

Códigos consonánticos en 196 lenguas trans-neoguineanas

El suelo del método, y qué se ve desde ahí

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

Resumen

Presentamos un catálogo de 3.929 códigos consonánticos que agrupan 33.781 formas léxicas de 196 lenguas del Nuclear Trans New Guinea. Un código es un esqueleto consonántico compartido por palabras del mismo concepto en lenguas distintas, construido sin etimología, protoformas ni juicios de cognación.

Es la familia donde el instrumento rinde menos de toda la serie, y ése es el motivo de hacerla. Su precisión supera en 13 veces lo esperable por azar —frente a $216\times$ del turco y $104\times$ del indoeuropeo— y su agrupamiento apenas 1,4 veces: 51,9 % real contra 37,4 % del nulo. Corregido por azar, $\kappa = 0,232$.

La causa está medida y es una sola: la palabra corta. Con 2,41 consonantes por forma, el espacio de esqueletos se satura y barajar los conceptos dentro de cada lengua ya agrupa más de un tercio de las formas. El 35 % de los códigos resultantes son de una sola consonante, y catorce de los quince de mayor alcance también: *n* agrupa cien lenguas para *comer*, y otras tantas para *tú*, *yo*, *madre* y *dormir*.

Con esta familia la serie llega a siete y aparece una regularidad que con seis no se veía: el nulo se dispara por debajo de tres consonantes por palabra. Las dos familias con aridez menor que tres —ésta, 2,41, y el austronesio, 2,79— tienen nulos del 37 % y del 44 %; las cinco por encima de 3,03, entre el 4,7 % y el 15 %. Ocho veces de diferencia a los dos lados de un escalón estrecho.

Y no todo cae. La señal geográfica es la mejor de la serie: las lenguas que comparten veintiún códigos o más están al 9 % de la distancia esperable por azar, con una curva monótona y sin la inversión que estropea la del afroasiático. Las clases emergentes salen limpias —toda la serie labial y fricativa— y los 1.383 pares de códigos hermanos disjuntos triplican con creces los del afroasiático.

1. Qué se busca, y por qué esta familia

El procedimiento —qué es un código, en qué orden se construye, qué nulos usa y qué autoriza a afirmar— está descrito una sola vez para toda la serie en «El catálogo de códigos consonánticos» y no se repite aquí. Lo imprescindible: un código es un esqueleto consonántico que aparece para el mismo concepto en tres lenguas distintas o más, y no se afirma que sus miembros sean cognados, ni que descendan de una protoforma, ni que el código signifique nada.

Esta familia se eligió por su rendimiento bajo, no a pesar de él. El perfil del corpus la daba en 13 veces el azar, la cifra más baja de las sesenta familias evaluables con material suficiente. Un instrumento se entiende igual de bien viendo dónde deja de funcionar que viendo dónde luce, y un suelo descrito vale más que un suelo intuitivo.

También trae algo que el afroasiático no tenía: diez ramas, seis de ellas con siete lenguas o más, y la mayor —Madang— al 51 % en vez del 72 % del chádico. La estructura genealógica que allí faltaba, aquí está.

2. La población

196 lenguas · 94.306 formas · 895 conceptos · 10 ramas.

Rama	Lenguas	Formas
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

El 100 % de las formas obtiene esqueleto. Una sola fuente, sin la pérdida del 7 % que costó IDS en afroasiático ni el 21 % del cirílico turco. Es la población más limpia de la serie.

Una sola macroárea: la lista de anomalías del §6.1-bis no existe aquí, y se dice antes de los resultados.

2.1 Ninguna lengua se excluye, y las dos propuestas revelaron defectos del chequeo

La fase 0 propuso dos exclusiones. Las dos se rechazan.

Danaru, 634 formas. Se proponía excluirla «por marcar tono en dígitos». Tiene dos formas con un superíndice, y las dos son la misma palabra: səg¹ ro-. Es el 0,32 %, y el informe lo redondeaba a «0 %», así que la propuesta se contradecía en su propia línea.

East Kewa frente a **West Kewa**, 73 % de formas idénticas. La regla mecánica que esta serie publica es *mismo glotocódigo, o más del 80 %*. Los glotocódigos difieren (east2516 / west2599) y el 73 % está por debajo del umbral. Son dos dialectos distintos y entran los dos.

Los dos chequeos se corrigieron: el de tono exige ahora un 5 % antes de proponer, y el de duplicados solo propone por encima del umbral que la serie publica. Es el tercer chequeo de la fase 0 que proponía más de lo que su evidencia autoriza —el primero fue el geográfico en afroasiático, que quería borrar el semítico entero—, y el patrón merece nombre: un chequeo que propone una acción drástica con evidencia fina es peor que no tenerlo, porque su propuesta se acepta por inercia.

2.2 Un bug de segmentación, encontrado al construir la tabla de normalización

Las prenasalizadas escritas como dígrafo plano —nd, mb, ng— no se partían en dos posiciones, mientras que las escritas con superíndice —ⁿd, ^mb, ⁿg— sí. Y tres lenguas usan las dos escrituras a la vez: el enga tiene 35 formas con dígrafo y 57 con superíndice, el malgache merina 80 y 57, el guaraní 15 y 107.

El resultado era que el mismo fonema recibía dos tratamientos dentro de la misma lengua: monda-daba m·n·d y k.e.nd.e daba k·n, perdiendo la oclusiva entera. Corregido con una tabla explícita de los quince dígrafos atestiguados —sin generalizar a secuencias no vistas—, sobre 478 formas del corpus. Aislado sobre 20.000 formas, cambian dos y ninguna sin dígrafo.

3. Qué cuesta el instrumento aquí, y es lo que define el paper

Aridad media: 2,41 consonantes por forma. La más baja de la serie, por debajo del austronesio (2,79) y muy por debajo del indoeuropeo y el pama-ñungano (3,35).

Colisión: 57,2 %. Más de la mitad de las formas comparten esqueleto con otro concepto dentro de su propia lengua.

Las dos cifras son la misma cosa dicha dos veces, y de ellas se sigue todo lo demás.

3.1 El tercer tipo de colisión, y con éste la serie los tiene todos

El turco enseñó dos. La colisión de la lengua: *kül* «ceniza» y *kül-* «reír» son homófonas ya en turco antiguo, así que ni conservando la vocal se separarían. Y la colisión que fabrica nuestra tabla, cuando la normalización escribe *n* la velar η .

Aquí aparece la tercera: la colisión de la lista de conceptos. Diez lenguas dan una sola palabra para las seis maneras de cortar que el cuestionario pregunta por separado —CHOP, CUT, CUT (WITH AXE), CUT (WITH KNIFE), CUT (WITH SCISSORS), CUT DOWN—: *san-* en watiwa cubre ocho conceptos, *o-* en kare siete.

No es de la lengua ni nuestra: es de quien diseñó el cuestionario. Y se comporta distinto de las otras dos, porque cambia entre familias según cuánto colexifique cada una.

3.2 Y hay que medirla antes de descontarla, porque no es lo que parece

Dos lecturas opuestas dan el mismo síntoma. (a) La lista pide seis cosas casi iguales y cualquier lengua respondería con una palabra: sería artefacto de la elicitación. (b) Estas lenguas colexifican de verdad, y entonces es un rasgo tipológico que se reporta, no se descuenta.

Se separan haciendo las mismas preguntas en otras familias. Sobre los 89 racimos de conceptos que comparten cabeza de glosa:

Familia	Racimos con ≥ 2 conceptos	Con esqueleto repetido
Nuclear Trans New Guinea	3.055	41,8 %
Indoeuropeo	4.004	23,8 %
Afroasiático	1.808	9,9 %
Pama-ñungano	1.756	9,6 %
Austronesio	7.036	8,4 %

Si fuera la lista, todas darían lo mismo. No lo dan: esta familia colapsa cuatro o cinco veces más. La hipótesis (a) sola no explica el dato.

Pero el control invierte la lectura fácil. Sus palabras son cortas, así que dos formas cualesquiera coinciden más por azar. Comparando dentro del racimo contra fuera, en la misma lengua:

Familia	Dentro	Fuera	Razón
NTNG	38,8 %	0,8 %	47×
Austronesio	29,4 %	0,5 %	58×
Pama-ñungano	18,4 %	0,2 %	80×
Indoeuropeo	15,8 %	0,3 %	51×
Afroasiático	14,4 %	0,2 %	83×

En absoluto colexifica más que ninguna; en razón sobre su propia línea base, menos que ninguna. Su colisión fuera del racimo ya es cuatro veces la del afroasiático. El rasgo es real y el motor es la longitud de la palabra, no una semántica inusual.

Y pesa poco: de las colisiones observadas, solo el 9,9 % ocurre enteramente dentro de un racimo. Descontarlo bajaría la colisión de 29,1 % a 26,3 % —2,8 puntos— y escondería un rasgo real detrás de un supuesto artefacto. Se declara y no se descuenta.

4. El catálogo

3.929 códigos sobre 895 conceptos, que agrupan 33.781 formas de 92.608 (36,5 %). De ellos, 2.464 son binarios o ternarios y 1.375 son unarios: el 35,0 %.

Aridad	Códigos	%
1	1.375	35,0 %
2	1.880	47,8 %
3	584	14,9 %
4	83	2,1 %
5+	7	0,2 %

Precisión: 3,79 % frente a 0,294 % del azar = $13 \times [13 - 13 \times \text{según semilla}]$. Enriquecimiento mediano de los binarios y ternarios: $16 \times$, contra $100 \times$ del afroasiático y $155 \times$ del urálico.

Y sobre el agrupamiento sin umbral de tres lenguas: 51,9 % real frente a 37,4 % del nulo —ventaja de 14,5 puntos, $1,4 \times$ el azar, $\kappa = 0,232$.

4.1 Los códigos de mayor alcance son de una sola consonante

Los quince de más alcance, sin filtrar:

Código	Concepto	Lenguas	Aridad
n	<i>comer</i>	100	1
n	<i>tu</i> (sg.)	87	1
n	<i>ti</i>	75	1
n	<i>tú</i>	65	1
n	<i>yo</i>	65	1
n	<i>madre</i>	60	1
m	<i>taro</i>	59	1
m	<i>morir</i>	53	1
m	<i>piojo</i>	51	1
j	<i>mi</i>	51	1
m • m	<i>padre</i>	48	2
n	<i>dormir</i>	46	1
m	<i>pecho</i>	43	1
m	<i>ojo</i>	41	1
n	<i>su</i>	38	1

Catorce de los quince son unarios, y n sola aparece ocho veces para conceptos sin relación entre sí: comer, tu, ti, tú, yo, madre, dormir, su.

Esto no significa que sean falsos. La revisión del paper afroasiático mostró que un código unario puede ser un etimón perfectamente real —el $\mathfrak{l}$ | vaca chádico es la lateral $\mathfrak{t}$ de *ta*—. Significa algo

peor de manejar: el catálogo no puede distinguir un etimón monoconsonántico verdadero de una coincidencia por disponibilidad, porque a los dos les da la misma evidencia, y en esta familia esa indistinción se lleva el techo de la lista.

El único binario del grupo, **m·m | padre** en 48 lenguas, es el par de reduplicación labial que aparece en todas las familias de la serie, y sobre el que el paper austronesio ya avisaba: su forma se explica por la crianza antes que por la historia.

5. El hallazgo de la serie: el nulo se dispara por debajo de tres consonantes

Con siete familias medidas aparece una regularidad que con seis no se veía. Ordenadas por aridez:

Familia	Aridad	Agrupar	Nulo	Ventaja	κ
Nuclear Trans New Guinea	2,41	51,9 %	37,4 %	14,5	0,232
Austronesio	2,79	64,8 %	44,0 %	20,8	0,371
Turco	3,03	56,9 %	15,0 %	41,9	0,493
Afroasiático	3,16	29,7 %	4,7 %	25,0	0,262
Urálico	3,26	33,9 %	5,9 %	28,0	0,298
Indoeuropeo	3,35	44,3 %	12,9 %	31,4	0,361
Pama-ñungano	3,35	40,2 %	14,0 %	26,2	0,305

Las dos familias por debajo de tres consonantes por palabra tienen nulos del 37 % y del 44 %; las cinco por encima de 3,03, entre el 4,7 % y el 15 %. Ocho veces de diferencia a los dos lados de un escalón de un cuarto de consonante.

Hasta este trabajo el austronesio era el único caso de palabra corta de la serie, y su nulo alto podía leerse como rasgo suyo. Con dos casos deja de ser anécdota. Con siete familias no es una ley —y no se presenta como tal—, pero sí una regularidad que la serie tiene que declarar.

La consecuencia práctica es la que importa. Esta familia agrupa el 51,9 % de sus formas, más que el afroasiático (29,7 %) y más que el urálico (33,9 %). Leída sola, esa cifra parece un buen resultado. Contra su nulo vale 1,4 veces el azar, frente a 6,3 del afroasiático. La cifra bruta de agrupamiento no es comparable entre familias sin su nulo al lado, y este trabajo es la demostración más clara de por qué.

5.1 El despiece automático, y por qué cae más aquí

El descubrimiento automático de afijos —la fase 2 del protocolo, que el catálogo de esta serie no usa— sube el agrupamiento de 51,9 % a 67,9 %, dieciséis puntos. Y sube el nulo veintiséis, de 37,4 % a 63,7 %: tras despiezar, dos de cada tres formas se agrupan por azar.

Paso	En grupo	Nulo	Ventaja	κ	Códigos de 4+ consonantes
Palabra entera	51,9 %	37,4 %	14,5	0,232	14,8 %
+ afijos descubiertos	67,9 %	63,7 %	4,2	0,116	2,4 %

$\Delta\kappa = -0,118$, la segunda caída mayor de la serie tras el austronesio ($-0,151$).

Y aquí hay que decir algo que la propia serie desmintió después. La primera versión de esta sección afirmaba que el orden de las caídas sigue el mismo eje que el nulo —cuanto más corta la palabra, más destruye el despiece—. El sino-tibetano, medido después, lo falsa: con aridez 2,22, la más baja de las ocho familias, su caída es de $-0,020$, la tercera más pequeña.

Familia	Aridad	$\Delta\kappa$
Afroasiático	3,16	−0,004
Turco	2,85	−0,009
Sino-tibetano	2,22	−0,020
Urálico	3,27	−0,035
Indoeuropeo	3,35	−0,094
Pama-ñungano	3,35	−0,102
Nuclear Trans New Guinea	2,41	−0,118
Austronesio	2,78	−0,151

Las dos familias de palabra más corta están en los extremos opuestos de la tabla, y las dos de palabra más larga en medio. La asociación no existe. Lo que sí queda en pie es el hallazgo de §5, que habla del nulo del catálogo y no del despiece: ahí el sino-tibetano confirma —aridad 2,22, nulo 41,5 %—.

Se deja escrito porque el error es instructivo: con siete familias se veía un patrón que la octava rompe, y la única razón de haberlo visto es que las siete cabían en una tabla.

6. Lo que la familia sí da, y da mejor que ninguna

6.1 La geografía, sin curva en U

Códigos compartidos	Pares	Distancia media	vs azar
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

Monótona hasta el final. Las lenguas que comparten veintiún códigos o más están al 9 % de la distancia esperable por azar, y la curva no se invierte en ningún tramo.

Conviene decir por qué, porque explica también el fracaso ajeno. En afroasiático esta misma sección subía otra vez a partir de once códigos compartidos, y la causa era el reparto desigual de las ramas: el chádico compacto, el semítico de Marruecos a Omán. Aquí no hay ninguna rama dispersa —la familia entera tiene una mediana de 95 km al centro— y la curva baja limpia. Compacidad y señal geográfica son la misma propiedad vista dos veces, y una familia extensa no puede dar esta sección por mucho que su método funcione.

6.2 Pares de sustitución: 1.383 candidatos disjuntos

De 1.493 pares de códigos del mismo concepto que difieren en una sola posición, 1.383 son disjuntos en lenguas —ninguna lengua está en los dos— y solo esos pueden proponerse como isoglosa. Triplícan con creces los 450 del afroasiático, y la razón es la estructura: diez ramas contra cuatro.

Y hay que leerlos con un aviso. Cuatro de los siete pares de mayor alcance son TOBACCO: $k \cdot b$, $k \cdot h$ y $k \cdot p$ frente a $k \cdot s$, que alcanza 33 lenguas. El tabaco llegó a Nueva Guinea en el siglo XVII por comercio. Es un préstamo moderno que el catálogo agrupa exactamente igual que a cualquier herencia, como la mandioca en afroasiático, y encabeza la lista. De 274 formas en 131 lenguas.

Los que no son tabaco: $g \cdot n$ $m \cdot n$ en *piojo* (10 y 26 lenguas, cruzando Madang, Finisterre-Huon y Mek), $n \cdot n$ $s \cdot n$ en *nosotros*, $m \cdot l$ $m \cdot r$ en *bejuco*, $n \cdot n$ $t \cdot n$ en el oblicuo de *tú*.

6.3 Clases emergentes: la serie labial y fricativa

Las 196 lenguas se parten en dos mitades, una para descubrir y otra para comprobar. Los pares que sobreviven en las dos:

Par	Veces (descubre)	Razón	Veces (comprueba)	Razón
$f \leftrightarrow \phi$	80	9,2×	91	10,3×
$j \leftrightarrow z$	580	7,7×	714	9,3×
$v \leftrightarrow w$	745	6,2×	958	7,8×
$b \leftrightarrow f$	522	2,4×	1.315	5,9×
$v \leftrightarrow \beta$	69	5,5×	60	4,7×
$f \leftrightarrow w$	337	2,1×	676	4,1×
$b \leftrightarrow \phi$	94	2,2×	171	4,0×

Seis de los siete son labiales, y el conjunto se lee como una sola serie: oclusiva, fricativa, aproximante y sus variantes bilabiales frente a labiodentales. Es el resultado más coherente de la serie —el afroasiático daba una mezcla de sonoridad y modo repartida por varios lugares de articulación— y no son correspondencias regulares: falta entorno y falta dirección. Son candidatas, con el sitio donde buscarlas.

7. Límites

Una sola macroárea. La lista de anomalías del §6.1-bis no existe, y no por un fallo del método sino porque la familia ocupa un área continua.

El 35 % del catálogo es de una sola consonante, y el catálogo no puede distinguir el etimón monoconsonántico verdadero de la coincidencia por disponibilidad. Los unarios se cuentan y no se leen: el paper trabaja con los 2.464 binarios y ternarios, el 62,7 %.

Las labiovelares **kp** y **gb** se funden en **k** y **g**. Son características de esta zona lingüística, y la tabla les quita el componente labial. La primera versión de este paper lo dejaba «sin medir» como primer trabajo pendiente; la revisión adversarial lo midió y cuesta poco: 95 formas en 11 lenguas, que tocan 16 de los 3.929 códigos del catálogo —el 0,4 %—, y en 15 de esos 16 la fusión es portante, es decir que el código reúne formas con **kp** y formas con **k** simple. En el Anexo A hay un solo código afectado, $k \cdot n$ | izquierda, con **kp** en 4 lenguas y **k** en 9.

El aviso se mantiene y la cifra lo acota: la tabla borra un contraste real de esta zona, y lo hace en menos de medio punto porcentual del catálogo. No es lo mismo declarar un límite que dejarlo sin número.

Madang es el 51 % de las lenguas. Es mejor reparto que el 72 % chádico, pero sigue siendo una rama que domina, y toda cifra de la familia está ponderada hacia ella.

El tabaco encabeza los pares de sustitución y es un préstamo del siglo XVII. No hay manera, dentro de este método, de separar lo que llegó de lo que se heredó.

8. Conclusión

Esta familia se hizo por su rendimiento bajo, y lo ha entregado con precisión: 13 veces el azar, $\kappa = 0,232$, un nulo del 37,4 % y el 35 % del catálogo en códigos de una sola consonante. Es el suelo de la

serie, y ahora está descrito en vez de intuido.

Lo que ese suelo enseña vale para las otras seis. El nulo se dispara por debajo de tres consonantes por palabra, y con dos familias a ese lado del escalón deja de poder atribuirse a una de ellas. La consecuencia es una regla de lectura para toda la serie: esta familia agrupa el 51,9 % de sus formas, más que el afroasiático y más que el urálico, y vale mucho menos que cualquiera de las dos. La cifra bruta de agrupamiento no dice nada sin su nulo al lado.

Y el suelo no es un agujero. La señal geográfica es la mejor de las siete —9 % de la distancia esperable, sin inversión—, las clases emergentes salen como una serie labial coherente, y los 1.383 pares disjuntos triplican los de la familia anterior. La misma compacidad que hunde la precisión, porque comprime el espacio de esqueletos, es la que hace legible la geografía.

Un instrumento no se entiende por su mejor caso. El austronesio parecía el caso difícil disfrazado de fácil; éste es el caso difícil sin disfraz, y es el que dice dónde está el límite y por qué.

Referencias

- Brown, C. H., Holman, E. W., Wichmann, S. y 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. y Bank, S. *Glottolog*. Max Planck Institute for Evolutionary Anthropology.
- Hubert, L. y Arabie, P. (1985). Comparing partitions. *Journal of Classification* 2(1). — [la corrección por azar que este paper usa como κ]
- 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. En *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. y Forkel, R. (2016). Concepticon: a resource for the linking of concept lists. *LREC*.
- Pawley, A. y Hammarström, H. (2018). The Trans New Guinea family. En B. Palmer (ed.), *The Languages and Linguistics of the New Guinea Area*. De Gruyter Mouton. — [la clasificación de referencia]
- Toledo Martínez, A. (2026). *El catálogo de códigos consonánticos: un método comparativo sin genealogía, sin cognación y sin reconstrucción*. — [el método de la serie, descrito una sola vez]
- Toledo Martínez, A. (2026). *Corpus Integrativo: una base léxica multi-capas de 225 macrosistemas, con procedencia por fila y su aparato de verificación*. — [los datos] · <https://doi.org/10.13140/RG.2.2.31244.68486>
- Turchin, P., Peiros, I. y Gell-Mann, M. (2010). Analyzing genetic connections between languages by matching consonant classes. *Journal of Language Relationship* 3.

Datos y reproducibilidad

Catálogo	familias/nuclear-trans-new-guinea/output/ catalogo-nuclear-trans-new-guinea.tsv — 3.929 filas
Detalle por forma	...-lenguas.tsv — 21.817 filas
Higiene (§2.1)	pipeline/fases/fase0_higiene.py → runs/fase0-higiene. md
Decisiones de datos (§2.1, §3)	runs/fase0-decisiones.md
Prenasalizadas (§2.2)	corpus_integrativo/ingest/recompute_prenasal_digrafo. py
Tabla de normalización (§3)	pipeline/fases/tabla_normalizacion.py → runs/ normalizacion.md
Carga, población, avisos	cod/familia.py
Catálogo (§4)	pipeline/fases/catalogo_familia.py
Agrupamiento y nulos (§4, §5)	pipeline/fases/agrupamiento_familia.py → runs/ agrupamiento.md
κ (§5)	cod/metricas.py::kappa
Despiece automático (§5.1)	pipeline/fases/fase2_despiece.py → runs/ fase2-despiece.md
Geografía (§6.1)	pipeline/fases/codigos_vs_parentesco.py → runs/ codigos-parentesco.md
Códigos hermanos (§6.2)	pipeline/fases/codigos_hermanos.py → runs/ codigos-hermanos.md
Clases emergentes (§6.3)	pipeline/fases/fase6_clases_emergentes.py → runs/ fase6-clases.md
Sustituciones por rama	pipeline/fases/fase6_por_rama.py → runs/ fase6-por-rama.md

Todas las cifras salen de esos scripts sobre la misma población declarada en §2, con las mismas exclusiones —que aquí son ninguna—. Ninguna se escribió a mano.

Anexo A · Cien códigos comentados

El catálogo tiene 3.929 códigos y los datos completos están en el directorio `output/` de la familia. Lo que sigue es una muestra legible: los cien códigos binarios y ternarios de mayor alcance, cada uno con hasta dieciséis lenguas y sus palabras.

Tres cosas sobre la presentación. Las lenguas están repartidas a lo largo de la lista alfabética, no tomadas de su cabeza. Los nombres largos se recortan; las formas no se alteran. Y la agrupación temática es un orden de lectura declarado a mano, no una clasificación semántica.

Este anexo no lleva columna de código emergente. Las clases de §6.3 salen como una serie labial — $f \leftrightarrow \Phi$, $v \leftrightarrow w$, $b \leftrightarrow f$, $v \leftrightarrow \beta$ — y no como una partición cerrada del inventario: el cierre a $\geq 2\times$ mete dieciséis segmentos en una sola clase. §6.3 las publica como candidatas, que es lo que son, y `ci/clases.py` declara para esta familia un diccionario vacío.

Qué selecciona este anexo, medido

Ordenar por alcance no es neutral. Medido con `pipeline/tools/audita_anexo.py`:

	catálogo	este anexo
códigos binarios	47,8 %	93 %
códigos ternarios	14,9 %	7 %
de rama dominante madang	66,3 %	85 %
de rama dominante asmat-awyu-ok	10,3 %	0 %
que cruzan cuatro ramas o más	1,4 %	5 %
enriquecimiento medio	69×	43×

El madang pasa del 66 % al 85 %, y la segunda rama del catálogo, el asmat-awyu-ok, no tiene ni un código aquí. El §7 del paper avisa de que «toda cifra de la familia está ponderada hacia el madang»; en este anexo esa ponderación llega a cinco de cada seis códigos.

Y el enriquecimiento baja un tercio. Alcance y enriquecimiento se compran uno con otro: un código que llega a muchas lenguas suele estar hecho de consonantes frecuentes, que es exactamente lo que el enriquecimiento penaliza.

Lo que este anexo enseña y el paper no vio: dieciséis códigos de pronombre

Dieciséis de los cien códigos son pronombres, y su reparto por esqueleto es este:

Esqueleto	Códigos	Conceptos
$n \cdot n$	8	ellos, su (pl.), oblicuo de <i>tú</i> , nosotros, oblicuo de <i>ti</i> , tú (2), vuestro
$n \cdot g$	5	él o ella, oblicuo de <i>ti</i> , oblicuo de <i>tú</i> , tú (2)
$s \cdot n$	2	nosotros, nuestro
$t \cdot n$	1	oblicuo de <i>tú</i>

Esto merece leerse con las dos lecturas encima de la mesa, porque las dos son defendibles.

(a) La clasificación trans-neoguineana de Ross se construye precisamente sobre paradigmas pronominales, y la serie *na* / *ni* / *nu* es su material central. Que dieciséis de los cien códigos de mayor alcance de esta familia sean pronombres, y que trece de ellos empiecen por *n*, es la misma observación por otro camino — y este procedimiento no consultó ningún árbol.

(b) n es la consonante más frecuente del inventario de esta familia: 34.037 posiciones, más que ninguna otra. n·n es, por construcción, el esqueleto más disponible que existe aquí. Un pronombre corto compuesto de la consonante más común es exactamente lo que la disponibilidad produce sin necesidad de historia.

El catálogo no separa (a) de (b), y este anexo tampoco. Lo que sí se puede decir es lo estrecho: los enriquecimientos de esos dieciséis van de 8× a 30×, que es la mitad baja de la lista. Un pronombre no es un buen candidato genealógico *para este instrumento* aunque lo sea para el método comparativo.

Y hay un tercer aviso, más incómodo. Nueve de los dieciséis dependen de que la tabla escriba n la velar ŋ: n·n|e|llos es n en 18 lenguas y ŋ en 8; n·n|su es 15 y 9; n·n|oblicuo de tú es ŋ en 11 y n en 9. La evidencia pronominal de este anexo descansa sobre una fusión de la normalización.

La auditoría de la tabla, y por qué aquí sale barata

Veintitrés de los cien códigos tienen al menos una posición donde la normalización funde dos segmentos que tres lenguas o más escriben distinto. En el anexo sino-tibetano eran sesenta y ocho. La diferencia es de familia: allí s reúne once símbolos sibilantes y africados; aquí la tabla es corta y veintiuno de los veintitrés casos son la misma fusión, n/ŋ.

Los otros dos: g·r|piel, con r en 16 lenguas y ñ en 4, y k·n|izquierda, con k en 9 y kp en 4. Ése es el único código del anexo tocado por la fusión de las labiovelares, el límite que §7 del paper declara y que la revisión adversarial acotó en el 0,4 % del catálogo.

Y una cosa que el anexo sino-tibetano no puede decir

Ninguna de las 1.545 formas que este anexo cita lleva una sola marca que el esqueleto ignore. Cero por ciento. En el anexo sino-tibetano son el 75 %, casi todo tono.

No es una virtud de esta familia ni un defecto de la otra: es que aquí lo que se ve escrito es lo que el código usa, y allí no. Cuando en este anexo dos formas se parecen, se parecen en lo mismo en que el procedimiento las juntó. Es el único sitio de la serie donde se puede decir eso sin matices.

El enriquecimiento alto es local, y aquí se ve en estado puro

Los diez códigos de mayor enriquecimiento de este anexo:

Código	Concepto	Lenguas	Enriq.	Ramas
g·n·m	zarigüeya	13	309×	1
w·g·j	cocodrilo	11	278×	1
g·l·m	tambor	16	262×	1
g·r·m	tambor	13	190×	1
s·p·r	hacha	12	170×	1
b·r·r	hacha	11	168×	1
f·n	arpón de pesca	12	156×	1
g·g·r	falda de hierba	10	155×	1
b·m	taparrabo	23	127×	1
b·b	hermano mayor del mismo sexo	15	122×	1

Diez de diez están dentro de una sola rama, y nueve son cultura material. Zarigüeya, cocodrilo, tambor, hacha, arpón, falda de hierba, taparrabo: objetos y animales de un sitio, con nombres largos y raros dentro del inventario de esa rama.

Eso es lo que mide el enriquecimiento: distintividad, no profundidad. Un objeto local con un nombre específico marca el máximo. El paper sino-tibetano lo encontró por la vía del préstamo —un cacahuete persa a 1086×—; aquí sale por la vía de lo endémico, y es el mismo fenómeno: el enriquecimiento premia lo que no se parece a nada, y no tiene opinión sobre desde cuándo.

Y cuatro códigos están contados dos veces

s · m sale con «olla» y con «olla de cocinar», doce lenguas cada uno y las doce son las mismas. n · g sale con «hembra» y «persona de sexo femenino», diecisiete y diecisiete, las mismas diecisiete. k · k con «cargar» y «cargar a la espalda», trece de catorce compartidas. m · n con «pelo» y «pelo (del cuerpo)», las diez compartidas.

Ocho de las cien entradas de este anexo son cuatro códigos escritos dos veces, porque el cuestionario pregunta dos veces por la misma cosa. Es la colisión de la lista que §3.2 del paper mide sobre las formas, aquí vista en el catálogo de códigos. §3.2 concluye que pesa poco y se declara en vez de descontarse; este anexo enseña cuánto pesa cuando se lee de arriba abajo.

Índice inverso: códigos que reaparecen en varios conceptos

Dieciocho códigos cubren setenta y cuatro de las cien entradas del anexo; la tabla de abajo enseña los diez que reaparecen tres veces o más. Tres cuartas partes de los bloques que siguen repiten un esqueleto ya visto, y es lo que el 61 % de colisión de §3 significa puesto en una lista.

m · n se lleva doce conceptos: piojo, pecho, ser, serpiente, piedra, lengua, hacha, pelo (del cuerpo), chico, madre, casa y pelo. No hay lectura semántica que sostenga eso, y no se intenta: m y n son las dos consonantes más frecuentes de la familia, y sus enriquecimientos aquí van de 5× a 12×, la mitad baja de todo el anexo.

n · n se lleva otros doce, y ocho son pronombres: ellos, su, oblicuo de *tú*, nosotros, oblicuo de *ti*, tú, vuestro, tú. Los otros cuatro —madre, hermana, pulgar, comer— no lo son. Ésos son los dos extremos con el mismo formato de fila: un código que parece reunir un paradigma y un código que reúne lo que cabe en dos consonantes frecuentes. Y el mismo código hace las dos cosas a la vez.

Código	Conceptos con los que aparece, y su enriquecimiento
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×)

Partes del cuerpo

Doce códigos, y seis de los doce vienen en pareja para el mismo concepto: $m \cdot n$ y $n \cdot m$ para «pecho», $g \cdot r$ y $g \cdot n$ para «piel», $g \cdot m$ y $k \cdot m$ para «hígado». No son duplicados de la lista —son códigos distintos para el mismo concepto—, y por eso van a §6.2 como pares candidatos a isoglosa.

Los enriquecimientos son bajos: $m \cdot n$ | lengua 6×, $m \cdot n$ | pelo 8×, $m \cdot n$ | pecho 11×. $d \cdot h$ | oreja, con 65×, es la excepción y es de una sola rama.

$m \cdot n$ | BREAST · binario · 23 lenguas · enriquecimiento 11×

Alekano amina	Amaimon amun	Anam amaj	Atemble aman
Bepour mēna	Faita amēna	Gende ami-ne	Kesawai amne
Kuman amu-na	Kāte moŋ-	Mape moŋ-	Moere mēna
Pal meŋ	Sausi amene	Tauya amena	Tokano amina

$g \cdot m$ | liver · binario · 22 lenguas · enriquecimiento 43×

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

$g \cdot r$ | skin · binario · 20 lenguas · enriquecimiento 26×

Anjam gaŋa	Bagupi garΛ-	Baimak gara-	Bongu gare
Lemio geŋ	Matepi gare-	Mawan gari-	Mosimo gara-
Nake gara-	Nobonob gaŋa-	Rapting gara-	Rempi gaŋa
Saruga gara-	Siroi garo	Utu gara	Wagi gara

$n \cdot m$ | BREAST · binario · 18 lenguas · enriquecimiento 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-

$m \cdot n$ | tongue · binario · 14 lenguas · enriquecimiento 6×

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

$d \cdot h$ | ear · binario · 12 lenguas · enriquecimiento 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-

$g \cdot t$ | head · binario · 12 lenguas · enriquecimiento 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 · binario · 12 lenguas · enriquecimiento 21×

Amowe gon–a	Bilakura guan	Gumalu ganau–	Isebe gʌna–
Panim gana–	Parawen goʌna	Ukuriguma guʌ:n	Usan goʌn
Wagi gone–	Waskia goaŋ	Yaben guʌ:nu	Yarawata guana

k·m|liver · binario · 11 lenguas · enriquecimiento 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 · binario · 10 lenguas · enriquecimiento 18×

Doghoró '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 · binario · 10 lenguas · enriquecimiento 14×

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

m·n|HAIR · binario · 10 lenguas · enriquecimiento 8×

Apali mənɐ	Korak –munɐ	Mala –imun	Miani –umun
Mum mənɐ	Parawen muna	Sileibi minə	Utarmbung mini–
Waskia mo:nɐ	Yarawata munia		

Parentesco y persona

Diecisiete códigos, la sección más poblada del anexo, y ocho de ellos son pronombres. La cabecera de este anexo los discute juntos porque no se leen por separado.

Los dos que encabezan la sección son los que hay que leer con más cuidado. **m·m|padre** en 48 lenguas es el código de mayor alcance de toda la familia, y §4.1 del paper ya avisa: una nasal labial reduplicada para «padre» se explica por la crianza antes que por la historia, y aparece en familias sin relación. **m·m|madre**, doce lenguas, es el mismo caso. Y hay tres códigos más para «madre» —**n·n**, **m·n**, **n·m**— con enriquecimientos de 5× a 13×.

Cuatro códigos para «madre» en el mismo anexo no es un resultado sobre la familia: es lo que pasa cuando el concepto se dice con dos consonantes muy comunes.

m·m|FATHER · binario · 48 lenguas · enriquecimiento 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 · binario · 26 lenguas · enriquecimiento 23×

Anam nɛ-ŋ	Anamgura nɛŋ	Asas nɛnɛ	Danaru nɛn
Ganglau naŋ	Kesawai nɛnɛ	Kube iŋeŋ	Migum nɛnɛ
Saep naŋ	Sausi nɛnɛ	Siroi na-nɛ	Somba-Siawari yeŋân
Urigina nɛnɛ	Uya nɛn	Wadaginam ni-ŋ	Yabong nɛ-ŋ

w·n|name · binario · 25 lenguas · enriquecimiento 32×

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

n·m|name · binario · 23 lenguas · enriquecimiento 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 upim	Yangulam ñem

n·n|MOTHER · binario · 22 lenguas · enriquecimiento 13×

Amaimon nɛnia	Bargam -nen	Bilakura niña	Gavak naŋ
Jilim ñaŋ	Kare nan-	Mauwake nene	Nakai nãŋ
Parawen nain	Rerau ñaŋ	Setaman nein	Ukuriguma nɛ:n
Usan nɛni	Waskia naŋ	Yabong nono-	Yarawata naini

n·n|WE · binario · 17 lenguas · enriquecimiento 12×

Amaimon ini-ŋi	Bilakura ep-ina	Borong nono	Dedua nini
Eipomek nun	Ketengban nun	Komba nen	Kube nini
Kuman nono	Mape ŋaŋu	Nabak nin	Parawen ina-na
Selepet nen	Somba-Siawari nini	Timbe nen	Tobo nini

s·n|WE · binario · 17 lenguas · enriquecimiento 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 · binario · 17 lenguas · enriquecimiento 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 nɛ-g°	Wagi nug	Wamas nugo

d·n|man · binario · 15 lenguas · enriquecimiento 35×

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

n·g|THOU · binario · 14 lenguas · enriquecimiento 16×

Angal nge	Bagupi neg	Garus neg	Matepi neg
Mosimo nɔg	Murupi naga	Nake nɔg	Rapting nag
Samosa nɔge	Saruga nɔga	Silopi neg	Utarmbung na-gə
Wamas nage	Yoidik nɔg		

m·m|MOTHER · binario · 12 lenguas · enriquecimiento 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 · binario · 11 lenguas · enriquecimiento 16×

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

n·n|YOU · binario · 11 lenguas · enriquecimiento 11×

Anam na-ŋə-	Anamgura na-ŋ-	Dedua ŋeni	Faita na-nɛ-
Kâte ŋoŋe	Mape ŋoŋai	Ono ŋine	Somba-Siawari iŋiŋi
Utarmbung na-ŋi-	Wadaginam na-ŋi-	Waskia nV-ŋa-	

m·n|MOTHER · binario · 10 lenguas · enriquecimiento 5×

Anam -mna	Borong maŋ-	Mungkip mɛŋ	Nakame meŋi
Nek meɛŋ	Nuk meni	Paynamar amiŋ	Som miŋi
Utarmbung -mɛŋ	Yabong munu		

n·m|MOTHER · binario · 10 lenguas · enriquecimiento 8×

Atemble ŋam	Awa ʌnómá	Biyom -ainam	Borong namo
Danaru anɔm	Kalam -nm	Miani namu	Moere anɔm
Somba-Siawari nam-	Waskia niam		

n·n|SISTER · binario · 10 lenguas · enriquecimiento 19×

Awa nane	Awiaana nana	Borong naaŋ-	Fore noná
Gadsup nanoi	Inoke-Yate núna	Kamano núna	Kube naoŋ-
Somba-Siawari nen-	Tobo naaŋ-		

n·n|THOU · binario · 10 lenguas · enriquecimiento 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		

Naturaleza y cielo

Solo tres códigos —*m·n*|piedra, *j·g*|agua, *k·n*|luna—, y es una sección pequeña para lo que suele ser en esta serie. La naturaleza de esta familia entra al catálogo en códigos unarios, que este anexo no toma.

***m·n*|stone** · binario · 16 lenguas · enriquecimiento 8×

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

***j·g*|water** · binario · 13 lenguas · enriquecimiento 38×

Amowe jagɛ	Bilakura yuag	Ganglau yag	Ogea yɿge
Parawen yɿga	Rerau yag	Saep yɿg	Sam jag
Sop yɿg	Ukuriguma yɿ:g	Yaben yagu	Yabong yag
Yarawata yɿ:ga			

***k·n*|moon** · binario · 11 lenguas · enriquecimiento 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	

Animales

Nueve códigos, y tres son para «piojo» —*m·n* (26 lenguas), *n·m* (12) y *g·n* (10)—. §6.2 del paper usa justamente *g·n* *m·n* en piojo como su mejor par de sustitución que no es tabaco, cruzando madang, finisterre-huon y mek.

También hay dos códigos para «serpiente» y dos para «pájaro». La sección entera es material de isoglosa, que es lo que esta familia da mejor que ninguna otra de la serie.

***m·n*|louse** · binario · 26 lenguas · enriquecimiento 12×

Anjam mana	Apali iman	Bongu mana	Borong imoŋ
Jilim miŋi	Komba imen	Kube imin	Mape imen
Mungkip miŋ	Nabak emin	Nend eman	Nuk ɛmin
Rerau muŋi	Selepet imen	Somba-Siawari imin	Tobo imin

***m·n*|snake** · binario · 17 lenguas · enriquecimiento 8×

Amowe mani	Asmat mini	Asmat mini	Asmat mini
Bau man	Danaru man	Girawa me:n	Gumalu ma:ni
Isebe man	Kein man	Munit mani	Ogea mɿ:ni
Panim man	Sihan ma:ni	Urigina mɿ:ni	Usan ma:ni

***n·m*|bird** · binario · 12 lenguas · enriquecimiento 10×

Alekano nama	Benabena namá	Dano nama	Gadsup numi
Gimi ními	Inoke-Yate namá	Kamano nāmá	Kambaira numi
Siane nema	Tokano nama	Yagaria namá	Yaweyuha namá

n·m|louse · binario · 12 lenguas · enriquecimiento 10×

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

m·t|snake · binario · 11 lenguas · enriquecimiento 16×

Baimak mat	Gal mat	Mala mat	Mawan mat
Miani mwat	Nobonob mat	Sop mɔ:t ^h	Sumau mɔ:tu
Wagi mat	Wamas mat	Waskia mot	

m·r|CHICKEN · binario · 11 lenguas · enriquecimiento 22×

Bagupi meru	Mawan maru	Mosimo mañu	Murupi marua
Nake maru	Sam muru	Samosa marue	Saruga maru
Silopi meru	Utu meru	Wagi maru	

j·n|bird · binario · 10 lenguas · enriquecimiento 21×

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

m·m|dog · binario · 10 lenguas · enriquecimiento 12×

Amowe mamo	Biyom mɔm	Jilim mam	Rerau mɔ:mu
Sop mɔm	Sumau mam	Tauya mɔmo	Urigina mɔm
Uya mɔ:m	Waube ma:mu		

g·n|louse · binario · 10 lenguas · enriquecimiento 17×

Amaimon gunu	Bilakura gu:n	Brem gu:n	Gavak igun
Malas gunu	Parawen gu:nu	Ukuriguma gu:n	Usan gu:n
Yaben gunu	Yarawata gu:n		

Plantas y comida

Dos códigos: **n·m|árbol** y **n·n|comer**, con 13× y 6×. Los cultivos de la familia —taro, plátano, sagú, tabaco— están en «Otros» porque sus glosas no están en la lista temática, y ahí está el tabaco que §6.2 convierte en el aviso central del paper.

n·m|tree · binario · 15 lenguas · enriquecimiento 13×

Amaimon ɣɔmu	Anjam ɣam	Bepour nɔmɔ	Bilakura nuom
Bongu aɣam	Gavak ɣam	Kobol neim	Male ɣam
Miani nɔm	Moere namɔ	Pal nɔ:m	Parawen nɔ:ma
Usan nɔm	Yaben namu	Yarawata nɔ:ma	

n·n|eat · binario · 10 lenguas · enriquecimiento 6×

Awyu nin	Dani nan	Dani nan	Gadsup naano
Kalam ñŋ-	Kobon ñiŋ-	Mesem nin-	Mungkip nɔna
Tairora naana	Walak nan		

Verbos de acción y estado

Cinco códigos. **k·m|morir** en 25 lenguas y cuatro ramas es el mejor de la sección; **n·l|vomitar** (87×) y **d·z|quemar** (85×) son de enriquecimiento alto y una sola rama, es decir el patrón local que la cabecera describe.

Dos códigos distintos para «dormir» —**n·g** en 16 lenguas y **k·n** en 14—, y los dos cruzan dos ramas.

k·m|die · binario · 25 lenguas · enriquecimiento 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 · binario · 16 lenguas · enriquecimiento 15×

Bagupi nog-	Baimak unug-	Dano óng	Gal nig-
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 · binario · 14 lenguas · enriquecimiento 12×

Arawum kipi-	Biyom akin	Danaru kin-	Faita akun-
Jilim kin-	Lemio kipi-	Migum kipΛ-	Rerau kin-
Siroi kini-	Sop kin-	Tifal akaan	Urigina kina-
Watiwa kini-	Yangulam kip-		

n·l|vomit · binario · 12 lenguas · enriquecimiento 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 · binario · 10 lenguas · enriquecimiento 85×

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

Objetos y cultura

Cuatro códigos, y tres son para «hacha»: **m·n** (13 lenguas, 7×), **s·p·r** (12, 170×) y **b·r·r** (11, 168×).

Es el mejor ejemplo del anexo de lo que un enriquecimiento significa y de lo que no. Los tres nombran el mismo objeto en la misma familia. El primero es un binario de consonantes frecuentes y no distingue nada; los otros dos son ternarios raros que marcan su zona. La diferencia entre 7× y 170× no es antigüedad ni importancia: es cuánto se parece ese esqueleto al resto del inventario.

m·n|AXE · binario · 13 lenguas · enriquecimiento 7×

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

s·p·r|AXE · ternario · 12 lenguas · enriquecimiento 170×

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

b·r·r|AXE · ternario · 11 lenguas · enriquecimiento 168×

Bepour buñiñ	Brem buñiñ	Korak buñiñ	Kowaki buñiñ
Maia biñiñ	Mala buñiñ	Malas buñiñ	Mauwake buñiñ
Mawak buñiñ	Moere buñiñ	Waskia buñiñ	

m·n|house · binario · 10 lenguas · enriquecimiento 5×

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

Cualidades

Un solo código, **k·n|** izquierda, y resulta ser el único del anexo tocado por la fusión de las labiovelares: **k** en 9 lenguas y **kp** en 4. El límite que §7 del paper declaraba sin número tiene aquí su único caso visible.

k·n|left · binario · 14 lenguas · enriquecimiento 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		

Otros

Cuarenta y siete códigos. Como en el anexo sino-tibetano, no es una categoría residual: es la parte del vocabulario que la lista temática —escrita para vocabulario básico— no contempla, y en esta familia eso son sobre todo objetos y ocho pronombres oblicuos que la lista no reconoce como tales.

Aquí está **k·s|tabaco**, 33 lenguas, el segundo código de mayor alcance de toda la familia, que §6.2 del paper usa como su aviso central: el tabaco llegó a Nueva Guinea en el siglo XVII y el catálogo lo agrupa exactamente igual que a cualquier herencia.

Y aquí están los diez de mayor enriquecimiento de la tabla de la cabecera —zarigüeya, cocodrilo, tambor, hacha, arpón, falda de hierba, taparrabo—, todos de una sola rama. Léase esta sección sabiendo que su techo por alcance es un préstamo del siglo XVII y su techo por enriquecimiento es una lista de objetos de una sola rama. Ninguno de los dos es lo que el lector busca cuando mira un catálogo de códigos, y el catálogo no tiene con qué apartarlos.

k·s|TOBACCO · binario · 33 lenguas · enriquecimiento 74×

Amowe kaso	Arawum kʌs	Baimak kas	Bongu kas
Gumalu kas	Kare kʌ:si	Lemio kʰʌs	Matepi kʌs
Munit kasa	Panim kʌs	Rempi kʌs	Sam kas
Sausi kasi	Silopi kas	Urigina ka:si	Wagi kʌs

b·m|G–STRING · binario · 23 lenguas · enriquecimiento 127×

Amele bam	Bagupi ba:m	Baimak bam	Gal bam
Garus baim	Kein ba:m	Matepi bam	Mosimo ba:m
Nake ba:m	Nobonob bam	Rapting baim	Rempi baim
Saruga ba:me	Utu ba:m	Wamas ba:m	Yaben bamu

n·n|THEIR · binario · 23 lenguas · enriquecimiento 28×

Anam nɐ-ŋa-	Anamgura nɐ-ŋ	Asas nen-	Faita nɐ-nɐ-
Ganglau naŋ-	Lemio nʌn-	Migum nʌn-	Mum nuŋu-
Rerau nani-	Saep naŋ-	Sinsauru nenɛ-	Siroi nanɛ-
Tauya nɐ-nɐ-	Urigina nen-	Wadaginam ni-ŋa-	Waskia nu-na

m·g|BANANA · binario · 20 lenguas · enriquecimiento 39×

Bagupi mug	Baimak mu:g	Bongu moga	Gal mug
Mawan mug	Mosimo mugu	Murupi mugu	Nake mug
Rapting mug	Samosa mugu	Saruga mu:gu	Silopi mug
Utu mug	Uya mug	Wagi mug	Wamas mug

m·r|VINE · binario · 19 lenguas · enriquecimiento 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 mu:ri

b·m|SAGO · binario · 19 lenguas · enriquecimiento 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) · binario · 19 lenguas · enriquecimiento 14×

Amaimon ani-ŋi	Anam na-ŋ	Anamgura na-ŋ	Bilakura aŋ-ina
Dedua ŋeni	Faita na-nɐ	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 · binario · 19 lenguas · enriquecimiento 16×

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

n·g|FEMALE PERSON · binario · 17 lenguas · enriquecimiento 25×

Bagupi anɔg	Gal anag	Isebe anɔg	Kein ana:g
Matepi anɔg	Mawan anɔg	Murupi anɔ:gu	Nake anɔ:g
Panim anɔg	Rapting anɔg	Samosa anɔ:gu	Saruga anɔ:gu
Sihan anɔgu	Silopi anɔg	Utu anɔg	Wagi anig

n·g|FEMALE · binario · 17 lenguas · enriquecimiento 28×

Bagupi anɔg	Gal anag	Isebe anɔg	Kein ana:g
Matepi anɔg	Mawan anɔg	Murupi anɔ:gu	Nake anɔ:g
Panim anɔg	Rapting anɔg	Samosa anɔ:gu	Saruga anɔ:gu
Sihan anɔgu	Silopi anɔg	Utu anɔg	Wagi anig

g·l·m|DRUM · ternario · 16 lenguas · enriquecimiento 262×

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

w·t|FILL · binario · 16 lenguas · enriquecimiento 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) · binario · 16 lenguas · enriquecimiento 40×

Amele mete-	Baimak mete	Bau met ^h	Gal mete
Garus met ^h	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 · binario · 15 lenguas · enriquecimiento 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) · binario · 15 lenguas · enriquecimiento 17×

Bagupi neg	Garus neg	Mandobo Atas ɲgo	Matepi neg
Mosimo nɔg	Murupi naga	Nake nɔg	Rapting nag
Samosa nɔge	Saruga naga	Silopi neg	Utarmbung na-gə
Wamas nage	Wambon ɲgu	Yoidik nɔg	

b·b|SAME-SEX OLDER SIBLING · binario · 15 lenguas · enriquecimiento 122×

Amowe babo	Anjam babo	Bargam -bab	Ganglau baba-
Garus bab	Kobol bab	Maia -bab	Mala -bab
Malas baba	Miani baba	Ogea -babo	Parawen baba
Rempi baba	Yabong baba	Yarawata babo	

m·l|VINE · binario · 14 lenguas · enriquecimiento 34×

Amowe malu	Bagupi muli	Garus mul_i	Jilim mul
Migum moli	Mosimo muli	Ogea mΛ:lu	Rapting muil
Rempi mul_i	Rerau mul	Siroi muli	Waskia mΛil
Yangulam muli	Yoidik mul_i		

n·g|YOU (OBLIQUE CASE OF YOU) · binario · 14 lenguas · enriquecimiento 15×

Aeka inggo	Amaimon ani-g	Anjam -ng-	Asas -nVg-
Bongu -ing-	Ewage-Notu inggo	Ganglau -nag-	Kobol no-g
Male -ng-	Migum -ng-	Sam -ng-	Sinsauru -neg-
Sop -nig-	Waube -nigi-		

k·k|carry · binario · 14 lenguas · enriquecimiento 34×

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

k·m|BOW · binario · 14 lenguas · enriquecimiento 23×

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

m·l|G-STRING · binario · 14 lenguas · enriquecimiento 43×

Amowe ma:le	Anjam mal	Bongu mal	Jilim mΛl
Male mal	Migum mΛl	Ogea mΛ:le	Pulabu mal
Rerau mΛlu	Sam mal	Sinsauru mΛ:l	Siroi mal
Waube mΛ:le	Yangulam mal		

g·m|BASKET · binario · 13 lenguas · enriquecimiento 49×

Bagupi gΛ:m	Bau gam	Garus gΛm	Gumalu gΛm
Isebe gam	Mawan gΛm	Mosimo gΛm	Munit gama
Murupi gΛ:m	Nake gΛ:m	Panim gΛm	Rapting gam
Yoidik gΛ:m			

g·r·m|DRUM · ternario · 13 lenguas · enriquecimiento 190×

Anjam guruma	Bau giram	Danaru garΛm	Faita giroma
Isebe giram	Jilim gurumo	Silopi giram	Sumau giroma
Utu girom	Uya girom	Wagi girΛm	Yaben girΛ:m
Yangulam geram			

t·n|YOU (OBLIQUE CASE OF YOU) · binario · 13 lenguas · enriquecimiento 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) · binario · 13 lenguas · enriquecimiento 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 · binario · 13 lenguas · enriquecimiento 37×

Bagupi kuki-	Nake kuku-	Nobonob kuk-	Pal kɔki
Panim kuku-	Rapting kuku-	Rempi kuku-	Saruga kuki-
Sausi kakau-	Urigina kuku-	Utu koki-	Wagi kuku-
Waube koko-			

s·n|OUR · binario · 13 lenguas · enriquecimiento 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 · ternario · 13 lenguas · enriquecimiento 309×

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

m·n|HAIR (BODY) · binario · 12 lenguas · enriquecimiento 5×

Aisi mini	Apali mini	Kobon u'maŋɔ	Korak -munɛ
Mala -imun	Miani -umun	Mum mɛn	Parawen muna
Sileibi minə	Utarmbung mini-	Waskia mo:nɛ	Yarawata munia

w·g|CROCODILE · binario · 12 lenguas · enriquecimiento 108×

Amele waga	Bagupi wagɔi	Gal wagai	Gumalu wagɔi
Isebe wagi	Matepi waga	Mawan waga	Panim waga
Silopi wagɔ:	Utu wa:ga	Wagi wagə	Waube wagai

s·m|POT · binario · 12 lenguas · enriquecimiento 78×

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

s·m|COOKING POT · binario · 12 lenguas · enriquecimiento 89×

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

g·n|HOW MANY · binario · 12 lenguas · enriquecimiento 38×

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

s·m|VULVA · binario · 12 lenguas · enriquecimiento 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 · binario · 12 lenguas · enriquecimiento 156×

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

m·n|BOY · binario · 11 lenguas · enriquecimiento 6×

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

w·g·j|CROCODILE · ternario · 11 lenguas · enriquecimiento 278×

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

n·n|YOUR (PLURAL) · binario · 11 lenguas · enriquecimiento 13×

Anam na-ŋ-	Anamgura na-ŋ-	Faita na-nɤ-	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 · binario · 10 lenguas · enriquecimiento 8×

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

b·t|tail · binario · 10 lenguas · enriquecimiento 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 · binario · 10 lenguas · enriquecimiento 22×

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

t·m|NAPE (OF NECK) · binario · 10 lenguas · enriquecimiento 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 · binario · 10 lenguas · enriquecimiento 23×

Amaimon mu:ku	Bilakura muku	Ganglau muku	Parawen mu:ku
Telefol mook	Tifal mook	Ukuriguma mu:k	Yaben muku
Yabong mikε-	Yarawata muka		

t·t|SMALL (NOT TALL) · binario · 10 lenguas · enriquecimiento 19×

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

t·t|tie · binario · 10 lenguas · enriquecimiento 23×

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

k·n|FILL · binario · 10 lenguas · enriquecimiento 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 · ternario · 10 lenguas · enriquecimiento 155×

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