

Códigos consonánticos en 32 lenguas turcas

Un catálogo construido sin genealogía, y lo que la familia enseña sobre el instrumento

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

Resumen

Presentamos un catálogo de 1.517 códigos consonánticos que agrupan 24.517 formas léxicas de 32 lenguas turcas. Un código es un esqueleto consonántico compartido por palabras del mismo concepto en lenguas distintas. No se usa reconstrucción, ni cognación experta, ni ninguna hipótesis sobre el proto-turco.

El procedimiento acierta 216 veces más que el azar (25,5 % de precisión frente al 0,118 % esperable), frente a las 35 del austronesio recalculado con el mismo código. Pero el trabajo de este paper es descomponer esa cifra en vez de celebrarla: dos tercios de la diferencia viven en el denominador, que depende de la longitud de palabra de la familia y no de su historia. La precisión observada solo difiere 1,9×, y el 216× es ahora exacto por el lado del azar. La conclusión que se sostiene es que las razones sobre el azar de dos familias no son comparables entre sí, y que toda comparación se publica descompuesta.

De ahí salen cuatro resultados sobre el instrumento. El despiece morfológico rinde cuando la regla se declara y empeora cuando se descubre por parecido: el descubridor automático gana 6,3 puntos de agrupamiento y 13,9 de nulo, y encuentra 365 prefijos en una familia sin prefijos. La matriz de sustituciones, agregada sobre la familia, pierde la correspondencia que define su división primaria y solo la recupera al restringirse al par de ramas: de invisible a 6,1 veces sobre el azar. El catálogo separa ramas sin saber que existen — 297 pares de códigos hermanos, y los mayores agrupan el chuvasio con el siberiano y el oguz con el kipchak occidental. Y el enriquecimiento máximo vuelve a ser un préstamo, aquí ruso, por un mecanismo que se puede nombrar: lo produce el concepto raro, no el código raro.

Se declara lo que el instrumento cuesta: el 59 % de las formas turcas comparten esqueleto, dentro de su propia lengua, con una forma de otro concepto. El código $k \cdot l$ reúne diez conceptos, entre ellos *kül* «ceniza», *kel-* «venir», *köl* «lago» y *kül-* «reír».

1. Qué se busca, y por qué

Esta serie pregunta si se puede hacer algo útil renunciando a la reconstrucción, y el paper del método describe el instrumento que resulta: un catálogo de códigos consonánticos, más modesto y más barato que el método comparativo, que hace parte de su trabajo sin hacer sus afirmaciones.

Este paper hace otra pregunta: qué significan las cifras que aquel produjo. Para eso hace falta aplicar el instrumento a una familia cuya historia sí se conoce, y el turco sirve por tres razones: profundidad temporal conocida y baja —el proto-turco se sitúa hacia el cambio de era—, correspondencias diagnósticas bien establecidas contra las que contrastar, y morfología fuertemente aglutinante, que es donde el límite declarado del trabajo anterior tenía que doler.

1.1 El método, y dónde está descrito

El procedimiento —qué es un código, en qué orden se construye, qué nulos usa y qué se puede afirmar al batirlos— está descrito una vez para toda la serie en «El catálogo de códigos consonánticos», y no se repite aquí. Lo esencial, para poder leer lo que sigue:

El catálogo se construye sin genealogía. Los códigos salen del esqueleto consonántico de una forma, del concepto que expresa y de un umbral de tres lenguas distintas. Ninguna clasificación en ramas, ninguna protoforma, ninguna lista de cognados. El procedimiento no sabe que esta familia tiene ramas, y la genealogía entra en §5 solo para leer un catálogo ya cerrado.

No se reconstruye, no se afirma parentesco, no se distingue herencia de préstamo, y no se atribuye significado a un código: se escribe $b \cdot s | five$ y nunca $b \cdot s = five$.

Y para esta familia en particular: nada de lo que sigue tiene que ver con la hipótesis altaica, ni la apoya ni la refuta. No se sale de la familia, y allí donde el material toca las correspondencias que ese debate usa — $z \rightarrow r, \check{s} \rightarrow l$, la $-z$ pronominal— se describe la distribución sin asumir dirección de cambio ni estatus morféxico.

Este paper añade a esa descripción lo que el turco enseña sobre el instrumento, que es su razón de ser dentro de la serie.

2. Qué es un código, y qué cuesta en turco

2.1 El esqueleto consonántico

El esqueleto es la secuencia de consonantes de una palabra, con las vocales retiradas —no perdidas: quedan en una columna aparte— y cada consonante escrita con un representante canónico. El turco usa 25 símbolos canónicos, y ésta es la tabla completa. No está escrita a mano: se deriva alineando cada forma con su esqueleto (`analysis/fases/tabla_normalizacion.py`, sobre las 54.881 formas con esqueleto —54.875 tras los filtros del catálogo—, de las cuales 54.829 alinean sin ambigüedad: el 99,9 %. Las 52 restantes tienen más segmentos que posiciones —prenasales, africadas que ocupan dos— y se descartan del recuento para no adivinar).

Símbolo	Posiciones	Reúne, además de sí mismo
s	22.778	$\int \text{tʃ} \int \text{dʒ} \text{tɕ} \text{ɕ} \text{ɕ} \text{dʒ} \text{s:} \text{ʒ} \text{tʃ:} \text{ts} \dots$
r	17.150	$r \text{ ʁ} \text{ r} \text{ r:} \text{ r}^h \text{ r}^j \text{ r}^w$
t	16.397	$t: \text{t} \text{ t} \text{ t}^h \text{ t}^h$
l	14.082	$l: \text{ɭ} \text{ ɭ} \text{ ɭ}^h \text{ ɭ}^h \text{ ɭ:} \text{ ɭ:}$
k	13.482	$k \text{ k:} \text{ k}^h \text{ k}^w \text{ k}^j:$
n	13.418	$\eta \text{ n:} \text{ n} \text{ n} \text{ n:} \text{ n} \text{ n:} \text{ n}$
m	11.361	$m \text{ m:}$
q	10.650	$q:$
j	9.489	$j: \text{ j} \text{ j}^h$
b	7.380	$b \text{ b:} \text{ b}^h$
g	5.309	$g \text{ g:} \text{ g}^h \text{ g}^w$
x	5.240	$\chi \text{ x} \text{ x}^w \text{ x:}$
d	5.055	$\text{ɖ} \text{ d} \text{ d:} \text{ ɖ}$
z	4.219	$\text{z} \text{ z:} \text{ z:} \text{ z}^w$
p	3.453	$p: \text{ p} \text{ p}^h$
w	2.655	$w:$
h	1.987	$ʔ \text{ ʔ}$
v	757	$v: \text{ v} \text{ v}^h$
c	383	$\text{ç} \text{ c:} \text{ c}^h$

Símbolo	Posiciones	Reúne, además de sí mismo
f	374	f:
ʋ	359	ʋj
ð ɣ θ β	212 · 162 · 74 · 51	(nada; se escriben como se pronuncian)

Tres consecuencias que el lector encontrará en los datos:

Las sibilantes y las africadas se funden en **s**. *š*, *ž*, *č*, *dž* y *ts* se escriben todas *s*. Por eso el kazajo *bes*, el turco *beš*, el uzbeko *beʃ* y el jaladź *bi:ef*, todos «cinco», dan el mismo *b · s* — y por eso el bloque *s · s* | *hair* reúne *sač*, *šaš*, *čač* y el chuvasio *šüš*.

Lo que NO se funde importa tanto como lo que sí. *z* se escribe *z*, distinto de *s*; *ɣ* se escribe *ɣ*, distinto de *g*; y *ð* y *θ* —las interdentalas del baskir y el turkmeno— se escriben como suenan. Los tres aparecen en los resultados de §5.3 y no podrían si la tabla los hubiera fundido.

Y el coste, que en turco es mayor que en austronesio. Al fundir toda sibilante en *s*, la normalización hace invisible una de las correspondencias que definen la división primaria de la familia: el lambdacismo, chuvasio *l* frente a turco común *š*. Como *š* se escribe *s*, ese par queda diluido entre las 22.778 posiciones de *s* y §5.3 no puede verlo. Se declara aquí y se retoma allí.

2.2 Un código es un esqueleto que se repite entre lenguas

Un esqueleto que aparece, para el mismo concepto, en tres lenguas distintas o más. El umbral y la razón de exigir el concepto están en el paper del método; aquí basta el ejemplo: el kazajo *bes*, el turco *beš*, el uzbeko *beʃ* y el jaladź *bi:ef*, todos «cinco», dan *b · s* | *five*.

2.3 El coste del instrumento: la colisión

Aquí hay que decir algo que en austronesio se podía dejar en segundo plano y en turco no.

El 59 % de las formas turcas comparten esqueleto, dentro de su propia lengua, con una forma de otro concepto. No entre lenguas: dentro de una sola. Es la medida limpia de lo que el esqueleto no distingue, sin genealogía de por medio, y es alta porque las palabras turcas son cortas: 2,85 consonantes de media. (La unidad importa: es el 59 % de las formas; contado sobre tripletes de lengua, esqueleto y concepto sale 55 %, y son dos poblaciones distintas.)

El ejemplo que mejor lo enseña está en el índice inverso del anexo. El código *k · l* aparece con diez conceptos, de los cuales el anexo enseña los cuatro de mayor alcance: «ceniza», «venir», «lago» y «reír», que en turco son *kül*, *kel-*, *köl* y *kül-*. Tres son distinguibles por la vocal, que es exactamente lo que el esqueleto no registra; y el cuarto es más interesante todavía, porque **kül** «ceniza» y **kül-** «reír» son homófonas ya en turco antiguo: ni un procedimiento que conservara las vocales las separaría. Parte de la colisión no es un coste del instrumento sino un hecho de la lengua.

Y al lado hay un caso que parece igual y no lo es. *k · n* aparece con «día» y con «sol», pero ahí *kün* es la misma palabra usada para las dos cosas. Eso no es colisión sino colexificación, y es un hecho de la lengua, no un defecto del instrumento.

El código, por sí solo, no distingue los dos casos. Puede reunir una palabra usada para dos cosas o dos palabras que el instrumento confunde, y hace falta mirar las formas para saber cuál. Esta advertencia gobierna la lectura de todo el catálogo.

2.4 Aridad

Por qué los unarios no sostienen una afirmación y por qué un código de cuatro consonantes o más es motivo para dudar del análisis, en el paper del método. En turco esa sospecha resultó fundada y §3.2 la resuelve: antes del despiece los de cuatro o más eran el 29 %.

Aridad	Códigos	
1 (unario)	185	12,2 %
2 (binario)	693	45,7 %
3 (ternario)	457	30,1 %
4 o más	181	12,0 %

Las cifras que este paper defiende se restringen a los 1.152 binarios y ternarios.

Códigos por aridad

Turkic · 1,517 códigos con 3 lenguas o más

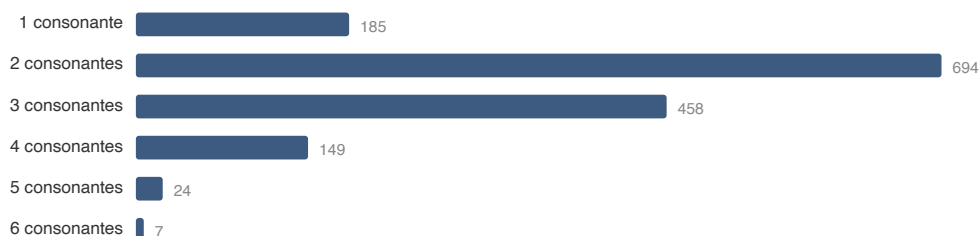


Figure 1: Reparto de los códigos por aridad. El 88 % tiene tres consonantes o menos — pero antes del despiece de §5.2 los de cuatro o más eran el 29 %.

3. Método

3.1 Datos

Corpus Integrativo, subconjunto turco: 32 lenguas, 54.875 formas con transcripción, concepto asignado y esqueleto. Las fuentes son listas léxicas estandarizadas —Lexibank, IDS y NEL— bajo licencias CC-BY, y el concepto es un nodo de Concepticon.

Tres precisiones sobre esas cifras, todas salidas de la auditoría de fase 0 (`runs/fase0-higiene.md`):

El corpus bruto tiene 70.276 formas y publicamos 54.875. La resta son 15.401, y se reparte en dos: 14.976 formas en alfabeto cirílico que el esqueletizador no reconoce y descarta en silencio, 272 del doculecto excluido, y 147 con esqueleto pero sin concepto asignado. Comprobamos que la pérdida cirílica no cuesta cobertura —cero conceptos existen solo en cirílico: todos tienen su forma latina, y el chuvasio, que es la lengua crítica de §5.3, no pierde ni una— pero el recuento que se publica es el de las formas que entran.

Este descarte no es exclusivo del turco: en todo el corpus son 147.999 formas cirílicas sin esqueleto, de las cuales 126.874 son del nakh-daguestano — el 55 % de esa familia. Se está construyendo un módulo de transliteración (`ci/translit/`) que ya recupera el 71 % del kumyko validado contra su gemelo latino. No se ha aplicado a este paper: las cifras de aquí son sin él.

Se excluye un doculecto. Crimean Tatar y CrimeanTatar comparten esqueleto en el 83 % de sus conceptos comunes: son el mismo lecto transcrito dos veces bajo dos glotocódigos. Se conserva el que

más formas aporta. No se excluyen Tatar/KazanTatar (64 %) ni los dos altái (59 %), que vienen de fuentes distintas y aportan datos distintos. El umbral del 80 % es una decisión declarada.

La familia tiene una sola macroárea, así que este paper no puede tener la sección de anomalías geográficas que el austronesio sí tenía.

Las 32 lenguas, su fuente y su rama. Se dan aquí porque de esta tabla dependen varias lecturas posteriores, y porque sin ella el lector no puede evaluar el argumento de §5.1.

Fuente	Lgs	Lenguas
lexibank	22	BarabaTatar, CrimeanTatar, Dolgan, Gagauz, KK-Halqabad 1 (<i>karakalpako</i>), KarachayBalkar, Karaim, KazanTatar, Khakas, Khalaj, Ki-Akmuz 1 (<i>kirguís</i>), North Altay, NorthAltai (<i>altái meridional</i> ; <i>el nombre contradice su glotocódigo sout2694</i>), OldTurkic, Salar, Shor, Tofa, Turkmen, Tuvan, Uyghur, Uzbek, West-Yugur
nel	7	Bashkir, Chuvash, Kazakh, Northern Uzbek, Sakha, Tatar, Turkish
ids	3	Azerbaijani, Kumyk, Nogai

Todas son turco común salvo el chuvasio, única superviviente del ogur y la otra mitad de la división primaria de la familia. De ahí que §5.3 tenga una rama de una sola lengua, con las consecuencias que allí se explican.

Tres advertencias sobre esta lista, y las tres van en contra del propio paper. La primera: **OldTurkic** es una lengua antigua, no una lengua hermana — es el turco rúnico del siglo VIII, ancestro o casi de buena parte de las otras. Cuenta como una lengua más en todas las cifras de alcance, entra en el nulo y entra en la matriz de sustituciones, y sus coordenadas (50,3° N, 137,9° E) son las del valle del Orjón, no las de una comunidad de hablantes viva. La segunda: hay pares de doculectos que comparten esqueleto entre el 47 % y el 66 % de sus conceptos comunes — Tatar/KazanTatar (65 %), KazanTatar/BarabaTatar (66 %), Uzbek/Northern Uzbek (60 %), los dos altái (60 %) —, todos por debajo del umbral del 80 % y todos con glotocódigos distintos y fuentes distintas, así que se conservan y cada uno cuenta como un doculecto. La tercera: tres lenguas vienen de una sola encuesta soviética (*ids*), y §5.4 muestra qué produce eso.

Las tres podrían inflar la precisión por razones que no son el tiempo, que es justamente lo que §5.1 quiere medir. Así que se mide en lugar de suponerlo: quitando OldTurkic la cifra de cabecera pasa de 216× a 217×; quitando además el doculecto menor de cada par solapado —criterio mecánico, no de opinión— queda en 215×. Es un 1 %, y en un sentido que ni siquiera es el temido: quitar la lengua antigua la SUBE. Ninguna corrección se aplica al catálogo publicado; las tres se declaran, y ahora se sabe lo que cuestan.

3.2 El despiece de la forma de cita, y por qué va antes que todo

Éste es el hallazgo de método más importante del trabajo y conviene ponerlo pronto.

Las fuentes no citan los verbos igual. *ids* los da casi siempre en infinitivo —*ичмакъ*, *ичм[?]к*, *ичмаз* «beber» —: el 85 % de sus verbos lleva -мАК. *lexibank* y *nel* están mezcladas y lo llevan solo en el 7,4 % y el 6,4 %, así que junto a *œlmæk* «morir» aparece *øl*.

(Una versión anterior atribuía el infinitivo a *nel* además de a *ids*. El reparto se calculaba agrupando las formas por *lect.source_id*, la fuente del registro de la lengua, cuando la forma de cita es propiedad de quien editó la forma: una misma lengua reúne formas de varias fuentes. Con *form.source_id*, que es la columna correcta, *ids* pasa de 25.739 formas a 11.831 y su tasa de infinitivo sube del 52 % al 85 %. El contrato de *ci/esquema.py* existe para que esa confusión no se pueda volver a escribir.) El sufijo turco -мАК aporta dos posiciones consonánticas al esqueleto, así que el mismo verbo salía como *j · f · m · q* en una lengua y *j · f* en otra, y no formaba código.

Son 4.974 formas, el 9,1 % de la familia, repartidas en 24 lenguas y 541 conceptos. No es morfología de la lengua: es lo que el editor del diccionario añadió al citarla.

El despiece se aplica con doble condición —tiene que coincidir el sufijo escrito y la cola del esqueleto— para que una raíz que acabe en `-mak` por casualidad sobreviva, y no se aplica si el sufijo es toda la palabra. La regla está declarada por familia en `ci/decisiones.py`, no descubierta automáticamente: recortar finales «que parecen sufijos» es la vía rápida a fabricar los códigos que uno quiere ver.

3.3 Enriquecimiento

Cobertura del código dividida por su frecuencia base; la definición completa y la advertencia de que no es comparable entre familias están en el paper del método. En turco, `b · s | f i v e` tiene alcance 25 y enriquecimiento $128\times$.

3.4 Nulos

El nulo baraja los conceptos dentro de cada lengua, conservando el inventario de cada una; qué rompe y qué conserva, en el paper del método. En turco el agrupamiento real cubre el 56,7 % de las formas y el nulo el 13,7 %, ambos con umbral de dos lenguas — el catálogo usa tres (§4.1).

4. Resultados

4.1 El catálogo

1.517 códigos sobre 793 de los 1.828 conceptos, agrupando 24.517 formas.

Magnitud	Valor	Sobre qué población
Formas dentro de un código	44,7 % (24.517 de 54.875)	umbral de 3 lenguas, el del catálogo
Formas en algún grupo	56,7 % (nulo: 13,7 %)	umbral de 2 lenguas, prueba de agrupamiento
Precisión	25,5 % (azar: 0,118 %)	todos los pares de formas de lenguas distintas
Veces sobre el azar	$216\times$ [216–216; el azar se calcula exacto]	ídem
Binarios y ternarios	1.152 (75,9 %)	de los 1.517 códigos
Enriquecimiento mediano	$185\times$	de los 1.152 binarios y ternarios

Corrección de 2026-09-09. Esta fila decía 26,2 % (14.350 de 54.875), y esa cifra no era la que su nombre promete: contaba pares (lengua, grafía) distintos —las filas del TSV de detalle, que deduplica— en vez de formas. La cobertura del catálogo sobre formas es 44,7 %. El error venía de leer el recuento del TSV como si fuera el de la población, y estaba también en el paper urálico; los otros seis usan la cifra correcta. Las demás filas de esta tabla no se mueven, y ninguna conclusión del paper depende de ésta.

Cada fila lleva su población porque son tres distintas, y confundirlas fue un error del borrador de este paper: la cobertura del catálogo (44,7 %) y la del test de agrupamiento (56,7 %) se citaban como si fueran la misma. La primera exige tres lenguas por código; la segunda, dos. La precisión es otra cosa: con qué frecuencia dos formas de lenguas distintas que comparten esqueleto comparten también concepto.

4.2 Los códigos de mayor alcance

Concepto	Código	Lenguas	Enriquecimiento
WHO	k · m	30	352×
ash	k · l	29	119×
long	z · n	28	471×
GRASP	t · t	28	278×
I	m · n	28	224×
see	k · r	27	168×
stone	t · s	27	102×

Códigos de mayor alcance

binarios y ternarios, por número de lenguas que los comparten

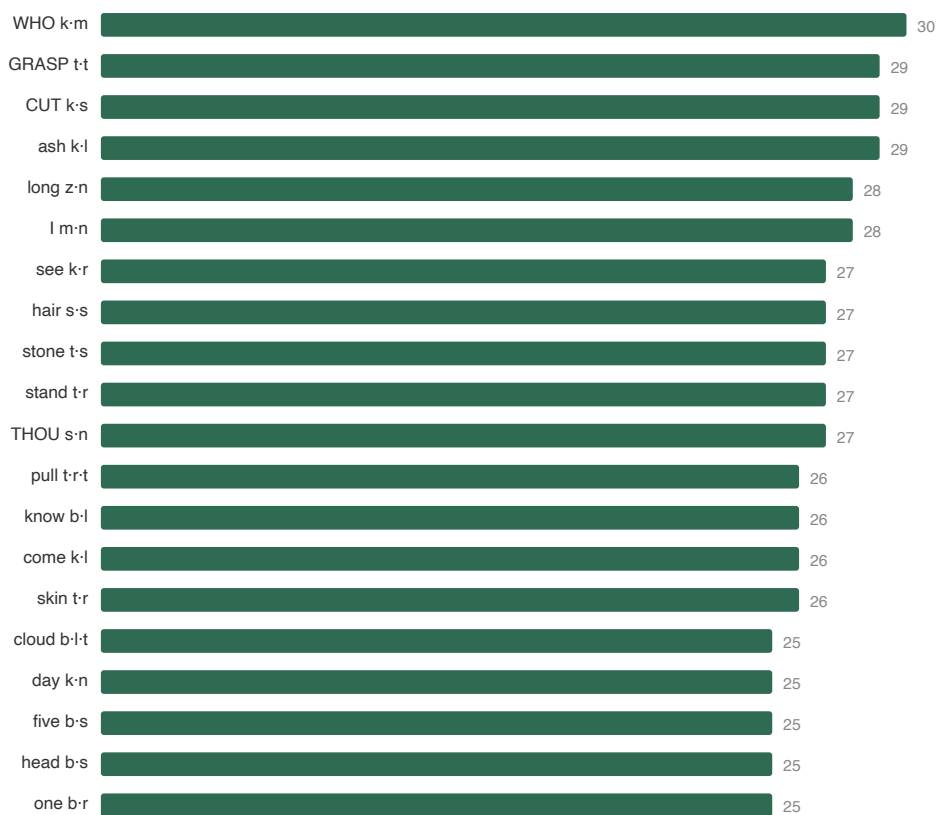


Figure 2: Los veinte códigos binarios y ternarios de mayor alcance, por número de lenguas.

Alcance de los códigos

cuántos códigos comparten cuántas lenguas



Figure 3: Distribución del alcance: cuántos códigos comparten cuántas lenguas. La mediana es de cuatro lenguas por código.

5. Lo que esta familia enseña sobre el instrumento

5.1 La cifra de cabecera es un cociente, y hay que publicar sus dos mitades

El turco rinde 216 veces sobre el azar; el austronesio, recalculado con el mismo código y la misma definición, 35. La tentación es leer eso como «el método funciona seis veces mejor en turco». Sería falso, y descomponer el cociente enseña por qué:

	Turco	Austronesio	Razón
Precisión observada	25,48 %	13,40 %	1,92×
Azar	0,118 %	0,383 %	3,35× (el denominador baja)
× azar	216	37	5,89×

Dos tercios de la diferencia viven en el denominador. Y el denominador no es una constante del universo: es la probabilidad de que dos formas cualesquiera de lenguas distintas compartan concepto, que depende de cuán específico es el esqueleto en esa familia — es decir, de la longitud de la palabra. El turco tiene 2,85 consonantes de media y el austronesio 2,79, pero su reparto es distinto, y en turco hay más formas cortas compartiendo pocos esqueletos.

Conviene descartar de entrada una explicación cómoda que no vale: no es el espacio conceptual. El turco tiene 1.828 conceptos y el austronesio 1.823. Casi idénticos.

Y hay que decir qué tan estable es la cifra. Una versión anterior de este paper publicaba 225× con un intervalo de [172–227] entre tres semillas, y atribuía esa horquilla a los datos. Era nuestra: el azar se estimaba con 80.000 sorteos, que a una tasa base del 0,1 % son unos ochenta aciertos, así que quitar seis formas de las 54.881 bastaba para mover el nulo un 14 %. El azar se calcula ahora exacto, en forma cerrada sobre el reparto de conceptos y lenguas, y solo queda como fuente de variación el submuestreo de los grupos grandes — que en esta familia es ninguna: [216–216].

Y hubo una segunda corrección, de reproducibilidad. La consulta que carga una familia no ordenaba sus filas, y PostgreSQL no promete ningún orden sin ORDER BY: la misma familia cargaba en orden distinto en cada proceso, así que el submuestreo de los grupos grandes sorteaba sobre una lista distinta y las cifras no se podían reproducir. Corriendo dos veces salían números distintos. Con la carga ordenada por una clave única, la cifra es estable en corridas sucesivas y quedó en 216× (25,48 % de precisión). Antes de ordenar oscilaba alrededor de 218×, que es lo que publicó la versión anterior — un valor de una sola tirada. Es una estimación por muestreo —80.000 sorteos para el azar, submuestreo de 60 por esqueleto— y publicarla a tres cifras significativas, como hicimos en el borrador de este paper, era darle una precisión que no tiene. El austronesio, con seis veces más datos, es mucho más estable: [35–38].

5.1-bis Lo que NO podemos concluir, y por qué se dice aquí

Un borrador de este paper sostenía que esa diferencia mide profundidad temporal: el turco tiene unos dos mil años y el austronesio más de cinco mil. La descomposición de arriba no permite sostenerlo, y conviene enumerar los candidatos que quedan sin controlar en vez de descartarlos por decreto:

- La longitud de la palabra, que mueve el denominador directamente.
- La armonía vocálica. En turco la vocal es en buena medida predecible desde la raíz, así que descartarla cuesta *menos información* que en austronesio. El instrumento tira justo la dimensión que en turco es redundante — una explicación puramente estructural de la precisión alta, sin que el tiempo intervenga.

- El eje del cambio. El cambio turco es sobre todo vocálico y morfológico; el austronesio es masivamente consonántico, con fusiones y pérdidas. El instrumento mide el eje que el turco menos mueve.
- El préstamo intrafamiliar, que en turco es enorme: chagatai sobre kipchak, otomano sobre balcánico, ruso soviético reinyectando léxico entre kazajo, karakalpako y nogayo.
- La composición de la muestra. 32 doculectos, entre ellos varios que comparten esqueleto en torno al 60 % y **OldTurkic**, del siglo VIII (§3.1). Medido, todo eso junto vale un 1 % de la cifra: $216 \times \rightarrow 215 \times$.
- El número de lenguas: 32 frente a 582. La precisión se define sobre pares de lenguas, y con 582 la inmensa mayoría de pares son remotos.

La conclusión honesta es más pequeña y más útil que la que teníamos: las razones sobre el azar de dos familias no se comparan entre sí como si midieran lo mismo. Cada familia se compara con su propio nulo, y toda comparación entre familias se publica descompuesta, con precisión y azar por separado.

5.2 Declarar el despiece rinde; descubrirlo, no

El turco es aglutinante, así que aquí la morfología tenía que doler. Duele, y de una forma que solo se ve publicando la columna del nulo:

Paso	En grupo	Nulo	Ventaja	κ	Códigos de 4+ consonantes
Palabra cruda	54,4 %	12,5 %	41,9	0,479	29,1 %
+ despiece declarado -mAK	56,7 %	13,7 %	43,0	0,498	23,8 %
+ afijos descubiertos	63,0 %	27,6 %	35,4	0,489	4,0 %

La columna κ es la corrección de Hubert-Arabie, $(\text{observado} - \text{nulo}) / (1 - \text{nulo})$: normaliza por el techo alcanzable y permite comparar entre familias lo que la ventaja en puntos no deja comparar — diez puntos sobre un nulo del 5 % y diez sobre uno del 44 % no valen lo mismo—. Aquí $\Delta\kappa = -0,009$. 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 $-0,151$.

(Esta lista tenía seis familias y las cifras de entonces. Se rehizo con las ocho el 2026-09-09 desde `pipeline/tools/tabla_serie.py`, que las deriva de una sola vez para que dos papers no puedan discrepar. 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. Lo dice §5.1 del paper sino-tibetano.)

El despiece declarado —la regla **-mAK** de §3.2— sube la ventaja sobre el azar. El descubrimiento automático de afijos la baja: gana 6,3 puntos de señal y 13,9 de ruido. Por el criterio que el propio procedimiento declara —que suba el agrupamiento *sobre el nulo*, no que baje la aridez— la fase automática fracasa, aunque su columna de cuaternarios sea la más vistosa de la tabla.

Y hay una segunda prueba, tipológica, que no necesita ningún número: el descubridor automático encuentra 365 prefijos en una familia que no tiene prefijos. No ha descubierto morfología: ha descubierto el inventario consonántico. Es exactamente el diagnóstico que el paper austronesio ya había hecho de su propio detector, y que este trabajo estuvo a punto de repetir sin citarlo.

Lo que el catálogo publicado usa es la fila del medio: despiece declarado, sin afijos descubiertos.

La lección para el protocolo no es «el despiece rinde en familias aglutinantes» sino algo más estrecho y más fiable: rinde lo que se declara y se justifica; lo que se descubre por parecido fabrica la morfología que uno quería ver.

5.3 La correspondencia diagnóstica se ahoga si se promedia la familia

El turco se divide entre el ogur, del que solo sobrevive el chuvasio, y el turco común, todo lo demás. Y tiene correspondencias diagnósticas conocidas: chuvasio *r* turco común *z* (PT **r*₂) y chuvasio *l* turco común *š* (PT **l*₂).

Sobre los nombres. La literatura las llama *rotacismo* y *lambdacismo* o bien *zetacismo* y *sigmatismo* según qué lado se considere primario, y esa elección es una toma de posición en un debate que este trabajo no aborda. Aquí se escriben como correspondencias —*z* *r*, *š* *l*— sin asumir dirección.

Agregando las sustituciones sobre toda la familia, la primera no aparece: sale a 1,4 veces sobre el azar, muy por debajo de *b* ↔ *p* (31,9×) o *d* ↔ *t* (19,5×). La razón es aritmética: el ogur tiene una lengua y el turco común treinta y una, así que los pares que cruzan la división son el 6,7 % del total y su señal queda sepultada por el 93,3 % restante.

Restringiendo al par de ramas, aparece. La columna que decide es la última: cuánto más concentrada está la correspondencia al cruzar el corte que sin cruzarlo.

Segmento	Principal	Casos	Cuota al cruzar	Cuota sin cruzar	Diagnóstico
f	v	26 de 30	87 %	0 %	solo al cruzar
b	p	2.887 de 3.385	85 %	43 %	2,0×
d	t	862 de 1.088	79 %	85 %	0,9×
w	ʋ	230 de 409	56 %	0 %	solo al cruzar
z	r	550 de 1.006	55 %	9 %	6,1×
ð	s	32 de 64	50 %	9 %	5,3×
j	ʃ	1.602 de 3.305	48 %	71 %	0,7×
g	x	64 de 136	47 %	7 %	6,6×

z *r* pasa de invisible a 6,1×, y la tabla explica por qué es diagnóstica y *d* ↔ *t* no: cuando dos lenguas del turco común difieren en esa posición, *z* corresponde a *r* solo el 9 % de las veces; cuando el corte es el del ogur, el 56 %.

Tres filas más merecen comentario, y dos van en contra nuestra:

- *w* *ʋ* y *f* *v* no existen sin cruzar el corte (cuota 0 %). Por el criterio de concentración que esta misma sección declara son las más limpias de la tabla, más que *z* *r*. Corresponden a la prótesis labial del chuvasio (*vut* 'fuego' turco común *ot*).
- *g* *x* da 6,6×, por encima del rotacismo, y no sabemos qué es. Con una rama de una sola lengua, cualquier peculiaridad de transcripción de esa lengua sale como diagnóstica. Lo decimos en vez de celebrar solo la fila que nos convenía.
- *ð* *s* (5,3×) es una innovación baskira, no ogur: mide una lengua del lado grande contra el chuvasio. La columna «Diagnóstico» no distingue eso, y es su límite.

El lambdacismo no aparece ni restringido, y el diagnóstico correcto no es el que dimos primero. Culpamos a la normalización de §2.1, que escribe *s* tanto la *s* como la *š*. Pero la causa principal es otra y se comprueba en un ejemplo: el étimo es PT **bēl*(k), cuyo reflejo ogur es chuvasio *pillēk* → *p* · *l* · *k*, frente al turco común *beš* → *b* · *s*. Longitudes distintas, y el procedimiento solo compara esqueletos de la misma longitud. La fusión sibilante agrava, pero lo que impide ver el lambdacismo es la restricción de longitud.

Y una advertencia sobre los «casos». Con una lengua a un lado y treinta y una al otro, cada forma chuvasia genera hasta 31 pares: los 1.006 casos de *z* *r* no son 1.006 observaciones independientes, sino unas decenas de étimos contados muchas veces. La cifra señala dónde mirar; no es un test.

5.4 El catálogo separa ramas sin saber que existen

Éste es el resultado que más tardamos en ver, porque estaba en el TSV y no en ningún informe.

Buscando en el catálogo pares de códigos del mismo concepto que difieren en una sola posición, aparecen 297, de los cuales 184 son disjuntos en lenguas —ninguna lengua está en los dos— y solo esos pueden proponerse como isoglosa. Las sustituciones que generan los disjuntos son j s (38 pares), d t (21), q x (13), k q (12), k x (10), g k (8).

Y conviene mirar al lado la lista de los 113 que solapan, porque las dos listas no coinciden: allí manda k q (17), b p (13) y g k (13). Una sustitución que domina la segunda lista y no la primera es casi seguro una convención de transcripción y no un hecho de la lengua.

Los disjuntos de mayor alcance son cortes de rama reconocibles:

Concepto	Qué separa		
one	b·r (25 lgs)	p·r (7)	chuvasio + siberiano: <i>per, pur, pir</i> frente a <i>bir, ber</i>
head	b·s (24)	p·s (6)	los mismos: <i>pas, paš</i> frente a <i>baş, bas</i>
five	b·s (25)	p·s (7)	siberiano, sin chuvasio (que dice <i>pillēk</i> , §5.3)
four	t·r·t (16)	d·r·t (14)	oguz + kipchak occidental: <i>dört</i> frente a <i>tört</i>
see	k·r (27)	g·r (6)	<i>k-</i> > <i>g-</i> oguz

Las siete lenguas de p·r|one son chuvasio, jakasio, shor, los dos altái, tártaro de Baraba y yugur occidental: el ensordecimiento inicial que agrupa el oguz con el siberiano, que ningún árbol da tal cual. Las catorce de d·r·t|four son azerí, gagauzo, kumyko, karaim, tártaro de Crimea, tártaro de Kazán y baskir: el oguz más el kipchak occidental.

El catálogo no sabe que existen esas ramas. Las separa igual, porque separa formas.

Esto obliga a matizar §5.3: allí $d \leftrightarrow t$ sale a $0,9 \times$ y se lee como «alternancia general de sonoridad, no correspondencia de rama». Con los pares de códigos delante, la lectura correcta es otra: $d \leftrightarrow t$ no se concentra en el único corte que el corpus admite —oguz frente a turco común— pero sí separa un grupo coherente. El corpus solo tiene dos etiquetas de rama (§3.1) y el corte oguz nunca se pudo probar.

La lección de protocolo es que esta búsqueda debe ser una fase estándar. Sale gratis del catálogo, está condicionada por concepto —que es lo que le falta a la matriz agregada— y no necesita etiquetas de rama.

5.5 El enriquecimiento máximo vuelve a ser un préstamo

En austronesio, el código de mayor enriquecimiento resultó ser un galicismo. Se repite, y con la misma forma:

Código	Concepto	Lenguas	Enriquecimiento	Formas	Qué es
n·s·r·t·r·k	BANYAN	3	13.720×	<i>nusrat-terek</i>	máximo del catálogo entero
f·r·m·r	FARMER	3	8.232×	<i>fermer</i>	ruso <i>фермер</i>
p·l·m	PALM TREE	3	7.840×	<i>palma</i>	ruso <i>пальма</i> · máximo entre binarios y ternarios
m·n·t	COIN	4	7.840×	<i>manat, mänet</i>	ruso <i>монета</i>

Los superlativos llevan su población pegada: 13.720× es el máximo del catálogo; 7.840× lo es solo entre binarios y ternarios, que es la población que este paper defiende (§2.4).

Alcance frente a enriquecimiento

cada punto es un código binario o ternario; eje vertical logarítmico

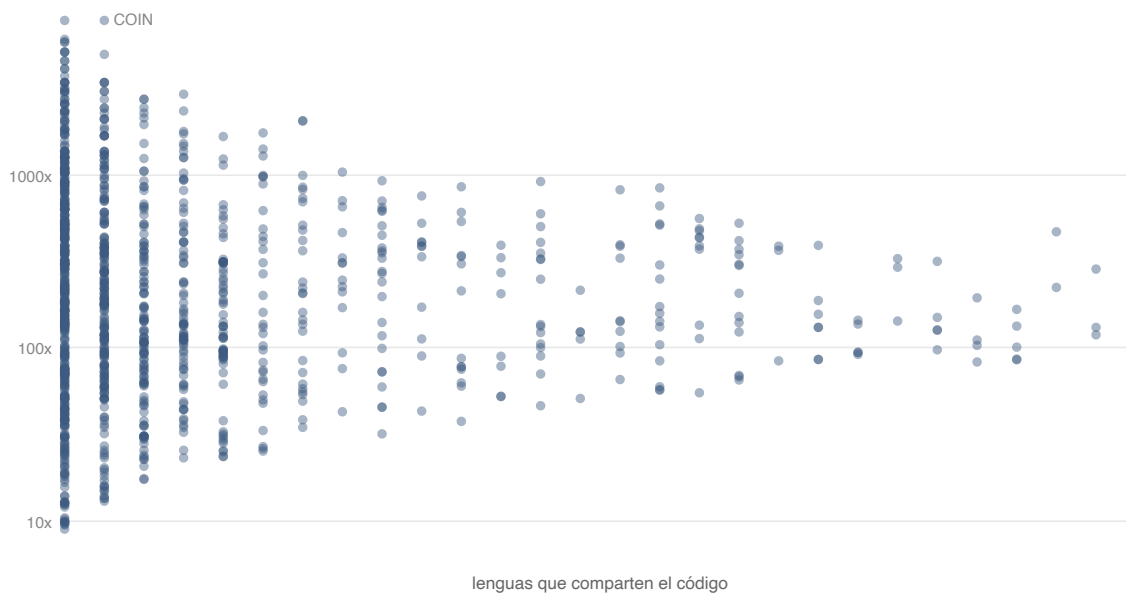


Figure 4: Alcance frente a enriquecimiento. Cada punto es un código binario o ternario; el eje vertical es logarítmico. Los códigos de enriquecimiento extremo son pocos y de alcance corto — y como muestra la tabla de arriba, préstamos recientes cuyo concepto está atestiguado en tres o cuatro lenguas.

Confirma en una segunda familia lo que el paper anterior estableció en la primera: un enriquecimiento alto no indica profundidad histórica, indica que una secuencia rara se repite, y una secuencia rara repetida entre vecinos es lo que produce un préstamo reciente.

Y el mecanismo se ve mejor aquí que allí. Estos conceptos —«baniano», «granjero», «palmera», «moneda»— están atestiguados en solo tres o cuatro lenguas del corpus. Como el enriquecimiento es cobertura ÷ frecuencia base, y la cobertura de un código presente en *todas* las lenguas que atestiguan su concepto es 1,000 por construcción, el enriquecimiento extremo lo produce el concepto raro, no el código raro. Es un artefacto del denominador, y conviene decirlo antes de que alguien lo descubra.

5.6 La geografía dice menos, no dice nada

Un borrador de este paper afirmaba que en turco no existe relación entre compartir códigos y estar cerca. Era falso: usábamos solo la primera de las dos partes del análisis.

La primera parte, por pares de lenguas, no discrimina. Con 32 lenguas de una familia joven, todos los pares comparten muchos códigos: 467 de los 496 pares posibles caen en el cubo «21+», y los cubos que el austronesio necesitaba están vacíos. De un diseño sin rango dinámico no se puede concluir nada, y menos un cero.

La segunda parte, por grupos, sí discrimina, y da un gradiente monótono:

Tamaño del grupo	Grupos	Dispersión ÷ azar
3–5 lenguas	763	0,64
6–10	224	0,72
11–25	149	0,85
26–60	14	0,98

Los códigos turcos sí forman grupos más compactos de lo esperable, y la compacidad se pierde al crecer

el alcance — el mismo patrón que en austronesio, donde el rango iba de 0,40 a 0,65. La diferencia es de intensidad, no de existencia: en turco la señal es más débil.

Una explicación turcológica que no habíamos considerado y que probablemente pesa más que cualquier «modo de diversificación»: las coordenadas modernas de varias de estas lenguas son producto de deportaciones del siglo XX —karaim en Lituania, tártaro de Crimea deportado a Uzbekistán, karachayo-balkaro en 1943— y no de la diversificación. El proxy geográfico está envenenado en esta familia por razones históricas concretas.

5.7 Las clases emergentes no salen, y eso también se publica

Cerrar transitivamente las sustituciones que aguantan en la mitad de validación no da nada útil: a cualquier umbral practicable, quince de los diecisiete segmentos que alcanzan el umbral caen en un solo grupo, y solo se desprende {d, t}. (El alfabeto completo del turco son 25 símbolos, §2.1; diecisiete es cuántos llegan al mínimo de pares.)

Por eso el Anexo A no lleva columna de código emergente, y los TSV publicados la llevan vacía. Trasplantar las clases austronesias habría sido cómodo y falso: el valor de una clase es relativo a la familia que la produjo. Lo decimos porque estuvo mal hecho: durante parte de este trabajo, el generador del catálogo escribió en esa columna las clases del austronesio.

6. Límites

El instrumento no distingue colisión de colexificación, y en una familia con el 59 % de colisión es el límite que más pesa (§2.3).

El lambdacismo es invisible, y no por lo que creíamos: la causa principal es que sus étimos dan esqueletos de longitudes distintas (*pillēk* frente a *beš*) y el procedimiento solo compara esqueletos de igual longitud. La fusión sibilante de §2.1 agrava, pero no es la causa (§5.3).

No hay reflejos cero. Es un límite de la matriz de sustituciones, no del catálogo entero: éste agrupa por identidad de esqueleto, donde la longitud no se compara.

Una sola macroárea, luego no hay lista de anomalías geográficas.

El corpus tiene dos etiquetas de rama —ogur y turco común— cuando la clasificación de referencia distingue oguz, kipchak, karluk, siberiano y arghú además del ogur. Así que §5.3 no probó «pares de ramas» en plural: probó el único corte que el corpus admite. Los resultados de §5.4 sugieren que con etiquetas completas habría más.

El despiece resuelve la forma de cita, no la morfología. Quitar el infinitivo -mAK es quitar lo que puso el diccionario. La segmentación de verdad sigue sin resolver, y §5.2 muestra que descubrirla automáticamente empeora el resultado.

La muestra mezcla doculectos registrados con siglos de distancia. Incluye OldTurkic, del siglo VIII, junto a variedades contemporáneas, y ninguna medida de similitud léxica modela eso; nosotros tampoco.

Lo que no se declara aquí es que unos doculectos «no sean lenguas de verdad». La unidad de este trabajo es el doculecto con lista léxica, y decidir cuáles cuentan como una sola lengua es un juicio ajeno al método —el mismo tipo de juicio que §1 se compromete a no hacer—. Lo que sí es un problema de datos es un registro duplicado, y para eso hay una regla mecánica declarada: mismo glotocódigo, o más del 80 % de esqueletos idénticos. Los pares próximos de §3.1 no la cruzan, y su efecto sobre la cifra de cabecera está medido: 2 %.

No calibramos contra cognación experta, habiéndola. Savelyev y Robbeets (2020) publicaron una filo-

genia turca con cognación experta sobre una muestra parecida a la nuestra, distribuida por Lexibank —la fuente que este pipeline ya lee— y no está ingerida en el Corpus Integrativo. El paper austronesio sí calibró (contra IE-CoR y ABVD); éste no, y es una carencia, no una imposibilidad.

Una sola familia. Todo lo que este paper dice sobre el significado de las cifras se apoya en dos puntos. Y docs/PERFIL-FAMILIAS.md ya tiene sesenta familias perfiladas con la misma métrica, donde el nakh-daguestano —mucho más profundo que el turco— empata con él. Esa comparación está por hacer.

7. Conclusión

El catálogo de códigos turcos agrupa el 26,2 % de las formas en códigos de tres lenguas o más, con una precisión 216 veces superior al azar, y reproduce sin buscarlo el vocabulario que la turcología conoce: los numerales, los pronombres, las partes del cuerpo.

Pero lo que este trabajo aporta no es ese catálogo. Es lo que la familia enseña sobre el instrumento, y casi todo son correcciones de lo que creíamos:

Que la cifra de cabecera es un cociente, y que dos tercios de la diferencia entre familias vienen del denominador — así que las razones sobre el azar de dos familias no se comparan entre sí. Que el despiece rinde cuando se declara y empeora cuando se descubre, y que un descubridor que encuentra 365 prefijos en una familia sin prefijos ha encontrado el inventario, no la morfología. Que la matriz de sustituciones hay que calcularla por par de ramas, porque agregada borra justamente lo que define la genealogía. Que la geografía dice menos, no dice nada. Y que el enriquecimiento máximo señala préstamos por un mecanismo nombrable: lo produce el concepto raro.

Y un resultado que no buscábamos y tardamos en ver, porque estaba en los datos y no en ningún informe: el catálogo separa ramas sin saber que existen. Trescientos cincuenta y siete pares de códigos hermanos, y los mayores agrupan el chuvasio con el siberiano y el oguz con el kipchak occidental. Eso no lo aprendió de ningún árbol.

Un instrumento que afirma poco puede aplicarse donde no hay bibliografía. Pero para saber qué afirma, hay que aplicarlo donde sí la hay — y estar dispuesto a que la respuesta sea que afirmaba de más.

Referencias

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

Clauson, G. (1972). *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford University Press.

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

Erdal, M. (2004). *A Grammar of Old Turkic*. Brill.

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

Holman, E. W., Wichmann, S., Brown, C. H. *et al.* (2008). Explorations in automated language classification. *Folia Linguistica* 42.

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

- Johanson, L. (2021). *Turkic*. Cambridge University Press.
- Johanson, L. y Csató, É. Á. (eds.) (1998). *The Turkic Languages*. Routledge.
- Holman, E. W., Brown, C. H., Wichmann, S. *et al.* (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52(6). — [el programa de leer la similitud léxica como profundidad temporal, que §5.1 no consigue sostener]
- 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., Greenhill, S. J. y Gray, R. D. (2017). The potential of automatic word comparison for historical linguistics. *PLOS ONE* 12(1).
- Róna-Tas, A. (1998). The reconstruction of Proto-Turkic and the genetic question. En Johanson y Csató (eds.).
- Róna-Tas, A. y Berta, Á. (2011). *West Old Turkic: Turkic Loanwords in Hungarian*. Harrassowitz. — [las correspondencias *z* → *r* y *š* → *l* del *ogur*]
- Savelyev, A. y Robbeets, M. (2020). Bayesian phylolinguistics infers the internal structure and the time-depth of the Turkic language family. *Journal of Language Evolution* 5(1). — [filogenia turca con cognación experta; data el proto-turco en 2.100 AP. No usada aquí, ver §6]
- Schönig, C. (1997–98). A new attempt to classify the Turkic languages. *Turkic Languages* 1–2. — [por qué el árbol turco no es limpio]
- Tekin, T. (1979). Once more zetacism and sigmatism. *Central Asiatic Journal* 23. — [el debate sobre la dirección de *z* → *r*]
- Toledo Martínez, A. (2026). *Corpus Integrativo: una base léxica multi-capa 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	families/turkic/output/catalogo-turkic.tsv — 1.517 filas
Detalle por forma	families/turkic/output/catalogo-turkic-lenguas.tsv — 14.383 filas
Higiene (§3.1)	analysis/fases/fase0_higiene.py → runs/fase0-higiene.md
Decisiones de datos (§3.1, §3.2)	ci/decisiones.py
Construcción	analysis/fases/catalogo_familia.py
Agrupamiento y nulos (§3.4, §4.1)	analysis/fases/agrupamiento_familia.py → runs/agrupamiento.md
Despiece morfológico (§5.2)	analysis/fases/fase2_despiece.py → runs/fase2-despiece.md

Sustituciones agregadas (§5.3, §5.6)	<code>analysis/fases/fase6_clases_emergentes.py → runs/fase6-clases.md</code>
Sustituciones por rama (§5.3)	<code>analysis/fases/fase6_por_rama.py → runs/fase6-por-rama.md</code>
Códigos y geografía (§5.5)	<code>analysis/fases/codigos_vs_parentesco.py → runs/codigos-parentesco.md</code>
Anexo A	<code>analysis/fases/anexo_codigos.py</code>
Cifras de cabecera (§4.1, §5.1)	<code>ci/metricas.py · ci/familia.py</code>
Tabla de normalización (§2.1)	<code>analysis/fases/tabla_normalizacion.py → runs/normalizacion.md</code>
Pares de códigos hermanos (§5.4)	consulta sobre el catálogo, en <code>runs/</code>
Auditoría anti-mezcla de familias	<code>analysis/tools/audita_familias.py</code>

Todos los informes de **runs/** declaran su población en la cabecera, y son idénticas. Es la comprobación que evita el fallo que este paper tuvo en su borrador: cuatro secciones calculadas sobre poblaciones distintas sin que nadie lo notara. `analysis/tools/verifica_paper.py` falla si difieren.

Fuentes de datos: Lexibank (List, Forkel, Greenhill et al. 2022, CC-BY-4.0), IDS (Key & Comrie, CC-BY-4.0), NEL, Concepticon y Glottolog (CC-BY-4.0). Todas redistribuibles con atribución.

Anexo A · Cien códigos comentados

El catálogo tiene 1.517 códigos y los datos completos están en `families/turkic/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, salvo los puntos con que lexibank separa los segmentos en 296 formas turcas (`b.i.r` se lee `bir`), que son puntuación de la fuente. Y la agrupación temática es un orden de lectura declarado a mano, no una clasificación semántica.

Este anexo no lleva columna de código emergente, y la ausencia es un resultado. En austronesio, las sustituciones se cerraban en tres clases limpias que se escribían junto a cada código. En turco no: a cualquier umbral útil el cierre transitivo mete dieciséis de los dieciocho segmentos en un solo grupo (§5.7). Una clase de dieciséis segmentos no informa de nada, así que la columna no está. Trasplantar aquí las clases austronesias habría sido cómodo y falso: el valor de una clase es relativo a la familia que la produjo.

Índice inverso: códigos que reaparecen en varios conceptos

Esta tabla es, en turco, la más importante del anexo — y conviene leerla antes que los bloques, porque explica el resto.

k·l aparece con diez conceptos, de los que esta tabla enseña los cuatro de mayor alcance: «ceniza», «venir», «lago» y «reír» — en turco *kül*, *kel-*, *köl* y *kül-*. Tres se separan por la vocal, que es justo lo que el esqueleto no registra. El cuarto es mejor todavía para entender el coste: **kül** «ceniza» y **kül-** «reír» son homófonas ya en turco antiguo, así que ni conservando la vocal se separarían. Parte de la colisión es de la lengua, no del instrumento. **k·n** enseña los tres casos a la vez: cubre «día», «sol» y «ancho». *kün* «día» y *kün* «sol» son la misma palabra, y eso es colexificación real. Pero *keñ* «ancho» entra por una tercera puerta que no es ni la vocal ni la lengua: la normalización de §2.1 escribe *n* la velar *ŋ*. Es colisión fabricada por nuestra propia tabla, y conviene tenerla como categoría aparte.

Ésa es la diferencia que hay que sostener al leer todo el anexo: un mismo código puede reunir la misma palabra usada para dos cosas, o dos palabras que el instrumento no distingue, y el código por sí solo no dice cuál de las dos. El 53 % de colisión que declara §2.3 es esto, contado.

Código	Conceptos con los que aparece, y su enriquecimiento
k·l	ash (29, 119×) · come (26, 104×) · lake (23, 95×) · laugh (23, 95×)
t·r	stand (27, 86×) · skin (26, 83×) · narrow (20, 68×) · sit (16, 51×)
s·r	yellow (19, 55×) · suck (18, 60×) · rotten (18, 58×) · ask (request) (15, 47×)
s·n	thou (27, 86×) · count (20, 66×) · true (18, 57×)
t·s	stone (27, 101×) · tooth (23, 92×) · breast (17, 66×)
k·n	day (25, 150×) · wide (20, 124×) · sun (17, 102×)
b·r	one (25, 98×) · give (22, 86×) · go (22, 86×)
j·t	say (18, 143×) · lie down (17, 144×) · swallow (15, 123×)

Numerales

Los numerales son el mejor bloque del anexo, y también el que mejor enseña lo que el anexo esconde.

Cada numeral está partido en dos códigos por la alternancia de sonoridad inicial, y aquí solo se ve la mitad grande de cada par: «cuatro» es *t·r·t* (16 lenguas) y *d·r·t* (14); «cinco», *b·s* (25) y *p·s* (7); «uno», *b·r* (25) y *p·r* (7). §5.4 muestra que esas mitades son cortes de rama reales —el oguz frente al resto, el chuvasio con el siberiano— así que la sección elegida para enseñar que «el esqueleto no se mueve» es también donde más visiblemente se mueve. El criterio de alcance del anexo deja fuera la mitad pequeña.

La vocal varía y el esqueleto no se mueve — mientras la consonante inicial no cambie. «Uno» es *bir* en azerí y uigur, *ber* en baskir, *bır* y *bir* en kazajo, *bi:r* en dolgano y sajá: cinco timbres vocálicos sobre *b·r*. «Cinco» recorre *beš*, *bes*, *beʃ*, *bif*, *bäš* y *bi:ef* sin salirse de *b·s*.

t·r·t|four tiene la primera y la última posición iguales, y merece atención por eso: *tört*, *tört*, *tyört*, *tort*. Un ternario con *t...t* tiene menos espacio de posibilidades que uno con tres consonantes distintas, y por tanto un enriquecimiento algo más barato — 129× frente a los 139× de *b·s*, que es binario. Conviene no leer el enriquecimiento sin mirar de qué está hecho el código.

b·s|five · binario · 25 lenguas · enriquecimiento 127×

Azerbaijani beš	Bashkir beš	Dolgan bies	Gagauz bulaʃwq
KarachayBalkar beš	Karaïm beš	KazanTatar biš	Khalaj bi:ef, be:ef
Kumyk beš	NorthAltai beš	Northern Uzbek beʃ	Sakha bies
Tatar biʃ	Turkish beʃ	Turkmen bāš	Uyghur bāš

b·r|one · binario · 25 lenguas · enriquecimiento 98×

Azerbaijani bir	Bashkir ber	Dolgan bi:r	Gagauz bir
KarachayBalkar bir	Karaïm bir	KazanTatar ber	Ki-Akmuz 1 bir
Nogai bir	Northern Uzbek bir	OldTurkic bir	Salar bir
Tatar ber	Turkish bir	Turkmen bir	Uyghur bir

t·r·t|four · ternario · 16 lenguas · enriquecimiento 124×

BarabaTatar tört	Dolgan tyört	KK-Halqabad 1 tört	KarachayBalkar tört
Kazakh tört	Khakas tört	Khalaj tø:ört, ty:ört	Ki-Akmuz 1 tort
North Altay tört	NorthAltai tört	Northern Uzbek tort	OldTurkic tört
Sakha tyört	Shor tört	Uyghur tört	Uzbek tört

Partes del cuerpo

s·s|hair reúne el rastro de una correspondencia diagnóstica sin declararla. Las formas van de *sač* y *šaš* a *čač* y *cac*, y entre ellas está el chuvasio *šüs*. El instrumento las agrupa porque la normalización canónica escribe *s* todas esas sibilantes; lo que no puede es decir que *šüs* frente a *sač* sea el mismo cambio que separa el oguz del turco común. Eso lo dice §5.3, y solo cuando se restringe al par de ramas.

Los binarios dominan y son cortos: *b·s|head* (*baš/bas*), *t·l|tongue* (*til/tel*), *k·z|eye* (*köz/küz*), *t·s|tooth*. Es la aridez media de 2,85 hecha lista. Y explica de dónde viene el coste del instrumento en esta familia: con dos consonantes hay poco donde equivocarse, pero también poco donde distinguir.

s·s|hair · binario · 27 lenguas · enriquecimiento 134×

Azerbaijani sač	BarabaTatar cac	Chuvash šüs	Gagauz sač
KK-Halqabad 1 sas	Karaïm sač	KazanTatar čäč	Khakas sas
Ki-Akmuz 1 čač	Nogai šaš	North Altay tcatç	OldTurkic sač
Shor šaš	Tofa čeʃ	Turkmen sač	Uzbek soč

t·r|skin · binario · 26 lenguas · enriquecimiento 83×

BarabaTatar तरी	Bashkir tire	CrimeanTatar teri	Dolgan tiri:
KarachayBalkar teri	Kazakh teri	KazanTatar tire	Khalaj teri
Kumyk teri	Nogai teri	NorthAltai tere	Northern Uzbek teri
Sakha tiri:	Shor tere	Tatar tire	Uzbek teri

b·s|head · binario · 25 lenguas · enriquecimiento 127×

Azerbaijani baş	Bashkir baş	Dolgan bas	Gagauz baş
KarachayBalkar baş	Karaim baş	KazanTatar baş	Khalaj ba:f
Kumyk baş	North Altay ba:f	NorthAltai baş	OldTurkic baş
Sakha bas	Tofa ba ^ʃ š	Turkish ba:f	Uyghur baş

t·l|tongue · binario · 24 lenguas · enriquecimiento 143×

BarabaTatar til	Bashkir tel	Dolgan tıl	KK-Halqabad 1 til
Karaim til	Kazakh til	Khakas təl	Khalaj til
Kumyk til	Nogai til	NorthAltai til	Northern Uzbek til
Sakha tıl	Shor til	Uyghur til	Uzbek til

t·s|tooth · binario · 23 lenguas · enriquecimiento 92×

Bashkir teš	CrimeanTatar tiš	Dolgan ti:s	KarachayBalkar tiš
Karaim tis	KazanTatar teš	Khakas təs	Ki-Akmuz 1 tiš
Kumyk tiš	Nogai tis	NorthAltai tiš	Northern Uzbek ti:f
Sakha ti:s	Salar tič	Tatar te:f	Uzbek tiš

q·n|blood · binario · 22 lenguas · enriquecimiento 393×

BarabaTatar qan	Bashkir qan	CrimeanTatar qan	KarachayBalkar qan
Kazakh qan	KazanTatar qan	Ki-Akmuz 1 qan	Kumyk qan
North Altay qan	NorthAltai qan	Northern Uzbek qon	Salar qan
Shor qan	Tatar qan	Uyghur qan	Uzbek qon

r·n|PUS · binario · 21 lenguas · enriquecimiento 389×

Azerbaijani irin	BarabaTatar äräj	Bashkir eren	CrimeanTatar irin
KK-Halqabad 1 irin	KarachayBalkar irin	Karaim irin	KazanTatar eren
Khakas erəŋ	Ki-Akmuz 1 irin	Nogai irin	NorthAltai irin
Sakha iriŋe	Tofa irin	Turkish irin	Turkmen irin

q·l|ARM OR HAND · binario · 20 lenguas · enriquecimiento 152×

BarabaTatar qol	Bashkir qul	CrimeanTatar qol	KK-Halqabad 1 qol
Kazakh qol	KazanTatar qul	Khalaj qol	Ki-Akmuz 1 qol
Nogai qol	North Altay qol	NorthAltai qol	OldTurkic qol
Shor qol	Tofa qol	Uyghur qol	Uzbek qol

q·l·q|ear · ternario · 20 lenguas · enriquecimiento 301×

BarabaTatar ҡолаҡ	Bashkir qоlaq	CrimeanTatar ҡулаҡ	KK-Halqabad 1 qulaq
Kazakh qulaq	KazanTatar ҡолаҡ	Khalaj qula:q	Ki-Akmuz 1 qulaq
Nogai qulaq	North Altay qulaq	NorthAltai qulaq	Northern Uzbek quloq
Tofa ҡулаҡ	Tuvan qulaq	Uyghur qulaq	Uzbek ҡулаҡ

b·r·n|nose · ternario · 20 lenguas · enriquecimiento 528×

Azerbaijani burun	CrimeanTatar burun	Gagauz burnu	KarachayBalkar burun
Kazakh burun	KazanTatar borın	Khalaj burun	Ki-Akmuz 1 burun
Nogai burın	North Altay burun	NorthAltai burun	Northern Uzbek burun
Tatar borın	Turkish burn	Turkmen burun	Uyghur burun

q·n·t|wing · ternario · 20 lenguas · enriquecimiento 347×

BarabaTatar ҡанат	Bashkir qanat	CrimeanTatar ҡанат	KK-Halqabad 1 qanat
Kazakh qanat	KazanTatar ҡанат	Ki-Akmuz 1 qanat	Kumyk qanat
North Altay qanat	NorthAltai qanat	Northern Uzbek qanot	OldTurkic ҡанат
Shor qanat	Tatar qanat	Uyghur qanat	Uzbek ҡанат

t·z|knee · binario · 19 lenguas · enriquecimiento 136×

BarabaTatar tizi	CrimeanTatar tiz	KK-Halqabad 1 tiz ^{je}	Karaim tyz, tiz, kiz
Kazakh tiz ^{je}	KazanTatar tez	Ki-Akmuz 1 tize	Kumyk tiz
Nogai tiz	North Altay tize	NorthAltai tize	OldTurkic tiz
Shor tize	Tatar tez	Uyghur tiz	Uzbek tiz

k·z|eye · binario · 18 lenguas · enriquecimiento 303×

Azerbaijani köz	BarabaTatar köz	CrimeanTatar köz	KK-Halqabad 1 köz
KarachayBalkar köz	Karaim köz	Kazakh köz	KazanTatar küz
Ki-Akmuz 1 köz	Nogai köz	Northern Uzbek koz	OldTurkic köz
Salar koz	Tatar kuz	Uyghur köz	Uzbek qoz

t·s|BREAST · binario · 17 lenguas · enriquecimiento 66×

BarabaTatar түш	Bashkir tyʃ	Dolgan tyøs	KK-Halqabad 1 ? tøs
KarachayBalkar tös	Karaim tös	Kazakh tös	KazanTatar түш
Kumyk tös	North Altay tøj	NorthAltai tös	OldTurkic tös
Sakha tyøs	Shor tös	Tatar tøj	Tofa tös

r·q|back · binario · 17 lenguas · enriquecimiento 331×

Bashkir arqa	CrimeanTatar arqa	KK-Halqabad 1 arqa	Kazakh arqa
KazanTatar arqa	Khalaj arqa	Ki-Akmuz 1 arqa	Kumyk arqa
Nogai arqa	North Altay arqa	NorthAltai arqa	Northern Uzbek orqa
OldTurkic arqa	Turkmen arqa	Uyghur arqa	Uzbek orqa

b·t|UPPER LEG (THIGH) · binario · 16 lenguas · enriquecimiento 124×

Bashkir bot	CrimeanTatar but	Dolgan bu:t	Gagauz but
Karaim but	KazanTatar bot	Khalaj bu:ut	Kumyk but
Nogai ? but	North Altay but	NorthAltai but	Sakha ? bu:t
Tatar bot	Tofa but	Turkish but	Turkmen but

q·r·n|belly · ternario · 15 lenguas · enriquecimiento 504×

CrimeanTatar qarın	KK–Halqabad 1 ? qarwın	KarachayBalkar qarwın	Kazakh qarən
KazanTatar qarın	Khalaj qa(:)rwın	Ki–Akmuz 1 qwrın	North Altay qarwın
NorthAltai qarın	Northern Uzbek qwrın	OldTurkic qarın	Shor qarwı
Uyghur qerın	Uzbek qwrın	WestYugur qarwın	

s·s·k|flower · ternario · 15 lenguas · enriquecimiento 354×

Azerbaijani tʃɪtʃæk	BarabaTatar çiçäk	Bashkir säskä	Chuvash tɕɛ'ɕɛk
CrimeanTatar çiçek	Gagauz çiçek	KazanTatar čäčäk	Kumyk čecek
North Altay ɬɛɬek	NorthAltai čecek	OldTurkic čecek	Tatar tɕæɕkæ
Turkish tʃɪtʃek	Tuvan čecek	Uyghur ? ɬɛɬek	

Parentesco y persona

El sistema pronominal completo aparece en el catálogo, y aparece como sistema:

m·n|I (*men, min, men*) · **s·n|THOU** (*sen, sin, sän*) · **b·z|WE** (*biz, bez, bız*) · **s·z|YOU** (*siz*) · **l·r|THEY** (*olar, alar, ular, ılar*).

Obsérvese la pareja de arriba y la de abajo: la 1.^a y 2.^a de singular acaban en **n** y sus plurales en **z**. Tres precisiones, porque la formulación fácil sería incorrecta. La regla vale para esas cuatro casillas y no para **l·r|THEY**, que acaba en **r** y no pertenece al mismo sistema (es el demostrativo *ol* más el plural productivo *-lar*). Falta la 3.^a de singular. Y que esa **–z** sea un morfema es una hipótesis de la turcología, no un dato: el plural turco es *-lar*. El catálogo no la resuelve — solo pone las cuatro formas contiguas y con sus cifras, que es lo que sabe hacer.

n·d|THERE alcanza 562×, el enriquecimiento más alto de esta sección. Con 19 lenguas y dos consonantes frecuentes, esa cifra dice más de lo rara que es la secuencia **n·d** en turco que de la antigüedad del demostrativo.

k·m|WHO · binario · 30 lenguas · enriquecimiento 352×

Azerbaijani kim	BarabaTatar kim	Chuvash kam	Dolgan kim
KK–Halqabad 1 kim	Karaim kim	KazanTatar kem	Khalaj kim
Kumyk kim	Nogai kim	NorthAltai kem	OldTurkic kem
Salar kem	Tatar kem	Turkmen kim	Uzbek kim

m·n|I · binario · 28 lenguas · enriquecimiento 224×

Azerbaijani men	BarabaTatar min	CrimeanTatar men	KK–Halqabad 1 men
Karaim men	Kazakh men	Khakas min	Ki–Akmuz 1 men
Nogai men	North Altay men	Northern Uzbek men	Salar men
Tatar min	Tofa men	Tuvan men	Uzbek men

s·n|THOU · binario · 27 lenguas · enriquecimiento 86×

Azerbaijani sen	BarabaTatar sin	Gagauz sän	KarachayBalkar sen
Karaim sen	KazanTatar sin	Khalaj sen	Ki–Akmuz 1 sen
Nogai sen	NorthAltai sen	Northern Uzbek sen	Salar sen
Tatar sin	Tofa sen	Turkmen sen	Uyghur sän

l·r|THEY · binario · 20 lenguas · enriquecimiento 305×

Bashkir ular	CrimeanTatar olar	KK-Halqabad 1 olar	Karaim alar
KazanTatar alar	Khakas olar	Khalaj ollar	Ki-Akmuz 1 alar
Nogai olar	North Altay ular	NorthAltai olor	Northern Uzbek ular
Shor ilar	Turkmen olar	Tuvan olar	Uyghur ular

b·z|WE · binario · 20 lenguas · enriquecimiento 208×

Azerbaijani biz	BarabaTatar biz	CrimeanTatar biz	Gagauz biz
KarachayBalkar biz	Karaim biz	Kazakh biz	KazanTatar bez
Ki-Akmuz 1 biz	Kumyk biz	Nogai biz	Northern Uzbek biz
Tatar bez	Turkish biz	Turkmen biz	Uyghur biz

n·d|THERE · binario · 19 lenguas · enriquecimiento 562×

BarabaTatar anda	Bashkir anda	Chuvash anda	CrimeanTatar anda
KarachayBalkar anda	Karaim anda	KazanTatar anda	Khakas anda
Ki-Akmuz 1 anda	Kumyk anda	Nogai anda	NorthAltai anda
Shor anda	Tatar anda	Tofa ğnda	Tuvan ğnda

q·s·n|WHEN · ternario · 19 lenguas · enriquecimiento 489×

BarabaTatar qačan	Bashkir qasan	KK-Halqabad 1 qasan	KarachayBalkar qačan
Kazakh qačan	KazanTatar qačan	Ki-Akmuz 1 qačan	Kumyk qačan
Nogai qačan	North Altay qačan	NorthAltai qačan	Salar qačan
Shor qačan	Tuvan qačan	Uyghur qačan	Uzbek qačan

k·s|PERSON · binario · 18 lenguas · enriquecimiento 85×

BarabaTatar kiři	Bashkir keře	Karaim kiři	Kazakh kiři
KazanTatar keře	Ki-Akmuz 1 kiři	Nogai kiři	North Altay kiři
OldTurkic kiři	Salar kiři	Shor kiři	Tofa kiři
Turkish kiři	Tuvan kiři	Uyghur kiři	Uzbek kiři

s·z|YOU · binario · 17 lenguas · enriquecimiento 125×

Azerbaijani siz	CrimeanTatar siz	Gagauz siz	KK-Halqabad 1 siz
KarachayBalkar siz	Karaim siz	Kazakh səz	KazanTatar sez
Khalaj siz	Kumyk siz	Nogai siz	Northern Uzbek siz
OldTurkic siz	Tatar sez	Turkish siz	Turkmen siz

b·l|CHILD (DESCENDANT) · binario · 15 lenguas · enriquecimiento 71×

Bashkir bala	CrimeanTatar bala	KK-Halqabad 1 bala	KarachayBalkar bala
Karaim bala	Kazakh bala	KazanTatar bala	Khalaj bala
Ki-Akmuz 1 bala	Kumyk bala	Nogai bala	North Altay bala
NorthAltai bala	Uyghur bala	Uzbek bala	

m·n·d|HERE · ternario · 15 lenguas · enriquecimiento 919×

BarabaTatar mında	CrimeanTatar mında	Kazakh mında	KazanTatar monda
Khakas mında	Ki-Akmuz 1 mında	Kumyk munda	Nogai munda
North Altay munda	NorthAltai mında	Shor mında	Tatar monda
Tofa mında	Tuvan mında	WestYugur mında	

Naturaleza y cielo

k·n con «día» y con «sol» es colexificación, no colisión — y es la distinción que el índice inverso pide sostener. En turco *kün* significa las dos cosas; es una palabra usada para dos conceptos, no dos palabras confundidas por el instrumento. Que ambos entren en el catálogo por separado es correcto: son dos conceptos de Concepticon, y el catálogo indexa por concepto.

t·t·n | smoke (*tütün*) es de los pocos ternarios con consonante repetida del anexo, y **q·m** | sand (*qum*) de los binarios más cortos posibles. Entre uno y otro está casi todo el rango de aridez de la familia.

k·l | ash · binario · 29 lenguas · enriquecimiento 119×

Azerbaijani kyl	Bashkir köl	CrimeanTatar kül	Gagauz kyl
KarachayBalkar kyl	Kazakh kyl	KazanTatar köl	Khalaj kil
Kumyk kul	North Altay kyl	Northern Uzbek kul	OldTurkic kül
Salar kul	Tatar kæl	Turkmen kyl	Uzbek kul

t·s | stone · binario · 27 lenguas · enriquecimiento 101×

BarabaTatar taş	Bashkir taş	Dolgan ta:s	KK-Halqabad 1 tas
KarachayBalkar taş	Kazakh tas	Khakas tas	Khalaj ta:af
Kumyk taş	North Altay taf	NorthAltai taş	OldTurkic taş
Salar taş	Shor taş	Turkish taf	Uzbek tash

b·l·t | cloud · ternario · 25 lenguas · enriquecimiento 318×

Azerbaijani bulut	Bashkir bolot	Dolgan bilit	Gagauz bulut
KarachayBalkar bulut	Karaim bulut	KazanTatar bolit	Khalaj bulwt
Kumyk bulut	NorthAltai bulut	Northern Uzbek bulut	Sakha bilit
Tatar bolwt	Turkish bulut	Turkmen bulut	Uyghur bulut

k·n | day · binario · 25 lenguas · enriquecimiento 150×

BarabaTatar kOn	Bashkir kön	CrimeanTatar kün	Dolgan kyn
KarachayBalkar kyn	Karaim kyn, kin	KazanTatar kön	Khakas kyn
Ki-Akmuz 1 kyn	North Altay kyn	NorthAltai ? kyn	OldTurkic kün
Sakha kyn	Shor kyn	Tatar kæn	Uzbek kun

k·l | lake · binario · 23 lenguas · enriquecimiento 95×

BarabaTatar köl	Bashkir kyl	Chuvash kylə	KK-Halqabad 1 köl
KarachayBalkar köl	KazanTatar kül	Khakas köl	Kumyk köl
Nogai köl	North Altay köl	Northern Uzbek kol	OldTurkic köl
Shor köl	Tatar kæl	Uyghur köl	Uzbek qol

q·r | snow · binario · 22 lenguas · enriquecimiento 132×

BarabaTatar qar	Bashkir qar	CrimeanTatar qar	KarachayBalkar qar
Kazakh qar	KazanTatar qar	Ki-Akmuz 1 qar	Kumyk qar
North Altay qar	NorthAltai qar	Northern Uzbek qor	Salar qar
Shor qar	Tatar qar	Uyghur qar	Uzbek qar

t·m·n | fog · ternario · 21 lenguas · enriquecimiento 368×

BarabaTatar tOman	Bashkir toman	CrimeanTatar tuman	Dolgan tuman
Karaim tuman	Kazakh tuman	KazanTatar toman	Kumyk tuman
Nogai tuman	NorthAltai tuman	OldTurkic tuman	Sakha tuman
Salar tuman	Tofa tuman	Tuvan tuman	Uyghur tuman

t·s·k|HOLE · ternario · 18 lenguas · enriquecimiento 664×

BarabaTatar tēšik	Bashkir tišek	CrimeanTatar teşik	KK-Halqabad 1 tesik
KarachayBalkar teşik	Karaim teşik	Kazakh tesik	KazanTatar tišek
Kumyk teşik	Nogai tesik	North Altay tezic	NorthAltai težik
Northern Uzbek teʃik	Shor težik	Tatar tifek	Uyghur töšük

x·s|WOOD · binario · 18 lenguas · enriquecimiento 159×

Azerbaijani aḡač	BarabaTatar aḡač	Bashkir aḡas	CrimeanTatar aḡač
KK-Halqabad 1 aḡaš	KarachayBalkar aḡač	Karaim aḡac	Kazakh aḡaš
Khakas aḡas	Kumyk aḡač	Nogai aḡaš	North Altay aḡaʃ
NorthAltai aḡaš	OldTurkic iḡač	Salar aḡač	Shor aḡaš

t·n|night · binario · 18 lenguas · enriquecimiento 105×

BarabaTatar tön	Bashkir tön	Dolgan ty:n	KK-Halqabad 1 tyn
Kazakh tyn	KazanTatar tön	Khakas tyn	Ki-Akmuz 1 tyn
North Altay tyn	NorthAltai tyn	Northern Uzbek tun	OldTurkic tün
Sakha ty:n	Shor tyn	Tatar tən	Uyghur tyn

t·t·n|smoke · ternario · 17 lenguas · enriquecimiento 389×

BarabaTatar tötin	Bashkir tötön	CrimeanTatar tütün	Gagauz tytyn
KK-Halqabad 1 tytin	KarachayBalkar tytyn	Karaim tytyn, tytsy etc.	Kazakh tytyn
KazanTatar tötön	Khalaj tityn	Ki-Akmuz 1 tytyn	Kumyk tutun
Nogai tytin	Northern Uzbek tutun	OldTurkic tütün	Tatar tæten

k·n|sun · binario · 17 lenguas · enriquecimiento 102×

Azerbaijani kün	Dolgan kyn	KK-Halqabad 1 kyn	KarachayBalkar kyn
Kazakh kyn	Khakas kyn	Khalaj kyn, kin	Ki-Akmuz 1 kyn
Nogai kyn	North Altay kyn	NorthAltai kyn	OldTurkic kün
Sakha kyn	Salar kun	Shor kyn	Uyghur kyn

j·r|earth · binario · 16 lenguas · enriquecimiento 113×

Azerbaijani jer	BarabaTatar yer	Bashkir jer	CrimeanTatar yer
Gagauz jer	Karaim jer	Khalaj jer 'Ort; Erde'	Ki-Akmuz 1 jer
Kumyk jer	Nogai jer	OldTurkic yer	Turkish yer
Turkmen jer	Uyghur jer	Uzbek yer	WestYugur ? jer

Animales

b·t|louse (*bit*), q·s|bird (*quš*), q·r·t|worm (*qurt*) y b·l·q|fish (*balıq*): dos binarios y dos ternarios. q·r·t es el mejor ejemplo de ambigüedad del anexo y por eso está aquí: *kurt* es 'gusano' y 'lobo', y

el propio bloque lo delata — el nogayo escribe *qurt* (*jer qurtu*), «gusano de tierra», porque la palabra sola no basta.

El tamaño del bloque no dice nada del léxico animal: lo fija el corte de los cien de mayor alcance.

b·t|louse · binario · 22 lenguas · enriquecimiento 188×

Azerbaijani bit	Bashkir bet	CrimeanTatar bit	KarachayBalkar bit
Karaim bit	Kazakh bit	Khalaj bit	Ki-Akmuz 1 bit
Nogai biyt	NorthAltai biyt	Northern Uzbek bit	Sakha bīt
Tatar bet	Tofa bīʿt	Turkmen bit	Tuvan bīt

q·s|bird · binario · 20 lenguas · enriquecimiento 140×

BarabaTatar kuš	Bashkir qoʃ	CrimeanTatar kuš	KK-Halqabad 1 qus
KazanTatar koš	Ki-Akmuz 1 kuš	Kumyk quš	Nogai qus
NorthAltai quʃ	Northern Uzbek quʃ	OldTurkic kuš	Salar quʃ
Tatar qoʃ	Tuvan quʃ	Uyghur quʃ	Uzbek kuš

b·l·q|fish · ternario · 19 lenguas · enriquecimiento 374×

BarabaTatar balıq	Bashkir balıq	CrimeanTatar balıq	KK-Halqabad 1 balıq
Kazakh balıq	KazanTatar balıq	Kumyk baliq	Nogai balıq
North Altay baluq	NorthAltai balıq	Northern Uzbek baliq	Tofa balıq
Turkish balıq	Turkmen balıq	Tuvan balıq	Uyghur beliq

q·r·t|worm · ternario · 19 lenguas · enriquecimiento 434×

BarabaTatar k0rt	CrimeanTatar kurt	KK-Halqabad 1 qurt	KarachayBalkar qurt
Kazakh qurt	KazanTatar k0rt	Ki-Akmuz 1 qurt	Kumyk qurt
Nogai qurt (jer qurtw)	North Altay qurt	NorthAltai qurt 'worm'	OldTurkic kurt
Shor qurt	Tatar qort	Tofa kuʿrt	Tuvan qurt

Plantas y alimento

t·m·r|root (*tamir*) es un ternario limpio y de buen alcance. **t·z|salt** (*tuz*) es el caso opuesto: dos consonantes muy frecuentes, alcance alto y poca especificidad. La sección enseña bien el compromiso: más consonantes es más difícil de acertar por azar, pero también más difícil de conservar. Y **s·j·n|CHEW**, con 823×, es el enriquecimiento más alto del bloque — un ternario con tres consonantes poco frecuentes, que es exactamente la combinación que §3.3 predice.

t·z|salt · binario · 18 lenguas · enriquecimiento 133×

CrimeanTatar tuz	Gagauz tuz	KK-Halqabad 1 tuz	KarachayBalkar tuz
Karaim tuz	Kazakh tuz	KazanTatar toz	Khalaj tu(:)z
Kumyk tuz	Nogai tuz	Northern Uzbek tuz	OldTurkic tuz
Tatar toz	Turkish tuz	Uyghur tuz	Uzbek tuz

s·j·n|CHEW · ternario · 17 lenguas · enriquecimiento 823×

Azerbaijani čeynemag	Bashkir säynä	Gagauz țijne	KK-Halqabad 1 fajna
KarachayBalkar țajna	Karaim țajna	Kazakh fajna	KazanTatar čäynä
Khalaj țejne, țejle	Ki-Akmuz 1 țajna	Kumyk čaynamaq	Nogai fajna
North Altay fajna	NorthAltai țajna	Salar čäynä	Turkmen țejne

Verbos de acción y estado

Es la sección donde el despiece de la fase 0 se nota, y conviene decirlo aquí. Las fuentes nel e ids citan los verbos en infinitivo (*jaſamɒq*, *auuamakə*) y lexibank en raíz (*yaša*). Sin quitar ese –mAK, el mismo verbo salía con dos esqueletos y no formaba código: son 4.974 formas, el 9 % de la familia. Buena parte de los códigos de este bloque no existirían sin esa decisión, que está declarada en §3.2 y medida.

k·l con «venir» y con «reír» es el par que el índice inverso anuncia: *kel-* y *kül-*, dos verbos que solo la vocal separa. Aquí el catálogo agrupa de más, y no tiene con qué darse cuenta.

k·s | **CUT** · binario · 29 lenguas · enriquecimiento 132×

Azerbaijani kesmag	BarabaTatar kēs–	CrimeanTatar kes	KK–Halqabad 1 kes
Karaim kes	KazanTatar kis	Khakas kis	Ki–Akmuz 1 kes
Nogai kes	NorthAltai kes	OldTurkic kes	Salar kes
Tatar kisə	Turkish kes	Tuvan kes	Uzbek kes

t·t | **GRASP** · binario · 29 lenguas · enriquecimiento 287×

Azerbaijani tut	BarabaTatar t0t–	Chuvash tüt	Dolgan tut
KK–Halqabad 1 tut	Karaim tut	KazanTatar tot	Khalaj tut
Kumyk tut	North Altay tut	OldTurkic tut	Sakha tut
Shor tut	Turkish tut	Tuvan tut	Uzbek tut

k·r | **see** · binario · 27 lenguas · enriquecimiento 168×

BarabaTatar kör–	Bashkir kyr	CrimeanTatar kör	KK–Halqabad 1 kör
KarachayBalkar kör	Kazakh kör	Khakas kör	Khalaj ker
Nogai kör	NorthAltai kör	Northern Uzbek kormoq	Sakha kör
Shor kör	Tatar kürə	Tuvan kör	Uzbek kør

t·r | **stand** · binario · 27 lenguas · enriquecimiento 86×

BarabaTatar t0r	Bashkir tor	CrimeanTatar tur	KK–Halqabad 1 trø:
KarachayBalkar tur	Kazakh tru	Khakas tur	Khalaj tur
Kumyk tur	North Altay tur	NorthAltai tur	OldTurkic tur
Shor tur	Tatar toru	Tuvan tur	Uzbek tur

k·l | **come** · binario · 26 lenguas · enriquecimiento 104×

BarabaTatar kël	Bashkir kil	CrimeanTatar kel	Dolgan kel
KarachayBalkar kel	Kazakh kel	KazanTatar kil	Khalaj kël
Nogai kel	North Altay kel	Northern Uzbek këlmoq	OldTurkic kel
Shor kel	Tofa kel	Tuvan kel	Uzbek kel

b·l | **know** · binario · 26 lenguas · enriquecimiento 111×

Azerbaijani bil	Bashkir bel	Dolgan bil	Gagauz bil
KarachayBalkar bil	Kazakh blø	KazanTatar bel	Ki–Akmuz 1 bil
Nogai bil	NorthAltai bil	OldTurkic bil	Sakha bil
Tatar belə	Turkish bil	Turkmen bil	Uyghur bil

t·r·t | **pull** · ternario · 26 lenguas · enriquecimiento 196×

BarabaTatar tart-	Bashkir tart	CrimeanTatar tart	Dolgan tart
KarachayBalkar tart	Kazakh tart	KazanTatar tart	Ki-Akmuz 1 iteret
Nogai tart	North Altay tart	Northern Uzbek tortmoq	OldTurkic tart
Shor tart	Tofa ti ^r rt	Tuvan tirt	Uzbek tort

t·p|FIND · binario · 24 lenguas · enriquecimiento 294×

Azerbaijani tap	BarabaTatar tap-	Chuvash tup	CrimeanTatar tap
Karaim tap	Kazakh tap	Khakas tap	Ki-Akmuz 1 tap
Nogai tap	North Altay tap	Northern Uzbek topmoq	OldTurkic tap
Tofa ti ^r p_	Turkmen tap	Uyghur tap	Uzbek tɔp

t·k|sew · binario · 24 lenguas · enriquecimiento 329×

Azerbaijani tik	BarabaTatar tik-	CrimeanTatar tik	Dolgan tik
KarachayBalkar tik	Karaim tik	KazanTatar tek	Khakas tək
Ki-Akmuz 1 tik	Kumyk tik	North Altay tik	NorthAltai tik-
Sakha tik	Shor tik	Uyghur tik	Uzbek tik

k·l|laugh · binario · 23 lenguas · enriquecimiento 95×

BarabaTatar köl-	Bashkir köl	Chuvash kul	KK-Halqabad 1 kyl
KarachayBalkar kyl	Kazakh kyl	KazanTatar köl	Khalaj kil
Ki-Akmuz 1 kyl	Kumyk kulemaq	North Altay kyl	NorthAltai kül-
Sakha kyl	Salar kulli	Uyghur kyl	Uzbek kul

j·n|play · binario · 23 lenguas · enriquecimiento 145×

Azerbaijani ojna	Bashkir ujna	Dolgan o:jn(')o:	KK-Halqabad 1 ojna
KarachayBalkar ojna	Kazakh ojna	Khakas ojna	Kumyk ojna
Nogai ojna	North Altay ojna	Northern Uzbek ojnamoq	Salar ojna
Tatar ujnau	Turkish ojna	Tuvan ojna	Uyghur ojna

q·l|REMAIN · binario · 22 lenguas · enriquecimiento 157×

BarabaTatar qal-	Bashkir qal	CrimeanTatar qal	KarachayBalkar qal
Kazakh qal	KazanTatar qal	Ki-Akmuz 1 qal	Kumyk qal
North Altay qal	NorthAltai qal-	Northern Uzbek qolmoq	Salar qal
Shor qal	Tatar qalu	Uyghur qal	Uzbek qol

b·r|give · binario · 22 lenguas · enriquecimiento 86×

BarabaTatar ber-	Bashkir bir	CrimeanTatar ber	KK-Halqabad 1 ber
KarachayBalkar ber	Karaim ber	KazanTatar bir	Ki-Akmuz 1 ber
Nogai ber	NorthAltai ber	Northern Uzbek bermoq	Sakha bier
Tatar birə	Tofa ber	Tuvan ber	Uyghur bär

b·r|go · binario · 22 lenguas · enriquecimiento 86×

BarabaTatar bar	Bashkir bar	CrimeanTatar bar	KK-Halqabad 1 ? bar
KarachayBalkar bar	Karaim bar	KazanTatar bar	Ki-Akmuz 1 bar
Nogai bar	NorthAltai bar	Northern Uzbek bormoq	Sakha bar
Salar bar	Tofa bar	Tuvan bar	Uyghur bar

s·t|hear · binario · 21 lenguas · enriquecimiento 85×

Azerbaijani ešit	BarabaTatar išet-	Bashkir išet	CrimeanTatar ešit
KK-Halqabad 1 esit	KarachayBalkar ešit	Karaim esit	KazanTatar išet
Khalaj ifyt	Kumyk ešit	Northern Uzbek æfitmoq	OldTurkic ešit
Salar išti	Turkish ifit	Turkmen ešit	Uyghur išet

s·n|count · binario · 20 lenguas · enriquecimiento 66×

KK-Halqabad 1 sana	KarachayBalkar sana	Kazakh sana	KazanTatar sana
Khalaj sa:ana	Ki-Akmuz 1 sana	Kumyk sana	Nogai sana
NorthAltai sana-	OldTurkic sana	Salar sana'	Tatar sanau
Turkmen sana	Tuvan sana	Uyghur sana	Uzbek sana

q·r·q|fear · ternario · 20 lenguas · enriquecimiento 373×

BarabaTatar qorq-	Bashkir qurq	CrimeanTatar qorq	KK-Halqabad 1 qorq
Karaim qorq	Kazakh qorq	KazanTatar qurqu	Khalaj qorq
Kumyk qorq	Nogai qorq	North Altay qoruq	NorthAltai qoruq-
Shor qoruq	Tatar qurqu	Uyghur qorq	Uzbek qorq

t·k·r|SPIT · ternario · 19 lenguas · enriquecimiento 390×

BarabaTatar tökür-	Bashkir tökör	CrimeanTatar tükür	Gagauz tykyr
KK-Halqabad 1 tkörö	KarachayBalkar tykyr	Kazakh tköru	KazanTatar töker
Khakas tykyr	Ki-Akmuz 1 tykyr	Kumyk tukurmaq	North Altay tykyr
NorthAltai tykyr	Shor tykkyr	Tofa tükür	Turkish tykyr

j·s|BE ALIVE · binario · 18 lenguas · enriquecimiento 174×

Azerbaijani jafa	BarabaTatar yešä-	Bashkir jəfəy	CrimeanTatar yaša
Gagauz jafa	KazanTatar yäšä	Ki-Akmuz 1 jafɔ:	Kumyk jafa
Northern Uzbek jafamoq	OldTurkic yaša	Salar jafa	Tatar jafæ
Turkish jafamak	Turkmen jafa	Uyghur jafa	Uzbek yaša

s·r|SUCK · binario · 18 lenguas · enriquecimiento 60×

Azerbaijani sor	KK-Halqabad 1 sor	Kazakh sor	KazanTatar ?_suir
Khakas sor	Khalaj suor	Ki-Akmuz 1 sor	Kumyk sor
North Altay sor	NorthAltai soor	OldTurkic sor	Shor sor
Tofa sor	Turkmen ? sor	Tuvan sor	Uyghur šora

q·z|dig · binario · 18 lenguas · enriquecimiento 251×

CrimeanTatar qaz	KK-Halqabad 1 qaz	KarachayBalkar qaz	Kazakh qaz
KazanTatar qaz	Khalaj qaz	Ki-Akmuz 1 qaz	Kumyk qaz
Northern Uzbek qazimoq	OldTurkic qaz	Salar qazw	Tatar qazu
Tofa qaz	Tuvan qaz	Uyghur qaz	Uzbek qazi

j·t|say · binario · 18 lenguas · enriquecimiento 143×

BarabaTatar äyt	Bashkir äyt	KK-Halqabad 1 ajt	KarachayBalkar ajt
Karaim ajt	Kazakh ajt	KazanTatar äyt	Ki-Akmuz 1 ajt
Nogai ajt	North Altay ajt	NorthAltai ajt	Northern Uzbek ajtmoq
Shor ajt	Tatar æjtə	Turkmen ajt	Uyghur ejt

j·t|LIE DOWN · binario · 17 lenguas · enriquecimiento 144×

Azerbaijani yatmag	BarabaTatar yat-	Bashkir jat	CrimeanTatar yat
Gagauz jat	Karaim jat	KazanTatar yat	Khalaj ja:t
Kumyk jat	Nogai jat	OldTurkic yat	Salar jat
Turkish yat	Turkmen jat	Uyghur jat	Uzbek yot

s·s|swell · binario · 17 lenguas · enriquecimiento 94×

Azerbaijani šiš	Bashkir šeš	Chuvash šiś	CrimeanTatar šiš
Gagauz šiš	Karaim sis	KazanTatar šeš	Khakas səs
Khalaj ʃwʃ	Ki-Akmuz 1 šiš	Kumyk šiš	Nogai sis
OldTurkic siš	Shor šiš	Turkish šiš	Turkmen čiš

q·s|vomit · binario · 17 lenguas · enriquecimiento 142×

CrimeanTatar qus	KK-Halqabad 1 qus	KarachayBalkar qus	Kazakh qüş
KazanTatar qos	Khalaj qus	Ki-Akmuz 1 qus	Kumyk qus
Nogai qus	North Altay qus	NorthAltai qus	OldTurkic qus
Tofa quʻs	Tuvan qus	Uyghur qus	Uzbek qus

t·r|sit · binario · 16 lenguas · enriquecimiento 51×

Azerbaijani otur	BarabaTatar otir-	CrimeanTatar otur-	Gagauz otur
KK-Halqabad 1 otir-	Kazakh otir-	KazanTatar utir-	Ki-Akmuz 1 otur-
North Altay otur	NorthAltai otur-	Northern Uzbek otirmoq	Salar otir-
Tatar utxru	Turkish otur	Turkmen otur	Uzbek otir-

s·r|ASK (REQUEST) · binario · 15 lenguas · enriquecimiento 47×

BarabaTatar sora-	CrimeanTatar sora	KK-Halqabad 1 sora	Kazakh süra
KazanTatar sora	Khakas sur	Ki-Akmuz 1 sura	Nogai sora
North Altay sura	NorthAltai sura	Shor sura	Turkish sor
Turkmen sora	Uyghur sora	Uzbek sora	

j·t|SWALLOW · binario · 15 lenguas · enriquecimiento 123×

BarabaTatar yut-	Bashkir jot	CrimeanTatar yut	Gagauz jut
Karaim jut	KazanTatar yot	Khalaj ju:ut	Kumyk jut
Nogai jut	North Altay jyt	NorthAltai yüt-	Northern Uzbek jutmoq
Tatar jotu	Turkish jutmak	Uzbek yut	

j·r|WALK · binario · 15 lenguas · enriquecimiento 106×

Azerbaijani jeri	BarabaTatar yör-	Bashkir jør(ø)	CrimeanTatar yür
Karaim jyry	KazanTatar yör	Khalaj jorw	Kumyk jyry
Nogai jyr	Northern Uzbek jurmoq	OldTurkic yorī	Salar jyr, jer
Turkish jyrımek	Turkmen jøre	Uzbek yur	

Objetos y cultura

Un solo código, **t·j·q|stick** (*tayaq*), y conviene no leer en esa escasez más de lo que hay: el tamaño del bloque lo fija el corte de los cien de mayor alcance, no una propiedad del léxico material. Lo que sí se puede decir es que *tayaq* es transparentemente derivado —*taya-* «apuntalar» más *-q*—, como **t·s·k|HOLE** (*teš-ik*) y **q·n·t|wing** (*kanat*). En turco, la sospecha por aridez que §2.4 declara para cuatro consonantes empieza en tres.

t·j·q|stick · ternario · 18 lenguas · enriquecimiento 514×

BarabaTatar tayaq	Bashkir tajaq	CrimeanTatar tayaq	KK-Halqabad 1 tajaq
KarachayBalkar tajaq	Kazakh tajaq	KazanTatar tayaq	Ki-Akmuz 1 tajaq
Nogai tajaq	North Altay tajaq	NorthAltai tajaq	Northern Uzbek tajoq
Shor ? tajaq	Tofa tayaq	Turkmen tajaq	Uyghur tajaq

Cualidades

q·r|black (*qara*), **s·r|yellow** (*sari*), **q·z·l|red** (*qizil*), **j·s·l|green** (*jašil*). Los colores básicos aparecen casi completos, y los dos últimos con el mismo sufijo *-l* visible en el esqueleto: el catálogo lo pone contiguo sin saber qué es, igual que con el *-z* de los pronombres.

t·l|full (*tolı*) comparte código con **t·l|tongue** (*til*). Es la cuarta vez que esta advertencia aparece en el anexo, y esa insistencia es el punto: en una familia con el 59 % de colisión, dos consonantes no bastan para separar dos palabras, y el lector tiene que llevarlo puesto en cada bloque.

z·n|long · binario · 28 lenguas · enriquecimiento 471×

Azerbaijani uzun	BarabaTatar uzın	Gagauz uzun	KarachayBalkar uzun
Kazakh üzın	KazanTatar ozın	Khalaj uzu:n	Kumyk uzun
North Altay uzun	NorthAltai uzun	OldTurkic uzun	Shor uzun
Tofa uzun	Turkish uzun	Tuvan uzun	Uzbek uzun

q·r|black · binario · 23 lenguas · enriquecimiento 138×

BarabaTatar qara	Bashkir qara	CrimeanTatar qara	KarachayBalkar qara
Kazakh qara	Khalaj qara	Ki-Akmuz 1 qara	Nogai qara
North Altay qara	NorthAltai qara	OldTurkic qara	Salar qara
Tatar qara	Tofa qara	Uyghur qara	Uzbek qora

s·k|old · binario · 22 lenguas · enriquecimiento 132×

Azerbaijani əski	BarabaTatar ěski	CrimeanTatar eski	KK-Halqabad 1 eski
KarachayBalkar eski	Karaım eski	KazanTatar iske	Ki-Akmuz 1 eski
Nogai eski	North Altay eski	NorthAltai eski	OldTurkic eski
Salar eski	Shor eski	Turkish eski	Uyghur äski

s·l|left · binario · 20 lenguas · enriquecimiento 70×

Azerbaijani sol	BarabaTatar sol	CrimeanTatar sol	KK-Halqabad 1 sol
Kazakh sol	KazanTatar sul	Khakas sol	Ki-Akmuz 1 sol
Nogai sol	North Altay sol	NorthAltai sol	Northern Uzbek sol
Shor sol	Tatar sul	Turkish sol	Uyghur sol

t·r|narrow · binario · 20 lenguas · enriquecimiento 68×

Bashkir tar	CrimeanTatar tar	KK–Halqabad 1 tar	KarachayBalkar tar
Kazakh tar	KazanTatar tar	Khakas tar	Khalaj ta:ar
Kumyk tar	Nogai tar	NorthAltai tar	OldTurkic tar
Tofa tar	Tuvan tar	Uyghur tar	Uzbek tər

q·z·l|red · ternario · 20 lenguas · enriquecimiento 418×

BarabaTatar qizil	KK–Halqabad 1 qizil	KarachayBalkar qwzwl	Kazakh qəzəl
Khalaj qwzwl	Ki–Akmuz 1 qizil	Kumyk qizil	Nogai qwzwl
NorthAltai qwzwl	Northern Uzbek qizil	OldTurkic qizil	Shor qwzwl
Tofa qizil	Tuvan qwzwl	Uyghur qizil	Uzbek qizil

k·n|wide · binario · 20 lenguas · enriquecimiento 124×

BarabaTatar kəj	Bashkir kiŋ	CrimeanTatar keŋ	KK–Halqabad 1 keŋ
Karaim ken	Kazakh keŋ	KazanTatar kiŋ	Khalaj ki:en
Nogai keŋ	North Altay keŋ	NorthAltai keŋ	Northern Uzbek keŋ
Sakha kien	Tatar kiŋ	Uyghur kəj	Uzbek keŋ

t·r·n|DEEP · ternario · 19 lenguas · enriquecimiento 477×

BarabaTatar tärəŋ	Bashkir tärän	Chuvash tarən	CrimeanTatar teren
KK–Halqabad 1 tereŋ	KarachayBalkar teren	Kazakh tereŋ	KazanTatar tirän
Khakas tireŋ	Ki–Akmuz 1 tereŋ	Kumyk teran	North Altay tereŋ
NorthAltai tereŋ	OldTurkic teriŋ	Salar tiroŋ	Shor tereŋ

k·p|MANY · binario · 19 lenguas · enriquecimiento 438×

BarabaTatar köp	Bashkir kyp	KK–Halqabad 1 köp	Karaim köp
Kazakh köp	KazanTatar küp	Ki–Akmuz 1 köp	Kumyk kop
Nogai köp	North Altay көp	NorthAltai köp	Shor köp
Tofa köp	Turkmen köp	Uyghur köp	Uzbek köp

t·l|full · binario · 19 lenguas · enriquecimiento 114×

BarabaTatar tol–	Bashkir tulı	Chuvash tuli	CrimeanTatar tolu
KK–Halqabad 1 tolɜ	KarachayBalkar tolu	Kazakh tolı	KazanTatar tulı
Khalaj tu(:)ola	Kumyk tolu	Nogai tolı	Northern Uzbek tola
OldTurkic tolu	Salar tolu	Tatar tulɣ	Uzbek tōla

s·r|yellow · binario · 19 lenguas · enriquecimiento 55×

Azerbaijani saru	BarabaTatar sarı	Chuvash 'sarə	CrimeanTatar sarı
Gagauz sarı	KK–Halqabad 1 sarı	Karaim sarı	Kazakh sarı
KazanTatar sarı	Ki–Akmuz 1 sarı	Kumyk sarı	North Altay sarw
NorthAltai sarı	Salar sarı	Tatar sarɣ	Turkish sarw

s·n|TRUE · binario · 18 lenguas · enriquecimiento 57×

BarabaTatar čin	Chuvash t̪çən	KK-Halqabad 1 ? ʃwn	Kazakh šin
KazanTatar čin	Khakas s̪in	Ki-Akmuz 1 čin	Nogai ušin
NorthAltai čin	OldTurkic čin	Salar čen	Shor šin
Tofa šin	Turkmen čin	Tuvan šin	Uyghur čin

j·s·l|green · ternario · 18 lenguas · enriquecimiento 523×

Azerbaijani jaʃwl	BarabaTatar yišil	Bashkir jæfel	CrimeanTatar yešil
Gagauz jefil	Karaim jesil	KazanTatar yäšel	Kumyk jaʃwl
Northern Uzbek jaʃil	OldTurkic yašil	Salar ja:ʃwl	Tatar jafel
Turkish jefil	Turkmen jaʃwl	Uyghur jefil	Uzbek yašil

s·r|rotten · binario · 18 lenguas · enriquecimiento 58×

Azerbaijani čürü	Bashkir ser	Chuvash śer	CrimeanTatar čürü
Gagauz čürü	KK-Halqabad 1 šir	KarachayBalkar čir	Karaim ciri
KazanTatar čer	Khalaj ʈirry	Ki-Akmuz 1 čir	Kumyk čiri
Nogai širi	NorthAltai čiri	Salar čür	Turkish čürü

q·l·n|thick · ternario · 18 lenguas · enriquecimiento 844×

Bashkir qalwn	CrimeanTatar ʔal̪in	KK-Halqabad 1 qalən	KarachayBalkar qalwn (2D)
Kazakh qalən	KazanTatar ʔal̪in	Ki-Akmuz 1 qalwn	Kumyk qalwn
NorthAltai qalwn	Northern Uzbek qalin	OldTurkic ʔal̪in	Shor qalwn
Tatar qal̪n	Tuvan qalwn (2D)	Uyghur qelin	Uzbek ʔalin

q·s·q|short · ternario · 17 lenguas · enriquecimiento 396×

Bashkir qwsqa	CrimeanTatar ʔis̪ka	KK-Halqabad 1 qisqa	Kazakh qsqa
KazanTatar ʔis̪ka	Ki-Akmuz 1 qisqa	Kumyk qisqa	Nogai qwsqa
North Altay qwsqa	NorthAltai ʔis̪ka	Northern Uzbek qisqa	Shor qwsqa
Tatar q̪s̪qa	Tofa ʔi̪ʃ̪ka	Tuvan qwsqa	Uyghur qisqa

t·p|BOTTOM · binario · 16 lenguas · enriquecimiento 216×

BarabaTatar tüp	Bashkir töp	Chuvash töp	CrimeanTatar tüp
KK-Halqabad 1 typ	Karaim typ, kip etc.	Kazakh typ	KazanTatar töp
Khakas typ	Ki-Akmuz 1 typ	Kumyk typ	Nogai typ
North Altay typ	NorthAltai typ	Uyghur typ	WestYugur t̪ip

Other

s·k·r|JUMP · ternario · 15 lenguas · enriquecimiento 408×

CrimeanTatar sekir	KK-Halqabad 1 sekir	KarachayBalkar sekir	Karaim sekir
Kazakh sek̪ir	KazanTatar siker	Ki-Akmuz 1 sekir	North Altay sekir
NorthAltai sekir	Northern Uzbek sakramoq	OldTurkic sekri	Salar seker
Tatar siker̪	Uyghur ? sekre	Uzbek sakra	

k·j|PUT ON · binario · 15 lenguas · enriquecimiento 326×

Azerbaijani kej	BarabaTatar käy-	Bashkir kej	KK-Halqabad 1 kij
KarachayBalkar kij	Karaim kij	Kazakh kij	Ki-Akmuz 1 kij
Nogai kij	North Altay kij	NorthAltai kij	Northern Uzbek kijmoq
Salar kij	Tatar kij̪	Uyghur kij	