

Códigos consonánticos en 111 lenguas afroasiáticas

La familia donde la raíz consonántica ya era una teoría

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

Resumen

Presentamos un catálogo de 2.121 códigos consonánticos que agrupan 11.082 formas léxicas de 111 lenguas afroasiáticas. Un código es un esqueleto consonántico compartido por palabras del mismo concepto en lenguas distintas. Su construcción no usa etimología, protoformas ni juicios de cognación: solo la forma transcrita, el concepto que expresa y el requisito de que la coincidencia se dé en al menos tres lenguas.

El procedimiento agrupa el 17,4 % de las formas disponibles y su precisión supera en 44 veces lo esperable por azar (5,26 % frente a 0,119 %). Sobre el agrupamiento sin umbral de tres lenguas, la ventaja es de 25,0 puntos —29,7 % real frente a 4,7 % del nulo, 6,3 veces el azar—.

Esta familia se eligió por una razón que ninguna otra de la serie ofrece: es la única donde la raíz consonántica ya era una teoría antes que nosotros. La raíz trilítera semítica —*k-t-b* «escribir»— es exactamente nuestro objeto, descrito por otra vía y hace más de mil años. Por primera vez el instrumento tiene contra qué contrastarse, y no contra sí mismo.

El contraste se preregistró antes de correr nada. Su resultado principal se confirma: el semítico concentra el 50,6 % de sus formas en esqueletos de exactamente tres consonantes frente al 31,3 % del chádico (+19,3 puntos, $p = 0,0001$ con nulo permutando la rama sobre lenguas). Y no es que sus palabras sean más cortas: las medias de consonantes por forma son $2,99 \cdot 3,13 \cdot 3,02 \cdot 3,02$ en las cuatro ramas. Es concentración, no longitud. Un procedimiento ciego a la gramática reencuentra la forma de la raíz trilítera sin haberla usado.

Dos hipótesis del mismo preregistro no se confirman, y diagnostican el instrumento en vez de la familia. Se publican con el mismo detalle.

Los códigos de mayor alcance reproducen el vocabulario que la comparación afroasiática ya había establecido — $m \cdot t$ para *morir* en 38 lenguas de dos ramas, $d \cdot m$ para *sangre* en 29, $s \cdot m$ para *nombre* en 25— sin haber reconstruido nada.

1. Qué se busca, y por qué el afroasiático

Este trabajo continúa una serie que aplica el mismo instrumento a familias distintas. 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, para que lo que sigue se lea:

Un código es un esqueleto consonántico que aparece para el mismo concepto en tres lenguas distintas o más. No se afirma que sus miembros sean cognados, ni que desciendan de una protoforma, ni que el código signifique nada. Un grupo es un grupo: dos hechos observables —forma y concepto— y nada más.

Y esta familia se eligió por una razón que ninguna otra de la serie ofrece. En austronesio, en turco, en urálico, en pama-ñungano y en indoeuropeo, el esqueleto consonántico era un objeto nuestro: lo definíamos, lo medíamos y lo contrastábamos contra su propio nulo. Aquí no. La raíz trilítera semítica es un esqueleto consonántico, descrito por la filología árabe y hebrea hace más de mil años, y descubierto por una vía que no tiene nada que ver con la nuestra.

Eso convierte al afroasiático en el único sitio del corpus donde el instrumento puede converger o discrepar con una descripción externa. Es la prueba más informativa de la serie, y también la más arriesgada: si el esqueleto no reencuentra la raíz allí donde la raíz existe y está descrita, el problema es del esqueleto.

1.1 Lo que este paper NO hace

No reconstruye el protoafroasiático ni lo consulta. No usa los juicios de cognación que existen para esta familia. No propone etimologías. Y no afirma que los códigos de mayor alcance sean cognados, aunque muchos lo sean: el catálogo no puede distinguir herencia de contacto, y no se va a fingir que sí.

Lo que sí hace: construir el catálogo como en las cinco familias anteriores, y después medir el contraste con la raíz trilítera como una hipótesis preregistrada, con su nulo, su dirección declarada y sus falsadores escritos antes de mirar.

2. La población, y las decisiones que la fijan

2.1 Lo que entra

111 lenguas · 63.768 formas con esqueleto · 2.031 conceptos · 4 ramas.

Rama	Lenguas	Formas con esqueleto
Chadic	80	47.202
Semitic	27	12.162
Berber	2	2.262
Cushitic	2	2.420

El 72 % de las lenguas es chádico, y dos ramas tienen dos lenguas cada una. Esto gobierna lo que el paper puede sostener y se dice aquí, antes de los resultados, no después.

Faltan omótico y egipcio. El corpus no los trae para esta familia. El catálogo no cubre el afroasiático: cubre las cuatro ramas que el corpus contiene, y así se le debe citar.

2.2 Una lengua excluida, y veintinueve que no

La fase 0 excluye **Bara**, declarada chádica. El corpus tiene dos lectos con ese nombre, ambos con 414 formas y ambos con coordenadas en el sur de Madagascar: uno austronesio (bara1369) y otro afroasiático (bara1395). Sobre los 207 conceptos comunes, el 66 % de las formas son idénticas, y las que difieren lo hacen como dos encuestas de la misma lengua. La muestra no deja duda: biby «animal», zaza «niño», velo «vivo», satria «porque», manapaky «cortar». Es malgache. Una wordlist austronesia enganchada a un glotocódigo chádico.

Dos precisiones sobre esta exclusión, porque el 66 % invita a leerla mal. La primera: *no* es un duplicado por la regla mecánica de la serie —mismo glotocódigo, o más del 80 % de esqueletos idénticos—, que

Bara no cumple: los glotocódigos difieren y el 66 % está por debajo del umbral. Se excluye por mal_ asignado, que es otra clase con otra evidencia: el vocabulario y las coordenadas. El 66 % corrobora, no decide.

La segunda: excluirla no cambia los resultados. Con Bara dentro, la precisión de la familia es 43× en vez de 44×, y nada más se mueve. La exclusión es correcta y es inmaterial, y las dos cosas conviene decir las: un catálogo que solo documenta las decisiones que cambian algo esconde las que no.

Y aquí hay que informar de un chequeo que fallamos a propósito. El mismo procedimiento de fase 0 propuso excluir 30 de 112 lenguas por distancia al centro geográfico de la familia. Entre ellas: árabe estándar, hebreo moderno, amárico, ge'ez, maltés, tigríña, tarífit. Es decir, casi todo el semítico y todo el bereber.

El chequeo no está roto: mide lo que se le pidió, y lo que se le pidió no tiene sentido aquí. El centro de la familia cae en 10,8 °N · 13,8 °E —el corazón chádico— porque el 72 % de las lenguas del corpus lo son. Medida contra ese centro, una familia que se extiende de Marruecos a Omán declara intrusa a su propia mitad norte. En urálico, familia compacta, ese mismo chequeo cazó una lengua tai-kadai a 6.000 km; aquí 6.000 km son el árabe de Bujará, que es legítimo.

Se rechazan las 29. Y se deja escrito, porque vale para toda la serie: la distancia al centro es informativa en familias compactas y no dice nada en familias extensas. El chequeo debe declarar la dispersión de la familia junto a la lista, y abstenerse de proponer cuando la mediana es alta.

Aplicar esas 29 exclusiones habría borrado el semítico, que es la razón por la que se eligió esta familia.

2.3 Las formas que no entran, y no cuestan cobertura

De 68.823 formas, 4.777 no obtienen esqueleto (7 %), y son exactamente las 4.777 de la fuente IDS: verificado fuente a fuente, lexibank 62.239 de 62.239 con esqueleto, nel 2.221 de 2.221, ids 0 de 4.777. No es una caída silenciosa del pipeline: es la decisión de alcance ya vigente en la serie —IDS trae ortografía y no IPA— manifestándose.

Cero conceptos se pierden del todo. Tres lenguas dependen de IDS en parte —arameo antiguo 50 %, polci 37 %, hausa 28 %— y ninguna queda a cero.

Una consecuencia práctica: el chequeo de forma de cita, que en turco destapó el hallazgo más caro de la fase 0 —el infinitivo -mAK que unas fuentes ponen y otras no—, aquí señala terminaciones (-tā, -rā, -nā, -lā, entre el 2 % y el 6 %) que son todas de IDS. No entran al catálogo. Es ruido de una fuente que no participa.

2.4 Dos artefactos de transcripción declarados

La labialización etiosemitica escrita como dígrafo. En mesqan, ge'ez y kistane, gwəbbe = /g^wəbbe/, kwəll = /k^wəll/ «todo», čəkkwəɾə: la consonante labializada se escribe Cw y el esqueleto cuenta la w como consonante aparte. Son unas veinte formas. Se declara y no se corrige, porque distinguirla del grupo /kw/ genuino —el maltés 'gwerre, préstamo italiano; el árabe lwara:ʔ, artículo más raíz— exigiría una decisión por lengua que el corpus no sostiene.

El modificador glotal contado como consonante. ʔ y ʕ son letras modificadoras —glotalización, faringalización— y no consonantes. El segmentador de las fuentes las convierte en un segmento ʔ/ʕ propio cuando siguen a vocal y las absorbe cuando siguen a consonante. 248 formas afroasiáticas en 17 lenguas caían en ese caso. Se corrigió en el corpus antes de construir este catálogo, dejando intactos los ataques glotales a principio de palabra, que en chádico pueden ser consonante de pleno derecho.

3. Qué cuesta el instrumento aquí

Colisión: 38,1 %. Es el porcentaje de formas cuyo esqueleto, *dentro de su propia lengua*, cubre más de un concepto. Es el coste puro del instrumento, sin genealogía de por medio: cuánta información se pierde solo por retirar las vocales.

Aridad media: 3,16 consonantes por forma. La familia está en la mitad barata de la serie: por encima del sino-tibetano (2,22) y del austronesio (2,79), por debajo del pama-ñungano y el indoeuropeo (3,35).

Es coherente con lo que la propia familia es. Una morfología cuya raíz canónica tiene tres consonantes produce palabras con tres consonantes, y tres consonantes son suficientes para distinguir: el 85 % de los códigos que salen tiene tres o menos.

Nueve cadenas cubren cinco conceptos o más dentro de una misma lengua, y no todas son defecto del dato. *kwam* en polci cubre ocho conceptos que son todos «suegro/suegra/nuera» —colexificación real de un sistema de parentesco—; *ḍib̥kidi* en miya cubre siete decenas; *ḍáukàa* en hausa cubre siete maneras de llevar algo. Se informan para que la colisión se lea sabiendo cuánta viene de ahí.

4. El catálogo

2.121 códigos sobre 2.031 conceptos, que agrupan 11.082 formas de 63.768 (17,4 %). De ellos, 1.732 son binarios o ternarios (81,7 %), con enriquecimiento mediano de $100\times$ y mediana de 3 lenguas por código.

Todas esas cifras dependen de un umbral que es convención y no hallazgo, y conviene enseñar cuánto:

Umbral de lenguas	Códigos	Formas agrupadas	Cobertura
2	5.651	18.960	29,7 %
3	2.121	11.082	17,4 %
4	1.032	7.439	11,7 %
5	620	5.606	8,8 %
6	424	4.509	7,1 %

Cada escalón divide el catálogo aproximadamente por la mitad, y no hay codo. Ningún umbral se descubre en los datos: el tres es la convención de la serie —dos lenguas no distinguen coincidencia de repetición— y se mantiene por comparabilidad entre familias, no porque estos datos lo señalen.

Lo que conviene retener es qué depende de él y qué no. La cobertura, el número de códigos y todo el §6 dependen. El resultado principal del paper, H1, no: se mide sobre formas y su esqueleto, sin pasar por el catálogo, así que es indiferente al umbral.

Aridad	Códigos	%
1	157	7,4 %
2	830	39,1 %
3	902	42,5 %
4	191	9,0 %
5+	41	1,9 %

Precisión: 5,26 % frente a 0,119 % del azar = $44\times$ [44–44× según semilla]. La razón se da descompuesta a propósito: la precisión observada es baja en términos absolutos, y lo que la hace informativa es contra qué se compara.

Y hay que descomponerla una vez más, porque la cifra no es homogénea. El plan de la serie avisa de que un $\times$ azar alto puede medir la redundancia del corpus —muchas variedades próximas del mismo material— antes que el acierto del instrumento. Aquí ocurre:

	Precisión	$\times$ azar
Familia entera	5,26 %	44 $\times$
Solo chádico (80 lgs)	5,08 %	34 $\times$
Solo semítico (27 lgs)	18,68 %	175 $\times$
Familia sin los 13 lectos árabes	4,40 %	36 $\times$

El semítico solo rinde 175 $\times$, cuatro veces más que la familia, y quitar los trece lectos árabes baja el titular de 44 $\times$ a 36 $\times$: el 18 % de la cifra de cabecera viene de ese bloque. Ocho dialectos árabes con una lista de conceptos común comparten esqueleto porque son variedades de lo mismo, no porque el método acierte más con ellos.

El 44 $\times$ es correcto y es la cifra de la familia tal como el corpus la trae. Pero no se lee sola: se lee junto al 36 $\times$ sin los árabes, que es la parte de la cifra que no depende de cuántas variedades próximas haya recogido la fuente.

Y sobre el agrupamiento sin umbral de tres lenguas, que es la medida que permite comparar con el nulo directamente: 29,7 % real frente a 4,7 % del nulo —ventaja de 25,0 puntos, 6,3 $\times$ el azar.

4.1 Los códigos de mayor alcance reproducen los comparanda clásicos

Éstos son los quince códigos de más alcance del catálogo, sin filtrar, porque la lista sin filtrar dice más que una selección:

Código	Concepto	Lenguas	Aridad	Rama dominante
m·t	<i>morir</i>	38	2	Chadic
m	<i>agua</i>	33	1	Chadic
l	<i>vaca</i>	31	1	Chadic
d·m	<i>sangre</i>	28	2	Semitic
m·b·j	<i>mandioca</i>	27	3	Chadic
b·b	<i>padre</i>	26	2	Chadic
s·m	<i>nombre</i>	25	2	Semitic
s	<i>beber</i>	25	1	Chadic
s·n	<i>diente</i>	22	2	Semitic
l·m	<i>nombre</i>	22	2	Chadic
t·r	<i>luna</i>	22	2	Chadic
l·m	<i>oreja</i>	21	2	Chadic
j·m	<i>agua</i>	20	2	Chadic
f·d	<i>cuatro</i>	20	2	Chadic
v·j	<i>estación lluviosa</i>	20	2	Chadic

Hay tres clases distintas ahí dentro, y confundirlas sería el error fácil.

Comparanda establecidos. m·t es el *mwt* «morir» pan-afroasiático, y es el código de mayor alcance de todo el catálogo; d·m es el *dam* «sangre» semítico; s·m el *šm* «nombre» —árabe *ism*, hebreo *šem*—; s·n el *sinn* «diente». El procedimiento no los buscó: salieron de exigir que un esqueleto se repitiera para el mismo concepto en tres lenguas.

Códigos unarios, y aquí hay que ser preciso porque la respuesta fácil es falsa. m|agua en 33 lenguas, l|vaca en 31, s|beber en 25. Lo cómodo sería decir que una sola consonante coincide por disponibilidad y no por historia. Pero l|vaca son *sla*, *lâ*, *lâ*, *slâ*: la lateral fricativa *ɬ* que la normalización

escribe **l**, y ese *la* «vaca» es un etimón chádico de pleno derecho. El código es real; lo que pasa es que tiene una sola consonante.

El problema, entonces, no es que los unarios sean falsos: es que el catálogo no puede distinguir un etimón monoconsonántico verdadero de una coincidencia por disponibilidad, porque a los dos les da la misma evidencia. Por eso el paper trabaja con binarios y ternarios —el 81,7 % del catálogo— y los unarios se cuentan pero no se leen. No se descartan por falsos, se apartan por indistinguibles.

Y un préstamo moderno, que el catálogo no distingue de lo demás. **m·b·j** |mandioca alcanza 27 lenguas. La mandioca es un cultivo americano introducido en África en época colonial: ese código no es herencia ni contacto antiguo, es un préstamo reciente con su palabra. El catálogo lo agrupa exactamente igual que a **m·t**, y no tiene con qué separarlos. Es el mejor ejemplo del límite que §1.1 declara, y por eso se deja en la tabla en vez de quitarlo.

Lo que la sección sí sostiene, entonces, es más estrecho: cuando hay una correspondencia establecida y de alcance amplio, el catálogo la contiene. Lo que no sostiene es lo contrario —que todo lo de alcance amplio sea una correspondencia—, y la mandioca está ahí para recordarlo.

5. La espina: el contraste con la raíz trilítera

Todo lo que sigue estaba preregistrado antes de correr nada (`runs/preregistro-espina-semitica.md`), con su dirección declarada y sus falsadores escritos. Se hizo así porque el modo de fallo diagnosticado en trabajos anteriores fue «titular primero, nulo después», y aquí la tentación es máxima: la filología semítica lleva mil años diciendo qué tendría que salir.

Y el preregistro tiene dos debilidades que conviene decir aquí y no dejar que las encuentre el lector.

La primera: sus falsadores no fijan magnitud. El de H1 dice «si el semítico no concentra en 3 más que el chádico, el contraste no existe». Escrito así, cualquier diferencia positiva confirma: un 33 % frente a un 31 % habría pasado. Un preregistro sin efecto mínimo declarado no puede fallar contra un resultado pequeño pero del signo correcto, y eso es la mitad de para lo que sirve un preregistro. Aquí no importó porque la diferencia fue de 19,3 puntos con separación completa entre lenguas, pero no importó por suerte, no por diseño.

La segunda: H1 era la menos arriesgada de las tres. Que la raíz semítica tenga tres consonantes lo sabe cualquiera que haya abierto una gramática árabe. Lo que estaba en duda no era el semítico sino si nuestro esqueleto lo recuperaría, y eso es un contraste sobre el instrumento, no sobre la lengua —el paper lo dice en §8, pero el aparato de preregistro le presta a H1 un aire de riesgo que no tenía. H2 y H3 sí eran inciertas de verdad, y las dos fallaron. Registrar tres hipótesis y confirmar una suena a disciplina; conviene añadir que la que sobrevivió era la que menos se jugaba.

5.1 La predicción, que se invierte respecto a lo obvio

La lectura ingenua sería «el instrumento funcionará mejor en semítico». La predicción registrada es la contraria, y por una razón concreta: la morfología semítica es templática. Sus afijos son vocales y patrones interdigitados, no consonantes pegadas al borde. Si la teoría nativa es correcta, entonces el esqueleto consonántico de una palabra semítica ya es casi su raíz, y debería concentrarse en tres.

5.2 H1 · confirmada, y es el resultado del paper

Medida	Semitic	Chadic	dif.	p
Formas con esqueleto de exactamente 3 consonantes	50,6 %	31,3 %	+19,3	0,0001

El nulo permuta la etiqueta de rama sobre doculectos —27 lenguas frente a 80, no 12.017 formas frente a 47.162—, con 10.000 permutaciones. Es la lección de la pseudorreplicación: la unidad de intercambiabilidad es la lengua, no la forma.

Y ese nulo todavía supone demasiado, así que conviene una cifra que no suponga nada. Permutar sobre lenguas trata las 27 semíticas como intercambiables, y no lo son: once son etiosemiticas con una lista de conceptos idéntica y ocho son dialectos árabes. El n efectivo es menor que 27, y con un n bastante menor el valor p se ablandaría.

La respuesta es mirar el nivel de la lengua directamente. Las 27 lenguas semíticas están por encima de la mediana chádica. Las veintisiete. El rango semítico va de 40,3 % (harari) a 62,4 % (árabe egipcio); el chádico, de 23,1 % a 40,9 %. Sobre los 2.160 pares lengua-a-lengua, el semítico concentra más en 2.158, un AUC de 0,999. Los dos pares que van al revés son el mismo caso: el harari, la semítica más baja, contra el mbazla, la chádica más alta.

Ese número no supone independencia entre lenguas, ni normalidad, ni nada: dice que si se toma una lengua semítica y una chádica al azar, la semítica concentra más en tres consonantes el 99,9 % de las veces. Aunque el n efectivo fuera de tres o cuatro grupos independientes, la separación seguiría siendo completa.

5.2-bis · Y no es un efecto del proyecto que recogió los datos

Queda una objeción más grave que las anteriores, y hay que responderla con datos. El chádico y el semítico del corpus vienen de proyectos de campo disjuntos: kraftchadic (61 lenguas) y gravinachadic (30) por un lado; ratcliffearabic (14) y felekesemitic (11) por el otro. Cada proyecto transcribe con sus convenciones. El contraste entre ramas podría ser un contraste entre equipos.

La respuesta está en que tres proyectos cubren las dos ramas, y sirven de control interno:

Proyecto	Chádico	Semítico	Diferencia
gravinachadic	31,1 % (17.231 formas)	45,8 % (753)	+14,8
idssegmented	31,7 % (1.983)	55,1 % (1.924)	+23,5
polyglottaaficana	29,6 % (270)	47,1 % (626)	+17,5

Dentro del mismo proyecto y con la misma práctica de transcripción, el semítico concentra entre 15 y 24 puntos más. Y el chádico es notablemente estable entre cinco proyectos independientes —29,6 · 31,1 · 31,2 · 31,7 · 32,3 %—: el efecto de proyecto ronda los 3 puntos y la brecha entre ramas los 19.

La objeción era correcta y la respuesta es que no explica el resultado.

Y la comprobación que hace legible el número: no son palabras más cortas.

consonantes	Semitic	Chadic	Berber	Cushitic
1	2,8 %	5,6 %	7,1 %	4,0 %
2	25,9 %	30,7 %	27,9 %	32,5 %
3	50,6 %	31,3 %	35,6 %	36,1 %
4	14,0 %	17,4 %	18,0 %	16,1 %
5	3,6 %	7,4 %	8,0 %	7,2 %
6+	3,0 %	7,6 %	3,3 %	4,1 %
media	2,99	3,13	3,02	3,02

Las medias son prácticamente idénticas en las cuatro ramas. Lo que distingue al semítico es la concentración: pone la mitad de su masa en un solo valor y las otras tres la reparten.

Bereber (35,6 %) y cushítico (36,1 %) quedan entre medias, en la dirección del gradiente de templaticidad que el preregistro anticipó. No se apoya nada en ellos: son dos lenguas cada uno, y así estaba declarado antes de mirar.

Y hay una segunda vía, que no estaba preregistrada y por eso se presenta como corroboración y no como prueba. H1 mide formas; se puede medir lo mismo sobre los códigos ya construidos, preguntando qué rama domina cada uno según su aridad:

Aridad del código	Chádicos	Semíticos	% semítico
1	190	14	6,8 %
2	868	150	14,7 %
3	476	237	33,2 %
4	116	32	21,6 %
5	21	3	12,5 %

El máximo está exactamente en tres, y los ternarios son 2,3 veces más semíticos que los binarios. La curva sube hasta tres y baja después, que es la forma que H1 predice trasladada del nivel de la forma al del código.

Con dos cautelas. La primera: dos tercios de los códigos ternarios siguen siendo chádicos (476 frente a 237). La afirmación es sobre proporciones relativas, no sobre una partición. La segunda: esto no es independiente de H1 —los códigos se construyen desde las formas—, así que corrobora sin añadir un contraste nuevo.

Antes de leer esta cifra se comprobó, también según lo preregistrado, que la forma de cita no la fabrica: ninguna terminación supera el 2,9 % en ninguna de las dos ramas, y no hay nada parecido a la convención de infinitivo que en turco resultó decisiva. Queda declarada una asimetría: el semítico bebe de dos fuentes y el chádico de una.

5.3 H2 · falsada

Medida	Semitic	Chadic	dif.	p
Formas que el despiece automático modifica	65,6 %	59,6 %	+6,0	0,049

La predicción era semítico menor que chádico. Sale al revés, y débilmente: con dos contrastes y corrección de multiplicidad queda en 0,10. No se apoya nada en el signo; lo que se afirma es que la predicción no se cumple.

Y la lectura no es sobre las ramas sino sobre el instrumento: un despiece que modifica tres de cada cinco formas en las dos ramas no está detectando morfología.

5.4 H3 · no discrimina

La lista preregistrada de afijos consonánticos pansemíticos era m-, t-, n-, ʔ-, con ʔ- declarado intestable porque la normalización funde ʔ ɸ h en un solo símbolo.

Los prefijos que el descubrimiento automático halla a ciegas, por número de lenguas donde aparecen: k- 63 %, m- 63 %, d- 56 %, h- 56 %, s- 48 %, b- 44 %, r- 37 %, n- 33 %, g- 33 %, t- 30 %.

m- está arriba. Pero k-, d-, b-, s-, r- y g- aparecen al mismo nivel o por encima, y no son afijos semíticos: son consonantes de raíz. Que m- salga primero es compatible con que sea real y con que sea casualidad; con esta tabla no se puede decidir, y no se decide.

5.5 Lo que las dos hipótesis caídas enseñan

H2 y H3 dicen lo mismo por dos vías: el descubrimiento automático de afijos marca de más. Medido, marca como afijo el 69 % del inventario consonántico de cada lengua afroasiática.

Y ese hallazgo generaliza fuera de esta familia: la misma medición en las cinco anteriores da entre el 59 % y el 89 %. En una familia cuya raíz es un trío de consonantes, marcar siete de cada diez consonantes como afijo significa recortar la raíz.

Medido sobre el agrupamiento, que es el criterio que el protocolo declara:

Paso	En grupo	Nulo	Ventaja	κ	Códigos de 4+ consonantes
Palabra entera	29,7 %	4,7 %	25,0	0,262	29,9 %
+ afijos descubiertos	38,8 %	17,5 %	21,3	0,258	9,6 %

Ocho puntos de señal y trece de ruido. La columna κ es la corrección de Hubert-Arabie, (observado – nulo) / (1 – nulo), que normaliza por el techo alcanzable y permite comparar entre familias: aquí $\Delta\kappa = -0,004$. Las ocho familias de la serie dan el mismo signo: turco $-0,009$, afroasiático $-0,004$, sino-tibetano $-0,020$, urálico $-0,035$, indoeuropeo $-0,094$, pama-ñungano $-0,102$, trans-neoguineano $-0,118$, austronesio $-0,151$. (*Lista rehecha con las ocho el 2026-09-09 desde pipeline/tools/tabla_serie.py. La magnitud NO crece con la ganancia bruta, que es lo que este párrafo decía: el sino-tibetano, con la palabra más corta de las ocho, cae menos que el indoeuropeo.*) La columna de cuaternarios se desploma del 29,9 % al 9,5 % y es espectacular, pero el criterio no es que baje la aridez —eso se consigue despiezando de más— sino que suba el agrupamiento sobre el nulo, y no sube.

Por eso el catálogo de este paper no usa el descubrimiento automático de afijos, igual que no lo usan los de la serie. La fase queda como diagnóstico, y su resultado se publica porque es informativo, no porque se aplique.

6. Lo que esta familia permite, y lo que no

6.1 Los candidatos genealógicos son pocos, y se dice por qué

Con cuatro ramas y el 72 % de las lenguas en una sola, un código que cruce tres ramas es en la práctica un código chádico-semítico que además asome en una de las cuatro lenguas restantes. Los 349 candidatos que dio el pama-ñungano con quince ramas equilibradas no se repiten aquí, y no por culpa del método.

Lo que sí funciona es el corte principal, chádico frente a semítico, que reparte 80 y 27 lenguas: 14.328 sustituciones cruzan ese corte, el 15,7 % del total, y dan candidatos de correspondencia con su cuota de concentración.

6.2 La lista de anomalías sí existe: dos macroáreas

Es una de las cinco familias del corpus con más de una macroárea —las otras son indoeuropeo, austronesio, arawak y chibcha—. 393 códigos (7,0 %) cruzan macroáreas y son los que merecen una segunda mirada: coincidencia de forma y sentido entre lenguas que no comparten vecindario.

6.3 Pares de sustitución: 513, de los cuales 450 utilizables

Pares de códigos del mismo concepto que difieren en una sola posición. De los 513, 450 son disjuntos en lenguas —ninguna lengua está en los dos— y solo esos pueden proponerse como isoglosa. Los 63 que solapan no son erróneos: una lengua puede tener dos palabras, la misma raíz con dos sufijos, o dos transcripciones.

Los de más alcance: *l* *s* en *nombre* (*l*·*m* chádico 22 lenguas frente a *s*·*m* 25, semítico y chádico), *c* *t*, *s* *t* y *r* *t* en *morir* frente al *m*·*t* mayoritario, *l* *r* en *aceite*, *b* *h* en *padre*.

Ninguno se buscó. Salieron de contar diferencias en pares ya alineados por concepto y código.

6.4 Las clases emergentes, con validación en lenguas no miradas

Las 111 lenguas se parten en dos mitades: 55 para descubrir y 56 para comprobar. Los pares que sobreviven en las dos, con su razón sobre el azar:

Par	Veces (descubre)	Razón	Veces (comprueba)	Razón
f ↔ v	71	6,9×	266	18,3×
c ↔ t	158	6,4×	281	8,0×
l ↔ ð	114	8,7×	143	7,7×
c ↔ s	287	8,0×	386	7,6×
b ↔ v	95	4,7×	202	7,1×
f ↔ p	118	7,0×	165	6,9×
h ↔ x	184	4,8×	290	5,4×
b ↔ p	201	6,0×	244	5,2×

Son alternancias de sonoridad y de modo dentro del mismo lugar de articulación. No son correspondencias regulares —falta entorno y falta dirección— sino candidatas, con el sitio donde buscarlas.

6.5 La geografía funciona

Las lenguas que comparten más códigos están más cerca, y la relación no es plana:

Códigos compartidos	Pares	Distancia media	vs azar
1	1.369	1.827 km	1,16
2	701	1.139 km	0,72
3–5	1.018	608 km	0,39
6–10	463	469 km	0,30
11–20	176	766 km	0,49
21+	216	907 km	0,58

El mínimo está en 6–10 códigos compartidos, a menos de un tercio de la distancia esperable por azar.

Que la curva suba otra vez a partir de 11 tiene una explicación concreta, y no es la que parece. No son los códigos pan-familiares apareciendo en todas partes: es qué rama llena cada tramo. Contando la composición de los pares:

Códigos compartidos	% de pares semítico-semítico	% chádico-chádico
1	1,9 %	36,4 %
2	4,5 %	54,8 %

Códigos compartidos	% de pares semítico-semítico	% chádico-chádico
3–5	6,5 %	77,2 %
6–10	6,7 %	89,5 %
11–20	17,8 %	77,7 %
21+	41,3 %	56,5 %

El mínimo de distancia coincide con el tramo más chádico (89,5 %), y el chádico ocupa un área compacta entre Nigeria y Chad. El tramo de más solapamiento se llena de pares semíticos (41,3 %, veinte veces la proporción del primer tramo), y el semítico del corpus va de Marruecos a Omán y a Etiopía.

Dicho de otro modo: la curva no mide una relación entre parecido y distancia, sino el reparto geográfico desigual de las dos ramas grandes. Que el chádico sea compacto y el semítico disperso basta para producir esta forma sin que haya nada que interpretar sobre los códigos. La sección se queda porque la primera mitad —más códigos compartidos, menos distancia, hasta 6–10— sí es informativa dentro de una rama; la segunda mitad es un artefacto del muestreo de la familia.

7. Límites

El catálogo no cubre el afroasiático. Cubre chádico, semítico, bereber y cushítico. Faltan omótico y egipcio, y bereber y cushítico entran con dos lenguas cada uno, que es presencia y no muestra.

La espina corre sobre 27 lenguas y 12.162 formas, la quinta parte de la familia. Toda afirmación sobre el contraste con la raíz trilítera lleva ese denominador.

El semítico del corpus tiene 139 conceptos por lengua de mediana, la cifra más baja de las seis familias de la serie —el urálico tiene 970—. Once lenguas etiosemiticas comparten una lista idéntica de 139 conceptos y ocho dialectos árabes traen unos 100. Eso no es un filtro nuestro: es la longitud de las listas de origen, y no hay nada que recuperar —IDS aporta cero conceptos nuevos y las dos entradas semíticas de kaikki están vacías—.

La normalización funde consonantes que en semítico son distintas, y son dos fusiones, no una.

La primera, $\text{ʔ} \ \text{ʕ} \ \text{ħ} \ \text{h}$ en un solo símbolo h . Pesa el 5,5 % de los segmentos semíticos frente al 1,9 % de los chádicos: es asimétrica entre las dos ramas que se contrastan y favorece al semítico en cualquier medida de agrupamiento.

La segunda pesa más y es la que faltaba declarar: **s** reúne **ʃ**, **ʒ**, **tʃ**, **dʒ**, **ts** y las demás sibilantes, el 6,9 % de los segmentos semíticos frente al 3,9 % de los chádicos. En semítico la oposición *s* / *ʃ* no es un detalle fonético: es la distinción del propio *shibboleth*, y separa raíces. Medido sobre el catálogo, 63 de los 418 códigos semíticos —el 15,1 %— reúnen dentro del mismo código formas con **s** y formas con **ʃ**. Uno de cada siete existe porque la tabla los escribe igual.

Ninguna de las dos afecta a H1, que cuenta posiciones y no identidades — y ésta es la razón por la que el resultado principal del paper es el que menos depende de la normalización. Las dos afectan al catálogo, al enriquecimiento y a los pares de sustitución, y se declaran aquí en vez de en una nota.

El esqueleto no captura el tono. Sobre las formas que entran al catálogo no aparece ninguna marca, ni en dígitos ni en diacríticos, pese a que el perfil global daba 47 % de tono en diacríticos para esta familia: esa cifra viene de las formas que no entran. Se declara y no se finge medirlo.

8. Conclusión

El afroasiático era, en el diseño de la serie, la familia donde el instrumento podía romperse contra una descripción externa. No se rompió: el esqueleto consonántico reencuentra la forma de la raíz trilitera sin usarla, con una concentración en tres consonantes que dobla la de la rama hermana y que no es un efecto de longitud de palabra.

Es una convergencia, no un descubrimiento. La raíz trilitera estaba descubierta. Lo que este trabajo añade es que un procedimiento que no sabe nada de gramática semítica, corriendo sobre transcripciones de listas de vocabulario, llega a la misma forma. Para un instrumento cuyo problema declarado es que se apoya solo en sí mismo fuera del indoeuropeo, eso es exactamente el apoyo externo que le faltaba.

Y las dos hipótesis que cayeron valen tanto como la que se sostuvo. El descubrimiento automático de afijos marca como afijo el 69 % del inventario consonántico, aquí y en las cinco familias anteriores. En una familia cuya raíz es un trío de consonantes, eso significa recortar la raíz — y sin esta familia, sin una teoría externa contra la que medirlo, el fallo habría seguido pasando por una ganancia.

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.

Frisch, S. A., Pierrehumbert, J. B. y Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22(1). — [la cuantificación gradiente de la restricción sobre raíces semíticas]

Greenberg, J. H. (1950). The patterning of root morphemes in Semitic. *WORD* 6(2). — [la restricción clásica sobre la composición de la raíz trilitera]

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

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

Key, M. R. y Comrie, B. (eds.). *Intercontinental Dictionary Series*.

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*.

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/afro-asiatic/output/catalogo-afro-asiatic.tsv — 2.121 filas
Detalle por forma	familias/afro-asiatic/output/ catalogo-afro-asiatic-lenguas.tsv
Higiene (§2.2)	pipeline/fases/fase0_higiene.py → runs/fase0-higiene. md
Decisiones de datos (§2.2)	runs/fase0-decisiones.md · ci/decisiones.py
Segmentos falsos (§2.4)	pipeline/fases/fase0_segmentos_falsos.py → runs/ fase0-segmentos-falsos.md
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 nullos (§4)	pipeline/fases/agrupamiento_familia.py → runs/ agrupamiento.md
Preregistro de la espina (§5)	runs/preregistro-espina-semitica.md — escrito antes de correr nada
Contraste por rama (§5.2–5.3)	pipeline/fases/contraste_por_rama.py → runs/ contraste-semitico-chadico.md
Verdad de referencia (§5.4)	runs/verdad-afijos-semiticos.md — escrita antes del barrido
Resultados de la espina (§5)	runs/resultados-espina.md
Despiece automático (§5.5)	pipeline/fases/fase2_despiece.py → runs/fase2-despiece. md
Códigos hermanos (§6.3)	pipeline/fases/codigos_hermanos.py → runs/ codigos-hermanos.md
Clases emergentes (§6.4)	pipeline/fases/fase6_clases_emergentes.py → runs/ fase6-clases.md
Sustituciones por rama (§6.1)	pipeline/fases/fase6_por_rama.py → runs/fase6-por-rama. md
Geografía (§6.5)	pipeline/fases/codigos_vs_parentesco.py → runs/ codigos-parentesco.md

Todas las cifras de este paper salen de esos scripts sobre la misma población declarada en §2.1, con las mismas exclusiones. Ninguna se escribió a mano.

Anexo A · Cien códigos comentados

El catálogo tiene 2.121 códigos y los datos completos están en familias/afro-asiatic/output/. 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 la fase 6 salen aquí como pares — $f \leftrightarrow v$, $c \leftrightarrow t$, $l \leftrightarrow \delta$, $b \leftrightarrow p$ — y no como una partición cerrada del inventario, así que escribirlas junto a cada código daría una precisión que no tienen. §6.4 las publica como candidatas, que es lo que son.

Y hay una cosa que este anexo enseña, con una advertencia que hay que leer antes. Recorriendo los bloques, muchos códigos ternarios son raíces trilíteras reconocibles — $l \cdot s \cdot n$ para *lengua*, $b \cdot t \cdot n$ para *vientre*, $r \cdot k \cdot b$ para *rodilla*, $k \cdot l \cdot b$ para *perro*, $s \cdot m \cdot s$ para *sol*, $k \cdot b \cdot r$ para *grande*—. No se buscaron ni se anotaron: son códigos que el procedimiento produjo.

La advertencia: este anexo no es una muestra representativa del catálogo, y sesga en la dirección que más favorece esa lectura. Sus cien códigos se eligen por alcance, y el alcance favorece al semítico porque los dialectos árabes del corpus comparten una lista de conceptos común y por tanto llegan a más lenguas por código. Medido:

	catálogo completo	este anexo
códigos de rama dominante semítica	21 %	41 %
entre los ternarios	33,2 %	62,9 %

Es decir: el anexo dobla la proporción semítica. La afirmación que el catálogo sí sostiene es más estrecha y está medida en §5: los ternarios son más semíticos que los binarios —33,2 % frente a 14,7 %, un factor 2,3, con el máximo exactamente en tres consonantes—, y dos tercios de los ternarios del catálogo siguen siendo chádicos. Quien lea solo el anexo se llevará una impresión más fuerte de la que los datos autorizan, y por eso se dice aquí y no en una nota al pie.

Auditoría del anexo: qué códigos dependen de la tabla, y cuántos son préstamos

Antes de los bloques conviene el resultado de auditar los cien códigos uno a uno, porque cambia cómo se leen.

Diecinueve de los noventa y uno reúnen dentro de sí símbolos que la normalización funde —s con j, o h con ʔ ʕ ħ—. Pero hay que separar dos casos que no son el mismo:

En diez, la fusión solo reetiqueta. Todas o casi todas sus formas llevan el mismo símbolo fundido: $s \cdot m \cdot s$ | $soʔ$ tiene sibilante en 19 de 21 formas, $l \cdot h \cdot m$ | carne gutural en 13 de 14. Deshacer la fusión los dejaría igual de agrupados, solo escritos de otra manera.

En nueve, la fusión es portante: el código existe porque la tabla escribe igual dos consonantes que las lenguas distinguen.

Código	Concepto	Fusión	Reparto	Ejemplo
$s \cdot n$	diente	sibilante	9 / 20	jɛn frente a siɲ
$p \cdot s$	sol	sibilante	7 / 11	pats frente a pɪs
$w \cdot h \cdot d$	uno	gutural	13 / 6	weeħ_di frente a wáħidu
$h \cdot s \cdot r$	piedra	gutural	10 / 5	ħaḍʒar frente a háʃar

Código	Concepto	Fusión	Reparto	Ejemplo
s·n	soñar	sibilante	6 / 7	šinè frente a sunà
s·t·n	demonio	sibilante	6 / 7	shetan frente a setan
s·t	pimienta	sibilante	12 / 4	šìta frente a dšítà
p·l·s	caballo	sibilante	5 / 13	pəlìšà frente a plís
h·m	agua	gutural	8 / 4	a' am frente a hāmə

El caso de **s·n|diente** merece leerse despacio, porque no es un fallo: la alternancia s š en «diente» es una correspondencia afroasiática real, y un comparatista uniría esas formas igual que nosotros. La diferencia es cómo: él lo haría demostrando la correspondencia; nosotros lo hacemos porque la tabla las escribe igual antes de mirar. El resultado coincide y el fundamento no, y esa distinción es el paper entero.

Y hay un segundo hallazgo de la auditoría que no buscábamos: el anexo tiene más préstamos de los que §4.1 señala. A la mandioca hay que añadirle, entre estos nueve, s·t·n|demonio —que es *šayṭān*, de difusión islámica—, s·t|pimienta y p·l·s|caballo, palabras de cultivo y de cultura material que viajan con la cosa. El catálogo las agrupa exactamente igual que a **m·t|morir**, y no tiene con qué separarlas. Léase el anexo sabiendo que al menos cuatro de sus cien códigos son cosas que llegaron, no cosas que se heredaron.

Índice inverso: códigos que reaparecen en varios conceptos

Solo tres códigos del anexo aparecen con más de un concepto, y es una cifra baja para la serie — el turco tenía diez conceptos bajo un solo código. Es coherente con la aridez de la familia: con tres consonantes de media, el esqueleto discrimina más que en una familia de palabras cortas.

Los tres se leen distinto, y conviene separarlos porque la diferencia recorre todo el anexo.

m·n cubre «quién», «qué» y «cuántos». Eso no es colisión del instrumento: es el radical interrogativo semítico, el mismo *man* que se lee en amárico, arameo antiguo, argobba y los dialectos árabes del bloque. Una lengua que usa el mismo material para las tres preguntas está colexificando, y el código lo refleja con razón.

l·m cubre «nombre», «oreja» y «mercado», y aquí sí hay que sospechar. Son tres campos sin relación, y el código es binario sobre un inventario donde l y m están entre los símbolos más frecuentes. Es el caso típico del binario de alcance amplio: agrupa por disponibilidad tanto como por historia.

k·m cubre «cara», «diez» y «oreja», con enriquecimientos bajos —42×, 37×, 30×— que son los más bajos del anexo. Un enriquecimiento bajo en un código de alcance alto es exactamente la señal de que el código está recogiendo lo frecuente, no lo compartido.

Esa es la distinción que hay que sostener al leer todo lo que sigue: un mismo código puede reunir la misma palabra usada para varias cosas, o varias palabras que el instrumento no distingue, y el código por sí solo no dice cuál. El 38 % de colisión que declara §3 es esto, contado.

Código	Conceptos con los que aparece, y su enriquecimiento
l·m	name (22, 90×) · ear (21, 71×) · market (12, 75×)
m·n	who (18, 66×) · what (12, 45×) · how many (10, 52×)
k·m	face (11, 42×) · ten (10, 37×) · ear (10, 30×)

Numerales

Los numerales parten la familia en dos, y lo hacen a la vista.

Los chádicos van en binario y los semíticos en ternario. «Cuatro» aparece dos veces: $f \cdot d$ en veinte lenguas —*fúdu*, *fidí*, *faḏ*, *fôdu*, *fudû*— y $f \cdot w \cdot d$ en trece —*fwad*, *fwədà*, *fwòdu*, *fwad*—, y las dos listas son enteramente chádicas. La diferencia entre los dos códigos es la *w*, es decir la labialización de la primera consonante, escrita unas veces como segmento y otras no. Es el mismo numeral partido por una convención de transcripción, y es la misma clase de artefacto que §2.4 declara para el etiosemitico.

Frente a eso, $w \cdot h \cdot d$ para «uno» y $h \cdot s \cdot r$ para «diez» son trílteras semíticas de manual —*wāḥid*, *ṣašar*— con doce lenguas cada una.

k·m|diez merece leerse al lado de **h·s·r|diez**. El mismo concepto sale con dos códigos, uno binario de enriquecimiento $37\times$ y otro ternario de $164\times$, y no compiten: son las dos mitades de la familia contando de maneras distintas.

f·d|four · binario · 20 lenguas · enriquecimiento $173\times$

Bade fúdu	Banana fidí	Banana fidi	Cuvok faḏ
D̩ ai fúḏū	Hausa fódu	Hwona faḏà	Karekare fěḏu
Lame fúḏī	Mandara úfade	MargiS foḏu	Matakam faḏ
Misme fidi	Miya fidì	Ngizim fuḏu	Tangale faḏəu

f·w·d|four · ternario · 13 lenguas · enriquecimiento $569\times$

Bata fwad	Boka fwədà	Cibak fwòdu	Fali Gili fwad
Fali Mucella fwəḏ	Gaʔanda fwədà	Gudu fwád	Higi Ghye fwade
Kamwe-Futu fwadu	Margi fwoḏu	MargiS fwōḏu	Psikye fwade
West Margi fwodo			

h·s·r|TEN · ternario · 12 lenguas · enriquecimiento $164\times$

Amharic ʔassir	Argobba ʔassir	Ge'ez ʔəṣru	Harari ʔassir
Kistane ʔasir	Mesqan ʔasir	Silt'e ʔassir	Standard Arabic ʕafara
Tarifit ʔəšr'a	Tigre ʔasir	Wolane ʔassir	Zway ʔassir

t·f|five · binario · 12 lenguas · enriquecimiento $207\times$

Bacama tuf	Bata túf	Cibak tufù	Fali Mucella tuf
Fali Bwagira tuf	Gudu tùf	Hwona tufù	Kilba tûfù
MargiS tufu	Ngwaxi tufù	Njanye tuf	West Margi tufù

w·h·d|one · ternario · 12 lenguas · enriquecimiento $297\times$

(Arabic) Ádirar w̤ hid	(Arabic) B̤ rān wāhid	BuxariArabic waaḥid
DamasceneArabic wa:h_ed	Egyptian Spoken A wa:hid	IraqiArabic wa:hi_d
Maltese wi:h_ed	MoroccoArabic wa_hed	Shoa Arabic wa:hid
Standard Arabic wa:hid	United Arab Emira waaḥ_id	YemeniticArabic wa:hi_d

m·k·r|three · ternario · 12 lenguas · enriquecimiento $258\times$

Bura makùř	Cibak makùř	Kilba mákrù	Margi makùř
MargiS maakər	Matakam màkâar	Mofu-Gudur maakar	Moloko makar
Muyang makar	Ngwaxi makùř	Ouldeme màkàr	West Margi makùr

k·m|TEN · binario · 10 lenguas · enriquecimiento 37×

Boka kum	Bura kuma	Cibak kuma	Gaʔanda kum
Hwona kùṃ	Kilba kùṃà	Margi kumu	MargiS kumou
Ngwaxi kuma	West Margi kuma		

Partes del cuerpo

Es el bloque más denso del anexo y el que mejor enseña la tesis del paper.

Siete de sus once códigos son ternarios, y todos son trilitéras reconocibles: *l·s·n lisān* «lengua», *b·t·n baṭn* «vientre», *r·k·b rukba* «rodilla», *s·h·r šaʔr* «pelo». El procedimiento no sabía que existieran; salieron de exigir que un esqueleto se repitiera para el mismo concepto en tres lenguas.

Y **l·s·n|lengua** tiene una lengua chádica dentro: el karekare *lusùŋ*, entre árabes y arameos. El catálogo lo agrupa porque comparte forma y concepto, y no puede decir si es herencia, préstamo o coincidencia. Es exactamente el tipo de caso que §1.1 declara fuera de alcance, y se deja a la vista en vez de esconderlo.

b·t·n|vientre tiene el enriquecimiento más alto del bloque, 666×, y es un ternario de once lenguas: la combinación b, t, n en ese orden y para ese concepto es rarísima por azar. Compárese con **l·m|oreja**, binario de veintiuna lenguas y solo 71×. El alcance y el enriquecimiento no miden lo mismo, y el anexo los da separados a propósito.

d·m|blood · binario · 28 lenguas · enriquecimiento 100×

(Arabic) Ádirar dāmu	(Arabic) Bā rān dam	Ancient Aramaic dmā	Bole(wa) domì
CypriotArabic d-m-_y	DamasceneArabic dam	Ge'ez dām	IraqiArabic Dam
Kistane dām	Maltese 'dēmm	Modern Hebrew dām	Ngamo dōm
Silt'e dām	Standard Arabic da_m	Tigrigna dām	YemeniticArabic Dam

s·n|tooth · binario · 22 lenguas · enriquecimiento 47×

(Arabic) Ádirar sin	(Arabic) Bā rān sin	Ancient Aramaic šennā	Banana siino
DamasceneArabic sinn	Dira šinè	Egyptian Spoken A sinn	Geji sìŋ
Maltese 'sinne	Mazera sine	Mbara siŋ	Mpade shán
Polci shen	Seya šin	Standard Arabic sin:	United Arab Emira sinn

l·m|ear · binario · 21 lenguas · enriquecimiento 71×

Afade slīm	Bura līm	Buwal zlam	Cibak līmà
Dugwor zlam	Fali Kiria līmu	Gava līmà	Gudu līm
Higi Ghye līme	Kamwe-Futu līmu	Malgbe slīm	Margi līmì
Mbara slumo	Merey zlam	Moloko zlam	Ngwaxi līm

h·n|eye · binario · 18 lenguas · enriquecimiento 83×

Amharic ʔayn	Argobba ʔeen	BuxariArabic ʔayn	DamasceneArabic ʔeen
Egyptian Spoken A ʔe:n	Ge'ez ʔayn	Harari ʔin	IraqiArabic ʔeen
Kistane ʔin	Mesqan ʔen	Silt'e ʔiin	Standard Arabic ʔai_n
Tigre ʔin	Tigrigna ʔayni	United Arab Emira ʔeen	Wolane ʔin

r·s|head · binario · 14 lenguas · enriquecimiento 104×

(Arabic) Ádirar rásu	(Arabic) Bā rān rās	Amharic ras
Ancient Aramaic rēšā	BuxariArabic ṛaaṣ	DamasceneArabic ra:s
Egyptian Spoken A r ^ʿ aa_s	IraqiArabic ra:s	Ki-nubi ras
Maltese ras	Modern Hebrew ṛoṣ	MoroccoArabic r ^ʿ as_ʿ
Shoa Arabic ra:s	United Arab Emira raas	

l·s·n|tongue · ternario · 14 lenguas · enriquecimiento 349×

(Arabic) Bā rān līsān	CypriotArabic l-s-n	DamasceneArabic lsa:n
Egyptian Spoken A lisaan	IraqiArabic lisa:n	Karekare lusùṅ
Ki-nubi *lisaan	Maltese lsi:n	Modern Hebrew lafon
MoroccoArabic lsa_n	Shoa Arabic lisa:n	Standard Arabic lisa_:n
United Arab Emira lisaa_n	YemeniticArabic lisaa_n	

b·t·n|belly · ternario · 11 lenguas · enriquecimiento 666×

(Arabic) Ádirar bátnu	BuxariArabic batin	DamasceneArabic bat ^ʿ n
Egyptian Spoken A bat ^ʿ _n*	IraqiArabic bat ^ʿ i_n	Ki-nubi batna
Modern Hebrew beten	Shoa Arabic bótana	Standard Arabic bat ^ʿ _n
United Arab Emira bat ^ʿ n	YemeniticArabic bat ^ʿ n	

r·k·b|knee · ternario · 11 lenguas · enriquecimiento 308×

(Arabic) Ádirar ríkib	(Arabic) Bā rān ríkibē	DamasceneArabic rəkb_e
Egyptian Spoken A ruk_ba*	IraqiArabic rukb_a	Ki-nubi *rukuba
Maltese rko_bba	MoroccoArabic rok_ba	Shoa Arabic rukuba
Standard Arabic rukba	United Arab Emira rikba	

m·j|m|mouth · binario · 11 lenguas · enriquecimiento 71×

Boka miya	Dugwor mey	Kamwe-Futu myæ	Margi mỳyà
MargiS miya	Mbara may	Mofu-Gudur mey	Moloko mey
Ngizim mỳya	Ouldeme mey	West Margi myà	

l·n|tooth · binario · 11 lenguas · enriquecimiento 72×

Cuvok zlen	Gaʔanda lēna	GizigaMarva slin	Gudu lln
Higi Ghye lline	Hwona lànà	Kamwe-Futu lline	Lagwan slini
Mbuko slan	Njanye line	Psikye slənə	

s·h·r|HAIR · ternario · 10 lenguas · enriquecimiento 165×

DamasceneArabic ʃaʃra	Egyptian Spoken A ʃaʃr	IraqiArabic ʃaʃra	Maltese ʃaʃa_r
Modern Hebrew ʃaʔaʔa	MoroccoArabic ʃʃer	Shoa Arabic ʃiʔir	Standard Arabic ʃaʃr
United Arab Emira ʃaʃar	YemeniticArabic ʃaʃra		

k·m|ear · binario · 10 lenguas · enriquecimiento 30×

Bole(wa) kúmo	Burma kìm	Cip kòm	Dera kumo
Dira kumè	Dwot kum	Gera kumi	Ngamo kùmà
Polci kum	Sura kom		

Parentesco y persona

b·b|padre con veintiséis lenguas y **m·m|madre** con trece son el par que aparece en todas las familias de la serie, y aquí también: reduplicación labial para el padre, nasal para la madre. §6.1 del paper austronesio ya avisaba de que estos códigos son los menos informativos del catálogo, porque su forma se explica por la crianza antes que por la historia.

h·n·t|tú tiene el enriquecimiento más alto de todo el anexo, 1.063×, y hay que leerlo con cuidado porque parte de esa cifra la fabrica nuestra tabla. Las formas son *ʔantʔ*, *ʔinta*, *ʔante*, *hint*: la oclusiva glotal inicial se escribe h por la fusión de guturales que §7 declara. Un código que empieza por h en una familia donde h reúne cuatro consonantes distintas es más frecuente de lo que su enriquecimiento sugiere. La cifra es correcta dentro de nuestra normalización, y la normalización es una decisión.

n·g·s|rey es el código más limpiamente ramificado del anexo: diez lenguas, todas etiosemiticas —amárico *nigus*, ge'ez *nigús*, harari *nʔgaaši*, tigrina, tigre, argobba, silt'e, kistane—. No cruza a chádico ni a bereber, y no debería: es una palabra de cultura política de un área concreta. Un catálogo sin genealogía que produce un código de rama es un catálogo que está midiendo algo.

b·b|FATHER · binario · 26 lenguas · enriquecimiento 244×

Bacama bàba	Banana bàba	Dera bàba	Dira bàba
Dugwor baba	Gaʔanda bàba	Gavar baba	Hwona bùba
Kamwe–Futu bàba	Karekare bába	Matakam bàbu	Mbudum baba
Misme bouba	Muyang baba	Ngizim bába	Ouldeme bàbá

s·m|name · binario · 25 lenguas · enriquecimiento 81×

Ancient Aramaic šmā	Angas sīm	Banana šēma	Banana sema
DamasceneArabic ism	Egyptian Spoken A ism	IraqiArabic əsəm	Ki–nubi *asma
Mazera sime	Misme sem	Modern Hebrew fēm	Mser sīm
Polci sīm	Standard Arabic ism	Sura sum	Tarifit isəm

l·m|name · binario · 22 lenguas · enriquecimiento 90×

Afade slīm	Boka līmà	Bura līm	Cibak līmà
Dugwor zlam	Fali Kiria līmu	Gaʔanda līma	Gavar zləm
Hwona līmè	Kilba lém	Lagwan slīmi	Margi līm
MargiS līmò	Mbara slim	Ngwaxi līm	Piðimdi līm

m·n|WHO · binario · 18 lenguas · enriquecimiento 66×

Amharic man	Ancient Aramaic man	Argobba mani
BuxariArabic miin	CypriotArabic m–n	DamasceneArabic ma:n
Egyptian Spoken A mi:n	Ge'ez mənnu	IraqiArabic minu
Ki–nubi munu	Maltese min	Silt'e maani
Standard Arabic ma_n	Tigre mən	Tigrigna mən
United Arab Emira min		

m·m|MOTHER · binario · 13 lenguas · enriquecimiento 106×

Bura mumà	Dugwor mama	Fali Kiria màma	Fali Bwagira màma
Gavar mōma	Kamwe–Futu mama	Matakam mòm	Mbudum mama
Moloko mama	Ngwaxi mamà	Psikye màma	Shoa Arabic ómma
Vame māmá			

m·n|WHAT · binario · 12 lenguas · enriquecimiento 45×

Amharic mìn	Argobba mìn	Bacama munò	Bata mìnò
Buduma mēni	Harari min	Hwona mani	Kistane mìn
Silt'e mìn	Tarifit min	Wolane mìn	Zway mìn

n·g·s|KING · ternario · 10 lenguas · enriquecimiento 240×

Amharic nīgus	Argobba nīgus	Ge'ez nīgus	Harari nəgaaši
Kistane nīgus	Silt'e nīgus	Tigre nəgus	Tigrigna nīgus
Wolane nīgus	Zway nīgus		

n·d|PERSON · binario · 10 lenguas · enriquecimiento 78×

Cibak ñdà	Cuvok nda	Fali Gili ñdu	Fali Mucella ñda
Kilba ñdú	Lamang ùndù	MargiS ndū	Ngwaxi ndà
Shoa Arabic na:di	Sukur nda		

h·n·t|THOU · ternario · 10 lenguas · enriquecimiento 1063×

Amharic ?antə	BuxariArabic hint	CypriotArabic ?-n-t
DamasceneArabic ?ante	Egyptian Spoken A ?inta	Ge'ez ?antə
IraqiArabic ?inta	Standard Arabic ?an_ta	Tigre ?inta
YemeniticArabic ?ant		

Naturaleza y cielo

El bloque donde el mismo concepto sale más veces con códigos distintos, y eso es informativo.

«Sol» aparece tres veces: **p·s** (18 lenguas), **s·m·s** (15) y **f·t** (13). **s·m·s** es *šams*, semítica sin discusión. Las otras dos son chádicas. Tres códigos para un concepto en una familia de cuatro ramas no es ruido: es el reparto.

«Agua» sale dos veces, **j·m** en veinte lenguas y **h·m** en once, y «noche» dos, **l·l** —*layl*— y **v·d**.

n·f·s|viento (358×) es *nafas* «aliento», y su presencia en el bloque de naturaleza en vez de en el del cuerpo es un recordatorio de que la agrupación temática de este anexo es un orden de lectura y no una clasificación: el concepto lo fija Concepticon, no nosotros.

t·r|moon · binario · 22 lenguas · enriquecimiento 88×

Angas tar	Ankwe tàr	Bole(wa) tére	Daba tìrà
Dera tere	Fali Gili tĩr	Gera tèra	Higi Ghye tìra
Karekare tárei	Lamang tré	Mandara tìre	Miya tĩr
Ngamo tèrè	Ngizim téra	Psikye trə	Sura tar

j·m|water · binario · 20 lenguas · enriquecimiento 143×

Buwal yam	Cibak yìmi	Cuvok yam	Daba yim
Fali Gili yam	Fali Jilbu yımû	Fali Kiria yàmu	Ga?anda yèma
GizigaMarva yam	Kamwe-Futu yèmi	Mbudum jim	Merey yàm
Moloko yàm	Ngwaxi yìmi	Ouldeme yàm	Psikye yemú

p·s|sun · binario · 18 lenguas · enriquecimiento 127×

Angas pus	Ankwe pis	Buwal pes	Cip pùs
Cuvok pas	Dugwor pats	Gavar pish	Gera piši
Kilba píčì	Matakam pats	Mbazla pàs	Mbudum pis
Moloko pats	Ouldeme pats	Sukur pis	Sura puus

s·m·s|sun · ternario · 15 lenguas · enriquecimiento 268×

(Arabic) B□ rān sēms	Ancient Aramaic šemšā	BuxariArabic fams
DamasceneArabic fams	Egyptian Spoken A jam_s	IraqiArabic famis
Ki-nubi *semsi	Maltese jem_ſ	Modern Hebrew fεmεf
MoroccoArabic fe_mf	Shoa Arabic šāmiš	Standard Arabic fa_ms
Tigre šimiš	United Arab Emira fams	YemeniticArabic fams

f·t|sun · binario · 13 lenguas · enriquecimiento 96×

Banana fata	Bata fótó	Bole(wa) fùtî	Fali Jilbu fitî
Fali Mucella fitu	Karekare fètî	Lamang fítí	Lame fəta
Mandara fúti	Misme fata	Ngamo fɔtî	Njanye fète
Ouldeme fāt			

n·r|fire · binario · 12 lenguas · enriquecimiento 177×

(Arabic) Ádirar nāru	(Arabic) B□ rān nār	Ancient Aramaic nūrā
BuxariArabic naar	DamasceneArabic na:r ^ʿ	Egyptian Spoken A na:r
IraqiArabic na:r	Ki-nubi nari	Maltese nar
Shoa Arabic na:r	Standard Arabic na:_r	YemeniticArabic naar

l·l|night · binario · 11 lenguas · enriquecimiento 75×

(Arabic) Ádirar láilu	(Arabic) B□ rān leil	BuxariArabic layl
CypriotArabic l-y-l	DamasceneArabic leelee	Egyptian Spoken A le:l
IraqiArabic leela	MoroccoArabic lila	Shoa Arabic le:le
Standard Arabic laila	United Arab Emira leel	

v·d|night · binario · 11 lenguas · enriquecimiento 145×

Bata vùdó	Cuvok vađ	Daba vùđu	Dghwede vuđi
Fali Jilbu viđi	Fali Mucella vùđu	Kamwe-Futu viđi	Matakam vađ
Njanye vɛde	Psikye víđi	Sukur vəd	

h·m|water · binario · 11 lenguas · enriquecimiento 58×

Gera hāmə	Karekare ʔəmù	Margi ʔimi	MargiS ʔɪmì
Mazera a'ɪm	Mbara 'am	Mbukò a'am	Ngamo hām
Sura ʔām	Tangale hām	Vame āhwám	

b·r·d|cold · ternario · 10 lenguas · enriquecimiento 237×

(Arabic) B□ rān bárid	BuxariArabic bard	DamasceneArabic barəd
Egyptian Spoken A bar_d	IraqiArabic beerə_d	Maltese bi:r_ed
MoroccoArabic bar_ed	Standard Arabic ba:_rid	United Arab Emira baari_d
YemeniticArabic baari_d		

h·s·r|stone · ternario · 10 lenguas · enriquecimiento 119×

(Arabic) Ādirar hāṣar	(Arabic) Bī rān hāṣar	CypriotArabic h-ṣ-r
DamasceneArabic ḥaṣa_ra	IraqiArabic ḥaṣa_r	MoroccoArabic ḥeṣ_ra
Shoa Arabic hiṣer	Standard Arabic ḥaḍar	United Arab Emira ḥaṣa_r
YemeniticArabic ḥaṣa_r		

n·f·s|wind · ternario · 10 lenguas · enriquecimiento 358×

Amharic nīfas	Argobba nīfas	Ge'ez nāfas	Kistane nāfas
Mesqan nīfas	Silt'e nīfas	Tigre nīfas	Tigrigna nīfas
Wolane nīfas	Zway nīfas		

Animales

k·l·b|perro (*kalb*, 16 lenguas, 258×) y **k·r|perro** (15 lenguas, 36×) son el mejor par del anexo para entender el enriquecimiento. Mismo concepto, alcance casi idéntico, y un factor siete de diferencia: el ternario es raro por azar y el binario no. Leer el alcance sin el enriquecimiento habría dado los dos por igual de buenos.

p·l·s|caballo (344×) cruza ramas, y es de los códigos que §6.2 marca como candidatos: el caballo llega al continente por una ruta conocida y su nombre viaja con él. El catálogo no puede decir que sea préstamo —no distingue herencia de contacto— pero sí puede señalar dónde mirar.

k·l·f|fish · ternario · 20 lenguas · enriquecimiento 474×

Bura kīlfà	Cuvok kəlef	Daba kīlif	Dghwede klfe
Gava kilifà	GizigaMarva kilef	Gələvda kīlfà	Kilba kálfi
MargiS kalfi	Matakam klef	Mazera kīlfa	Mbudum kəl:if
Merey kəlef	Moloko kəlef	Muyang kilif	Ngwaxi kulfu

p·l·s|HORSE · ternario · 16 lenguas · enriquecimiento 344×

Daba piliš	Dghwede pīliši	Dugwor pəles	Gava pəlīsà
GizigaMarva piles	Gələvda pīlīšà	Lamang plís	Matakam pliš
Mbara pilis	Mbuko pəles	Merey pəles	Moloko pəles
Muyang pilis	Nakatsa pələse	Ouldeme pəlis	Vame pəlēsh

k·l·b|dog · ternario · 16 lenguas · enriquecimiento 258×

(Arabic) Ādirar kálbu	(Arabic) Bī rān kēlb	Ancient Aramaic kalbā	BuxariArabic kalb
CypriotArabic k-l-b	DamasceneArabic kalb	Egyptian Spoken A klb	Ge'ez kəlb
IraqiArabic kalb	Maltese kelb	MoroccoArabic kel_b	Shoa Arabic kalb
Standard Arabic kal_b	Tigre kəlīb	Tigrigna kəlbi	YemeniticArabic kalb

k·r|dog · binario · 15 lenguas · enriquecimiento 36×

Dugwor kəra	Fali Gili kīri	Fali Kiria kuři	GizigaMarva kre
Hausa kare	Higi Ghye kuře	Kamwe-Futu kīře	Lamang kərə
Mandara kīrè	Merey kəra	Moloko kəra	Muyang kəra
Ouldeme kərə	Psikye kuře	Sukur kəra	

t·m|SHEEP · binario · 12 lenguas · enriquecimiento 107×

Ankwe tīm	Bura tīmà	Cibak tīmà	Cip tīm
Dira tīmè	Dwot tum	Kamwe-Futu tīmè	Ngwaxi tīmà
Polci tīm	Seya tīm	Sura tum	West Margi tīmà

t·r|bird · binario · 11 lenguas · enriquecimiento 44×

(Arabic) Ádirar táer	(Arabic) Bī rān táer	BuxariArabic tayra
DamasceneArabic t ^ʿ eer	Egyptian Spoken A t ^ʿ ee_r*	IraqiArabic t ^ʿ eer
Ki-nubi ter	MoroccoArabic t ^ʿ ir	Shoa Arabic t ^ʿ e:ra
Standard Arabic t ^ʿ air	United Arab Emira t ^ʿ eer	

Plantas y alimento

l·h·m|carne (*lahm*, ternario, 289×) frente a **l·w|carne** (binario, 20 lenguas, 149×): otra vez el mismo concepto partido entre el ternario semítico y el binario chádico, y otra vez el ternario paga más enriquecimiento por menos alcance.

m·m|miel con quince lenguas es un caso de reduplicación como los del bloque de parentesco, y conviene la misma cautela.

l·w|meat · binario · 20 lenguas · enriquecimiento 149×

Ankwe lowa	Boka luwə	Cuvok slaw	Dghwede luwè
Fali Mucella lùwa	Gaʔanda lîwa	Gava luwà	Gələvda luwà
Lamang slùw	Matakam luwe	Merey zlaw	Miya liwi
Nakatsa luwa	Ngizim lùwəi	Njanye liwo	Sura luwa

m·m|HONEY · binario · 15 lenguas · enriquecimiento 129×

Cibak mīmà	Dugwor amam	Gava mamà	GizigaMarva amam
Gələvda mēmà	Malgbe mam	Margi mùmù	MargiS mumō
Mbuko umam	Moloko omom	Mpade mām	Mser mam
Nakatsa māmà	Sukur mam	West Margi mīmè	

s·r·b|drink · ternario · 11 lenguas · enriquecimiento 198×

BuxariArabic farab	CypriotArabic f-r-b	DamasceneArabic ifrib
Egyptian Spoken A frb	IraqiArabic firab	Maltese foro_b
MoroccoArabic fre_b	Shoa Arabic farab	Standard Arabic fari_ba
United Arab Emira farab	YemeniticArabic firib	

h·k·l|eat · ternario · 10 lenguas · enriquecimiento 365×

Ancient Aramaic ʔkal	BuxariArabic ʔukul	CypriotArabic ʔ-k-l
DamasceneArabic ʔakal	Egyptian Spoken A ʔkl	IraqiArabic ʔakal
Ki-nubi ʾakulu	Standard Arabic ʔak_ala	United Arab Emira ʔakal
YemeniticArabic ʔakal		

l·h·m|meat · ternario · 10 lenguas · enriquecimiento 289×

BuxariArabic ləħəm	CypriotArabic l-h-_m	DamasceneArabic laħm
Egyptian Spoken A laħ_ma	IraqiArabic laħa_m	Maltese laħa_m
MoroccoArabic lħa_m	Standard Arabic laħ_m	United Arab Emira laħa_m
YemeniticArabic laħm		

Verbos de acción y estado

m·t|morir es el código de mayor alcance de todo el catálogo: treinta y ocho lenguas, veintiocho chádicas y veintidós semíticas. Es el único del anexo que cruza las dos ramas grandes con esa fuerza. La comparación afroasiática lleva un siglo discutiendo ese *mwt*; el catálogo lo encuentra sin haber consultado nada.

El resto del bloque son trílteras: s·m·h *samiša* «oír», h·k·l *ʔakala* «comer», h·r·f *ʕarafa* «saber», s·r·b *šariba* «beber». Los cuatro tienen entre diez y doce lenguas y enriquecimientos de 198× a 365×.

m·s|andar (87×) desentona, y por lo mismo que l·m en el índice inverso: binario, alcance de diez lenguas, enriquecimiento bajo. Está en el anexo por alcance, no por calidad.

m·t|die · binario · 38 lenguas · enriquecimiento 164×

(Arabic) Ádirar mét	Angas mùt	Buduma matə	Egyptian Spoken A maa_t
Fali ʔwagira mti	Higi Ghye mte	Ki-nubi mutu	Lamang mta
Margi mtu	Misme mat	Modern Hebrew met	Mser mato
Ngizim mùtu	Ouldeme mət	Standard Arabic ma:_ta	United Arab Emira maat

s·m·h|hear · ternario · 12 lenguas · enriquecimiento 243×

BuxariArabic samaʕ	CypriotArabic s-m-_ʕ	DamasceneArabic ismiʕ
Egyptian Spoken A smʕ	IraqiArabic sima_ʕ	Maltese sam_aʼ
Modern Hebrew ʃamaʔ	MoroccoArabic sm_eʕ	Shoa Arabic simiʼ
Standard Arabic sa_miʕ_a	United Arab Emira sima_ʕ	YemeniticArabic simiʕ_*

m·s|WALK · binario · 10 lenguas · enriquecimiento 87×

CypriotArabic mifi	DamasceneArabic mafa	Egyptian Spoken A mifi	IraqiArabic mifa
Ki-nubi Mashi	Maltese meʃ_a	MoroccoArabic mʃ_a	Shoa Arabic mafa
Standard Arabic ma_ʃa:	United Arab Emira mifa		

h·r·f|know · ternario · 10 lenguas · enriquecimiento 341×

BuxariArabic ʔaraf	CypriotArabic ʕ-r-f	DamasceneArabic ʕarʕaf
Egyptian Spoken A ʕirif	IraqiArabic ʕiraf	Ki-nubi ʕarufu
MoroccoArabic ʕrʕ_e_f	Shoa Arabic ʕirif	Standard Arabic ʕar_afa
United Arab Emira ʕaraf		

Objetos y cultura

Es el bloque más pequeño, y era previsible: los objetos de cultura material son donde más pesa el préstamo y donde menos coinciden las listas de conceptos entre fuentes.

l·b·r|aguja alcanza 730× con diez lenguas, el segundo enriquecimiento del anexo. Un ternario para un objeto concreto y con esa concentración es exactamente el perfil de un candidato a isoglosa, y §6.3 lo recogería si su par estuviera en el catálogo.

l·b·r|NEEDLE (FOR SEWING) · ternario · 10 lenguas · enriquecimiento 730×

Afade libira	Buduma líbūra	Dᵛ ai líbira	Karekare alíbura
Lagwan libra	Malgbe libra	Maltese 'lebre	Mpade líbrà
Mser libra	Ngizim alíbera		

Cualidades

j·b·s|seco con 938× es el enriquecimiento más alto después de **h·n·t**, y a diferencia de aquél no tiene ninguna gutural fundida: es *yābis*, con j, b y s bien separados en la tabla de normalización. Es, de todo el anexo, el código cuyo enriquecimiento hay que creerse más.

k·b·r|grande (*kabīr*, 266×) y **t·w·l|largo** (*tawīl*, 259×) completan un bloque enteramente ternario y enteramente semítico. Que las cualidades salgan todas de una rama y los numerales se partan en dos dice algo sobre qué campos comparte esta familia y cuáles no — pero decir qué exactamente exigiría la genealogía que este trabajo no usa.

j·b·s|dry · ternario · 13 lenguas · enriquecimiento 938×

(Arabic) Bā rān yābes	Ancient Aramaic yabbīš	BuxariArabic yabs
DamasceneArabic ji:bis	Ge'ez yibus	IraqiArabic jeebā_s
Kándin yābūši	MoroccoArabic jab_es	Shoa Arabic ja:bis
Standard Arabic ja:b_is	Tigre yibus	United Arab Emira jaabi_s
YemeniticArabic jaabi_s		

k·b·r|big · ternario · 12 lenguas · enriquecimiento 266×

(Arabic) Ādirar kab ru	(Arabic) Bā rān ikbér	BuxariArabic ka'biir
CypriotArabic k-b-_r	DamasceneArabic kbi:r	Egyptian Spoken A kibi:r
Ki-nubi kbir	Maltese kibi_r	MoroccoArabic kbi_r
Shoa Arabic kabi:r	Standard Arabic kab_i:r	YemeniticArabic Kabii_r

t·w·l|long · ternario · 11 lenguas · enriquecimiento 259×

CypriotArabic t ^ʿ -w-_l	DamasceneArabic t ^ʿ awi:_l	Egyptian Spoken A t ^ʿ aw_i:l
IraqiArabic t ^ʿ -w-_l	Ki-nubi towil	Maltese twil
MoroccoArabic t ^ʿ wi_l	Shoa Arabic t ^ʿ awil	Standard Arabic t ^ʿ a_wi:l
United Arab Emira t ^ʿ awii_l	YemeniticArabic t ^ʿ awi:_l	

m·l|full · binario · 10 lenguas · enriquecimiento 129×

Amharic mulu	Ancient Aramaic mlē	Argobba mulu	IraqiArabic əmmal_a
Kistane mula	Mesqan mula	Modern Hebrew malē	Silt'e mulla
Wolane mulli	Zway muli		

Otros

Lo que no cupo en los bloques anteriores, y se deja para que el anexo no dé la impresión de estar más ordenado de lo que está. El orden temático es de lectura; los códigos no salieron ordenados.

m·b·j|CASSAVA · ternario · 27 lenguas · enriquecimiento 581×

Boka mbəy	Cuvok mbay	Dghwede mbàya	Fali Gili mbàyi
Fali Jilbu mbayà	Fali Mucella mbàya	Ga?anda mbèy	Gava mbàya
Gudu mbəy	Higi Ghye mbàya	Hwona mbèy	Mandara mbàya
MargiS mbàyu	Merey mbay	Muyang mbay	Ouldeme mbàya

v·j|RAINY SEASON · binario · 20 lenguas · enriquecimiento 309×

Bura viyà	Buwal vèya	Cibak viyà	Cuvok viya
Fali Kiria viya	Fali Mucella viyà	Gava viyà	Gavar viya
Kamwe-Futu viya	Kilba vùya	Mandara viyà	Margi viya
Matakam viya	Mbudum vija	Mbuko viya	Piðimdi viya

m·t|DEATH · binario · 19 lenguas · enriquecimiento 109×

Angas mut	Ankwe mut	Banana mita	Bole(wa) mòto
Bura mítà	Cibak mítì	Fali Gili m̀ti	Kamwe-Futu mto
Lagwan m̀iti	Lame mat	Margi m̀tù	Ngamo màtò
Ngizim mùtu	Psikye m̀tə	Standard Arabic maut	Tangale mutò

k·w·r|DONKEY · ternario · 18 lenguas · enriquecimiento 341×

Bata kwà:rìò	Boka kwarè	Bura kwəra	Cibak kwara
Fali Kiria kwarā	Fali ʔwagira kwəra	Gaʔanda kwari	Gudu kwára
Hwona kwara	Kamwe-Futu kwara	Kilba kwára	Margi kwara
MargiS kwara	Ngizim kwara	Ngwaxi kwəra	Psikye kwarà

m·r|OIL (ORGANIC SUBSTANCE) · binario · 17 lenguas · enriquecimiento 87×

Bata maare	Bole(wa) mořì	Burma mar	Cip mòr
Dira miri	Fali Mucella mara	Gera mori	Gudu mář
Karekare mərù	Lame mar	Moloko amar	Ngamo mōř
Njanye mare	Piðimdi mēř	Polci mır	Seya mır

m·l|OIL (ORGANIC SUBSTANCE) · binario · 16 lenguas · enriquecimiento 99×

Banana mula	Bura mêl	Buwal mel	Cibak màl
Cuvok mal	Fali Jilbu malí	Gavar mali	Geji mili
GizigaMarva mal	Kilba màl	Margi màl	MargiS mal
Mbara mal	Ngwaxi màl	Ouldeme âmāl	West Margi màl

s·r|TRUTH · binario · 16 lenguas · enriquecimiento 50×

Bade jĩřái	Bura jırì	Cibak jırì	Dera jirè
Gava jirà	Gələvda jirà	Kamwe-Futu jırì	Karekare jirè
Kilba jírí	Mandara jirè	Margi j̀rì	MargiS jiri
Nakatsa jira	Ngamo jirè	Piðimdi jirè	West Margi jırì

l·w|FLESH · binario · 16 lenguas · enriquecimiento 155×

Ankwe lowa	Boka łuwà	Cuvok slaw	Dghwede sluwe
Dugwor slaw	Fali Gili łìwu	Fali Mucella lùwa	Gaʔanda łìwa
Gava łuwà	Gera lèwi	Gələvda łuwà	Hwona łuwa
Miya łiwì	Nakatsa łuwa	Njanye liwe	Sura luwa

p·t|CAT · binario · 15 lenguas · enriquecimiento 104×

Cibak pàtu	Fali Gili patù	Kamwe-Futu pàtu	Mandara pàtu
Margi pàtu	MargiS patu	Matakam patu	Muyang potu
Ngamo pàtu	Ngizim pātu	Ngwaxi pàtu	Piðimdi pàtu
Sukur patu	Vame pātú	West Margi pàtu	

k·l|ALL · binario · 14 lenguas · enriquecimiento 45×

Ancient Aramaic kull	BuxariArabic kullu	CypriotArabic kull	DamasceneArabic kəll
Egyptian Spoken A kull	Harari kullu	IraqiArabic kull	Ki-nubi kulu
Maltese Koll_u	MoroccoArabic koll	Standard Arabic kull	Tigrigna kullu
United Arab Emira kil	YemeniticArabic kull		

s·n|DREAM (SOMETHING) · binario · 13 lenguas · enriquecimiento 41×

Angas sun	Bura sunì	Cibak šinì	Daba šini
Fali Mucella sinî	Gera sunə	Karekare sunà	Mandara šinè
MargiS šuni	Ngamo sùnâ	Ngwaxi šinì	Piðimdi jina
West Margi sìnì			

s·t|PEPPER · binario · 13 lenguas · enriquecimiento 87×

Angas šità	Ankwe šità	Bole(wa) dšíta	Cip šita
Daba šitè	D̩ ai dšíta	Fali Kiria šatà	Karekare dšíta
Kilba čitá	Ngizim šíta	Shoa Arabic šéta	Sura žita
Tangale sità			

g·m|RAM · binario · 13 lenguas · enriquecimiento 93×

Ankwe gàm	Bade g̃ ma	Bole(wa) gèmi	Bura gâm
Dera gəm	D̩ ai gam	Kamwe-Futu game	Karekare gəmu
Kilba gâm	MargiS gāmo	Ngamo gəm	Ngwaxi gəm
West Margi agəmə			

k·w·l|PENIS · ternario · 13 lenguas · enriquecimiento 228×

Bata kwáàló	Bura kwâl	Cibak kwalə	Fali Gili kwala
Fali Kiria kwəla	Higi Ghye kwala	Kamwe-Futu kwala	Kilba kwâl
Margi kwal	MargiS kwal	Ngwaxi kwəl	Psikye kwala
West Margi kwal			

s·t·n|DEMON · ternario · 12 lenguas · enriquecimiento 773×

(Arabic) Ádirar šéitān	(Arabic) B̩ rān séitān	Afade setan	Karekare šítan
Malgbe shetan	Maltese ʃɪ'te:n	Mandara sātən	Merey setene
Mpade sítán	Ngizim šítan	Polci sheetan	Shoa Arabic séitān

l·m|MARKET · binario · 12 lenguas · enriquecimiento 75×

Bata lú:mù	Boka lumu	Daba lumû	Dghwede lumà
Fali Jilbu lumà	Fali Kiria lumà	Fali Mucella lumà	Ga?anda lîmo
Gudu lú:mú	Lame lumo	Njanye lumu	Tangale lomò

z·v·n|GUINEA FOWL · ternario · 12 lenguas · enriquecimiento 625×

Buwal zavan	Daba zàvun	Fali Gili zùvuni	Fali Jilbu zàv̀nî
Fali Kiria z̀v̀n	Fali Mucella zàvunu	Gavar zəvən	Gudu zúvan
Kamwe-Futu zivəno	Lagwan zavan	Mbudum zavaŋ	Psikye z̀v̀v̀nə

k·d·k|SWEET POTATO · ternario · 12 lenguas · enriquecimiento 514×

Ankwe kudùkù	Buwal kwadakwa	Dira kudùku	Fali Gili kudàku
Gavar kwadakwa	GizigaMarva kudaaku	Kilba kúdàkú	MargiS kùdakù
Mbuko kwadakwa	Merey kodoko	Ngwaxi kudàku	Njanye kudàku

r·z|RICE · binario · 12 lenguas · enriquecimiento 1034×

Amharic ruz	Ancient Aramaic rōzā	Argobba ruz	Ge'ez ruz
Harari ruuz	Kistane ruz	Mesqan ruz	Silt'e ruz
Tigre ruz	Tigrigna ruz	Wolane ruz	Zway ruz

k·m|FACE · binario · 11 lenguas · enriquecimiento 42×

Bura kìmà	Cibak kumà	Fali Gili kuma	Fali Kiria kuma
Fali Mucella kuma	Higi Ghye kima	Kamwe–Futu kima	Ngwaxi kìmà
Psikye kuma	Sukur kəma	West Margi kìmà	

b·c|MAT · binario · 11 lenguas · enriquecimiento 450×

Buduma boci	GizigaMarva buce	Kamwe–Futu bùci	Karekare buci
Mandara bùce	Margi buci	MargiS buci	Mbuko buce
Muyang bæci	Ngizim buci	West Margi buci	

ḡ·l·m|YAM · ternario · 11 lenguas · enriquecimiento 407×

Bura ḡulùm	Cibak ḡilma	Dugwor ḡəlam	Margi àḡilim
MargiS ḡulùm	Mbuko ḡəlam	Merey ḡəlem	Muyang aḡəlum
Ngwaxi ḡilma	Ouldeme āḡəlām	West Margi ḡilma	

g·m·l|CAMEL · ternario · 11 lenguas · enriquecimiento 998×

Amharic ḡiməl	Ancient Aramaic gamlā	Argobba gamela	Ge'ez ḡəməl
Harari gaaməla	Kistane ḡiməl	Mesqan gamela	Silt'e gaameela
Tigre ḡeməl	Tigrigna ḡiməl	Wolane gaameela	

l·g·m|CAMEL · ternario · 11 lenguas · enriquecimiento 862×

Buduma loguəme	Dghwede zlugwama	Dugwor zlegweme	Malgbe logome
MargiS dlugwam	Mbuko zlugweme	Merey zləgweme	Moloko ezlegweme
Muyang ezligwemi	Ouldeme ázləgwèmə	Vame ázləgwèmə	

k·r|DONKEY · binario · 11 lenguas · enriquecimiento 35×

Bade koro	Bole(wa) koro	Kamwe–Futu kwara	Karekare koro
Lame koro	MargiS kwara	Mbara kuro	Mpade kóró
Ngamo korô	Piðimdi koro	Psikye kwará	

b·t|CHEAP · binario · 11 lenguas · enriquecimiento 157×

Bata bú:tù	Bura butù	Cibak butù	Fali Kiria butù
Fali Mucella butà	Kamwe–Futu butì	Mandara butù	Margi butù
MargiS butu	Ngwaxi bùtu	West Margi butu	

x·l|STEAL · binario · 10 lenguas · enriquecimiento 129×

Bura xula	Cibak xulæ	Daba àxɪl	Fali Gili ɣəli
Fali Kiria ɣùlù	Kamwe-Futu ɣùlì	MargiS xulà	Ngwaxi xɪlà
Psikye ɣùlì	West Margi xɪlɛ		

b·n|COFFEE · binario · 10 lenguas · enriquecimiento 434×

Amharic bunna	Argobba bunn	Ge'ez bunna	Harari bun
Kistane buna	Silt'e bunno	Tigre bun	Tigrigna bun
Wolane bunno	Zway buna		

k·k|GRANDPARENTS · binario · 10 lenguas · enriquecimiento 180×

Boka kàka	Bole(wa) kàka	Bura kɪka	Dira kàka
Gaʔanda kàka	Hausa kàkà	Ngwaxi kàka	Piðimdi kàka
Polci kaka	West Margi kàka		

m·n|HOW MANY · binario · 10 lenguas · enriquecimiento 52×

Bole(wa) munò	Burma mənà	Dira mùnè	Dwot mənî
Fali Mucella mənà	Gera munu	Karekare manà	Miya mɪnà
Ngamo mònô	Polci mèni		

d·f|MUSH · binario · 10 lenguas · enriquecimiento 227×

Fali Gili dàfa	Fali Kiria dèfa	Fali Mucella dèfa	Gudu dà:f
Higi Ghye dàfa	Kamwe-Futu dàfa	Mandara dèfa	Margi dufù
Psikye dàfa	West Margi dífɛ		