

The Endolinguistic Method of Access

Descent through genealogies: structure, scope, and limits of the enquiry

Alejandro Toledo Martínez

Framing note. This is a methodological study within the endolinguistics programme (Meulemans–Elias tradition), not a new theory of it. It describes one way of accessing endolinguistics — a descent through genealogical strata — and specifies what can and cannot be claimed along that route. It keeps three levels apart: (i) the method (defensible now); (ii) an interpretive premise about the endpoint of the descent, a *latent common substrate*; and (iii) the future programme. “Access” is used in one precise sense, defined in §0: an epistemic-descriptive instrument, not a psychological or ontological reaching, and not a loose metaphor. The tradition’s identification of the substrate with a *shared unconscious* is treated strictly as a deferred premise. A full Spanish version follows the English one; shared technical appendices (experimental apparatus, robustness, reproducibility, glossary) close the paper. Coined terms are defined in Appendix C.

Abstract

This is a methodological study within the endolinguistics programme (Meulemans–Elias tradition), not a new theory of it. Two empirical results motivate it — a 4,721-language corpus analysis and a Germanic–Italic study of consonantal codes (both reproduced in Appendix A). The strong hypothesis the programme began with — that a consonantal code carries meaning as a residue above genealogy — is falsified: the corpus shows no such residue, and in the Germanic–Italic data the code-to-meaning coherence concentrates in the *deepest* form-sharing pairs (cognition below the reconstruction floor). We take this negative result at face value and draw a narrower, positive proposal: endolinguistics is best understood not as a semantic theory but as one method of access — an epistemic-descriptive instrument (§0) that reduces a form to consonantal-class coordinates and descends the genealogical strata (subject → language → branch → family → deeper proto-stages) toward a posited latent common substrate. On this reading genealogy is not a rival but the ladder of the descent; the programme rides on historical linguistics and inherits its horizon, making deep macro-comparison (e.g. Nostratic) the natural frontier and language isolates a hard limit. What endolinguistics adds over reconstruction is a comparative descriptive layer — how one shared code-inventory is loaded with *different* meaning-fields across sister branches (the *temperament* hypothesis, §4). A full robustness analysis (Appendix A.6) revises the headline magnitudes honestly: the trans-lineage cross-family signal, computed exactly and pair-weighted, is weak ($1.9\times$, 95% CI 1.2–3.0) rather than the $19\times$ of an earlier skeleton-averaged estimate — which strengthens, not weakens, the conclusion that it is areal/universal, not deep inheritance. We locate the success criterion in the philosophy of science (Lakatos, Laudan): a research programme is judged by the novel, failable predictions it generates (§8), not by a corpus residue. The identification of the substrate with a *shared unconscious* is retained as the tradition’s premise but deferred to separate philosophical work.

Keywords. endolinguistics; consonantal codes; historical linguistics; comparative method; cognation; sound symbolism; areal diffusion; Nostratic; language isolates; research programme; latent substrate; access by depth.

0. What "access" means

Because *access* is the load-bearing term, we fix its sense before using it. It is used here in exactly one way: as an epistemic-descriptive instrument — a reproducible procedure with a defined input, operation, and output.

- Input: a lexical form (or a lexicon).
- Operation: reduce it to its consonantal-class coordinates (a *code*; Appendix C) and align it across genealogical depth by the comparative method (cognition, reconstruction).
- Output: the form’s code, and — across many forms — the *distribution* of meaning over codes at each stratum (the *temperament* profile, §4). This output is reproducible and, per §8, can be wrong.

Three senses we explicitly do not intend. *Access* is not psychological (no claim that a speaker “accesses” anything); not ontological (descending does not reach, know, or prove the substrate — the endpoint stays a premise, §5); and not a loose metaphor. It is an instrument in the sense a coordinate system or a staining protocol is: justified instrumentally by what it lets one describe and predict, not by naming a new entity. The procedure’s output at a given depth is an endolinguistic profile — a coherent form–meaning description. Where the text says a profile “appears,” read *the coordinate description becomes available at that depth* — a fact about the output, not a manifestation to any observer. (The tradition reifies this profile as an *endolanguage*, an object-reading deferred like the substrate, §5.)

This also answers the fair objection “is *access* a new scientific object or just a heuristic?” Neither: it is a new instrument over the existing objects of historical linguistics. Its novelty is methodological, and the division of labour is explicit:

	historical linguistics	endolinguistics (this paper)	psychophonology (companion)
object	proto-forms, sound laws, family trees	the differential loading of shared codes across strata; the coordinate procedure	the psychic orientation of classes/codes on a subject
produces	reconstructions, cognate sets	code coordinates; temperament typology (§4)	idiolectal / clinical reading
status here	load-bearing ladder (used, not revised)	the methodological proposal	deferred (companion, in preparation)
judged by	the comparative method	fecundity + novel predictions (§8)	(developed elsewhere)

1. The thesis — and what it adds

The thesis proposes one way of working with endolinguistics: a methodology of *access by depth* (§0). Applying the instrument to a subject — a language, a concrete system — yields an endolinguistic profile: the coherent form–meaning description available at that depth (the object the tradition calls an *endolanguage* is its deferred reading). Descending by cognition and reconstruction, the branch is describable, then the family, then the next proto-stage; each depth is a *face* of one posited *latent common substrate*. The descent passes through the genealogies; it does not sidestep them. It is *one* route of access, not the definition of the discipline.

What this adds over historical linguistics. The honest answer is that it is not a rival explanation but a complementary descriptive layer (see the table in §0) with two distinctive objects:

1. a comparative typology of semantic distribution — how one shared inventory of consonantal codes is loaded with *different* meaning-fields across sister branches (the *temperament / intention-drift* hypothesis, §4). Reconstruction supplies the codes and the tree; it does not systematise this differential loading.
2. an applied use on the individual subject — reading an idiolect through the same coordinates. This applied layer is named, not developed here: it is a companion programme (psychophonology; §0 table), and the "subject" node in Figure 1 marks where the instrument would be applied, not a result claimed in this paper.

Neither claims a mechanism above genealogy. The value offered is descriptive-comparative, judged by fecundity (§8).

In candour, this paper defines that comparative layer; it does not yet demonstrate that the layer is fecund. Whether it yields results a linguist would find substantive is an open question that only the replication study of §8 (prediction 1) can settle. We claim a well-specified instrument, not yet a body of findings.

Three shifts found the thesis. From proof to access: we do not seek a semantic residue surviving genealogy (we looked; it is not there, §2) but a way of descending — the codes are the *coordinates*, not the object. From rival to ladder: genealogy is the ladder one climbs down; without historical linguistics there are no rungs. From object to premise: the substrate is not demonstrated but assumed and accessed toward; its identification with a *shared unconscious* is the tradition's hypothesis, deferred (§5).

2. Why: a negative result, taken at face value

The reframing is the correct reading of two hard results (both reproduced, with exact figures, in Appendix A):

- The corpus study (4,721 languages) sought a semantic residue *above* phonology and frequency and found none — every regularity reabsorbed into known phonology.
- The Germanic-Italic study tested the strong hypothesis (*the code carries meaning beyond cognition*). When two roots of the same code are required to have distinct real consonants — so they cannot share an ancestral skeleton — the code→meaning association falls to baseline (Appendix A.2). The coherence survives only where roots share a real skeleton: cognition deeper than we reconstruct.

We do not dress this up. The strong hypothesis is falsified: there is no code-borne meaning independent of genealogy. By itself this is a clean negative result, not evidence *for* anything. It does *redirect* the enquiry, and the redirection is not an ad-hoc rescue, for two reasons. First, the narrower proposal (access-by-depth) is logically independent of the falsified one: it does not need the residue; it needs only that cognation be real, which the same data confirm (within-family same-skeleton→same-concept runs at $47\times$ over chance, pair-weighted; Appendix A.6). Second, the new proposal is not self-immunising: it stakes novel, failable predictions (§8). The tempting move — "the signal lives in the deep, therefore the theory is confirmed" — we explicitly retract: concentration-in-depth is *consistent with* access-by-depth but does not confirm it, since it is equally consistent with plain cognation. The test of the reframing is §8, not §2.

3. The strata of the descent

Access travels a ladder of depths; at each, an *endolinguistic profile* is available — the description of the substrate’s face at that depth (Figure 1).

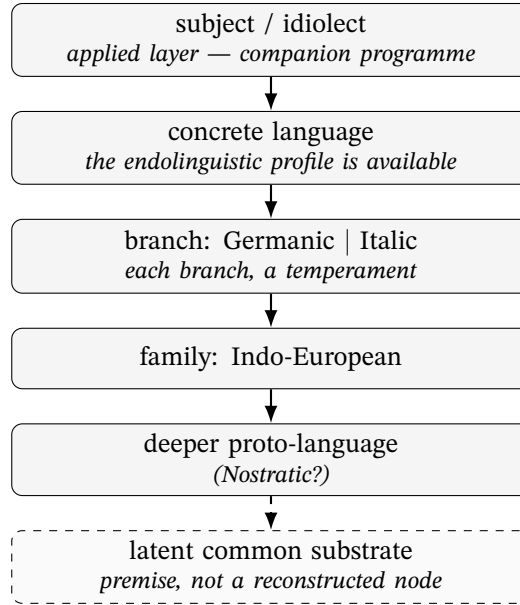


Figure 1: Access by depth: the descent through genealogical strata. Each step is deeper; the endpoint is a premise, not a reconstructed node, and the subject stratum marks where the applied (companion) layer would operate.

The endolinguistic profile is what the procedure describes at a given depth (the *endolanguage* is its deferred object-reading); the code (the consonantal dyad) is the coordinate by which forms are marked and compared across strata; the descent is passing from one stratum to the next by the genealogical route (cognition, reconstruction). It is neither a causal chain nor a hierarchy of truth: a series of depths of description over one posited origin.

4. The temperament hypothesis

The central descriptive claim — a hypothesis, not an established fact — is that the descent need not distribute meaning the same way at each stratum. Each branch’s profile may carry a distinctive branch-loading profile — what the programme calls, in its interpretive vocabulary, a *temperament*: its own way of loading meaning-fields onto the shared codes.

Operational definition. For a code k , let $D_{\text{fam}}(k)$ be the distribution of meaning-fields over the roots carrying k at the family level, and $D_b(k)$ the same distribution restricted to branch b . The branch’s *temperament* on k is the departure of $D_b(k)$ from $D_{\text{fam}}(k)$ — operationally, a divergence (e.g. Jensen–Shannon) between the two field-distributions. A branch *has* a temperament to the extent these departures are (a) larger than a permutation null that reassigns roots to branches at random, and (b) predictable: a blind coder, given a held-out meaning-field and its code, should assign it to the correct branch above chance. Both are testable (§8).

Current evidence — one study. The evidence so far is a single comparative study (Germanic–Italic); it is suggestive, not decisive, and the claim stands or falls on replication across further branch pairs. The observed pattern:

code	Indo-European	Germanic	Italic
K·R	hardness · shout · curve · seize · raw · rule	rule / rich / king · movement	heart / believe · field/grain · wheel
T·R	turn · access/boundary · tree→trust · water	access · movement	plant cluster · dry/heat
K·N	know · kinship · bend/angle	kinship · honey · bend	people (gens) · sing · sorrow · centre

Read cautiously: Italic *tends* toward the civil/inner (heart, believe, people, sorrow); Germanic toward the physical/social (rule, rich, movement). One study does not settle whether these departures beat the null or replicate; those are the §8 tests.

5. The status of the substrate — method vs. premise

The paper’s most delicate point, made explicit and kept apart from the method. The descent is demonstrated to work within a lineage: cognation is real (A.6). What is *not* demonstrated is the final step — that the descent’s vanishing-point is a *substrate* of any kind. That identification is an interpretive premise, not an empirical result, and we flag the jump as a jump.

A functional definition (its nature deferred). We define the *latent common substrate* only by its role in the method: it is the posited vanishing-point of the descent — the as-yet-unreconstructed limit toward which the genealogies of a form’s code appear to converge. We deliberately do not fix its nature: it may be read as cognitive, biological, phenomenological, informational, or metaphysical, and this paper commits to none. The method needs it only as a regulative posit (a placeholder endpoint that orients the descent), not as a characterised entity; every result in §§1–4 and §8 goes through under any of those readings, or under none. Two readings remain open:

- (a) the endpoint simply is the deepest genealogical stratum (pre-PIE, Nostratic...). Then endolinguistics is historical linguistics with a ladder and a premise — modest and coherent.
- (b) the genealogy points *through* itself to a principle it reveals but does not exhaust (why those root-families settled into those forms and those meanings). Then the substrate is not a genealogical object.

The residue that neither deep cognation nor semantic drift exhausts — “*why these forms and not others?*” — is where (b) would live. It is not resolved, and not resolvable by a corpus null.

The latent common substrate is a *premise*, assessed by fecundity (§8), not by a statistical residue. The tradition further identifies it with a shared unconscious; that identification is a philosophical commitment, treated here strictly as a deferred premise and developed elsewhere. Nothing in the method depends on it — a reader who rejects the identification, or leaves the substrate’s nature open, can still use every result in §§1–4 and §8.

Orientation and aims, stated openly. The *method* above is deliberately formal and interpretation-neutral: every empirical claim stands without any commitment about the substrate’s nature (§0, and the paragraph above). It would be dishonest, though, to present the *programme* as free of a horizon. Endolinguistics, in the Meulemans–Elias tradition, does have aims and affinities in ontology, psychology, and psychoanalysis — the *shared unconscious*, the subject stratum, the reading of an idiolect. We neither hide nor disown them: they are the programme’s motivation and its long-range horizon, held as aims and premises, not as results, and kept out of the method’s current claims. Separating the levels (framing note) is not a way of denying this orientation but the discipline that lets it be avowed without contaminating what the method can, today, defend.

6. The limit-phenomenon: non-cognate convergences (briefly)

One phenomenon the descent cannot, by construction, explain by genealogy: convergences between forms that are not cognate and are already separate at the deepest stratum we reconstruct. The stock example — Latin *terra* (< *ters-* 'dry', code *T·R*) and English *earth** (< a different root, code *R·T*) — converges in meaning and mirror-code without a common PIE ancestor. Two clarifications, both essential:

- Empirically, these are chance so far. The null over this class $\approx$ chance; same-code non-cognate pairs fall to baseline (A.2). Individual cases like *earth/terra* are almost certainly coincidence plus selection bias, and we keep them out of the findings.
- As a class, the question survives — but the effect is weak. We tested it at scale (A.3–A.6). Magnitude depends on how skeletons are weighted, and we report the more defensible pair-weighted exact figure as primary: same-skeleton $\rightarrow$ same-concept runs at $47\times$ within families (cognation) and only $1.9\times$ between unrelated families (95% CI 1.2–3.0; exact computation + family bootstrap, A.6). An earlier skeleton-averaged estimate inflated these to $317\times$ and $19\times$, but that over-weights rare, idiosyncratic skeletons (A.6). Stratifying by geography (A.5) the trans-lineage signal tracks areal distance, not the genealogical tree: it declines toward the universal floor with distance (sound-symbolism; Blasi et al. 2016).

A trans-lineage effect that is already weak ($2\times$) and shrinks with geographic distance is exactly what areal contact + universal convergence predict; neither is deep vertical inheritance, and the modest size leaves *less* room for it, not more. Whether a genuine deep-inheritance component *also* exists is undecidable at the reconstruction horizon: testing it would require a deep tree that dissociates genealogy from geography, and none exists — the available deep trees (Nostratic, Borean) are partly built from areal similarity, confounded with geography by construction. The hypothesis of an ancient genealogical tie is therefore neither confirmed nor refuted; it is the programme's limit-phenomenon (§7). This cuts both ways: within a lineage the descent works (A.6); only the crossing between deep lineages is undecidable — exactly where §5 located the premise.

7. The challenge of depth — historical linguistics as load-bearing

Universality, in principle. By construction, every human language admits the method of access: reduce it to its consonantal coordinates and descend by its genealogy. There is no language "without an endolinguistic profile."

The bottleneck: the reconstruction horizon. The descent reaches only as deep as reconstruction permits. Historical linguistics has a horizon (of the order of a few millennia); below it, reconstruction becomes uncertain and the ladder ends. Endolinguistic access inherits that horizon: we cannot descend to a stratum we cannot reconstruct.

Nostratic. Verifying or bounding macro-families such as Nostratic — or any proto-grouping deeper than IE — would extend the ladder one rung lower, so the programme has a direct interest in macro-comparison. In honesty, Nostratic is not accepted by the mainstream and is methodologically fragile (Campbell & Poser 2008); it is a frontier, not a foundation, but it is the natural frontier of the descent.

Language isolates. An isolate (Basque, Burushaski, and others) has no known relatives (Campbell 2010): no ladder, no intermediate strata to descend by. The method, which lives on genealogical descent, is left without rungs — an open problem: what is endolinguistic access for an isolate, beyond its language-stratum and idiolect?

In short, historical linguistics is neither a rival nor an ornament of endolinguistics but its load-bearing structure: it provides the ladder and fixes its limit. The programme advances with reconstruction and is bounded by its horizon and by isolates.

8. Falsifiability, and the criterion of a research programme

Endolinguistics is a research programme, not a single testable proposition, and is best judged by the standards developed for such programmes in the philosophy of science. Following Lakatos (1978), a programme is *progressive* only if it predicts novel facts, and *degenerating* if it merely accommodates known ones after the fact; following Laudan (1977), its worth lies in its problem-solving effectiveness; the individual empirical claims within it remain falsifiable in Popper's (1959) sense. On this yardstick the falsified strong hypothesis (§2) was a degenerating move; the access-by-depth reframing earns its keep only if it is progressive. It therefore commits to concrete, failable predictions:

1. Temperament replication. In an independent branch pair (e.g. Slavic vs Indo-Iranian, or Celtic vs Germanic), the field-distribution departures of §4 should exceed a branch-permutation null, and a blind coder should assign held-out fields to the correct branch above chance. No excess over null, or chance-level assignment, weakens the programme.
2. Directionality of drift. The civil/inner vs physical/social contrast (§4) predicts the *direction* of specific field assignments in a held-out branch; wrong-direction predictions count against it.
3. Subject-level structure (deferred). On an individual, descent by codes should reveal idiolectal structure that independent raters partly agree on and that predicts something outside the coding. This belongs to the companion applied programme (§0) and is stated here as a direction, not tested.
4. No corpus residue (already met, and it must stay met). The programme now *predicts* that no code-borne meaning survives a cognation control. Were such a residue to appear under proper controls, the reframing — not only the old hypothesis — would be in trouble.

The empirical anchors are in the accessible strata (temperament, and the deferred subject layer), not in a residue above genealogy. What stays open, and is *not* charged as demonstrated, is the nature of the endpoint (§5).

In one sentence

Endolinguistics is a method of access — an instrument that reduces a form to consonantal coordinates and descends *through* the genealogies, making the branch, the family and the proto-stages describable, face after face of one posited latent common substrate; historical linguistics is the load-bearing ladder that both enables and limits the descent, so its frontier (macro-comparison) and its limit (the reconstruction horizon, isolates) are the programme's own — and its worth is judged, as a research programme, by the novel predictions it makes (§8).

La Endolingüística como Metodología de Acceso

El descenso por las genealogías: estructura, alcance y límites de la investigación

(Versión en español del estudio anterior.)

Resumen

Estudio metodológico dentro del programa endolingüístico (tradición Meulemans–Elias), no una teoría nueva. Lo motivan dos resultados empíricos —el estudio de corpus de 4.721 lenguas y el germánico-itálico de códigos consonánticos (ambos reproducidos en el Apéndice A). La hipótesis fuerte con que empezó el programa —que un código consonántico carga sentido como residuo por encima de la genealogía— queda falsada: el corpus no muestra tal residuo, y en los datos germánico-itálicos la coherencia código→sentido se concentra en los pares que comparten forma más profunda (cognación bajo el piso de reconstrucción). Tomamos ese resultado negativo tal cual y extraemos de él una propuesta más modesta y positiva: la endolingüística se entiende mejor no como teoría semántica sino como un método de acceso —un instrumento epistémico-descriptivo (§0) que reduce una forma a coordenadas de clase consonántica y desciende por los estratos genealógicos (sujeto → lengua → rama → familia → protolenguas más profundas) hacia un substrato común latente postulado. Así la genealogía no es un rival sino la escalera del descenso; el programa cabalga sobre la lingüística histórica y hereda su horizonte, lo que hace del Nostrático el frente natural y de las lenguas aisladas un límite duro. Lo que la endolingüística aporta sobre la reconstrucción es una capa descriptivo-comparativa —cómo un mismo inventario de códigos se carga con campos de sentido *distintos* según la rama (la hipótesis del *temperamento*, §4). Un análisis de robustez (Apéndice A.6) corrige las magnitudes con honestidad: la señal trans-linaje entre familias, computada de forma exacta y pesada por pares, es débil ($1,9\times$, IC 95% 1,2–3,0) y no el $19\times$ de una estimación previa promediada-por-esqueleto —lo que refuerza, no debilita, la conclusión de que es areal/universal, no herencia profunda. Situamos el criterio de éxito en la filosofía de la ciencia (Lakatos, Laudan): un programa de investigación se juzga por las predicciones novedosas y falsables que genera (§8), no por un residuo de corpus. La identificación del substrato con un *inconsciente compartido* se conserva como premisa de la tradición, pero se difiere a un trabajo filosófico aparte.

Palabras clave. endolingüística; códigos consonánticos; lingüística histórica; método comparativo; cognación; simbolismo sonoro; difusión areal; nostrático; lenguas aisladas; programa de investigación; substrato latente; acceso por profundidad.

0. Qué significa «acceso»

Como *acceso* es el término portante, fijamos su sentido antes de usarlo. Se usa aquí de una sola manera: como un instrumento epistémico-descriptivo —un procedimiento reproducible con entrada, operación y salida definidas.

- Entrada: una forma léxica (o un léxico).
- Operación: reducirla a sus coordenadas de clase consonántica (un *código*; Apéndice C) y alinearla a través de la profundidad genealógica por el método comparativo (cognación, reconstrucción).

- Salida: el código de la forma y —sobre muchas formas— la *distribución* del sentido sobre los códigos en cada estrato (el perfil de *temperamento*, §4). Esa salida es reproducible y, según el §8, puede estar equivocada.

Tres sentidos que no intentamos. *Acceso* no es psicológico (no se afirma que un hablante “acceda” a nada); no es ontológico (descender no alcanza, conoce ni prueba el substrato —el punto final sigue siendo premisa, §5); y no es una metáfora vaga. Es un instrumento en el sentido en que lo es un sistema de coordenadas o un protocolo de tinción: se justifica instrumentalmente por lo que permite describir y predecir, no por nombrar una entidad nueva. La salida del procedimiento a una profundidad dada es un perfil endolingüístico —una descripción coherente de forma y sentido. Cuando el texto dice que un perfil “aparece”, léase *la descripción por coordenadas se hace disponible a esa profundidad* —un hecho sobre la salida, no una manifestación a ningún observador. (La tradición reifica ese perfil como un *endolenguaje*, una lectura-objeto diferida como el substrato, §5.)

Esto responde también a la objeción justa “¿es *acceso* un objeto científico nuevo o una heurística?”. Ninguna: es un instrumento nuevo sobre los objetos ya existentes de la lingüística histórica. Su novedad es metodológica, y el reparto de trabajo es explícito:

	lingüística histórica	endolingüística (este paper)	psicofonología (compañera)
objeto	protoformas, leyes fonéticas, árboles	la carga diferencial de códigos comunes entre estratos; el procedimiento de coordenadas	la orientación psíquica de clases/códigos sobre un sujeto
produce	reconstrucciones, conjuntos de cognados	coordenadas de código; tipología de temperamentos (§4)	lectura idiolectal / clínica
estatus aquí	escalera portante (usada, no revisada)	la propuesta metodológica	diferida (compañera, en preparación)
se juzga por	el método comparativo	fecundidad + predicciones novedosas (§8)	(desarrollado en otro lugar)

1. La tesis — y qué aporta

La tesis propone un modo de trabajar la endolingüística: una metodología de *acceso por profundidad* (§0). Al aplicar el instrumento a un sujeto —una lengua, un sistema concreto— resulta un perfil endolingüístico: la descripción coherente de forma y sentido disponible a esa profundidad (el objeto que la tradición llama *endolenguaje* es su lectura diferida). Al descender por cognación y reconstrucción se hacen describibles la rama, la familia, la siguiente protolengua; cada profundidad es una *cara* de un mismo *substrato común latente* postulado. El descenso pasa por las genealogías; no las esquiva. Es una vía de acceso, no la definición de la disciplina.

Qué aporta sobre la lingüística histórica. No es una explicación rival sino una capa descriptiva complementaria (véase la tabla del §0) con dos objetos propios:

1. una tipología comparativa de la distribución de sentido —cómo un mismo inventario de códigos consonánticos se carga con campos de sentido *distintos* según la rama (la hipótesis del *temperamento / deriva de intención*, §4). La reconstrucción da los códigos y el árbol; no sistematiza esa distribución diferencial.
2. un uso aplicado sobre el sujeto individual —leer un idiolecto por las mismas coordenadas. Esta capa aplicada está nombrada, no desarrollada aquí: es un programa compañero (psicofonología; tabla §0), y el nodo “sujeto” de la Figura 1 marca dónde se aplicaría el instrumento, no un resultado que este paper reclame.

Ninguno afirma un mecanismo por encima de la genealogía. El valor es descriptivo-comparativo, y se juzga por fecundidad (§8).

Con franqueza, este paper define esa capa comparativa; todavía no demuestra que la capa sea fecunda. Si arroja resultados que un lingüista halle sustantivos es una pregunta abierta que solo el estudio de replicación del §8 (predicción 1) puede zanjar. Reclamamos un instrumento bien especificado, aún no un cuerpo de hallazgos.

Tres desplazamientos fundan la tesis. De la prueba al acceso: no buscamos un residuo semántico que sobreviva a la genealogía (lo buscamos; no está, §2) sino un modo de descender —los códigos son las *coordenadas*, no el objeto. De rival a escalera: la genealogía es la escalera por la que se baja; sin lingüística histórica no hay peldaños. De objeto a premisa: el substrato no se demuestra sino que se asume y se accede hacia él; su identificación con un *inconsciente compartido* es la hipótesis de la tradición, diferida (§5).

2. Por qué: un resultado negativo, tomado tal cual

El reencuadre es la lectura correcta de dos resultados duros (ambos reproducidos, con cifras exactas, en el Apéndice A):

- El estudio de corpus (4.721 lenguas) buscó un residuo semántico *por encima* de la fonología y la frecuencia y no lo halló: cada regularidad se reabsorbió en fonología conocida.
- El estudio germánico-italico probó la hipótesis fuerte (*el código carga sentido más allá de la cognación*). Cuando se exige que dos raíces del mismo código tengan consonantes reales distintas —de modo que no puedan compartir un esqueleto ancestral— la asociación código→sentido cae a la línea base (Apéndice A.2). La coherencia solo sobrevive donde las raíces comparten esqueleto real: cognación más profunda de la que reconstruimos.

No lo maquillamos. La hipótesis fuerte queda falsada: no hay sentido portado por el código independiente de la genealogía. En sí mismo es un resultado negativo limpio, no evidencia *a favor* de nada. Pero *redirige* la investigación, y la redirección no es un rescate ad-hoc, por dos razones. Primero, la propuesta más modesta (acceso-por-profundidad) es lógicamente independiente de la falsada: no necesita el residuo; solo necesita que la cognación sea real, cosa que los mismos datos confirman (dentro de familia, mismo-esqueleto→mismo-concepto corre a $47\times$ sobre el azar, pesado por pares; Apéndice A.6). Segundo, la nueva propuesta no se autoinmuniza: apuesta predicciones novedosas y falsables (§8). El movimiento tentador —“la señal vive en lo profundo, luego la teoría queda confirmada”— lo retratamos explícitamente: la concentración-en-profundidad es *compatible* con el acceso-por-profundidad pero no lo confirma, pues es igualmente compatible con la mera cognación. La prueba del reencuadre es el §8, no el §2.

3. Los estratos del descenso

El acceso recorre una escalera de profundidades; en cada una un *perfil endolingüístico* está disponible —la descripción de la cara del substrato a esa profundidad (Figura 1). El perfil endolingüístico es lo que el procedimiento describe a una profundidad dada (el *endolenguaje* es su lectura-objeto diferida); el código (la díada consonántica) es la coordenada con que se marcan y comparan las formas a través de los estratos; el descenso es pasar de un estrato al siguiente por la vía genealógica (cognación, reconstrucción). No es una cadena causal ni una jerarquía de verdad: es una serie de profundidades de descripción sobre un mismo origen postulado.

4. La hipótesis del temperamento

La afirmación descriptiva central —una hipótesis, no un hecho establecido— es que el descenso no tiene por qué repartir el sentido igual en cada estrato. El perfil de cada rama puede portar un perfil de carga distintivo —lo que el programa llama, en su vocabulario interpretativo, un *temperamento*: su modo propio de cargar campos de sentido sobre los códigos comunes.

Definición operacional. Para un código k , sea $D_{\text{fam}}(k)$ la distribución de campos de sentido sobre las raíces que portan k a nivel de familia, y $D_b(k)$ la misma restringida a la rama b . El *temperamento* de la rama sobre k es la desviación de $D_b(k)$ respecto de $D_{\text{fam}}(k)$ —operacionalmente, una divergencia (p. ej. Jensen–Shannon) entre ambas distribuciones. Una rama *tiene* temperamento en la medida en que esas desviaciones son (a) mayores que un nulo de permutación que reasigna raíces a ramas al azar, y (b) predecibles: un codificador ciego, dado un campo reservado y su código, debería asignarlo a la rama correcta por encima del azar. Ambas son testables (§8).

Evidencia actual — un estudio. La evidencia hasta ahora es un solo estudio comparativo (germánico-itálico); es sugerente, no decisiva, y la afirmación se sostiene o cae según la replicación en más pares de ramas. El patrón observado:

código	Indoeuropeo	Germánico	Itálico
K·R	dureza · grito · curva · agarrar · crudo · regir	regir / rico / rey · movimiento	corazón / creer · campo/grano · rueda
T·R	giro · acceso/límite · árbol→confianza · agua	acceso · movimiento	racimo vegetal · seco/calor
K·N	conocer · parentesco · doblar/ángulo	parentesco · miel · doblar	gente (gens) · cantar · pena · centro

Con cautela: el Itálico *tiende* a lo civil/interior (corazón, creer, gente, pena); el Germánico a lo físico/social (regir, rico, movimiento). Un estudio no zanja si esas desviaciones batan el nulo ni si replican; esas son las pruebas del §8.

5. El estatus del substrato — método vs. premisa

El punto más delicado del paper, explícito y separado del método. El descenso está demostrado que funciona dentro de un linaje: la cognación es real (A.6). Lo que *no* está demostrado es el paso final —que el punto de fuga del descenso sea un *substrato* de ningún tipo. Esa identificación es una premisa interpretativa, no un resultado empírico, y señalamos el salto como salto.

Una definición funcional (su naturaleza diferida). Definimos el *substrato común latente* solo por su papel en el método: es el punto de fuga postulado del descenso —el límite aún no reconstruido hacia el que parecen converger las genealogías del código de una forma. Deliberadamente no fijamos su naturaleza: puede leerse como cognitiva, biológica, fenomenológica, informacional o metafísica, y este paper no se compromete con ninguna. El método lo necesita solo como posit regulativo (un punto final marcador que orienta el descenso), no como entidad caracterizada; todo resultado de §§1–4 y §8 se sostiene bajo cualquiera de esas lecturas, o bajo ninguna. Quedan dos lecturas abiertas:

- (a) el punto final simplemente *es* el estrato genealógico más profundo (pre-PIE, nostrático...). Entonces la endolingüística es lingüística histórica con una escalera y una premisa —modesta y coherente.
- (b) la genealogía apunta *a través* de sí misma a un principio que revela pero no agota (por qué esas familias-de-raíz se asentaron en esas formas y sentidos). Entonces el substrato no es un objeto genealógico.

El residuo que ni la cognación profunda ni la deriva semántica agotan —“¿por qué estas formas y no otras?”— es donde viviría (b). No está resuelto, y no se resolverá por un nulo de corpus.

El substrato común latente es una *premisa*, evaluable por su fecundidad (§8), no por un residuo estadístico. La tradición lo identifica además con un inconsciente compartido; esa identificación es un compromiso filosófico, tratado aquí estrictamente como premisa diferida y desarrollado en otro lugar. Nada del método depende de ella —quien rechace la identificación, o deje abierta la naturaleza del substrato, puede usar igualmente todos los resultados de §§1–4 y §8.

Orientación y fines, dicho abiertamente. El *método* anterior es deliberadamente formal y neutral en cuanto a interpretación: toda afirmación empírica se sostiene sin compromiso alguno sobre la naturaleza del substrato (§0, y el párrafo anterior). Sería deshonesto, sin embargo, presentar el *programa* como libre de un horizonte. La endolingüística, en la tradición Meulemans–Elias, sí tiene fines y afinidades en ontología, psicología y psicoanálisis —el *inconsciente compartido*, el estrato del sujeto, la lectura de un idiolecto. No los ocultamos ni los renegamos: son la motivación del programa y su horizonte de largo alcance, sostenidos como fines y premisas, no como resultados, y mantenidos fuera de las afirmaciones actuales del método. Separar los niveles (nota de encuadre) no es un modo de negar esta orientación, sino la disciplina que permite confesarla sin contaminar lo que el método puede, hoy, defender.

6. El fenómeno-límite: convergencias no-cognadas (breve)

Un fenómeno que el descenso, por construcción, no puede explicar por genealogía: convergencias entre formas que no son cognadas y ya están separadas en el estrato más profundo que reconstruimos. El caso típico —el latín *terra* (< *ters-* ‘seco’, código T·R) y el inglés *earth** (< una raíz distinta, código R·T)— converge en sentido y código en espejo sin ancestro PIE común. Dos aclaraciones, ambas esenciales:

- Empíricamente, hasta ahora son azar. El nulo sobre esta clase $\approx$ azar; los pares del mismo código no-cognados caen a la línea base (A.2). Casos individuales como *earth/terra* son casi con seguridad coincidencia más sesgo de selección, y los mantenemos fuera de los hallazgos.
- Como clase, la pregunta persiste —pero el efecto es débil. Lo probamos a escala (A.3–A.6). La magnitud depende de cómo se pesen los esqueletos, y reportamos como primaria la cifra pesada por pares, más defendible: mismo-esqueleto $\rightarrow$ mismo-concepto corre a $47\times$ dentro de familias (cognación) y solo $1,9\times$ entre familias no emparentadas (IC 95% 1,2–3,0; cómputo exacto + bootstrap sobre familias, A.6). Una estimación previa promediada-por-esqueleto los inflaba a $317\times$ y $19\times$, pero eso sobre-pesa esqueletos raros e idiosincráticos (A.6). Al estratificar por geografía (A.5) la señal trans-linaje sigue la distancia areal, no el árbol genealógico: decae hacia el piso universal con la distancia (simbolismo sonoro; Blasi et al. 2016).

Un efecto trans-linaje ya de por sí débil ($2\times$) y que mengua con la distancia geográfica es exactamente lo que predicen contacto areal + convergencia universal; ninguno es herencia vertical profunda, y su tamaño modesto deja *menos* sitio para ella, no más. Si además existe un componente de herencia profunda es indecible en el horizonte de reconstrucción: probarlo exigiría un árbol profundo que disocie genealogía de geografía, y no existe —los árboles profundos disponibles (Nostrático, Boreano) se construyen en parte de semejanzas areales, confundidos con la geografía por construcción. La hipótesis de una liga genealógica antiquísima queda, por tanto, ni confirmada ni refutada; es el fenómeno-límite del programa (§7). Y corta en ambos sentidos: dentro de un linaje el descenso funciona (A.6); solo el cruce entre linajes profundos es indecible —justo donde el §5 situó la premisa.

7. El reto de la profundidad — la lingüística histórica como estructura portante

Universalidad, en principio. Por construcción, toda lengua humana admite el método de acceso: se reduce a sus coordenadas consonánticas y se desciende por su genealogía. No hay lengua "sin perfil endolingüístico".

El cuello de botella: el horizonte de reconstrucción. El descenso llega solo tan hondo como permita la reconstrucción. La lingüística histórica tiene un horizonte (del orden de unos milenios); por debajo, la reconstrucción se vuelve incierta y la escalera se acaba. El acceso endolingüístico hereda ese horizonte: no podemos descender a un estrato que no podemos reconstruir.

El Nostrático. Verificar o acotar macro-familias como el Nostrático —o cualquier agrupación de protolenguas más profunda que el IE— extendería la escalera un peldaño más abajo, de modo que el programa tiene un interés directo en la comparación macro. En honestidad, el Nostrático no está aceptado por la corriente principal y es metodológicamente frágil (Campbell y Poser 2008); es un frente, no un cimiento, pero es el frente natural del descenso.

Las lenguas aisladas. Una lengua aislada (vasco, burushaski, y otras) no tiene parientes conocidos (Campbell 2010): no hay escalera, no hay estratos intermedios por los que descender. El método, que vive del descenso genealógico, se queda sin peldaños —un problema abierto: ¿qué es el acceso endolingüístico para un aislado, más allá de su estrato-lengua y su idiolecto?

En suma, la lingüística histórica no es un rival ni un adorno de la endolingüística sino su estructura portante: provee la escalera y fija su límite. El programa avanza con la reconstrucción y está acotado por su horizonte y por los aislados.

8. Falsabilidad, y el criterio de un programa de investigación

La endolingüística es un programa de investigación, no una proposición testable única, y se juzga mejor por los estándares desarrollados para tales programas en la filosofía de la ciencia. Siguiendo a Lakatos (1978), un programa es *progresivo* solo si predice hechos novedosos, y *degenerativo* si se limita a acomodar los ya conocidos a posteriori; siguiendo a Laudan (1977), su valor está en su eficacia para resolver problemas; las afirmaciones empíricas individuales siguen siendo falsables en el sentido de Popper (1959). Con esa vara, la hipótesis fuerte falsada (§2) fue un movimiento degenerativo, y el reencuadre acceso-por-profundidad se gana su lugar solo si es progresivo. Por eso se compromete con predicciones concretas y falsables:

1. Replicación del temperamento. En un par de ramas independiente (p. ej. eslavo vs indoiranio, o celta vs germánico), las desviaciones de distribución de campos del §4 deberían superar un nulo de permutación de ramas, y un codificador ciego debería asignar campos reservados a la rama correcta por encima del azar. Sin exceso sobre el nulo, o con asignación al nivel del azar, el programa se debilita.
2. Direccionalidad de la deriva. El contraste civil/interior vs físico/social (§4) predice la *dirección* de asignaciones de campo concretas en una rama reservada; las predicciones en dirección equivocada cuentan en contra.
3. Estructura a nivel del sujeto (diferida). En un individuo, el descenso por códigos debería revelar estructura idiolectal en la que evaluadores independientes coincidan en parte y que prediga algo fuera de la codificación. Esto pertenece al programa aplicado compañero (§0) y se enuncia aquí como dirección, no se prueba.
4. Sin residuo de corpus (ya cumplido, y debe seguir cumpliéndose). El programa ahora *predice*

que ningún sentido portado por el código sobrevive a un control de cognación. Si tal residuo apareciera bajo controles adecuados, el reencuadre —no solo la hipótesis vieja— estaría en problemas.

Los anclajes empíricos están en los estratos accesibles (temperamento, y la capa-sujeto diferida), no en un residuo por encima de la genealogía. Lo que queda abierto, y *no* se cobra como demostrado, es la naturaleza del punto final (§5).

En una frase

La endolingüística es un método de acceso —un instrumento que reduce una forma a coordenadas consonánticas y desciende *por* las genealogías, haciendo describibles la rama, la familia y las protolenguas, cara tras cara de un mismo substrato común latente postulado; la lingüística histórica es la escalera portante que a la vez posibilita y limita el descenso, de modo que su frente (la comparación macro) y su límite (el horizonte de reconstrucción, los aislados) son los del programa —y su valor se juzga, como programa de investigación, por las predicciones novedosas que hace (§8).

Appendices · Apéndices

The technical apparatus below is shared by both language versions. The numbers are the exact output of the scripts in Appendix B, reproducible under a fixed seed. · El aparato técnico siguiente es compartido por ambas versiones. Las cifras son la salida exacta de los scripts del Apéndice B, reproducibles con semilla fija.

Appendix A — Experimental apparatus: the discriminant battery

The reframing in §2 and §6 rests on a battery of null tests. They were not designed to prove endolinguistics; they were designed to discriminate *what kind of thing* the observed code→meaning coherence is, among four rival explanations: H1 deep cognation (within-lineage inheritance), H2 universal iconicity / sound-symbolism (Blasi et al. 2016), H4 areal diffusion (contact between geographic neighbours), and a hypothetical novel endolinguistic residue above all of these. Each test isolates one contrast. All report a ratio (observed ÷ null-expected) as the honest effect size. A.1–A.5 are Monte-Carlo estimates under a *skeleton-averaged* sampler; A.6 recomputes the headline exactly, pair-weighted, with a bootstrap and sensitivity analysis, and is the figure to cite.

Common method. Meaning is compared across forms/roots that share a consonantal *skeleton* or *code*; the null permutes the meaning↔code assignment (`random.seed(1)`) and re-measures. Data sources and exact commands are in Appendix B.

A.1 — Co-orientation null (is code → meaning above chance, within a lineage?). For reconstructed roots of each stratum, mean pairwise semantic similarity (Jaccard over gloss-bags) among roots sharing a binary code vs a shuffled null.

stratum	level	roots	pairs	observed	null	ratio
PIE	dyad (binary code)	428	4,775	0.0075	0.0061 ± 0.0002	1.23× (z +6.7)
PIE	family (all)	1,054	24,028	0.0067	0.0057 ± 0.0001	1.16×
Germanic	dyad	418	3,768	0.0073	0.0063 ± 0.0002	1.16×
Italic	dyad	399	4,299	0.0072	0.0064 ± 0.0003	1.12×

Reading: a real but weak association (10–23 % over chance). Stable across seeds (PIE dyad 1.23× under seeds 1–2). Codes exert only a faint semantic pull within a lineage — not a strong law.

A.2 — Deep-cognation discriminator (H1): does the signal need form-closeness? Among PIE roots (878), same-code pairs split by longest-common-subsequence of the real skeleton — form-close (LCS ≥ 2) vs form-far (LCS < 2) — vs a different-code baseline.

bin	semantic similarity	ratio	n (pairs)
baseline (different code)	0.0055	1.00×	≈188,763
same code · form-close (LCS ≥ 2)	0.0075	1.36×	8,323
same code · form-far (LCS < 2)	0.0058	1.06×	13,267

Correlation of form- with meaning-similarity among same-code pairs: $r = +0.083$. Across seeds 1–7 the ratios are stable (form-close 1.36–1.38×, form-far 1.06–1.07×). *Reading:* the signal concentrates in form-close pairs and evaporates to baseline when forms diverge → supports H1: the coherence is cognation, not a form-independent code.

A.3 — Cross-family discriminator (H1 vs H2): does it survive across unrelated families? Over the full corpus (1,225,467 forms; 4,721 languages; 207 families; 3,160 Concepticon concepts), P(same con-

cept | same skeleton) vs a random-pair baseline, skeleton-averaged (each distinct skeleton drawn uniformly).

condition	P(same concept)	ratio
baseline (random pairs)	0.0012	1.00×
same skeleton · within family	0.4067	338.95× (cognition, expected)
same skeleton · between unrelated families	0.0228	18.98×
within family, IE only	0.3600	299.99×
within family, non-IE only	0.4135	344.55×

Reading (with the A.6 caveat): within-family is cognation (trivial); the cross-family ratio is a *real* trans-lineage residue — above chance where cognation cannot operate — but its magnitude here is inflated by the skeleton-averaged weighting. A.6 recomputes it exactly and pair-weighted (47× within, 1.9× cross); the honest headline is A.6’s. It is either H2 (universal iconicity) or H4 (areal); discriminated in A.4–A.5.

A.4 — Nostratic-stratification test (the inheritance hypothesis). If the effect were deep inheritance, it should decay with genealogical distance along a macro-tree. Cross-family same-skeleton pairs stratified by Nostratic relatedness (skeleton-averaged).

stratum	P(same concept)	ratio	pairs
baseline (random cross-family)	0.0010	1.0×	—
A · same Nostratic branch	0.0249	24.4×	100,969
B · different Nostratic branch	0.0253	24.7×	308,655
C · outside Nostratic	0.0075	7.3×	6,800,677

Reading: A ≈ B (no monotone decay along the tree) but both above C — superficially “a Nostratic boost,” ambiguous between inheritance and geography (Nostratic families are also neighbours). Controlled in A.5.

A.5 — Inheritance vs areal control (geography). Among Nostratic↔Nostratic cross-family pairs, stratified by great-circle distance (haversine) between family centroids. Inheritance should persist regardless of distance; areal contact should collapse with distance (skeleton-averaged; the *relative* collapse is the point, not the absolute magnitudes — cf. A.6).

condition	P(same concept)	ratio	pairs
baseline	0.0010	1.0×	—
universal floor (outside Nostratic)	0.0075	7.3×	6,800,677
Nostr ↔ Nostr · close (< 3,000 km)	0.0267	26.1×	119,100
Nostr ↔ Nostr · mid (3–6k km)	0.0271	26.5×	249,893
Nostr ↔ Nostr · far (> 6,000 km)	0.0090	8.7×	40,631

Distant Nostratic pairs, individually: Indo-European↔Koreanic 0.5×, Afro-Asiatic↔Mongolic 0.7×, Afro-Asiatic↔Japonic 2.2×, Indo-European↔Japonic 3.8×. *Reading:* the boost tracks geography, not the tree — distant Nostratic pairs collapse to the universal floor. → H4 (areal) over the H2 universal floor (sound-symbolism), not deep inheritance.

A.6 — Robustness of the headline (exact computation, bootstrap, sensitivity). The A.3 ratios are Monte-Carlo estimates under a *skeleton-averaged* sampler (each distinct skeleton weighted equally). We recompute them without sampling and under a more defensible weighting (script `src/gi/robustness.py`, seed 1).

check	within-family	cross-family
A.3 Monte-Carlo (skeleton-averaged)	338.95×	18.98×
exact, skeleton-averaged (confirms A.3)	317×	19.0×
exact, pair-weighted (each pair equal)	47.0×	1.87×
family bootstrap (B = 200, 50 % of 207 families), pair-weighted	44.0× [95 % CI 28–70]	1.72× [95 % CI 1.23–2.99]
sensitivity: drop the 20 most frequent skeletons	55.0×	1.98×

Reading — two lessons. (i) The dramatic 339×/19× are an artifact of the skeleton-averaged weighting, which gives a rare, idiosyncratic skeleton the same weight as a ubiquitous one; under the more defensible pair-weighted exact computation the effects are far smaller — within-family cognation 47×, cross-family only 1.9×. (ii) The cross-family effect is nonetheless real, not noise: its family-bootstrap 95 % CI [1.23, 2.99] excludes 1×, and dropping the 20 commonest skeletons leaves it unchanged (1.98×). The §6 conclusion is thus *reinforced*: the trans-lineage signal is a weak (2×) universal/areal floor, with no room for a large deep-inheritance component. Wherever a headline number is quoted, the pair-weighted figure is the honest one; the skeleton-averaged figures (A.3–A.5) are retained for comparability and because the geographic *collapse* in A.5 is a within-estimator relative contrast.

Synthesis of the battery. The trans-lineage “binary code” decomposes into three known mechanisms, none a novel endolinguistic layer and none demonstrable deep vertical inheritance across lineages:

1. Within-lineage cognation — 47× pair-weighted (A.6), the ladder of the descent, entirely expected.
2. An areal component — the boost that follows geography and collapses with distance (A.5).
3. A universal floor — the weak cross-family residue (1.9×, A.6), consistent with sound-symbolism (Blasi et al. 2016).

Limits and threats to validity (honest list).

- *Effect size over significance, and weighting matters*: the headline is the pair-weighted exact ratio with a bootstrap CI (A.6); the skeleton-averaged Monte-Carlo figures (A.3–A.5) inflate magnitudes and are kept only for comparability and for A.5’s relative contrast.
- *The undecidability result*: testing deep inheritance requires a deep tree that dissociates genealogy from geography; none exists (below the reconstruction horizon), and available deep trees (Nostratic, Borean) are partly built from areal similarity → confounded with geography by construction. A deep-inheritance component is therefore not refuted — only undecidable (§6).
- *Skeleton coarseness*: OAS class-skeletons collapse many phonemes, so “same skeleton” is a coarse match; hence the base-rate null.
- *Gloss-bag semantics*: A.1–A.2 use English gloss-bags as a proxy for meaning — a coarse similarity.
- *Centroid geography*: family centroids and distance bins are approximate.
- *Seeds*: the permutation tests (A.1, A.2) accept a seed argument and are stable across seeds; the headline (A.3) is additionally recomputed exactly, removing seed dependence (A.6).

Appendix B — Reproducibility and data availability

Data. data/db/endolang.db (SQLite, 317 MB). Table lexeme — 1,225,467 forms; 4,721 languages; 207 families; 3,160 Concepticon concepts — columns: concept, concepticon_id, concepticon_gloss,

language, glottocode, family, form, skeleton, nucleus, ordered, arity. Built from Lexibank/CLDF wordlists and the NEL corpus; OAS class-skeletons derived by the class-mapping pipeline (`src/build_skeltree.py`); cross-corpus concept alignment via Concepticon IDs.

The Germanic-Italic reconstruction data (PIE / Germanic / Italic roots and their living reflexes) live in `docs/studies/germanic-italic/gi.*.js`, generated from Kaikki/Wiktionary PIE etymologies by `src/gi/build_findings_all.py`. Nostratic membership and macro-branch come from `src/nostratic_map.py` (`info(family)`); family centroids and the haversine distance live inside `src/gi/null_inherit_vs_areal.py`.

Determinism. Every test calls `random.seed(1)`; the two PIE permutation tests accept a seed argument (`python3 src/gi/null_h1_deepcognacy.py <seed>`) for multi-seed checks. Dependencies: Python 3 standard library + `sqlite3` only.

Commands (run from the repository root):

```
python3 src/gi/null_coorientation.py      # A.1 co-orientation null    (accepts a seed argument)
python3 src/gi/null_h1_deepcognacy.py    # A.2 deep-cognition (H1)    (accepts a seed argument)
python3 src/gi/null_noIE.py              # A.3 cross-family (H1 vs H2), skeleton-averaged Monte-Carl
python3 src/gi/null_nostratic_decay.py    # A.4 Nostratic stratification
python3 src/gi/null_inherit_vs_areal.py   # A.5 inheritance vs areal (geography)
python3 src/gi/robustness.py              # A.6 exact + family bootstrap + sensitivity (headline)
```

Each script prints its table to `stdout`; no external outputs are required to reproduce the figures in Appendix A.

Companion studies. The corpus study (`docs/endolinguistic-structure.md`) and the psychophonological note (`docs/oas-psychophonology.md`) are companion manuscripts (in preparation); their load-bearing results are reproduced here in Appendix A so that this paper’s argument is self-contained.

Provenance and licences. Lexibank datasets (List et al. 2022; CC-BY variants), Wiktionary via Wikitextract / Kaikki (Ylönen 2022; CC BY-SA), Concepticon (List et al. 2023; CC-BY), CLDF (Forkel et al. 2018), and the CLTS transcription systems (Anderson et al. 2018) underlying the OAS skeletons; the NEL corpus is internal to the programme. Users reproducing the corpus should cite the underlying Lexibank datasets and Concepticon.

Document build. This PDF is produced by `bash docs/build-endolinguistica-acceso.sh` (a tailored `docs/md2tex.py` converter → `xelatex`, two passes), typeset in STIX Two Text with Arial as fallback for the OAS symbol **M**.

Appendix C — Glossary of coined terms · Glosario de términos acuñados

término (ES)	term (EN)	definición · definition
acceso (por profundidad)	access (by depth)	Instrumento epistémico-descriptivo (§0): procedimiento reproducible que reduce una forma a coordenadas de clase consonántica y la desciende por los estratos genealógicos. No es acceso psicológico ni ontológico ni metáfora. · An epistemic-descriptive instrument (§0): a reproducible procedure, not a psychological/ontological reaching or a metaphor.

término (ES)	term (EN)	definición · definition
perfil endolingüístico (endolenguaje)	endolinguistic profile (endolanguage)	La salida del instrumento a una profundidad: la descripción coherente de forma y sentido a ese estrato. Es el término de trabajo; <i>endolenguaje</i> nombra el objeto que la tradición reifica a partir del perfil (lectura diferida, §5). · The instrument's output at a depth: the coherent form-meaning description at that stratum. The working term; <i>endolanguage</i> names the object the tradition reifies from it (deferred, §5).
código (binario)	(binary) code	Coordenada consonántica —una díada de dos clases OAS— con que se marca y compara una raíz a través de los estratos. No es la letra, ni la historia, ni el significado. · A consonantal coordinate (a two-OAS-class dyad) used to mark and compare a root across strata.
esqueleto	skeleton	La secuencia de consonantes canónicas (por valor cuálico OAS) de una forma; base de la comparación de "mismo código". · The canonical consonant sequence of a form; basis of same-code comparison.
clase OAS	OAS class	Una de las 7 clases de orientación psíquica ($\Phi \Theta X \Sigma \Lambda M \Xi$) a las que se reduce una consonante; orienta, no denota. · One of the 7 psychic-orientation classes a consonant reduces to; it orients, it does not denote.
temperamento	temperament	Hipótesis (§4): la desviación de la distribución de campos-de-sentido de una rama respecto de la línea base familiar, por código (divergencia p. ej. Jensen-Shannon); testable por nulo de permutación y predicción ciega. · Hypothesis (§4): a branch's field-distribution departure from the family baseline per code; testable.
deriva de intención	intention drift	El cambio de la distribución de sentido de un mismo código entre estratos (Indoeuropeo → Germánico → Itálico). · The shift in a code's meaning-distribution across strata.
co-orientación	co-orientation	Eje relacional que empareja raíces por código + campo, sean cognadas o no; probado (débil) en A.1. · Relational axis pairing roots by code + field, cognate or not; tested (weak) in A.1.
substrato común latente	latent common substrate	Definido solo por su papel: el punto de fuga postulado del descenso (el límite aún no reconstruido hacia el que convergen las genealogías). Su naturaleza (cognitiva/biológica/...) queda diferida; la tradición lo identifica hipotéticamente con un <i>inconsciente compartido</i> (§5). · Defined only by its role: the posited vanishing-point of the descent; its nature deferred; hypothesised by the tradition to be a <i>shared unconscious</i> (§5).

término (ES)	term (EN)	definición · definition
programa de investigación	research programme	El estatus epistemológico del conjunto (Lakatos/Laudan): se juzga por predicciones novedosas y fecundidad, no como proposición verdadera/falsa única (§8). · The epistemic status of the whole (Lakatos/Laudan): judged by novel predictions and fecundity (§8).
horizonte de reconstrucción	reconstruction horizon	El límite (del orden de milenios) bajo el cual la reconstrucción —y por tanto el descenso— se detiene. · The limit below which reconstruction, and hence the descent, stops.

This paper draws on: the Germanic-Italic study (docs/studies/germanic-italic/, codes × three endolanguages, and the H1 deep-cognition test), the corpus study (docs/endolinguistic-structure.md), and the psychophonological note (docs/oas-psychophonology.md) — the latter two companion manuscripts in preparation. None of those results is used here as proof of the substrate: they show that coherence lives in depth, and that access to it passes, necessarily, through genealogy. · Se apoya en el estudio germánico-italico, el de corpus y la nota psicofonológica (los dos últimos, manuscritos compañeros en preparación); ninguno se usa como prueba del substrato, sino para mostrar que la coherencia vive en la profundidad.

References · Bibliografía

- Anderson, C., Tresoldi, T., Chacon, T. C., Fehn, A.-M., Walworth, M., Forkel, R., & List, J.-M. (2018). A cross-linguistic database of phonetic transcription systems. *Yearbook of the Poznań Linguistics Meeting*, 4(1), 21–53.
- Benveniste, É. (1935). *Origines de la formation des noms en indo-européen*. Paris: Adrien-Maisonneuve.
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidence across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818–10823.
- Bomhard, A. R. (2018). *A comprehensive introduction to Nostratic comparative linguistics* (2nd ed.). Charleston, SC.
- Campbell, L. (2010). Language isolates and their history, or, what’s weird, anyway? *Proceedings of the Berkeley Linguistics Society*, 36(1), 16–31.
- Campbell, L., & Poser, W. J. (2008). *Language classification: History and method*. Cambridge: Cambridge University Press.
- Forkel, R., List, J.-M., Greenhill, S. J., Rzymiski, C., Bank, S., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G. A., & Gray, R. D. (2018). Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, 5, 180205.
- Illich-Svitych, V. M. (1971–1984). *Opyt sravnenija nostraticheskix jazykov* [An attempt at a comparison of the Nostratic languages] (Vols. 1–3). Moscow: Nauka.
- Lakatos, I. (1978). *The methodology of scientific research programmes* (Philosophical Papers, Vol. 1; J. Worrall & G. Currie, Eds.). Cambridge: Cambridge University Press.
- Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. Berkeley: University of California Press.

- List, J.-M., Forkel, R., Greenhill, S. J., Rzymiski, C., Englisch, J., & Gray, R. D. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*, 9, 316.
- List, J.-M., Rzymiski, C., Greenhill, S., Schweikhard, N., Pinykh, K., Tjuka, A., Wu, M.-S., & Forkel, R. (Eds.). (2023). *Concepticon 3.1: A resource for the linking of concept lists*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://concepticon.clld.org>
- Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.
- Wichmann, S., Holman, E. W., & Brown, C. H. (Eds.). (2022). *The ASJP database* (version 20).
- Ylönen, T. (2022). Wiktextextract: Wiktionary as machine-readable structured data. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC)* (pp. 1317–1325). Marseille: ELRA.