'mother' — n·n, m·n, n·m — with enrichments from 5× to 13×.

Four codes for 'mother' in the same appendix is not a result about the family: it is what happens when the concept is said with two very common consonants.

**m·m|FATHER** · binary · 48 languages · enrichment 58×

```
Aeka mama    Bagupi meme–  Bau meme–    Ewage–Notu mama
Ganglau mʌm  Gumalu meme–  Kare momo    Matepi memu–
Munit meme–  Nobonob mam   Panim meme–  Rempi mam–
Sihan meme–  Suena mama    Wagi mame–   Yekora mama
```

**n·n|THEY** · binary · 26 languages · enrichment 23×

```
Anam nʉ–ŋ     Anamgura nʉŋ  Asas nɛnɛ       Danaru nɛn
Ganglau naŋ   Kesawai nɛnɛ  Kube iŋeŋ       Migum nʌne
Saep naŋ      Sausi nɛnɛ    Siroi na–nɛ     Somba–Siawari yeŋân
Urigina nɛnɛ  Uya nɛn       Wadaginam nɨ–ŋ  Yabong nʌ–ŋ
```

**w·n|name** · binary · 25 languages · enrichment 32×

```
Asas weni      Baimak wani–  Biyom wa:ni   Faiwol win
Gumalu weni–   Kein wani–    Lemio weñi    Matepi wani–
Nobonob wani–  Setaman win   Silopi wani–  Sumau wen–
Tangko wi'ni   Tifal win     Uya woni–     Waube we:ni
```

**n·m|name** · binary · 23 languages · enrichment 18×

```
Amaimon unim    Anjam ñam    Bepour unum  Brem unim
Kobol unim      Maia –inim   Mala –num    Mauwake unum
Mawak unim      Miani –inum  Musar unim   Pamosu unim
Ukuriguma unim  Usan unum    Yaben uŋim   Yangulam ñem
```

**n·n|MOTHER** · binary · 22 languages · enrichment 13×

```
Amaimon nʌnia  Bargam –nen  Bilakura niña  Gavak naŋ
Jilim ñaŋ      Kare nan–    Mauwake nene   Nakai nãŋ
Parawen nain   Rerau ñaŋ    Setaman nein   Ukuriguma nʌ:n
Usan nʌni      Waskia naŋ   Yabong nono–   Yarawata naini
```

**n·n|WE** · binary · 17 languages · enrichment 12×

```
Amaimon ini–ŋi  Bilakura eɲ–ina     Borong nono  Dedua nini
Eipomek nun     Ketengban nun       Komba nen    Kube nini
Kuman nono      Mape ŋâŋu           Nabak nin    Parawen ina–na
Selepet nen     Somba–Siawari nini  Timbe nen    Tobo nini
```

**s·n|WE** · binary · 17 languages · enrichment 22×

```
Arawum sine   Asas sene      Biyom sina   Danaru sen
Kesawai sene  Lemio sine     Migum sine   Rerau si–ni
Sausi sene    Sinsauru sene  Siroi si–ne  Sop sin
Sumau sini    Urigina seno   Waube sini   Yabong si–ŋ
```

**n·g|HE OR SHE** · binary · 17 languages · enrichment 29×

```
Bagupi nug   Gal nug          Ganglau ngu  Garus nug
Matepi nug   Mosimo nugo      Murupi nuga  Nake nug
Nobonob nug  Rapting nugo     Samosa nugo  Saruga nuga
Silopi nug   Utarmbung nʉ–gᵊ  Wagi nug     Wamas nugo
```

**d·n|man** · binary · 15 languages · enrichment 35×

```
Amele dana   Bau dana     Gumalu dana  Isebe dʌna
Matepi dʌna  Mosimo dʌna  Murupi danʌ  Nake dan
Panim dana   Samosa dana  Saruga dʌna  Sihan dana
Silopi danʌ  Utu dana     Wamas dana
```

**n·g|THOU** · binary · 14 languages · enrichment 16×

```
Angal nge    Bagupi neg   Garus neg   Matepi neg
Mosimo nʌg   Murupi naga  Nake nʌg    Rapting nag
Samosa nʌge  Saruga nʌga  Silopi neg  Utarmbung na-gə
Wamas nage   Yoidik nʌg
```

**m·m|MOTHER** · binary · 12 languages · enrichment 14×

```
Bepour imem  Enga mamea    Komba mam-  Korak -mam
Kowaki -mem  Kuman mam     Maia mama   Mesem mam-
Nabak mam-   Selepet momo  Tauya mɛmɛ  Timbe meme-
```

**n·g|YOU** · binary · 11 languages · enrichment 16×

```
Amaimon ani-g   Anjam -ŋg-  Asas -nVg-    Bongu -iŋg-
Ganglau -nag-   Male -ŋg-   Migum -ng-    Sam -ŋg-
Sinsauru -neg-  Sop -nig-   Waube -nigi-
```

**n·n|YOU** · binary · 11 languages · enrichment 11×

```
Anam na-ŋʷ-       Anamgura na-ŋ-    Dedua ŋeni     Faita na-nʉ-
Kâte ŋoŋe         Mape ŋoŋai        Ono ŋine       Somba-Siawari iŋiŋi
Utarmbung na-ŋi-  Wadaginam na-ŋɨ-  Waskia nV-ŋa-
```

**m·n|MOTHER** · binary · 10 languages · enrichment 5×

```
Anam -mna       Borong maŋ-  Mungkip mēŋ    Nakame meŋi
Nek meɛŋ        Nuk menɨ     Paynamar amɨŋ  Som miiŋi
Utarmbung -mʉŋ  Yabong munu
```

**n·m|MOTHER** · binary · 10 languages · enrichment 8×

```
Atemble ŋam         Awa ʌnómʌ́    Biyom -ainam  Borong namo
Danaru anʌm         Kalam -nm    Miani namu    Moere anʌm
Somba-Siawari nam-  Waskia niam
```

**n·n|SISTER** · binary · 10 languages · enrichment 19×

```
Awa nane            Awiyaana nanoa   Borong naaŋ-  Fore noná
Gadsup nanoi        Inoke-Yate núna  Kamano núna   Kube naoŋ-
Somba-Siawari nen-  Tobo nauŋ-
```

**n·n|THOU** · binary · 10 languages · enrichment 8×

```
Amaimon ne-ŋi    Bilakura na-na  Brem ne-ne  Gavak ne-ŋ
Kobol no-ŋ       Moere ne-ne     Pal nɛ-ŋ    Parawen na-na
Ukuriguma ne-na  Yarawata na-na
```

## Nature and sky

Only three codes — `m·n|stone`, `j·g|water`, `k·n|moon` — a small section for what it usually is in this series. This family's nature vocabulary enters the catalogue as unary codes, which this appendix does not take.

**`m·n|stone`** · binary · 16 languages · enrichment 8×

```
Amele meen   Baimak men   Bau men      Gal men
Ganglau miŋ  Gumalu men   Isebe men    Jilim mʌ:nu
Munit meni   Nobonob men  Panim men    Rerau mʌ:nu
Saep muŋ     Sihan me:n   Waskia mʌŋa  Watiwa mun
```

**`j·g|water`** · binary · 13 languages · enrichment 38×

```
Amowe jɑgɛ      Bilakura yuag   Ganglau yag  Ogea yʌge
Parawen yʌga    Rerau yag       Saep yʌg     Sam jag
Sop yʌg         Ukuriguma yʌ:g  Yaben yagu   Yabong yag
Yarawata yʌ:ga
```

**`k·n|moon`** · binary · 11 languages · enrichment 8×

```
Abaga ikenai  Alekano ikani  Borong koiŋ         Dano iké'ni
Enga kana     Gadsup ikona   Isabi kine          Komba kain
Kyaka kana    Siane ikana    Somba-Siawari kain
```

## Animals

Nine codes, and three are for 'louse' — `m·n` (26 languages), `n·m` (12) and `g·n` (10). §6.2 of the paper uses precisely `g·n` `m·n` in louse as its best substitution pair that is not tobacco, crossing Madang, Finisterre-Huon and Mek.

There are also two codes for 'snake' and two for 'bird'. The whole section is isogloss material, which is what this family gives better than any other in the series.

**`m·n|louse`** · binary · 26 languages · enrichment 12×

```
Anjam mana   Apali iman    Bongu mana          Borong imoŋ
Jilim miɲi   Komba imen    Kube imin           Mape imeŋ
Mungkip mīŋ  Nabak emin    Nend eman           Nuk ɛmɨn
Rerau muñi   Selepet imen  Somba-Siawari imin  Tobo imin
```

**`m·n|snake`** · binary · 17 languages · enrichment 8×

```
Amowe mɑni  Asmat mini   Asmat mini     Asmat mini
Bau man     Danaru man   Girawa mɛ:n    Gumalu ma:ni
Isebe man   Kein man     Munit mani     Ogea mʌ:ni
Panim man   Sihan ma:ni  Urigina mʌ:ni  Usan ma:ni
```

**`n·m|bird`** · binary · 12 languages · enrichment 10×

```
Alekano nama  Benabena namá    Dano nama     Gadsup numi
Gimi nímí     Inoke-Yate namá  Kamano nämá   Kambaira numi
Siane nema    Tokano nama      Yagaria namá  Yaweyuha namá
```

**n·m|louse** · binary · 12 languages · enrichment 10×

```
Benabena namá    Fore numa:    Gadsup numi    Gimi nimi
Inoke-Yate namá  Isabi nʌma    Kambaira numi  Mum ñima
Siane nema       Tairora numa  Yagaria namá'  Yaweyuha nama
```

**m·t|snake** · binary · 11 languages · enrichment 16×

```
Baimak mat  Gal mat      Mala mat    Mawan mat
Miani mwat  Nobonob mat  Sop mʌːtʰ   Sumau mʌːtu
Wagi mat    Wamas mat    Waskia mot
```

**m·r|CHICKEN** · binary · 11 languages · enrichment 22×

```
Bagupi mɛru  Mawan maru  Mosimo mař̆u   Murupi marua
Nake maru    Sam muru    Samosa maruɛ  Saruga maru
Silopi mɛru  Utu mɛru    Wagi maru
```

**j·n|bird** · binary · 10 languages · enrichment 21×

```
Asas yene    Kesawai yene  Lemio yen   Migum yen
Mungkip jāŋ  Nakame jɑŋ    Sausi yene  Sinsauru yene
Tauya yene   Uri jɑŋ
```

**m·m|dog** · binary · 10 languages · enrichment 12×

```
Amowe mɑmo  Biyom mʌm    Jilim mam   Rerau mʌːmu
Sop mʌm     Sumau mam    Tauya mʌmo  Urigina mʌm
Uya mʌːm    Waube maːmu
```

**g·n|louse** · binary · 10 languages · enrichment 17×

```
Amaimon gunu  Bilakura guːn  Brem guːn       Gavak igun
Malas gunu    Parawen guːnu  Ukuriguma guːn  Usan guːn
Yaben gunu    Yarawata guːn
```

## Plants and food

Two codes: n·m|tree and n·n|eat, at 13× and 6×. The family's crops — taro, banana, sago, tobacco — are under "Other" because their glosses are not on the thematic list, and that is where the tobacco sits that §6.2 turns into the paper's central warning.

**n·m|tree** · binary · 15 languages · enrichment 13×

```
Amaimon ŋʌmu  Anjam ŋam   Bepour nʌmʌ     Bilakura nuom
Bongu aŋam    Gavak ŋam   Kobol neim      Male ŋam
Miani nʌm     Moere namʌ  Pal nʌːm        Parawen nʌːma
Usan nʌm      Yaben namu  Yarawata nʌːma
```

**n·n|eat** · binary · 10 languages · enrichment 6×

```
Awyu nin       Dani nan    Dani nan    Gadsup naano
Kalam ñŋ-      Kobon ñɨŋ-  Mesem nin-  Mungkip nʌna
Tairora naana  Walak nan
```

## Verbs of action and state

Five codes. `k·m|die` in 25 languages and four branches is the best of the section; `n·l|vomit` (87×) and `d·z|burn` (85×) are high-enrichment and single-branch, which is the local pattern the header describes.

Two different codes for 'sleep' — `n·g` in 16 languages and `k·n` in 14 — and both cross two branches.

**`k·m|die`** · binary · 25 languages · enrichment 28×

```
Aisi kum–     Arawum kumo–  Borong komu–  Danaru kume–
Enga kumi     Faita kum–    Kalam kum–    Lemio kumo–
Mape kâme–    Mum kum–      Pulabu kum–   Sileibi kumu–
Siroi kumi–   Sop kum–      Tobo kâmu–    Wambon kim–
```

**`n·g|sleep`** · binary · 16 languages · enrichment 15×

```
Bagupi nog–   Baimak unug–  Dano óng      Gal nigi–
Garus nig–    Matepi nig–   Mawan nug–    Mosimo nig–
Murupi nigi   Rempi neg–    Samosa nigu–  Saruga nag–
Silopi nog–   Utu nog–      Wamas nig–    Yoidik nig–
```

**`k·n|sleep`** · binary · 14 languages · enrichment 12×

```
Arawum kiŋi–  Biyom akin     Danaru kin–  Faita akun–
Jilim kin–    Lemio kiŋi–    Migum kiŋʌ–  Rerau kin–
Siroi kini–   Sop kin–       Tifal akaan  Urigina kina–
Watiwa kini–  Yangulam kiŋ–
```

**`n·l|vomit`** · binary · 12 languages · enrichment 87×

```
Anjam ŋil–    Arawum ŋal–   Bargam nil–  Bongu ŋil–
Jilim ŋiel–   Male ŋil–     Rerau ŋel–   Sam ŋil–
Siroi ŋule–   Sop nil–      Waube nili–  Yangulam ŋali–
```

**`d·z|burn`** · binary · 10 languages · enrichment 85×

```
Borong dze–          Dedua dza–  Kube dze–   Kâte dza–
Mape dza–            Mesem dzi–  Nabak dzi–  Ono dzi–
Somba–Siawari dze–   Tobo dze–
```

## Objects and culture

Four codes, and three are for 'axe': `m·n` (13 languages, 7×), `s·p·r` (12, 170×) and `b·r·r` (11, 168×).

It is the appendix's best example of what an enrichment means and what it does not. All three name the same object in the same family. The first is a binary of frequent consonants and distinguishes nothing; the other two are rare ternaries marking their zone. The difference between 7× and 170× is not antiquity or importance: it is how much that skeleton resembles the rest of the inventory.

**`m·n|AXE`** · binary · 13 languages · enrichment 7×

```
Amele mana   Danaru mʌn   Girawa mʌ:n   Gumalu manau
Isebe manu   Kein mʌn     Panim manu    Sihan mana
Sop mʌn      Sumau mʌ:ni  Urigina mʌ:n  Uya mon
Waube mʌ:nu
```

**s·p·r|AXE** · ternary · 12 languages · enrichment 170×

```
Gal sapor        Gumalu sapɔr  Jilim sapor  Kein sapor
Male sapoř       Matepi sapɔr  Mawan sʌpor  Munit sapɔr
Sileibi sa:puri  Silopi sapor  Siroi sapor  Utu sapor
```

**b·r·r|AXE** · ternary · 11 languages · enrichment 168×

```
Bepour buřiř  Brem buřiř   Korak buřiř   Kowaki buřiř
Maia biřiř    Mala buřiř   Malas buřiř   Mauwake buřiř
Mawak buřiř   Moere buřiř  Waskia buřiř
```

**m·n|house** · binary · 10 languages · enrichment 5×

```
Amaimon mɔne  Bepour mun     Bilakura muñe  Brem amun
Malas amun    Parawen moʌna  Ukuriguma mon  Usan mon
Yaben muñi    Yarawata mɔno
```

## Qualities

A single code, k·n|left, and it turns out to be the appendix's only code touched by the labial-velar merge: k in 9 languages and kp in 4. The limit §7 of the paper declared without a number has its one visible case here.

**k·n|left** · binary · 14 languages · enrichment 16×

```
Amowe kena  Bongu kwain  Dedua kana    Komba kane
Kube kana   Kâte kpana   Male kwen     Mape kpana
Mesem kana  Ogea kena    Selepet kane  Somba–Siawari kani
Timbe kane  Tobo kpana
```

## Other

Forty-seven codes. As in the Sino-Tibetan appendix, this is not a residual category: it is the part of the vocabulary the thematic list — written for basic vocabulary — does not contemplate, and in this family that is mostly objects, plus eight oblique pronouns the list does not recognise as such.

Here is **k·s|tobacco**, 33 languages, the second widest-reaching code of the whole family, which §6.2 of the paper uses as its central warning: tobacco reached New Guinea in the seventeenth century and the catalogue groups it exactly as it groups any inheritance.

And here are the ten highest-enrichment codes of the header's table — possum, crocodile, drum, axe, fishing spear, grass-skirt, g-string — all from a single branch. Read this section knowing that its ceiling by reach is a seventeenth-century loan and its ceiling by enrichment is a list of objects from one branch. Neither is what a reader is looking for in a catalogue of codes, and the catalogue has nothing with which to set them aside.

**k·s|TOBACCO** · binary · 33 languages · enrichment 74×

| | | | |
|---|---|---|---|
| Amowe kaso | Arawum kʌs | Baimak kas | Bongu kas |
| Gumalu kas | Kare kʌːsi | Lemio kʰʌs | Matepi kʌs |
| Munit kasa | Panim kʌs | Rempi kʌs | Sam kas |
| Sausi kʌsi | Silopi kas | Urigina kaːsi | Wagi kʌs |

**b·m|G–STRING** · binary · 23 languages · enrichment 127×

| | | | |
|---|---|---|---|
| Amele bam | Bagupi baːm | Baimak bam | Gal bam |
| Garus baim | Kein baːm | Matepi bam | Mosimo baːm |
| Nake baːm | Nobonob bam | Rapting baim | Rempi baim |
| Saruga baːme | Utu baːm | Wamas baːm | Yaben bamu |

**n·n|THEIR** · binary · 23 languages · enrichment 28×

| | | | |
|---|---|---|---|
| Anam nʉ–ŋa– | Anamgura nʉ–ŋ | Asas nɛn– | Faita nʉ–nʉ– |
| Ganglau naŋ– | Lemio nʌn– | Migum nʌn– | Mum nuŋu– |
| Rerau nani– | Saep naŋ– | Sinsauru nɛnɛ– | Siroi nanɛ– |
| Tauya nʉ–nʉ– | Urigina nɛn– | Wadaginam nɨ–ŋa– | Waskia nu–na |

**m·g|BANANA** · binary · 20 languages · enrichment 39×

| | | | |
|---|---|---|---|
| Bagupi mug | Baimak muːg | Bongu moga | Gal mug |
| Mawan mug | Mosimo mugu | Murupi mugu | Nake mug |
| Rapting mug | Samosa mogu | Saruga muːgu | Silopi mug |
| Utu mug | Uya mug | Wagi mug | Wamas mug |

**m·r|VINE** · binary · 19 languages · enrichment 29×

| | | | |
|---|---|---|---|
| Asas muru | Baimak muri | Gal muri | Gavak miř |
| Isebe mʌr | Matepi muri | Munit muri | Murupi muri |
| Nake muri | Nobonob muri | Samosa muri | Silopi muri |
| Sop muri | Sumau muri | Utu muri | Uya muːri |

**b·m|SAGO** · binary · 19 languages · enrichment 88×

| | | | |
|---|---|---|---|
| Anjam bem | Arawum bomie | Bongu boam | Ganglau bum |
| Garus bom | Jilim bom | Migum bom | Nobonob bom |
| Ogea boma | Pulabu bome | Rempi bom | Sam bom |
| Siroi bem | Sop bomo | Sumau bomu | Waskia bom |

**n·n|YOU (OBLIQUE CASE OF YOU)** · binary · 19 languages · enrichment 14×

| | | | |
|---|---|---|---|
| Amaimon ani–ŋi | Anam na–ŋ | Anamgura na–ŋ | Bilakura añ–ina |
| Dedua ŋeni | Faita na–nʉ | Kâte ŋoŋe | Malas na–n |
| Mape ŋoŋai | Migabac ŋeŋe | Ono ŋine | Sop nin |
| Utarmbung na–ŋ | Wadaginam na–ŋ | Waskia nV–ŋa– | Yabong nɔ–ŋ |

**m·n|BE** · binary · 19 languages · enrichment 16×

| | | | |
|---|---|---|---|
| Amowe mini– | Arawum məna– | Asas mɛn– | Biyom min– |
| Danaru min– | Isabi men– | Lemio mən– | Migum min– |
| Rerau min– | Sausi mɛnɛ– | Sinsauru mɛn– | Tauya mɛn |
| Urigina mᵊn– | Watiwa man– | Waube men– | Yabong man– |

**n·g|FEMALE PERSON** · binary · 17 languages · enrichment 25×

| | | | |
|---|---|---|---|
| Bagupi enʌg | Gal anag | Isebe anʌg | Kein ana:g |
| Matepi anʌg | Mawan anʌg | Murupi anʌ:gu | Nake anʌ:g |
| Panim anʌg | Rapting anʌg | Samosa anʌ:gu | Saruga anʌ:gu |
| Sihan anʌgu | Silopi anʌg | Utu anʌg | Wagi anig |

**n·g|FEMALE** · binary · 17 languages · enrichment 28×

| | | | |
|---|---|---|---|
| Bagupi enʌg | Gal anag | Isebe anʌg | Kein ana:g |
| Matepi anʌg | Mawan anʌg | Murupi anʌ:gu | Nake anʌ:g |
| Panim anʌg | Rapting anʌg | Samosa anʌ:gu | Saruga anʌ:gu |
| Sihan anʌgu | Silopi anʌg | Utu anʌg | Wagi anig |

**g·l·m|DRUM** · ternary · 16 languages · enrichment 262×

| | | | |
|---|---|---|---|
| Amele gilem | Arawum gilʌm | Bagupi gilʌm | Bilakura gilamu |
| Ganglau gilam | Garus gilʌm | Matepi gilom | Murupi gilʌ:m |
| Parawen gilam | Rapting gilam | Rempi gi:lam | Saep gilam |
| Saruga gilama | Sop giloumo | Ukuriguma gilamu | Yarawata gilama |

**w·t|FILL** · binary · 16 languages · enrichment 87×

| | | | |
|---|---|---|---|
| Amele wete | Bagupi wete- | Baimak wete- | Bau wet- |
| Isebe weti- | Mosimo wete- | Munit wete- | Murupi wete- |
| Nake wete- | Panim wete- | Rapting wet- | Saruga wete- |
| Sihan wete- | Silopi wet- | Wamas wete- | Waskia wate- |

**m·t|SHARPEN (SOMETHING)** · binary · 16 languages · enrichment 40×

| | | | |
|---|---|---|---|
| Amele mete- | Baimak mete | Bau metʰ | Gal mete |
| Garus metʰ | Gumalu mati | Isebe meti | Mawan mete |
| Mosimo mete | Panim mete | Parawen imut- | Rapting met |
| Sihan mati | Wagi met- | Wamas mete | Yoidik met |

**m·m|TARO** · binary · 15 languages · enrichment 21×

| | | | |
|---|---|---|---|
| Amaimon ma:mu | Bilakura mwam | Brem mam | Gavak mom |
| Maia mam | Mala mam | Malas mamu | Mauwake moma |
| Miani mam | Pal mam | Parawen ma:ma | Tauya moma |
| Ukuriguma mam | Yaben ma:mu | Yarawata ma:mo | |

**n·g|THEE (OBLIQUE CASE OF YOU)** · binary · 15 languages · enrichment 17×

| | | | |
|---|---|---|---|
| Bagupi neg | Garus neg | Mandobo Atas ŋgo | Matepi neg |
| Mosimo nʌg | Murupi naga | Nake nʌg | Rapting nag |
| Samosa nʌge | Saruga nʌga | Silopi neg | Utarmbung na-gə |
| Wamas nage | Wambon ŋgu | Yoidik nʌg | |

**b·b|SAME-SEX OLDER SIBLING** · binary · 15 languages · enrichment 122×

```
Amowe bʌbo  Anjam babo   Bargam –bab    Ganglau bʌba–
Garus bab   Kobol bab    Maia –bab      Mala –bab
Malas baba  Miani baba   Ogea –babo     Parawen baba
Rempi bʌba  Yabong bʌba  Yarawata babo
```

**m·l|VINE** · binary · 14 languages · enrichment 34×

```
Amowe mɑlu     Bagupi muli   Garus mul_i  Jilim mul
Migum moli     Mosimo muli   Ogea mʌ:lu   Rapting muil
Rempi mul_i    Rerau mul     Siroi muli   Waskia mʌil
Yangulam muli  Yoidik mul_i
```

**n·g|YOU (OBLIQUE CASE OF YOU)** · binary · 14 languages · enrichment 15×

```
Aeka iŋgo    Amaimon ani–g    Anjam –ŋg–     Asas –nVg–
Bongu –iŋg–  Ewage–Notu iŋgo  Ganglau –nag–  Kobol no–g
Male –ŋg–    Migum –ng–       Sam –ŋg–       Sinsauru –neg–
Sop –nig–    Waube –nigi–
```

**k·k|carry** · binary · 14 languages · enrichment 34×

```
Awyu a=keku–  Bagupi kuki–  Nake kuku–     Nobonob kuk–
Pal kʌki      Panim kuku–   Rapting kuku–  Rempi kuku–
Saruga kuki–  Sausi kakau–  Urigina kuku–  Utu koki–
Wagi kuku–    Waube koko–
```

**k·m|BOW** · binary · 14 languages · enrichment 23×

```
Aisi kim        Amaimon ke:mi   Bilakura kami  Faita kim
Gende kimi      Kein kem        Maia kumu      Malas ki:mu
Mum kɨm         Parawen kami    Siane kimi     Sileibi kimi
Ukuriguma komi  Yarawata kamia
```

**m·l|G–STRING** · binary · 14 languages · enrichment 43×

```
Amowe mɑ:le  Anjam mal     Bongu mal      Jilim mʌl
Male mal     Migum mʌl     Ogea mʌ:le     Pulabu mal
Rerau mʌlu   Sam mal       Sinsauru mʌ:l  Siroi mal
Waube mʌ:le  Yangulam mal
```

**g·m|BASKET** · binary · 13 languages · enrichment 49×

```
Bagupi gʌ:m  Bau gam    Garus gʌm   Gumalu gʌm
Isebe gam    Mawan gʌm  Mosimo gʌm  Munit gama
Murupi gʌ:m  Nake gʌ:m  Panim gʌm   Rapting gam
Yoidik gʌ:m
```

**g·r·m|DRUM** · ternary · 13 languages · enrichment 190×

```
Anjam guruma    Bau giram     Danaru garʌm  Faita giroma
Isebe giram     Jilim gurumo  Silopi giram  Sumau giroma
Utu girom       Uya girom     Wagi girʌm    Yaben girʌ:m
Yangulam geram
```

**t·n|YOU (OBLIQUE CASE OF YOU)** · binary · 13 languages · enrichment 27×

```
Arawum tane    Asas tene   Danaru ten    Kesawai tene
Lemio tane     Migum tane  Rerau tani    Sausi tene
Siroi tane     Sumau tini  Urigina tono  Uya ten
Yangulam –teŋ
```

**n·n|THEE (OBLIQUE CASE OF YOU)** · binary · 13 languages · enrichment 10×

```
Amaimon ne–ŋ    Bilakura na–na   Brem ne–ne    Gavak ne–ŋ
Kobol no–ŋ      Migum –nan–      Moere ne–ne   Pal nɛ–ŋ
Parawen na–na   Ukuriguma ne–na  Watiwa –nan–  Yangulam –nɛŋ–
Yarawata na–n–
```

**k·k|CARRY ON BACK** · binary · 13 languages · enrichment 37×

```
Bagupi kuki–  Nake kuku–     Nobonob kuk–  Pal kʌki
Panim kuku–   Rapting kuku–  Rempi kuku–   Saruga kuki–
Sausi kakau–  Urigina kuku–  Utu koki–     Wagi kuku–
Waube koko–
```

**s·n|OUR** · binary · 13 languages · enrichment 30×

```
Asas sɛn–    Kesawai son–  Lemio sin–      Migum sin–
Rerau sini–  Sausi son–    Sinsauru sɛnɛ–  Siroi sinɛ–
Sop sin–     Sumau sini–   Tauya si–ni–    Urigina sɛn–
Yabong siŋ–
```

**g·n·m|POSSUM** · ternary · 13 languages · enrichment 309×

```
Baimak ganim  Bongu genim    Gal ganim     Isebe ganim
Kein ganim    Male ginim     Matepi ganim  Mawan ganim
Munit ganim   Nobonob ganim  Panim gʌnim   Utu ganim
Wagi ganim
```

**m·n|HAIR (BODY)** · binary · 12 languages · enrichment 5×

```
Aisi mɨnɨ     Apali mɨnɨ       Kobon u'maŋʌ  Korak –munɛ
Mala –imun    Miani –umun      Mum mɨn       Parawen muna
Sileibi mɨnə  Utarmbung mɨnɨ–  Waskia mo:nɛ  Yarawata munia
```

**w·g|CROCODILE** · binary · 12 languages · enrichment 108×

```
Amele waga    Bagupi wagʌi  Gal wagai   Gumalu wagʌi
Isebe wagi    Matepi wʌga   Mawan waga  Panim waga
Silopi wagʌ:  Utu wa:ga     Wagi wagæ   Waube wagai
```

**s·m|POT** · binary · 12 languages · enrichment 78×

```
Asas somu     Biyom suma  Danaru som     Isabi somo
Kesawai sema  Sausi sama  Sinsauru somo  Sop suma
Sumau suma    Tauya sɛma  Urigina somo   Watiwa som
```

**s·m|COOKING POT** · binary · 12 languages · enrichment 89×

```
Asas somu     Biyom suma  Danaru som     Isabi somo
Kesawai sema  Sausi sama  Sinsauru somo  Sop suma
Sumau suma    Tauya sɛma  Urigina somo   Watiwa som
```

**g·n|HOW MANY** · binary · 12 languages · enrichment 38×

```
Asas genau  Baimak ga:ni    Gal gʌne    Gumalu ganio
Isebe gʌni  Nobonob gane    Ogea agoni  Panim gani
Rempi gain  Sinsauru genau  Uya agon    Wagi gʌne
```

**s·m|VULVA** · binary · 12 languages · enrichment 38×

```
Arawum simi  Asas simi      Kesawai simi  Lemio simi
Migum seme   Sausi simi     Siroi sume    Sumau simi
Tauya sumu   Urigina si:mi  Watiwa simi   Yangulam sime
```

**f·n|FISHING SPEAR** · binary · 12 languages · enrichment 156×

```
Bagupi fon   Baimak fan     Bau fan   Gumalu fan
Matepi fon   Mawan fʌn      Nake fan  Panim fan
Saruga fana  Sinsauru fana  Utu fon   Wagi fan
```

**m·n|BOY** · binary · 11 languages · enrichment 6×

```
Bargam amun  Hunjara-Kaina Ke mene_  Mandobo Atas mun  Mian mĭn
Migum min    Orokaiva mene_          Telefol miin      Tifal màn
Wambon mun   Yekora mani             Zia mani
```

**w·g·j|CROCODILE** · ternary · 11 languages · enrichment 278×

```
Amowe wagʌy    Anjam wagay   Baimak wagay   Mosimo wagay
Murupi wagʌy   Nake wagay    Nobonob wagay  Ogea wagʌy
Rapting wagʌy  Saruga wagʌy  Uya wagay
```

**n·n|YOUR (PLURAL)** · binary · 11 languages · enrichment 13×

```
Anam na-ŋ-        Anamgura na-ŋ-   Faita na-nʉ-  Gavak nɛ-n-
Malas na-n-       Moresada na-ŋa-  Sop nin-      Utarmbung na-ŋ-
Wadaginam na-ŋa-  Waskia ni-na-    Yabong nɔŋ-
```

**n·n|THUMB** · binary · 10 languages · enrichment 8×

```
Bilakura unan  Brem nʌn        Lemio nuŋu  Musar unen
Parawen una:n  Ukuriguma unan  Usan unon   Wanambre unen
Yaben unan     Yarawata unan
```

**b·t|tail** · binary · 10 languages · enrichment 21×

```
Amele biti-     Baimak biti     Bilakura ebitua  Faita bitu
Isebe bitu      Mawan biti      Munit bitu       Panim biti
Parawen ibi:ta  Yarawata ibita
```

**p·r|GARDEN** · binary · 10 languages · enrichment 22×

```
Aeka pure              Baimak p_rau  Gal pair       Gumalu pɛrau
Hunjara-Kaina Ke pure  Mawan pair    Orokaiva pure  Silopi pair
Siroi piro             Utu pair
```

**t·m|NAPE (OF NECK)** · binary · 10 languages · enrichment 22×

```
Aeka tumo       Ewage-Notu tumo  Hunjara-Kaina Ke _'tumo  Kesawai tamau
Orokaiva tumo_  Sausi tamau      Suena tuma               Yegha tumo
Yekora tuma     Zia tumo
```

**m·k|SALIVA** · binary · 10 languages · enrichment 23×

```
Amaimon mu:ku  Bilakura muku  Ganglau muku    Parawen mu:ku
Telefol mook   Tifal mook     Ukuriguma mu:k  Yaben muku
Yabong mikɛ-   Yarawata mukʌ
```

**t·t|SMALL (NOT TALL)** · binary · 10 languages · enrichment 19×

```
Amowe tutu    Bagupi tutu  Bau tutu      Gal tutu
Matepi tutu   Nake tutu    Nobonob tutu  Ogea tutu
Rapting tutu  Rempi tutu
```

**t·t|tie** · binary · 10 languages · enrichment 23×

```
Amele tito   Anjam titə  Bagupi tati  Gal tate
Gumalu tite  Matepi tat  Mawan tate   Murupi tata
Silopi tat   Utu tat
```

**k·n|FILL** · binary · 10 languages · enrichment 14×

```
Arawum kiñe-   Biyom kwiña-  Danaru kuni-  Jilim kun-
Migum kun-     Rerau kun-    Siroi kuñ-    Sop kun-
Urigina kuno-  Watiwa kun-
```

**g·g·r|GRASS-SKIRT** · ternary · 10 languages · enrichment 155×

```
Bagupi goguř      Kare gegura      Matepi gogur   Murupi gogura
Nake gugur        Parawen goaguru  Samosa gegure  Saruga gogura
Ukuriguma goguro  Wamas gegure
```