The Repertoire of Nurture

Four consonant codes, the clusters they form and the ones they do not: NANA · MAMA · PAPA · TATA across 5,501 languages

Alejandro Toledo Martínez

Scope note. This work measures a phenomenon at the level of acquisition and articulation —the *nursery forms*—, not a deep genealogical signature. It was carried out, among other reasons, as a demonstration that the instrument recovers what is already known before being asked for anything new. Three results are offered here as this paper's own: the quantified catalogue of clusters, the negative result concerning the infant, and the finding that no comparative corpus of the affective register exists for the vast majority of the languages treated here.

Abstract

Vowels are stripped from 100,507 lexical forms in Lexibank (5,501 languages, 298 families) and each consonant is replaced by its Dolgopolsky sound class. On that skeleton we ask whether the concepts of the nurturing core —mother, father, baby, grandmother, grandfather— draw on a repertoire of four codes: `N` (nana, ana), `M` (mama, ama), `P` (papa, apa, baba) and `T` (tata, ata).

The answer is affirmative and by a wide margin. Among words built from a single consonant class, the quartet takes 80.9% [79.8, 82.0] in the nurturing block against 48.0% [47.3, 48.8] in ordinary vocabulary, where chance across ten classes would give 40%; the gap is 32.9 points [31.5, 34.1], Cohen's $h = 0.71$, and it recurs within 40 of 45 families compared against themselves (paired Wilcoxon $p = 8.4{\cdot}10^{-10}$, Cliff's $\delta = 0.78$). The quartet does not merely rise: it displaces. The velar class `K`, the most frequent single-class code in ordinary vocabulary (18.1%), falls to 6.7%.

The effect-size breakdown forces a correction to the headline, and we give it in the abstract rather than burying it. The quartet does not rise as a block. `M`, `N` and `P` go from 30.9% to 63.0% and carry almost all of the effect; `T` does not move at all (17.2% $\rightarrow$ 17.9%, $h = 0.02$). `T` is not a class the nurturing field prefers — it is a class the field does not expel, while it does expel `K`, `S` and `R`. What has to be explained is an avoidance, not a preference.

The result does not depend on the sound-class table. Repeated under seven readings — Dolgopolsky as Lexibank ships it, Dolgopolsky recomputed with LingPy, SCA in two readings, ASJP, articulatory features from PanPhon, and raw IPA with no merging at all (201 distinct consonants, where chance gives the quartet 8%) — the nurture-minus-control gap stays between +31 and +34 points. Nor does it depend on the large families: dropping Indo-European and Austronesian entirely leaves +30.7, and drawing one language per family so that every family weighs the same leaves +31.8 [26.1, 37.0], positive in 100% of 500 draws.

The repertoire is shared, not partitioned: all four codes appear across all five concepts, with differing leanings —the feminine toward nasals, the masculine toward stops— and profile overlaps of up to 0.92. Each cluster is catalogued with its size, its familial spread and its geographic footprint, and a single figure —the number of families it touches— separates the repertoire clusters (50–60 families) from clusters of branch inheritance (4–7 families).

Two results are negative and are reported as such. First, reading the four codes as two crossed axes (nasal/stop $\times$ labial/dental) yields a global odds ratio of 2.01 [1.52, 2.67] that collapses to 1.01 [0.65, 1.56] under stratification by family. Second, and against expectation, there is no cluster for the infant: with eleven concepts, 4,563 languages, the language as unit of count and attempts equalised, the infant falls ten points below its own ordinary vocabulary (28.6% against 38.7%, $h = -0.21$), in 24 of 27 families, and with no more reduplication than the control. The reason is structural and legible in code length: languages do not name the infant with a primitive word, they derive it.

One effect does survive, however: when the infant does use the repertoire, it tends to take the code of the parents of its own language (18.7% against 14.5% expected, p = 0.0005, $h = 0.11$ — real and small). The code does not belong to the infant; it is on loan.

Finally we document a corpus gap that conditions everything above. Lexibank records the term of reference, not the term of address. Of its 1,740,092 records, nine carry a vocative or address marking, and those nine indicate that the vocative is a different word and closer to the repertoire. The figures in this work are therefore a floor, not a ceiling.

1. The problem, and what was already known

That the word for 'mother' should resemble itself across unrelated languages is one of the oldest observations in comparative linguistics, and also one of the best explained. Murdock (1959) compiled the parallels among parental terms across hundreds of languages; Jakobson (1960) supplied the explanation that still stands: the labial nasal is the first thing an infant produces with its mouth occupied at the breast, and the adult picks up that babble and returns it as a name. Ferguson (1964) described the whole register —*baby talk*— as a subsystem with its own phonology, morphology, and a specialised lexicon of twenty-five to sixty items.

What is added here is not the explanation, which already exists, but three things that had not been measured.

First: the size and shape of the clusters. Saying "they resemble each other in many languages" is not the same as saying "this code groups 639 languages from 53 families and that one 136 languages from 7." The second formulation distinguishes a repertoire from an inheritance, and that distinction is made with a number.

Second: what happens with the infant. The literature describes the register *directed at* the child; there is no record of anyone measuring whether the word *for* the child participates in the same repertoire. Here it is measured, and it does not.

Third: where the corpus runs out. The affective register —what the mother says to the child in her arms— is absent from comparative databases, and that is worth stating with figures.

A methodological warning runs through the whole work. A phenomenon as conspicuous as *mama / papa* invites apophenia: there will always be languages that fit. Every claim is therefore tested against a comparison block fixed before computing, and those that depend on family size are redone taking the family —not the language— as the unit.

1.1 Three words, defined before they are used

Three terms carry the argument and each is used in more than one sense in the surrounding literature. They are fixed here, once, and not stretched afterwards.

Code. A purely operational object: the string that remains when the vowels are removed from a form and each consonant is replaced by its sound class. `mama` $\rightarrow$ `MM`, `ama` $\rightarrow$ `M`, `mater` $\rightarrow$ `MTR`. A code is a representation of a word, nothing more. It carries no claim about meaning, about history, or about

the speaker's mind.

Cluster. A set of languages that give the same code to the same being. Nothing else. It is a *statistical* object, defined by extension: `N`·feminine is the 639 languages, from 53 families, whose word for mother or grandmother has code `N`. A cluster is not a historical family, not a proto-form, and not a rule — the languages inside it may be unrelated, and the languages outside it are not exceptions to anything. Throughout this paper, whenever a cluster is reported, two numbers accompany it: how many languages, and how many independent families. The second number is what tells a cluster of convergence apart from a cluster of inheritance, and §6.1 makes that separation explicit.

Repertoire. The claim proper, and the only one of the three that could be false. A repertoire is a small inventory of codes that (i) a large number of mutually unrelated languages draws on for the same small set of concepts, (ii) is *shared* across those concepts rather than assigned one code per concept, and (iii) is *closed* — it not only over-uses its members, it under-uses the alternatives. Each of the three parts is measured separately below: (i) in §4.3, by family; (ii) in §4.4, by profile overlap; (iii) in §4.1, by what happens to `K` and `S`. A finding that nurturing concepts merely used more nasals than average would not establish a repertoire in this sense; a finding that each concept had its own private code would falsify part (ii).

1.2 What level of explanation this is

The reader is owed a plain statement of what kind of claim is being made, because the same data could be dressed as four different theories.

This is a claim about articulation and acquisition, read off the typological record. The mechanism is Jakobson's and is not in dispute: the infant's earliest closures produce a narrow set of consonants, the caregiver hears them as address, and the resulting form fossilises in the lexicon. Nothing here proposes a new mechanism. What is new is measurement — how large the resulting convergence is, which classes it recruits and which it refuses, and whether it is one inventory or five.

It is explicitly not a claim about deep genealogy. No cluster reported here is offered as evidence of common descent, and §6.1 exists precisely to strip out the clusters that *are* inherited. It is also not a claim about semantics: the codes do not mean anything, and no interpretation of `M` as "maternal" or `T` as "paternal" is intended or supported.

Where the phenomenon sits between phonetics and pragmatics is a genuine open question, and §9 argues that it cannot be settled with the corpora that currently exist, because the register in which the phenomenon actually lives — address, not reference — is the one no database records.

2. Data and method

2.1 The corpus

Lexibank (List *et al.* 2022) assembles standardised word lists for more than five thousand languages with uniform phonetic segmentation and links to Concepticon. The complete `forms.csv` was read — 1,740,092 records, of which 1,729,060 carry phonetic material— and 100,507 forms across 33 concepts in 5,501 languages and 298 families were extracted.

A note on the access path, since it bears on reproducibility. The derived tables this project maintains discard forms with fewer than two distinct classes, which removes precisely `mama`, `ama`, `nana` and `papa`. For the concept MOTHER that means 1,530 forms and 96 families in the derived table, against 4,420 forms and 189 families in the raw file. Everything that follows is computed on the raw file.

2.2 The code

Vowels are removed from each form and every consonant is replaced by its Dolgopolsky (1964) sound class, a grouping of sounds by probability of phonetic change that Lexibank already supplies, computed via LingPy:

class	sounds	class	sounds
P	p, b, f	M	m
T	t, d, θ	N	n, ŋ, ɲ
S	s, z, ʃ	R	l, r
K	k, g, x, q	W	w, v
H	h, ʔ	J	j

Thus mama → MM, ama → M, nene → NN, ana → N, mater → MTR.

Two assignments deserve warning because they are not what articulatory intuition would suggest: Dolgopolsky's criterion is not place of articulation but historical stability. The fricative v does not join f but goes to W, alongside w; and the affricates do not join the sibilants: ts, dz and tɕ fall in K, while tʃ and dʒ fall in T. The assignment was verified to be unambiguous in the corpus —each segment to a single class in 100% of its occurrences— and the affricates affect 3.69% of the forms for the five beings against 5.63% of the control forms, so the decision does not favour the hypothesis.

A published and external inventory is used deliberately rather than one of our own. The inventory is frozen before the data are examined and cannot be tuned to what one hopes to see, which is the cheapest available prosthesis against compositional apophenia. That defence is not sufficient by itself, however, since a published inventory can still be the wrong one; §7 therefore repeats the central test under five further encodings, including one with no merging whatsoever.

Where a coarser grain is needed we use the signature {…}: the set of classes present, without order or repetition. M, MM and MMM share the signature {M}.

2.3 The comparison blocks

Fixed before computing:

- Nurture (5 concepts): MOTHER, FATHER, BABY, GRANDMOTHER, GRANDFATHER.
- Other kinship (7): CHILD, BROTHER, SISTER, SON, DAUGHTER, UNCLE, AUNT. If the pattern belongs to nurture and not to kinship generally, it should drop here.
- Ordinary vocabulary (14): WATER, STONE, FIRE, SUN, EYE, HAND, TREE, TOOTH, BLOOD, NAME, MOON, DOG, HEAD, BONE. If the pattern also appeared here, it would be an artefact of the method.
- The infant field (11), for section 8: the above plus BOY, GIRL, GRANDCHILD, CHILD (DESCENDANT), DESCENDANTS, YOUNG MAN, YOUNG WOMAN.

2.4 Guardrails

Four precautions, all necessary and one of them decisive.

One row per language. Languages with several forms for the same concept are counted once.

The family as unit. Austronesian contributes 956 languages to the corpus and Indo-European 139. Counting languages over-represents the large families, so every strong claim is redone by family.

Length control. A short word is more likely to come out single-class. Wherever the comparison could be contaminated by this, the null is restricted to forms with the same number of consonants, or the measure is conditioned on those already single-class.

Stratification. An aggregate association can exist without existing within any stratum. This is checked with Mantel–Haenszel (1959) whenever the claim concerns pairs within a single language. In section 5 this precaution demolishes a result that would otherwise have passed.

2.5 Uncertainty and magnitude

Every headline proportion in this paper is reported with a 95% Wilson (1927) interval, and every contrast between two proportions with a Newcombe (1998) interval on the difference and with Cohen's (1988) h as effect size, on the usual reading of 0.2 as small, 0.5 as medium and 0.8 as large. Contingency tables carry Cramér's (1946) V. The paired comparison across families carries Cliff's (1993) δ, the matched-pairs rank-biserial correlation, and a bootstrap interval over families. Odds ratios carry Woolf intervals; stratified odds ratios carry Robins–Breslow–Greenland (1986) intervals and Tarone's (1985) homogeneity test. Full tables are in `out/effects.md`.

3. One concept, the whole world

Before the full repertoire, the simplest case: a single concept.

MOTHER across 4,420 forms, 3,787 languages and 189 families. With the vowels gone, the world distributes across 471 distinct codes, but two lumps hold most of it.

The signature {N} gathers 794 languages from 65 families and {M} 624 from 67. Together, 1,418 languages —37.4% of the corpus— across 106 independent families. The third signature, {NT}, already falls to 182.

That they are nasal is not an effect of nasals being common. Against the entire Lexibank lexicon, and comparing only against words with the same number of consonants:

code	% in 'mother'	% in the lexicon	enrichment	95% CI	at equal length
N	14.5%	1.57%	×9.2	[8.6, 10.0]	×3.3
M	11.2%	1.03%	×10.8	[9.9, 11.9]	×3.9
NN	6.1%	0.61%	×9.9	[8.8, 11.3]	×8.3
MM	5.3%	0.20%	×26.1	[22.8, 30.0]	×21.8
T	3.1%	2.08%	×1.5	[1.3, 1.8]	×0.5
P	2.1%	1.72%	×1.2	[1.0, 1.5]	×0.4
K	1.9%	2.84%	×0.7	[0.5, 0.9]	×0.2

The stops are not merely un-enriched: they fall below expectation, and K's interval excludes 1 from above. 'Mother' is not an ordinary word with extra nasals; it is a word from which the stops withdraw.

The internal control is as clean as possible: the word next door, in the same languages.

In those 3,656 languages the signature {N} is 20.4 times more frequent in 'mother' than in 'father' [14.8, 28.2]; {P} is 6.9 times more frequent in 'father' [5.7, 8.3] and {T} 4.2 times [3.5, 5.0].

Each family chooses, and chooses differently: Turkic goes 83% to {N} (*ana*, with a normalised entropy of 0.18, almost monolithic), Tungusic 84%, Dogon 75%; Sino-Tibetan goes 59% to {M} (*ama*); Indo-European, by contrast, has MTR as its most frequent code —*mater*, the compound and not the bare nasal— at only 29%; and Pama-Nyungan has no code above 6%.

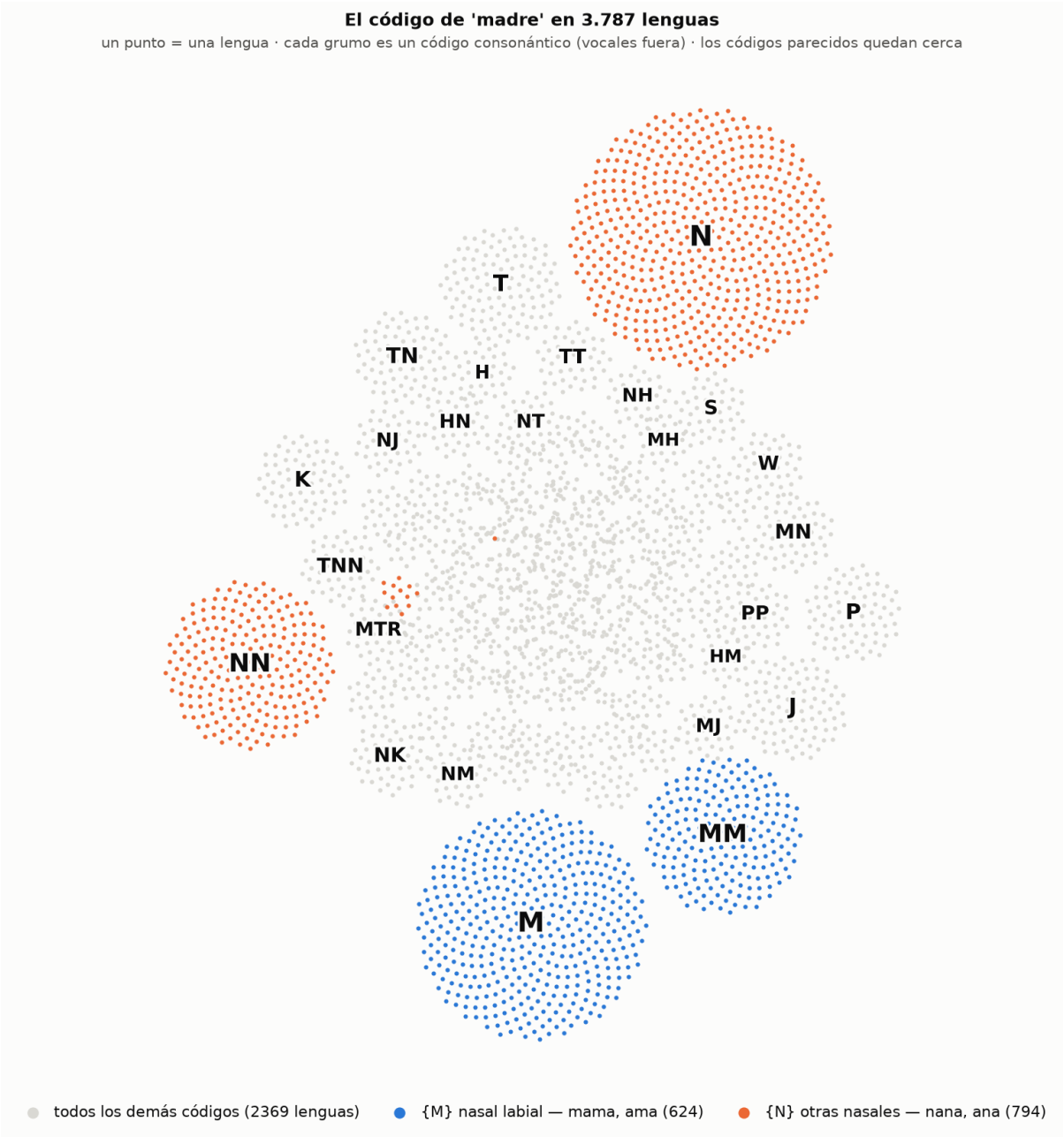

Figure 1: The cloud of codes for 'mother'. One dot is one language; each lump is a consonant code; similar codes lie close together (edit distance and multidimensional scaling). The four largest lumps are nasal.

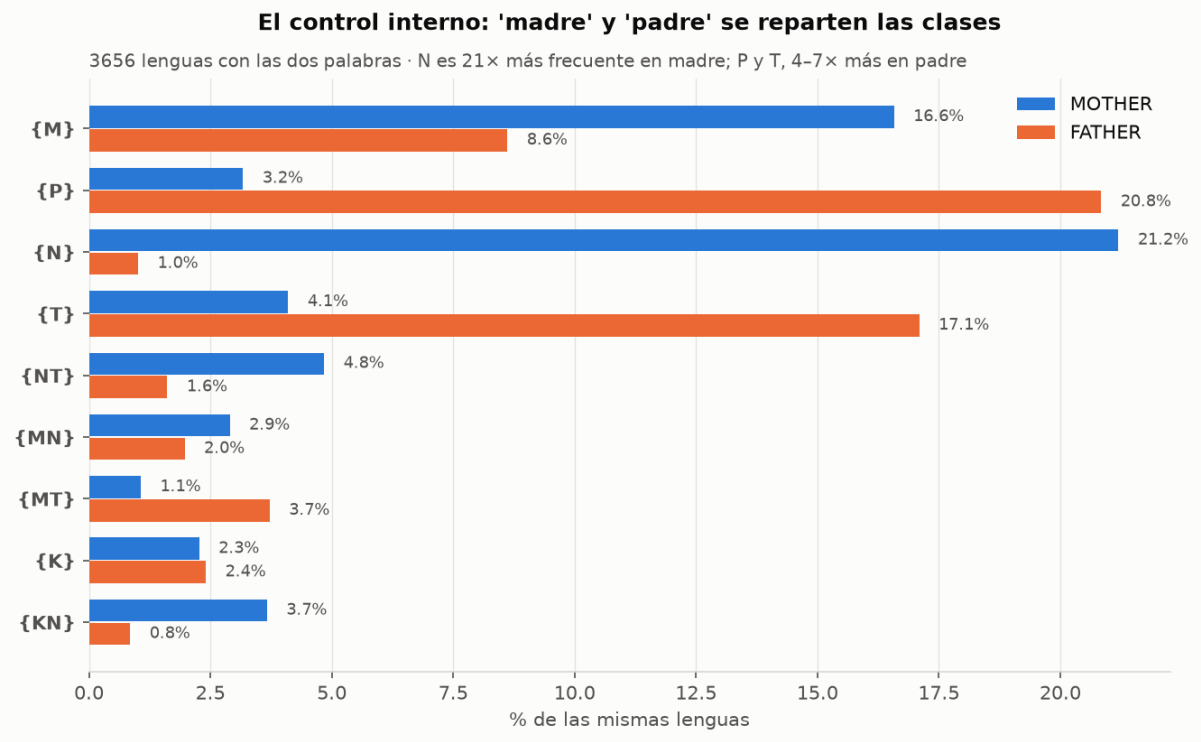

Figure 2: MOTHER against FATHER in the same 3,656 languages. If the procedure manufactured the pattern, both words would behave alike. The split is nearly complementary.

4. The quartet

The question is now generalised to the five nurturing concepts and the four classes. A code counts as a core code if its signature is of a single class and that class is `N`, `M`, `P` or `T`.

4.1 Among single-consonant words, which consonant?

This is the central test, and the one that neutralises length: conditioned on the word already being single-class, the only question left is *which* class it chooses.

The quartet takes 80.9% [79.8, 82.0] in nurture, 50.6% [48.1, 53.0] in the rest of kinship and 48.0% [47.3, 48.8] in ordinary vocabulary. The difference against ordinary vocabulary is 32.9 points [31.5, 34.1], with Cohen's $h = 0.71$ — a large effect on the conventional reading. And the displacement runs both ways: `K` falls from 18.1% to 6.7% and `S` from 10.7% to 2.4%. Nurture does not pick four classes at random out of ten: it picks these and pushes the others aside.

Class by class, among the single-class words of each block:

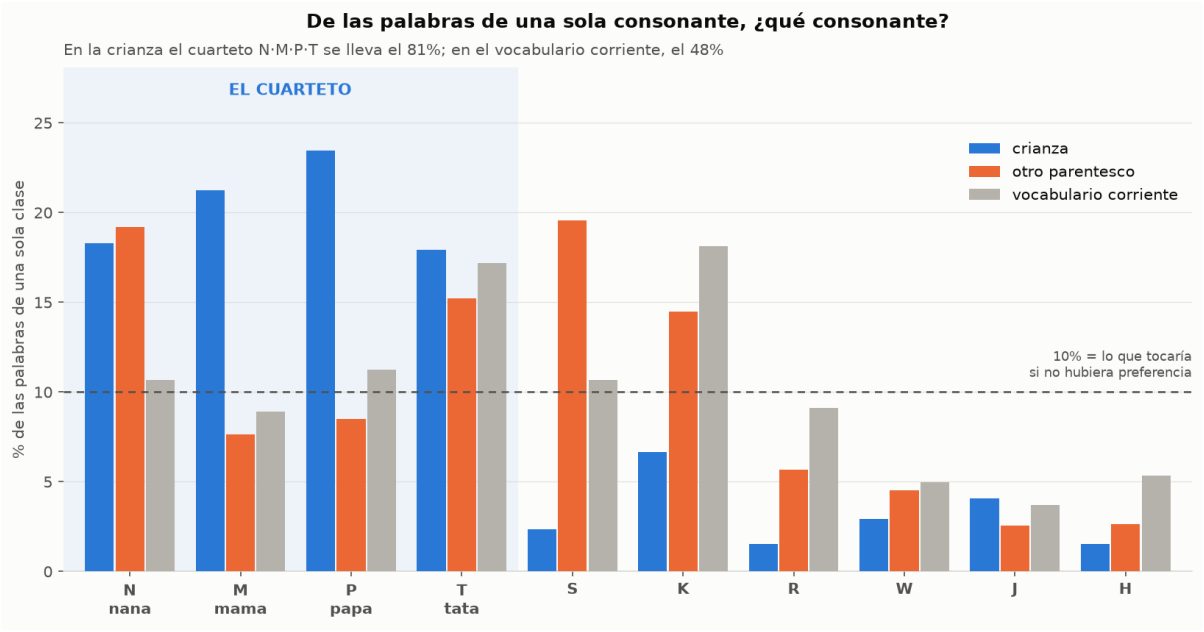

Figure 3: Among words formed from a single consonant class, which class it is. With ten classes, chance would give 10% to each and 40% to the quartet.

class	nurture	95% CI	ordinary	95% CI	Cohen's h
P	23.5%	[22.3, 24.6]	11.2%	[10.8, 11.7]	+0.33
M	21.2%	[20.1, 22.4]	8.9%	[8.5, 9.4]	+0.35
N	18.3%	[17.2, 19.4]	10.7%	[10.2, 11.2]	+0.22
T	17.9%	[16.9, 19.0]	17.2%	[16.6, 17.7]	+0.02
K	6.7%	[6.0, 7.4]	18.1%	[17.6, 18.7]	−0.36
J	4.1%	[3.6, 4.7]	3.7%	[3.5, 4.0]	+0.02
W	2.9%	[2.5, 3.4]	5.0%	[4.7, 5.3]	−0.11
S	2.4%	[2.0, 2.8]	10.7%	[10.2, 11.1]	−0.36
R	1.5%	[1.2, 1.9]	9.1%	[8.7, 9.6]	−0.37
H	1.5%	[1.2, 1.9]	5.3%	[5.0, 5.7]	−0.22

Across the full table of three blocks by ten classes, Cramér's $V = 0.231$ ($\chi^2 = 2{,}570$, df $= 18$): a moderate association for a table of this size, and one concentrated almost entirely in six of the ten rows.

Placed on a single scale, father and mother head the list.

A limit of the more naive metric is worth noting here. The proportion of words falling in the quartet *without conditioning* does not cleanly separate the blocks: FIRE reaches 24.1% and beats GRANDFATHER (21.5%), simply because 'fire' is also a short word. What does separate them is which class gets chosen: of FIRE's single-class words, 67% are from the quartet; of MOTHER's, 83%.

4.2 A trio that rises and a fourth that survives

The class-by-class table above forces a correction to the headline, and it is worth stating plainly because the aggregate figure hides it.

The quartet does not rise as a block. M, N and P go from 30.9% of ordinary single-class words to 63.0% of nurturing ones, and that is where nearly all of the 32.9 points come from. T does not move: 17.2% in ordinary vocabulary, 17.9% in nurture, $h = 0.02$. On this evidence T is not a class the nurturing field prefers. It is a class the field fails to expel, while it expels K ($h = -0.36$), S (−0.36), R (−0.37) and H (−0.22).

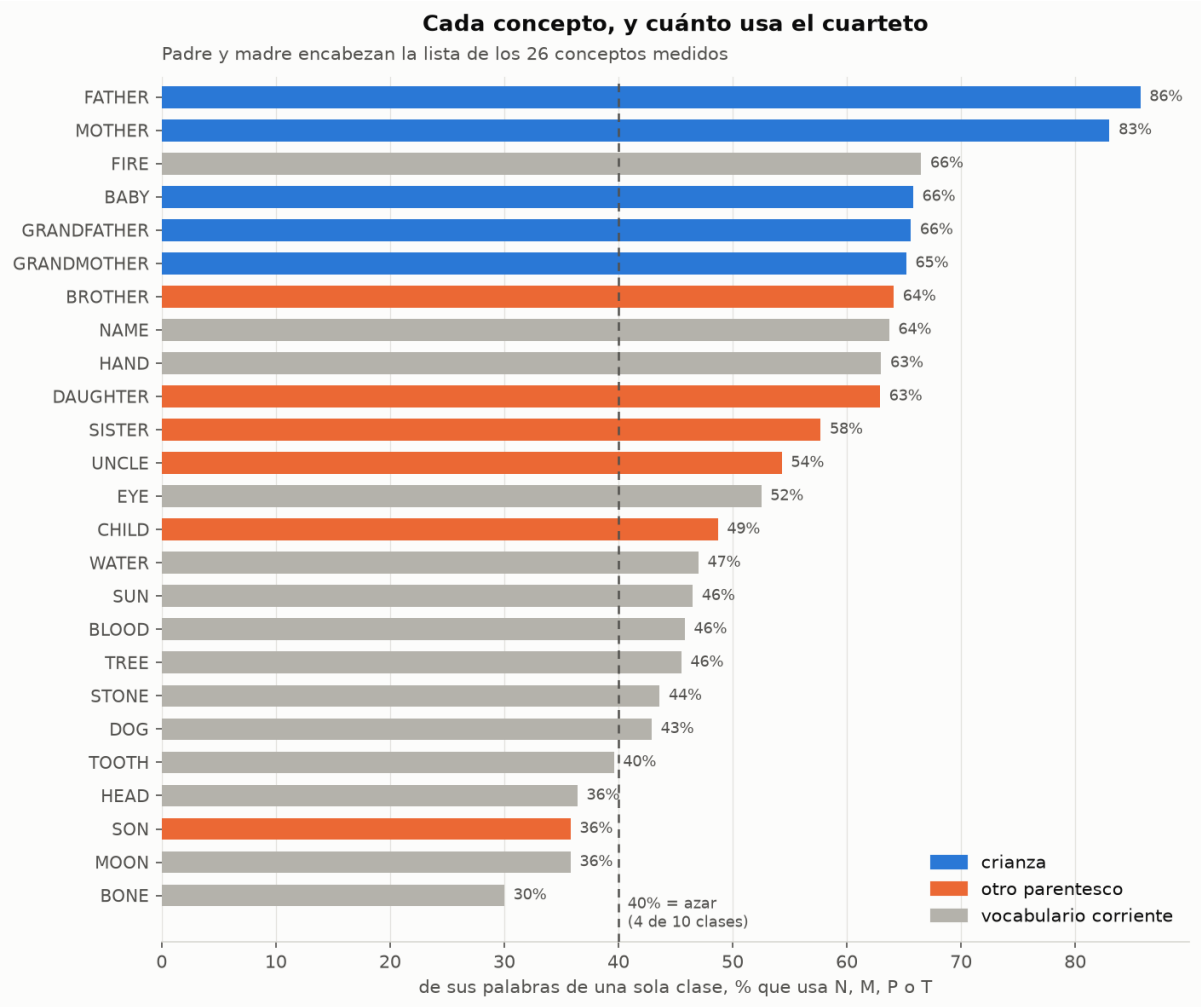

Figure 4: The 26 concepts of the three blocks, ordered by the proportion of their single-class words that use the quartet.

That reframing matters for what has to be explained. The question is not ”why does nurture choose T and not K” — it chooses neither — but ”why does nurture avoid K when it does not avoid T.” §10.3 returns to this, and rules out the cheapest candidate answer.

The quartet, then, is a real inventory with an internal seam: three classes that the field recruits and a fourth that it merely tolerates. Both facts are part of the finding.

4.3 The test that does not depend on family size

Each family is compared against itself: of its single-class words, what proportion uses the quartet in the nurturing concepts and what proportion in the control concepts.

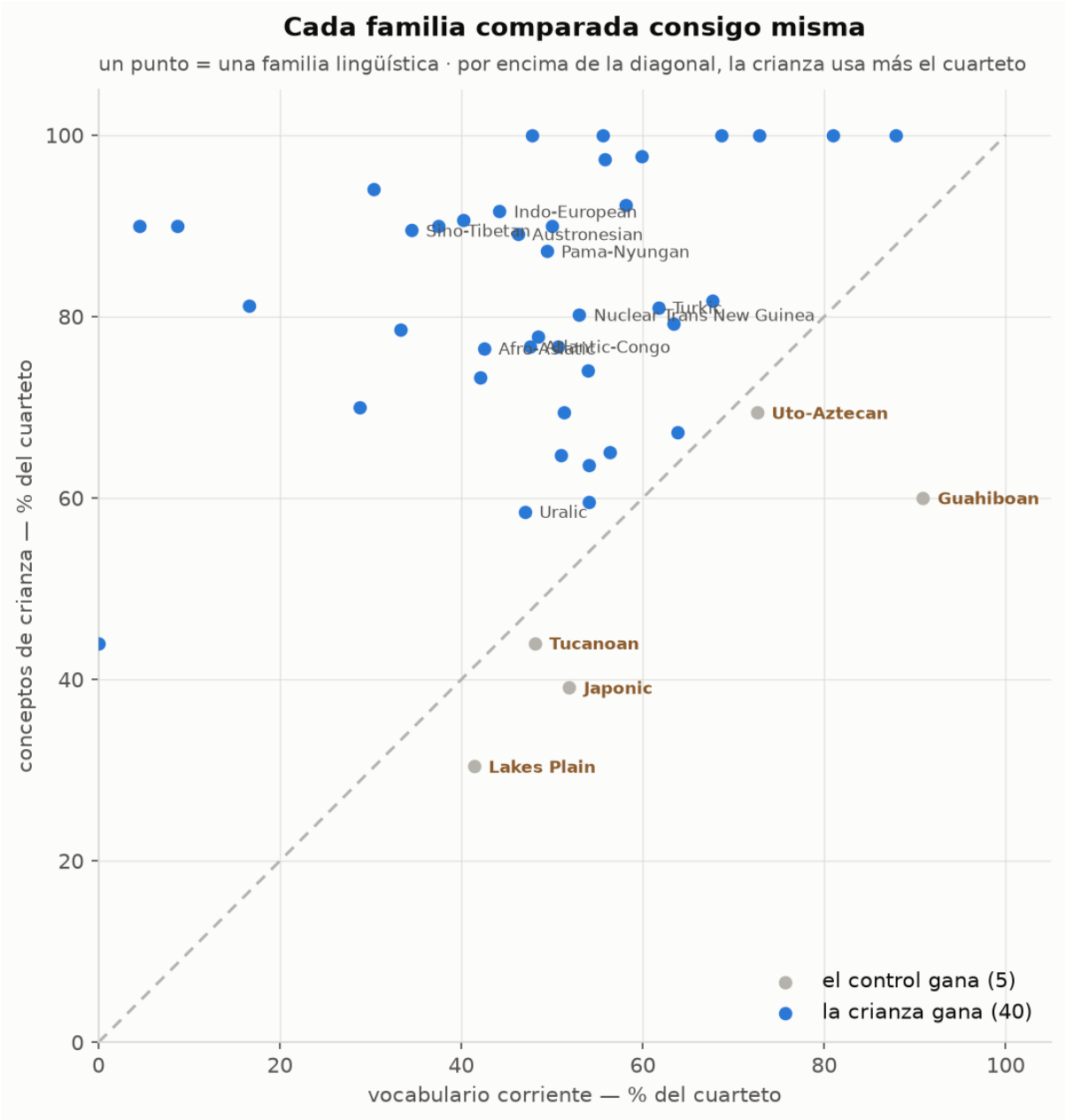

Figure 5: Each family compared against itself. One dot is one family; above the diagonal, nurture uses the quartet more than that same family’s ordinary vocabulary.

Across the 45 families with at least ten single-class forms in both blocks: 78.5% against 49.5%, a mean

difference of +29.0 points with a bootstrap interval over families of [+22.0, +35.8] and a median of +29.3. Nurture is ahead in 40 of 45 families (89%, CI 76–96%; sign test $p = 7.9 \cdot 10^{-8}$; paired Wilcoxon $p = 8.4 \cdot 10^{-10}$). The matched-pairs rank-biserial correlation is 0.91 and Cliff's δ is 0.78: it is not that a few families pull hard, it is that nearly all of them push the same way. No large family drives the result — §7.2 removes them one at a time to show it.

4.4 The repertoire is shared, not partitioned

This is an important point and it was a correction to the initial design. The hypothesis does not say that each concept gets a code of its own, but that these concepts draw on the same repertoire.

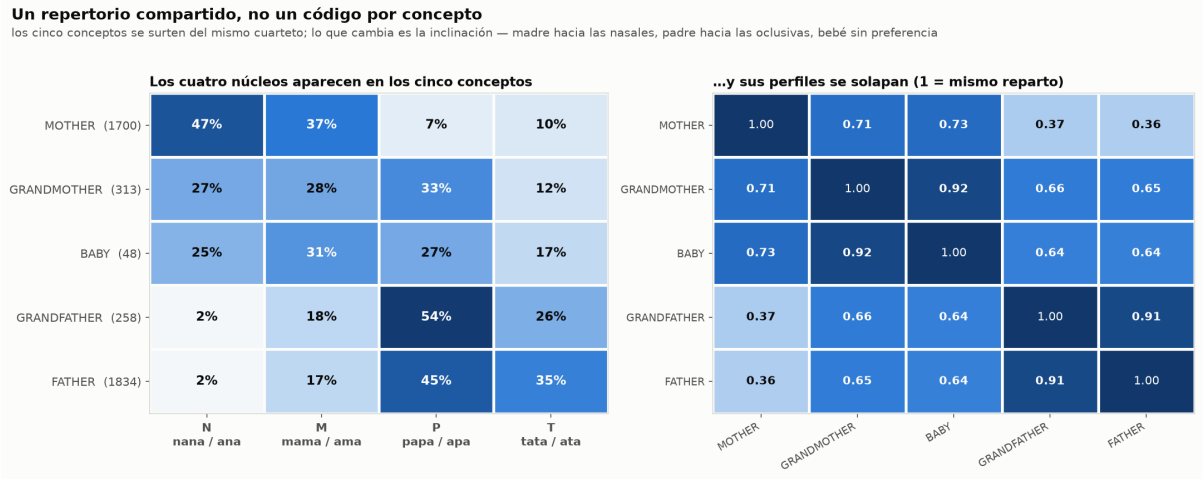

Figure 6: Left: all four core codes appear across all five concepts. Right: overlap of the profiles (1 = identical distribution).

What varies between concepts is the leaning, not the inventory: the feminine toward nasals (N+M = 84% in MOTHER), the masculine toward stops (P+T = 80% in FATHER), and baby with no appreciable preference (25/31/27/17), the purest case of a shared repertoire. Profiles overlap heavily — BABY·GRANDMOTHER 0.92, FATHER·GRANDFATHER 0.91— and the only genuinely contrasted pair is MOTHER·FATHER, at 0.36.

Within a single language the repertoire also works as a system of contrasts: in the 1,193 languages where mother and father both carry a core code, 94% [92.4, 95.1] use a different core for each. It is not a fixed assignment —17% of core fathers are `M`, *mama*— but a strong leaning within one inventory.

5. A negative result: the labial/dental axis

The four codes invite an elegant reading: two crossed axes, nasal/stop separating mother from father and labial/dental pairing them, so that a language saying *mama* would say *papa* and one saying *nana* would say *tata*.

Across languages with a nasal mother and a stop father, the aggregate table supports it:

	father `P` (papa)	father `T` (tata)
mother **`M`** (mama)	285	156
mother **`N`** (nana)	178	196

Odds ratio 2.01, Woolf interval [1.52, 2.67] ($p = 1.3 \cdot 10^{-6}$, Fisher). A language saying *mama* is twice as likely to say *papa* as *tata*.

The figure does not survive the control. Stratified across the twelve families with at least ten cases, the Mantel–Haenszel ratio drops to 1.01, with a Robins–Breslow–Greenland interval of [0.65, 1.56] and $p = 0.97$. The interval is not merely wide around a positive estimate; it sits almost centred on 1. The per-family odds ratios run from 0.08 to 3.38 with a median of 0.79 — spread on both sides of no association — and Tarone's homogeneity test over the seven strata with no empty cell gives $p = 0.02$: the families genuinely differ from one another.

That combination is the exact signature of a composition effect. There is no common association hiding behind between-family variation; there are families that prefer the *mama/papa* pair and families that prefer *nana/tata*, and pooling them produces a correlation that exists within almost none of them.

The nasal/stop axis is solid. The labial/dental axis is not established, and is recorded as such. The stratified interval excludes ratios above 1.56, so an effect the size of the pooled 2.01 can be ruled out; smaller effects cannot, and with twelve usable families the test has no power to try. The honest claim is not "it is false" but "this corpus does not support it, and could not detect a weak version of it."

6. The catalogue of clusters

The question is now inverted —not "which code belongs to this concept" but "this code, which beings does it group languages around"— and the repertoire is widened to include combinations: any code built solely from M, N, P and T, including `MN` (*mana*), `NT` (*nata*) or `PT`. The five concepts are grouped by the being they point to: feminine (mother + grandmother), masculine (father + grandfather) and infant (baby).

Universality is not what is sought. A cluster of 639 languages across 53 families is a fact even if the rest of the world does otherwise; what is catalogued is its size, its spread and its geography.

With combinations included, the repertoire covers 49.8% [48.9, 50.8] of the words for the five beings, against 22.8% [22.5, 23.1] of ordinary vocabulary and 17.0% of the entire lexicon. This is not a universal coverage rate: it is a density contrast, and it is what makes the clusters clusters rather than background noise.

Ordered by how much each code specialises, a clean gradient emerges:

code	leans toward	strength	languages	reading
`N`	feminine	94%	639	the most specialised in the repertoire
`NN`	feminine	93%	270	accompanies `N`
`TT`	masculine	86%	447	reduplicated and heavily marked
`P`	masculine	82%	716	the most transversal: touches 61 families
`PP`	masculine	76%	431	accompanies `P`
`M`	feminine	70%	619	nasal and labial at once
`T`	masculine	70%	453	marked, with real feminine presence
`MM`	feminine	58%	453	the hinge

The axis that orders the repertoire is not labial/dental —that one has already been shown not to hold— but nasal against stop. And there is a structural explanation for the ends of the gradient: `M` and `MM` are nasal, which pushes them toward the feminine, but labial, which allies them with `P`; that dual membership is precisely what leaves them in the middle, while `N`, which lacks the ambiguity, never drops below 94%.

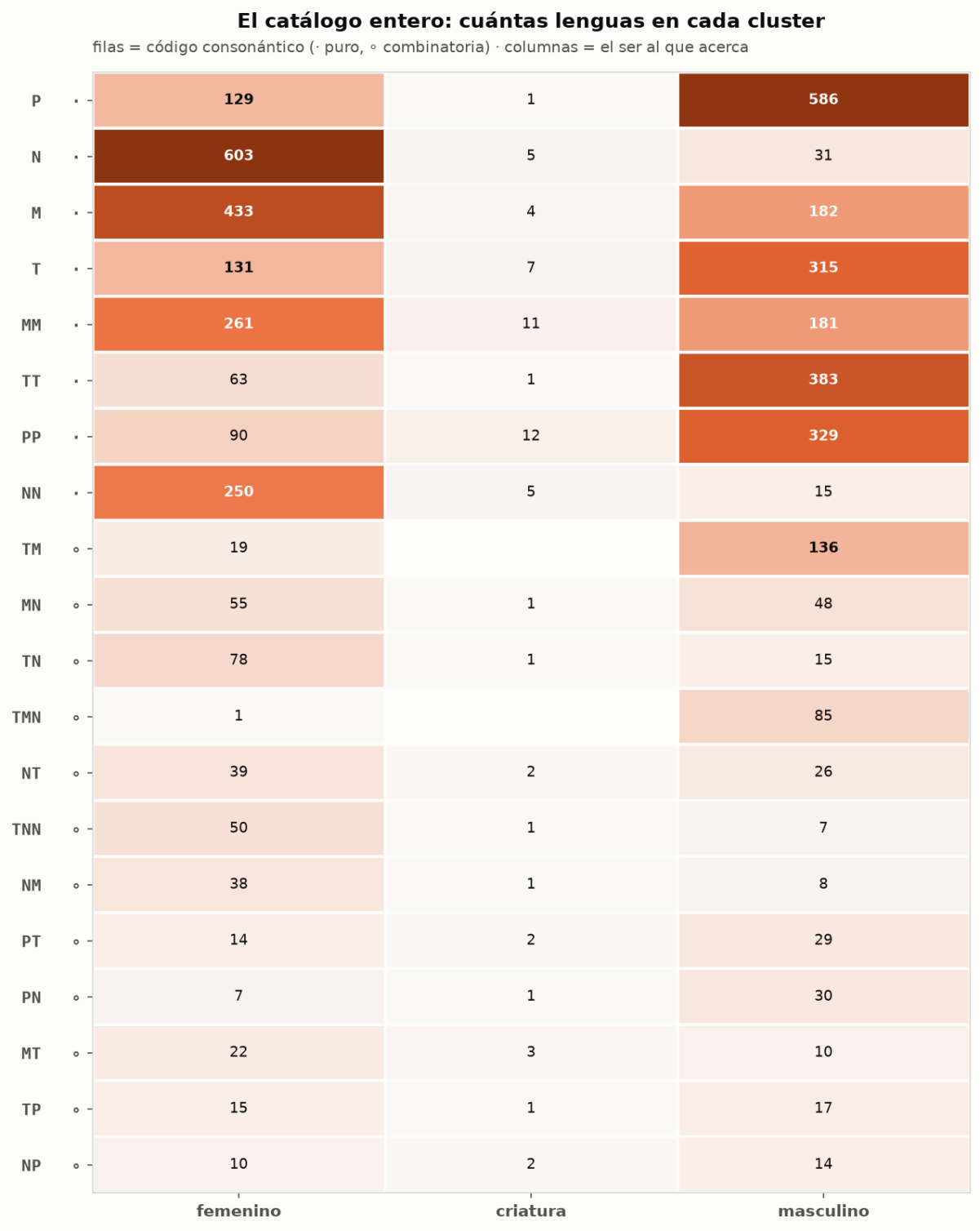

Figure 7: The complete catalogue: how many languages in each cluster. Rows, the consonant code; columns, the being it approaches.

The case of the labial for the feminine being deserves emphasis —the *baba* type, with its Slavic extension in *babushka*— because it is no isolated anomaly: P·feminine gathers 129 languages from 28 families and PP·feminine another 90 from 16. More than two hundred languages across some thirty families use the labial for mother or grandmother. The repertoire is not assigned.

6.1 Repertoire or inheritance: one figure separates them

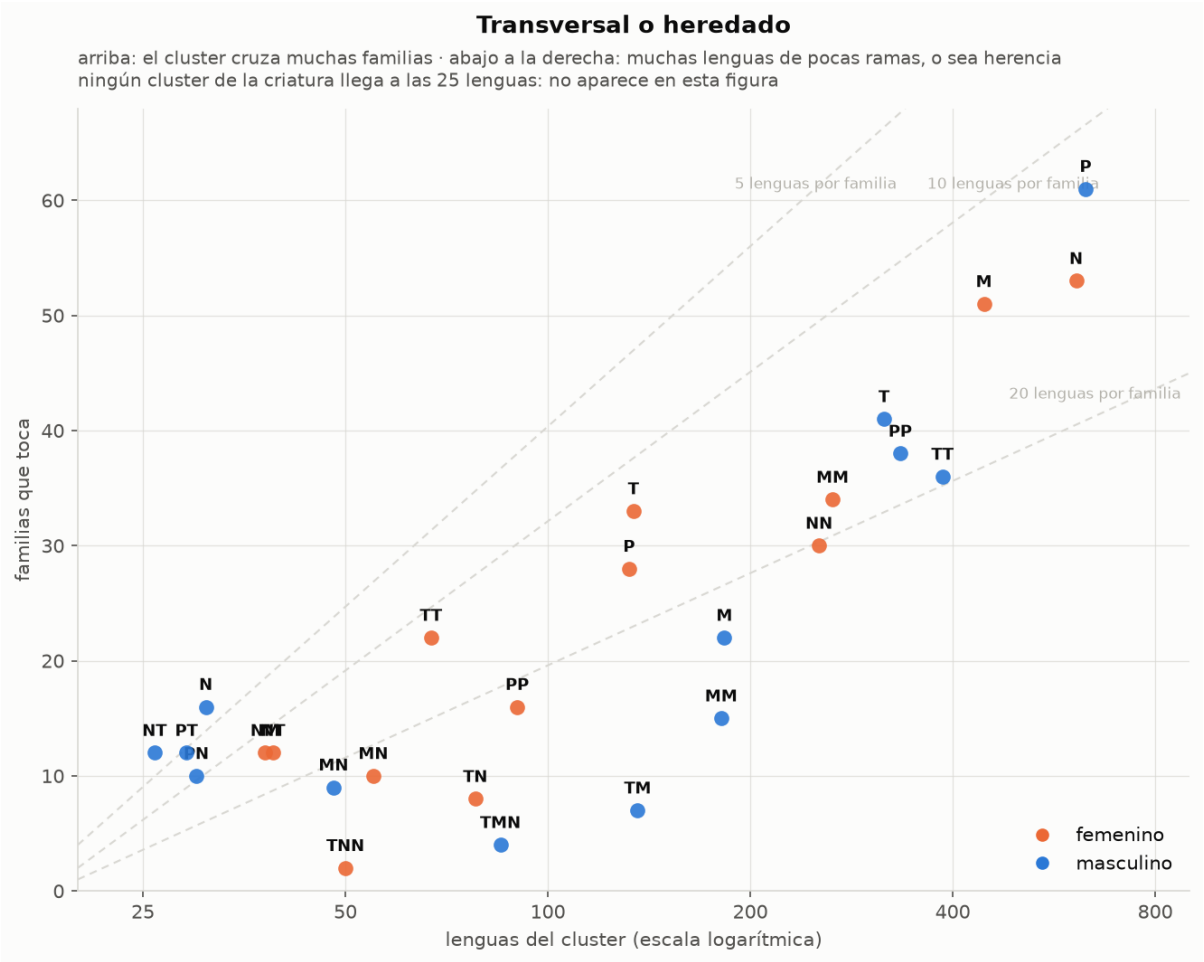

Figure 8: Transversal or inherited. Upper region: clusters crossing many families; lower right: many languages from few branches.

The repertoire clusters touch fifty or sixty families each: N·feminine 53, P·masculine 61, M·feminine 51. Against them, TM·masculine gathers 136 languages but only seven families —127 of them Austronesian, the reflex of *tama*—, TMN 85 languages in four families, and TNN·feminine 50 in two, the reflex of *tina*. Large in languages and minuscule in branches: that is not the repertoire acting, it is an inheritance travelling down one branch. The figure that separates them is the family count, and it is worth applying before interpreting any cluster.

7. Robustness: the encoding, the families, and what the reduction destroys

Everything above rests on one operation: strip the vowels, replace each consonant by a Dolgopolsky class. Three objections follow from that, and each is answered here with a number rather than an argument.

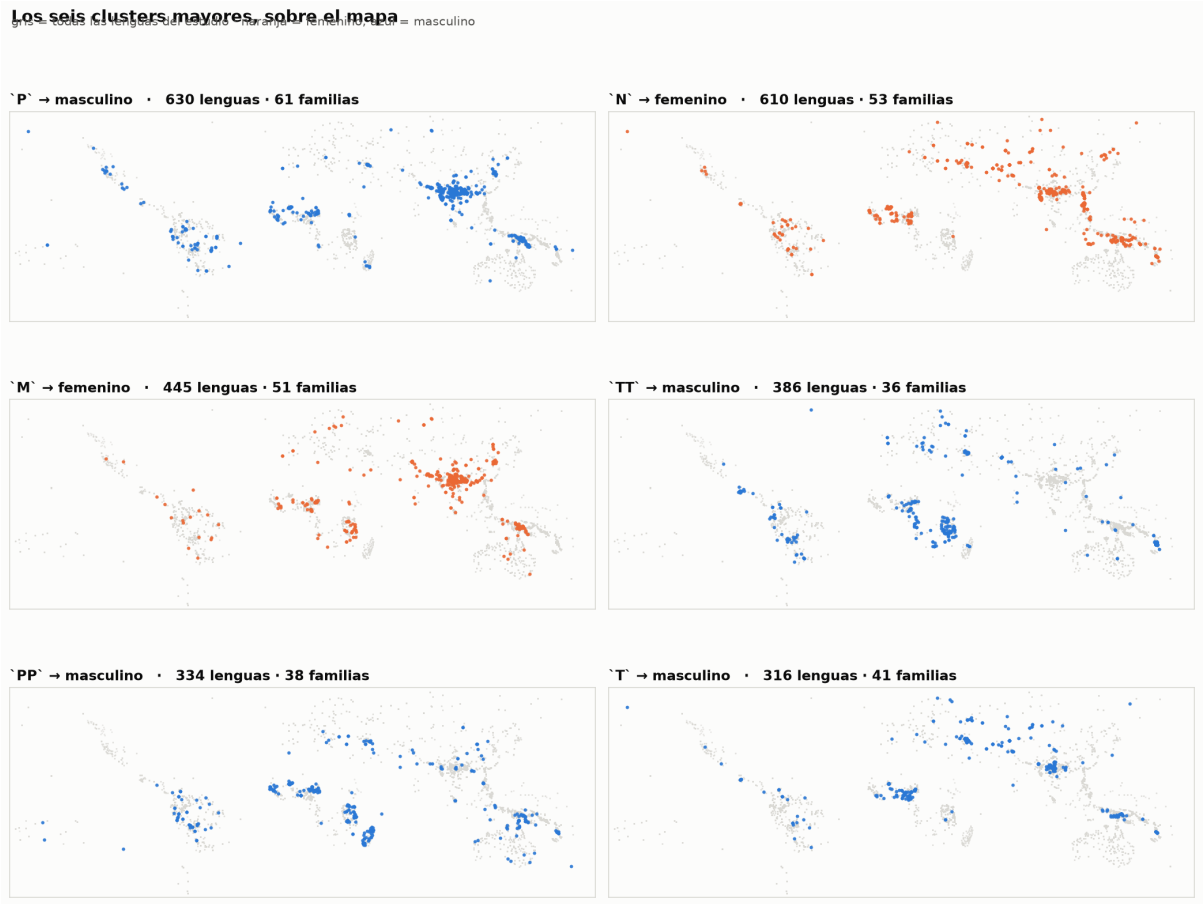

Figure 9: Where each of the six largest clusters lives. In grey, all the languages in the study.

7.1 Six alternative encodings, one result

The central test —among words already built from a single class, what share uses the quartet— was recomputed under six alternative readings: Dolgopolsky recomputed from IPA with LingPy, SCA in two readings, ASJP, articulatory features, and raw IPA with no merging. The quartet is defined the same way each time, by articulation rather than by letter: labial nasal, non-labial nasal, labial stop, coronal non-sibilant stop. Each system translates that into its own alphabet. The finer the encoding, the more classes it has and the less chance concedes.

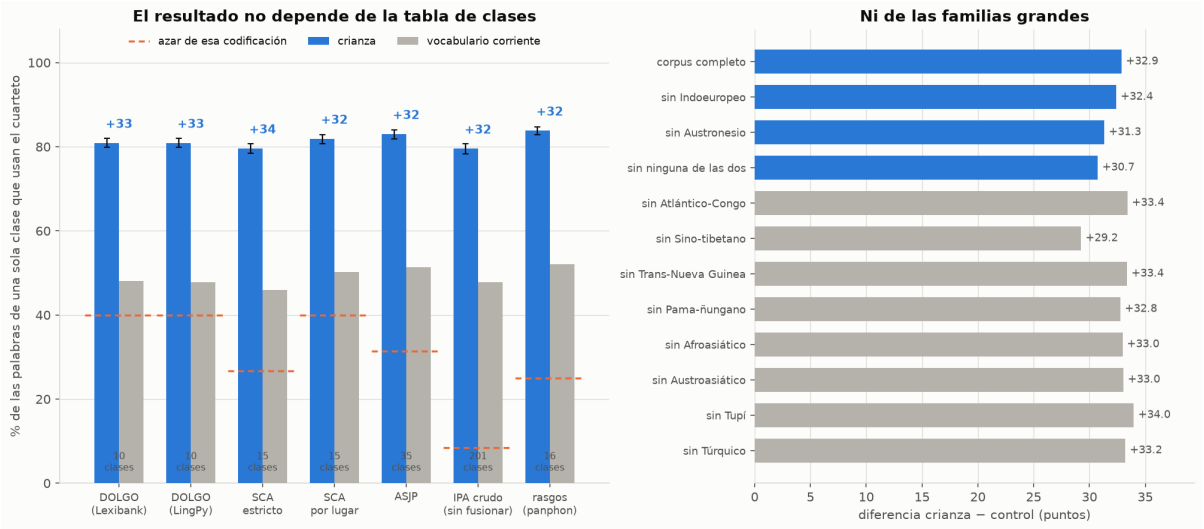

Figure 10: Left: the central figure under seven encodings, each with its own chance level; error bars are 95% Wilson intervals. Right: the same figure after removing whole families from the corpus.

encoding	classes	chance	nurture	kinship	ordinary	gap	families in favour	p
Dolgopolsky (Lexibank)	10	40%	80.9%	50.6%	48.0%	+32.9	40/45	$8.4{\cdot}10^{-10}$
Dolgopolsky (LingPy)	10	40%	80.9%	50.5%	47.9%	+33.0	40/45	$6.5{\cdot}10^{-10}$
SCA, strict	15	27%	79.6%	50.2%	45.9%	+33.7	40/45	$9.6{\cdot}10^{-8}$
SCA, by place	15	40%	81.9%	52.0%	50.3%	+31.6	40/45	$1.2{\cdot}10^{-9}$
ASJP	35	31%	83.0%	48.8%	51.4%	+31.6	42/46	$4.0{\cdot}10^{-9}$
raw IPA, no merging	201	8%	79.5%	46.8%	47.7%	+31.8	35/40	$1.4{\cdot}10^{-8}$
articulatory features	16	25%	83.8%	50.4%	52.1%	+31.7	46/49	$2.0{\cdot}10^{-11}$

Three of these deserve comment. SCA strict (List 2012) separates `f/v` from `p/b` and `θ/ð` from `t/d`, so the quartet is deliberately made smaller than Dolgopolsky allows; the gap widens rather than narrows. ASJP (Brown *et al.* 2008) is close to a transcription system, with 35 attested classes. Articulatory features are computed with PanPhon (Mortensen *et al.* 2016) directly from place and manner, with no historical table involved at all; under that encoding `ŋ` is a dorsal nasal and does *not* count toward the quartet, which is the strictest reading of the hypothesis and the one that scores highest.

The decisive row is the sixth. Under raw IPA with no merging whatsoever — 201 distinct base consonants, where chance would give the quartet 8% — the nurturing block still puts 79.5% of its single-consonant words on the quartet's sounds against 47.7% for ordinary vocabulary. No sound-class sys-

tem is involved in that number. Looked at without any table at all, the twelve most common single-consonant words show it directly:

#	nurture	%	ordinary	%
1	m	21.2%	t	10.5%
2	n	17.8%	m	9.0%
3	t	12.6%	n	8.7%
4	b	10.6%	k	8.5%
5	p	8.9%	s	6.9%
6	d	5.3%	p	5.3%
7	j	4.1%	l	4.5%
8	k	3.9%	d	4.2%
9	w	2.6%	h	4.1%
10	s	1.6%	j	4.1%
11	h	1.1%	w	4.0%
12	tʃ	1.0%	b	3.9%

The first six places in the nurturing column are m, n, t, b, p, d. Dolgopolsky's table did not put them there.

7.2 Without the big families

scenario	languages	families	nurture	ordinary	gap
full corpus	5,501	298	80.9% [79.8, 82.0]	48.0% [47.3, 48.8]	+32.9
no Indo-European	5,197	297	80.6%	48.2%	+32.4
no Austronesian	4,523	297	79.6%	48.3%	+31.3
neither	4,219	296	79.2%	48.5%	+30.7

Removing each of the ten largest families in turn leaves the gap between +29.3 (without Sino-Tibetan) and +34.0 (without Tupian). And the hardest control available: drawing one language per family, so that a family of 956 languages and a family of one weigh exactly the same, 500 times. The gap comes out at +31.8 points, 2.5–97.5 percentiles [+26.1, +37.0], positive in 100% of the draws.

7.3 How much information the reduction destroys

Stripping vowels and merging consonants brings together words that were not that close: *mater*, *mother* and *mutter* all end up sharing MTR. How much is lost can be answered with two numbers per step — how many bits of entropy survive, and how often two different words end up identical.

step	distinct forms	entropy (bits)	bits retained	collision, nurture	collision, ordinary	ratio
full IPA form	50,413	14.85	100%	0.12%	0.011%	×10.8
consonant skeleton (IPA)	19,153	11.43	77%	1.04%	0.262%	×4.0
Dolgopolsky code	6,388	8.81	59%	2.30%	0.871%	×2.6
signature (unordered)	665	6.83	46%	4.53%	1.593%	×2.8

The loss is real and large: 41% of the entropy is gone by the time a form becomes a Dolgopolsky code, and 50,413 distinct forms have become 6,388. But the two right-hand columns say something

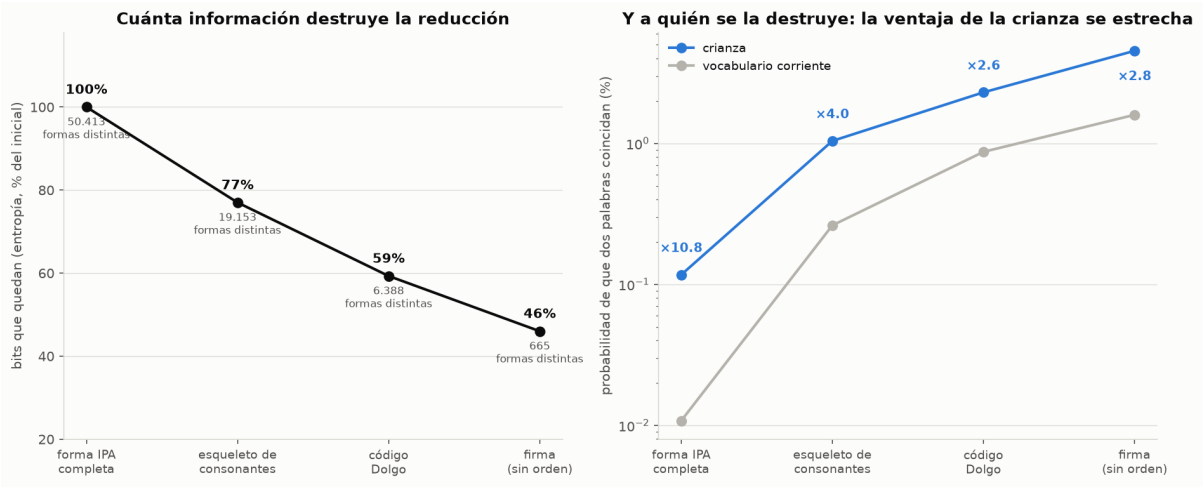

Figure 11: Left: entropy retained at each step of the reduction. Right: the probability that two words of the same block collide, at each step, in nurture and in ordinary vocabulary.

the objection does not anticipate. The reduction hurts the control block more. The ratio between nurture's collision rate and ordinary vocabulary's *falls* at every step — from $\times 10.8$ at full IPA to $\times 2.6$ at the Dolgopolsky code — because ordinary vocabulary is what gains most from being merged.

The same appears in the false-merge rate. Of every hundred pairs that the Dolgopolsky code declares identical, 55 have different IPA consonant skeletons in the nurturing block and 70 in the control. The merging is more generous with 'stone' than with 'mother'.

So the reduction does not manufacture the contrast; if anything it erodes it. At the level of the complete IPA form, before a single vowel is removed, two nurturing words drawn at random already collide ten times more often than two ordinary ones.

8. The infant: a negative and firm result

The most fragile point of the initial hypothesis was the third: that the infant in arms —*nene*, *bebé*— should participate in the repertoire. With the concept BABY alone it had no material: 601 languages and a mean code length of 3.50 consonants. It was redone with three design changes.

The unit of count becomes the language, not the concept. The right question is not "does BABY use the repertoire?" but "does this language have *any* word for the infant in the repertoire?" If 'baby' is a compound but *nene* exists as CHILD, the language has the word.

Attempts are equalised. A language with six words for the infant has six chances of landing in the repertoire; the control is given the same six, drawing six of its ordinary concepts, three hundred times. Without this correction the metric rises on its own.

The field is opened into three layers, because 'child' is an age and 'son' is a relation: infant in arms (BABY, CHILD, CHILD·DESC), boy/girl (BOY, GIRL, GRANDCHILD) and descent (SON, DAUGHTER, DESCENDANTS, YOUNG MAN, YOUNG WOMAN).

The result is negative in all three layers. Across 4,563 languages, 28.6% [27.3, 29.9] against 38.7% for the equalised control (2.5–97.5 percentiles of the resampling: 37.7–39.9): ten points below, Cohen's $h = -0.21$. The two intervals do not touch at any point; this is a firm negative, not a badly measured tie. And it is not for lack of data —BOY has 1,778 languages and 228 families, more than 'mother'.

Nor is it an Indo-European phenomenon diluted by the corpus: Indo-European ranks ninth of 27 families and also falls below its own control (45% against 49%). Only 3 of 27 families have more

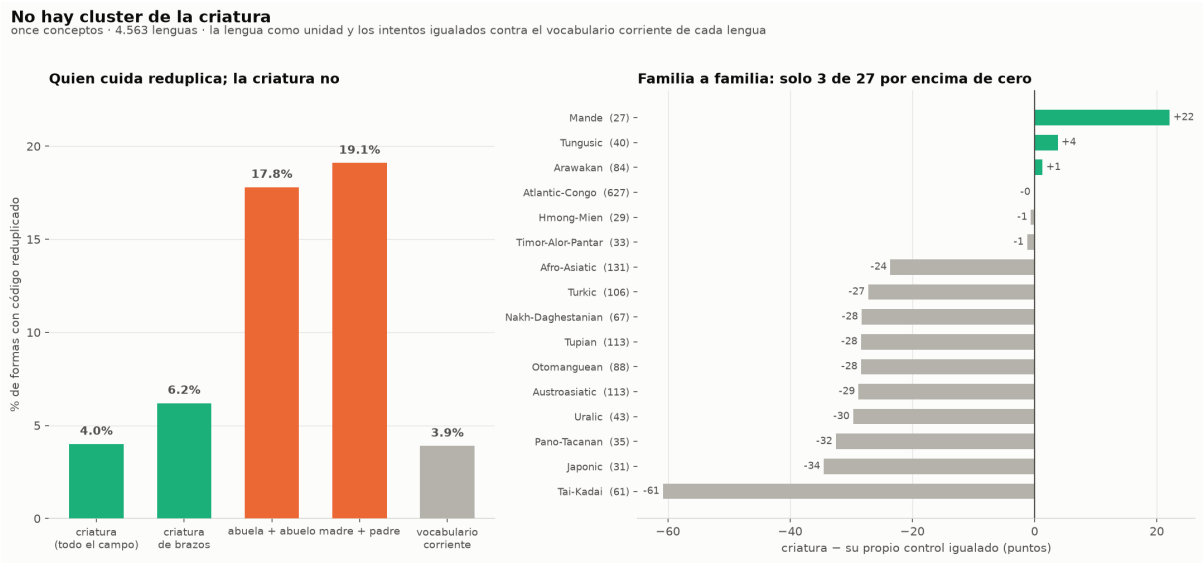

Figure 12: Left: reduplication by group. Right: each family against its own equalised control.

repertoire in the infant than in their ordinary vocabulary, and the only one with a clear effect is Mande (+22 points, 27 languages).

An alternative hypothesis was also tested: that what characterises the infant might not be the quartet but reduplication, with whatever consonant. It does not. Mother and father reduplicate 19.1% and grandparents 17.8% —five times more than ordinary vocabulary— while the infant sits at 4.0% against the control's 3.9%.

The reason lies in code length. Mother and father hover around 1.9 consonants; the infant runs from 2.6 to 4.2. Languages do not name the infant with a primitive word: they derive it —'the little one of', 'the son of', 'the one who is born'— and a derived word will not fit into a two-consonant code. Put differently: the repertoire of nurture belongs to the one who cares, not to the one cared for.

8.1 What does remain: transference

There is, however, a real effect, and precisely where the psychodynamic intuition predicts it —that parents recycle their own terms toward the child: *papi*, *papito*, *nene*.

When the infant does use the repertoire, it tends to use the same code as the parents of its own language: across 949 languages, 18.7% agreement [16.3, 21.3] against 14.5% expected when infant codes are shuffled among languages within each family (2,000 permutations, null 2.5–97.5 percentiles 12.9–16.1, $p = 0.0005$). Cohen's h is 0.11 — real, and small, and worth reading that way. The code does not belong to the infant: it arrives on loan from whoever names it. The transference exists; what does not exist is an autonomous cluster.

9. The gap: there is no corpus of the affective register

This section reports not a result but an absence, and it is probably the most important thing this work has to say about its own limits.

Everything measured so far is the term of reference —'the mother', 'the child'— which is what a comparative word list records by design. What the mother actually *says* to the infant in her arms belongs to another register, the vocative and affective, and it is not documented in comparable form

for the vast majority of the languages treated here.

The size of the gap can be given as a number. Of the 1,740,092 records in Lexibank —not of this study's sample, but of the entire repository— exactly nine carry any vocative or address marking. Nine out of one million seven hundred thousand. And those nine suffice to show that this is not a nuance but a different word:

language	term of reference	term of address
huntergatherer-258 (MOTHER)	*deni* → TN	"term of address is mama" → MM
huntergatherer-353 (FATHER)	*jum-me* → KM	*pa:* (vocative) → P
huntergatherer-353 (MOTHER)	*manko*	*ma:* (vocative) → M
huntergatherer-163 (MOTHER)	*niatija*	"originally vocative, reanalyzed as referential"

In the first three cases the term of address falls in the repertoire and the term of reference does not. The fourth is more eloquent still: a vocative frozen into a term of reference — that is, the very mechanism by which these words enter the lexicon.

The methodological consequence is direct: the figures in this work are a floor, not a ceiling. We have systematically measured the term that least favours the hypothesis, because it is the only one the corpus records.

We explored whether Wiktionary could fill the gap. The Kaikki dumps (Ylönen 2022) preserve the register in the text of the gloss —Spanish *mamá* appears with "a term of affection for a woman"— which allows a register pattern to be crossed with a referent pattern. A sweep of 105 local dumps returned 366 entries across 23 languages, of which 182 point to the infant.

It will not serve for typology, and it is worth saying exactly why. Of the 105 languages available locally, 104 are Indo-European —the dump had been downloaded for a study of that family— and English alone contributes 41% of the hits. They are moreover dictionary words (*poppet*, *honnepon*, *chiquitín*), not spontaneous speech. As a signal, and with no comparative null, the first consonant of those 182 terms falls in the quartet 60.9% of the time, with P dominant (41.9%); this is recorded as a lead, not a result.

The appropriate source exists and is a different one: CHILDES / TalkBank (MacWhinney 2000), with 48 languages and 436 publicly accessible corpora consisting of transcriptions of real speech between adults and children. The pending experiment is well defined: extract the vocatives the mother directs at the child, code them by the same procedure, and compare them against the term of reference for that same language in Lexibank, in a language-paired contrast. The four cases in the table above predict the direction of the result.

10. Discussion

What these data describe is neither a universal nor a kinship, but a repertoire in the sense fixed in §1.1: a small inventory of consonant codes that a large number of mutually unrelated languages deploys for the same beings, with statistical leanings rather than rules.

The three defining properties were each measured, and each holds.

It is a closed and expulsive inventory. Four classes out of ten, with K, S, R and H actively displaced, and the displacement is as large in effect size as the recruitment. This is what distinguishes the finding from "nurturing words happen to contain more nasals."

It is shared, not partitioned. All five beings draw on the same inventory, with profile overlaps up to 0.92. What distinguishes them is a leaning —nasal for the feminine, stop for the masculine— that admits massive exceptions: two hundred languages call the mother with a labial. A cluster is not a rule.

It discriminates by gradient, not by slot. From `N` at 94% to `MM` at 58% there is a continuum, and its low end has an internal phonological explanation: `M` participates in both nasality and labiality, and therefore does not distinguish.

10.1 Convergence, and why the finding is not simply Jakobson restated

The natural objection to all of the above is that it is a corollary of what was already known. If babbling produces labials and coronals, and caregivers hear babbling as address, then a cross-linguistic pile-up of `M`, `N` and `P` in parental terms is not a discovery but an inevitability — the phonological analogue of noting that heavy things fall.

The objection is partly right and it is worth conceding the part that is. The *direction* of the effect is predicted by Jakobson (1960) and by everything that followed: Ferguson's (1964, 1978) description of baby talk as a subsystem, Locke's (1983) and Vihman's (2014) accounts of what the infant can actually produce in the first year, MacNeilage and Davis's (2000) frame/content model in which mandibular oscillation with a stationary tongue yields closures at the lips and at the alveolar ridge as the default outputs of babbling. Nobody needed this paper to expect *mama*.

What was not predicted is the shape. Four points in the data are not corollaries of the articulatory account.

The refusal is stronger than the preference. An articulatory account predicts that the easy sounds should be over-represented. It does not predict that the *hard* ones should be driven out of a semantic field while remaining perfectly available in words of the same length and shape — see §10.3.

The inventory is shared rather than distributed. Convergent articulatory pressure acting independently on five concepts would produce five similar distributions, which is what we observe; but it gives no reason for the *within-language* contrast system of §4.4, where 94% of languages that use core codes for both parents use a *different* one for each. That is a lexical fact about paradigms, not an articulatory fact about mouths.

The one place where the articulatory account most clearly predicts the effect is the one place it fails. If the repertoire were simply babble captured, the infant — the being closest to the babble — should be its densest user. The infant is ten points below its own ordinary vocabulary (§8). The repertoire belongs to the one who cares, not to the one cared for, and that is an asymmetry an articulatory story alone does not generate.

The magnitudes are new. "Sounds similar in many languages" and "80.9% against 48.0%, holding in 40 of 45 families and under raw IPA" are different statements, and the second is the one that can be compared against the next phenomenon. Blasi *et al.* (2016) made the analogous move for sound–meaning biases in basic vocabulary, and Wichmann *et al.* (2010) for ASJP; this paper does it for the nurturing field, with the family as unit.

The wider frame this belongs to is the literature on non-arbitrariness — iconicity, systematicity and sound symbolism (Hinton, Nichols & Ohala 1994; Perniss, Thompson & Vigliocco 2010; Dingemanse *et al.* 2015). The nurturing repertoire is an unusual case within that frame, because the motivation is not perceptual resemblance between sound and referent but *the referent's own production*: the sound does not depict the infant, it quotes it. Whether that is best classified as iconicity, as indexicality, or as neither, is a question this data can inform but not settle.

10.2 What would falsify the repertoire reading

Since ”repertoire” and ”articulatory inevitability” make many of the same predictions, it is worth naming where they part company, so that the claim stays falsifiable.

If the repertoire is a *lexical* object, then it should behave like one: it should survive in the vocative register more strongly than in the referential one (§9 predicts this and CHILDES can test it); it should show the within-language contrast system in languages whose ordinary phonology gives it no help; and it should be recoverable in languages whose babbling-to-lexicon route is culturally different — sign languages with mouthings, or languages with heavily suppletive kin systems.

If, on the other hand, it is articulatory convergence and nothing else, then the vocative should show *no* enrichment over the referential term, the mother/father contrast should dissolve once frequency is controlled, and the K-avoidance of §10.3 should turn out to be an artefact of something we have not measured. Each of these is a real experiment, and none is run here.

10.3 Why not **K**

The paper’s most persistent loose end used to be stated as ”why T and not K”. §4.2 shows that the question was badly posed: T is not preferred at all ($h = 0.02$). The real question is why K is avoided, since it is the single most frequent one-class code in ordinary vocabulary (18.1%) and drops to 6.7% here — an h of -0.36, the same magnitude as S and R.

The cheapest candidate explanation is that the shape of the word does it. Nurturing words are short and often reduplicated, and if reduplication itself disfavoured dorsals, the whole pattern would follow from word shape rather than from the semantic field. That explanation fails, and it fails cleanly. Measured across the entire Lexibank lexicon, with nurture nowhere in sight:

class	of one-consonant codes	of reduplicated XX codes	ratio
N	11.3%	16.7%	×1.48
K	20.4%	25.7%	×1.26
T	14.9%	18.7%	×1.25
R	9.7%	9.4%	×0.97
P	12.3%	9.8%	×0.80
H	5.0%	3.9%	×0.78
M	7.4%	5.5%	×0.74
S	10.8%	6.6%	×0.61
W	4.7%	2.3%	×0.50
J	3.6%	1.3%	×0.36

In ordinary vocabulary a reduplicated word is *more* likely to be velar, not less, and K reduplicates exactly as well as T. *Kaka* is a perfectly ordinary shape for an ordinary word. The avoidance of K is specific to the semantic field, not a consequence of the phonological shape the field happens to favour.

Four hypotheses remain live, and they are separable by experiment rather than by argument.

Acquisition order. Velar closures require independent raising of the tongue body and appear later and less reliably in canonical babbling than labial and coronal closures, which fall out of the mandibular frame (MacNeilage & Davis 2000; Vihman 2014). If the repertoire is babble captured, whatever the babble does not reliably produce cannot be captured. This predicts that the strength of K-avoidance should track cross-linguistic variation in early velar production.

Perceptual and visual salience. Labial closure is visible; coronal closure is partly visible; dorsal closure is not. In face-to-face infant-directed interaction (Fernald & Simon 1984; Kuhl 2004; Saint-Georges *et al.* 2013), a name selected under mutual gaze is selected under a visibility constraint the rest of the

lexicon does not have. This predicts a labial-first ordering, which the data show (P and M are the two largest gains).

Register contrast. If the nursery register works partly by *not* sounding like ordinary vocabulary, then the most ordinary class is the worst candidate, and K's 18.1% share of ordinary one-class words is exactly what would disqualify it. This predicts that the avoided classes should be the frequent ones — which fits K and S, and fits R less well.

Sound-symbolic association. Dorsal and sibilant consonants recur in the harshness end of sound-symbolic scales (Ohala 1994; Hinton *et al.* 1994), which would disfavour them in an affective register independently of articulation. This predicts that the same classes should be avoided in other affective vocabularies, and that is a directly testable claim about a field we have not touched.

Nothing in these data chooses between the four, and the paper does not pretend otherwise. What the data do establish is that the fact to be explained is an avoidance and that it is not reducible to word shape.

10.4 The grandmother's asymmetry

One further question belongs to this work rather than being inherited. The grandfather inherits the father's profile (54% P, overlap 0.91) but the grandmother does not inherit the mother's: she spreads across three cores and her nearest neighbour is BABY (0.92), not MOTHER (0.71). Nothing available predicts that asymmetry, and it is left open.

11. Limitations

The code is computed over the whole word. Morphemes are not segmented, so a repertoire root with a suffix on top falls out of the count: Russian *babushka* yields PPSK instead of the PP of *baba*. Measured identically across the three blocks, this loss leaves out a further +6.5% in the five beings against +4.5% in the control, confirming that the debt is not symmetric and that the clusters are underestimated.

The corpus carries no prosody or stress. Nothing said here touches intonation, which in the nurturing register is a central component on Ferguson's account and the best-documented dimension of infant-directed speech (Fernald & Simon 1984). A repertoire of consonant codes is, at most, one layer of the register.

Family coverage is uneven. Austronesian contributes 956 languages and some families have fewer than twenty. Strong claims are therefore given by family, and §7.2 shows the result surviving both the removal of the large families and a one-language-per-family resampling. The power of the *stratified* tests remains limited, however — twelve usable families in the labial/dental test, and that is where the limitation bites.

The affective register is absent. This is the largest limitation and is developed in section 9.

The level of the phenomenon. Everything measured belongs to acquisition and articulation. It is not evidence of a deep genealogical signature and should not be read as such.

12. Conclusion

A repertoire of four consonant codes —N, M, P, T— exists and is deployed by a large number of mutually unrelated languages for the beings of the nurturing core. It takes 80.9% of the single-class words in that field against 48.0% in ordinary vocabulary, a gap of 32.9 points with Cohen's $h = 0.71$, and the

advantage holds within 40 of 45 families compared against themselves. It survives six alternative encodings of the sound inventory, including raw IPA with no merging at all, and it survives the removal of the two largest families and a resampling that gives every family equal weight.

Within that result there is a seam worth carrying forward: the quartet does not rise as a block. `M`, `N` and `P` are recruited; `T` is merely tolerated; and `K`, `S` and `R` are expelled with effect sizes as large as the recruitment. The thing to be explained is an avoidance.

The repertoire is shared among the concepts rather than partitioned, and it is ordered by a gradient of specialisation running from `N` at 94% to `MM` at 58%, with the nasal/stop axis separating the feminine being from the masculine.

Two attractive readings do not survive the controls and are recorded as negative: the four codes do not form two crossed axes —the labial/dental axis is a composition effect, with a stratified odds ratio of 1.01 [0.65, 1.56] and genuine heterogeneity between families— and the infant does not have a cluster of its own, but a code borrowed from its parents.

And one absence conditions the whole: no comparative corpus of the affective register exists for the vast majority of the languages treated. Nine forms out of 1,740,092. What has been measured here is the term of reference; the term of address —what the mother actually says— lies beyond the reach of the available data, and the few documented cases indicate it would fall further inside the repertoire, not further out.

References

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818–10823.

Brown, C. H., Holman, E. W., Wichmann, S., & Velupillai, V. (2008). Automated classification of the world's languages: A description of the method and preliminary results. *STUF – Language Typology and Universals*, 61(4), 285–308.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3), 494–509.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19(10), 603–615.

Dolgopolsky, A. B. (1964). Gipoteza drevnejšego rodstva jazykovyx semej Severnoj Evrazii s verojatnostnoj točki zrenija [A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia]. *Voprosy Jazykoznanija*, 2, 53–63.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, 66(6, part 2), 103–114.

Ferguson, C. A. (1978). Talking to children: A search for universals. In J. H. Greenberg (ed.), *Universals of Human Language*, vol. 1. Stanford: Stanford University Press, 203–224.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20(1), 104–113.

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. *Glottolog*. Max Planck Institute for Evolutionary Anthropology. https://glottolog.org

Hinton, L., Nichols, J., & Ohala, J. J. (eds.) (1994). *Sound Symbolism*. Cambridge: Cambridge University Press.

Jakobson, R. (1960). Why 'mama' and 'papa'? In B. Kaplan & S. Wapner (eds.), *Perspectives in Psychological Theory: Essays in Honor of Heinz Werner*. New York: International Universities Press.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5, 831–843.

List, J.-M. (2012). SCA: Phonetic alignment based on sound classes. In M. Slavkovik & D. Lassiter (eds.), *New Directions in Logic, Language and Computation*. Berlin: Springer, 32–51.

List, J.-M., Cysouw, M., & Forkel, R. (2016). Concepticon: A resource for the linking of concept lists. *Proceedings of LREC 2016*, 2393–2400.

List, J.-M., Forkel, R., Greenhill, S. J., Rzymski, C., Englisch, J., & Gray, R. D. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*, 9, 316. https://doi.org/10.1038/s41597-022-01432-0

Locke, J. L. (1983). *Phonological Acquisition and Change*. New York: Academic Press.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288(5465), 527–531.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719–748.

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., & Levin, L. (2016). PanPhon: A resource for mapping IPA segments to articulatory feature vectors. *Proceedings of COLING 2016*, 3475–3484.

Murdock, G. P. (1959). Cross-language parallels in parental kin terms. *Anthropological Linguistics*, 1(9), 1–5.

Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 17(8), 873–890.

Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (eds.), *Sound Symbolism*. Cambridge: Cambridge University Press, 325–347.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, 1, 227.

Robins, J., Greenland, S., & Breslow, N. E. (1986). A general estimator for the variance of the Mantel–Haenszel odds ratio. *American Journal of Epidemiology*, 124(5), 719–723.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.-C., & Cohen, D. (2013). Motherese in interaction: at the cross-road of emotion and cognition? *PLoS ONE*, 8(10), e78103.

Tarone, R. E. (1985). On heterogeneity tests based on efficient scores. *Biometrika*, 72(1), 91–95.

Vihman, M. M. (2014). *Phonological Development: The First Two Years* (2nd ed.). Chichester: Wiley-Blackwell.

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy*, 12(4), 844–858.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of*

the American Statistical Association, 22(158), 209–212.

Ylönen, T. (2022). Wiktextract: Wiktionary as machine-readable structured data. *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, 1317–1325. Data at https://kaikki.org

Reproducibility

All material is in `endolanguage/mother/`. The primary data come from `data/lexicon/lexibank/forms.csv`.

step	script	output
extraction of the 33 concepts, vowels stripped	`build.py`	`data/forms.csv`
the null over all of Lexibank	`baseline.py`	`data/baseline_codes.csv`
section 3	`analyze.py`, `plots.py`	`out/findings.md`, figures 1–5
sections 4–5	`nursery.py`, `plots_nursery.py`	`out/nursery.md`, figures 6–9
section 6	`groups.py`, `plots_groups.py`	`out/groups.md`, figures 10–12
section 7	`robustness.py`, `plots_robustness.py`	`out/robustness.md`, figures 14–15
section 8	`child.py`	`out/child.md`, figure 13
section 9	`endearment_scan.py`	`data/endearment_raw.csv`
intervals and effect sizes (all sections)	`effects.py`	`out/effects.md`, `out/effects.csv`

`robustness.py` requires LingPy (List *et al.* 2022) and PanPhon (Mortensen *et al.* 2016); `effects.py` requires SciPy and statsmodels. Computations involving randomisation use a fixed seed. The only step depending on resources external to the repository is that of section 9, which reads the Kaikki dumps in `data/lexicon/kaikki/dict/`.