

What Is an Endolinguistic Code?

Class, order, field, and the structural continuity that codes allow us to observe

Alejandro Toledo Martínez

Symbol key. The seven OAS classes are written as Latin letters and as their presentation symbols: P = Φ, T = Θ, K = X, S = Σ, L = Λ, M = **M**, N = Ξ. A *nucleus* is written {X, Y} (unordered); an *ordered code* X · Y · Z; a *skeleton* (x · y) in real consonants; IPA in slashes. The distinction between letters and symbols is not decoration: the letters belong to languages, the symbols belong to the model, and §9 explains why that difference carries the whole ontology.

Abstract

Asked what an endolinguistic code is, most readers — and most language models — answer that it is a sequence of consonants: *B–R, M–N–D, K–S*. That answer is not merely incomplete; it is wrong at the level of the object. A code is not a string of consonants, it is not a root, it is not a morpheme, it is not an etymology, and it does not mean anything. This paper gives the full definition and shows why every shorter one collapses.

The definition unfolds in three movements. The first is structural: a ladder of four levels — *cualo*, class, code, molecule — that must never be conflated, together with a bounded formal calculus in which order carries quality, and a continuous qualic field (vowels, prosody, endorhythm) that the discrete object presupposes but cannot contain. The second is methodological: a discipline of scales — *system* (a particular language), *system of languages* (a branch), *macrosystem* (a family) — organized around the coderivative, an endolinguistic object distinct from the historical linguist's cognate, and around the coderivative network as the basic unit of observation, in place of the word and in place of the code itself.

The third movement is ontological, and it is where the paper ends because it is where the theory currently stands. If a code is only an instrument of the observer, its readability across five or six thousand years of sound change would be an extraordinary coincidence. If it is a real structure of language, how can it have no meaning? The resolution defended here dissolves the dilemma instead of choosing a horn: codes are methodological instruments; what persists through genealogy is not the code, but a structural continuity that the codes allow us to observe. What exactly that continuity is — not a word, not a meaning, not a phoneme, and not the code — is left deliberately open, as the discipline's central question.

Keywords: endolinguistics, endolinguistic code, OAS classes, nucleus, qualic inversion, coderivative, semantic field, macrosystem, structural continuity

1. The question, and why the short answer is wrong

There is a sentence that recurs whenever endolinguistics is summarized by someone who has read it quickly: *an endolinguistic code is a pattern of consonants shared by words with related meanings*. Every content word in that sentence is either false or dangerously imprecise.

"Consonants" is false: the code is built of classes, and a class is not a consonant. "Pattern" is imprecise: a code is a bounded, ordered structure of two or three elements, and the boundedness and the order are load-bearing. "Shared by words" hides the hardest methodological problem in the discipline, which is deciding which words may legitimately be gathered together. And "related meanings" inverts the theory: a code does not have a meaning, does not confer one, and cannot be defined by one.

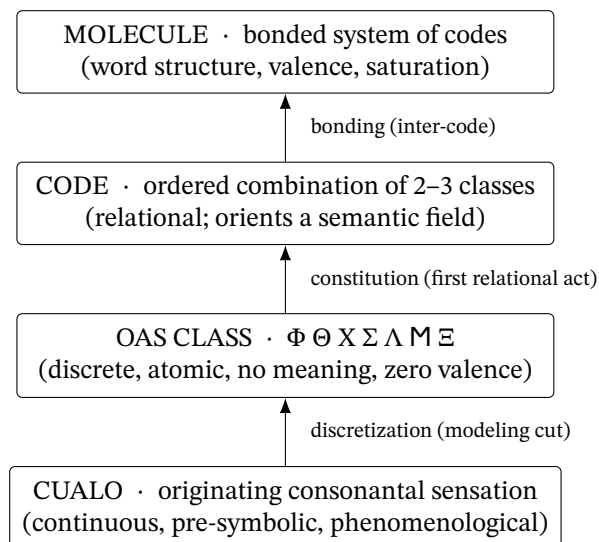
The error is understandable. The code has a visible face — the consonantal skeleton — and the visible face is the only part that survives casual reading. But a definition built on the visible face produces exactly the pathology the discipline was constructed to prevent: unlimited pattern-matching, in which any two words sharing two consonants can be declared related, and any relation can be narrated after the fact.

This paper therefore does something slightly unusual. It defines the code from the inside out — from the sensation that precedes it to the molecule that contains it — and it spends as much space on what a code is *not* as on what it is. The negative definitions are not rhetorical. Each one corresponds to a failure mode observed repeatedly in practice, including, systematically, in artificial-intelligence systems asked to reason with codes.

Insight. The complexity of the definition is not ornamental. Each layer exists because removing it destroys the object. A code without classes is phonetic coincidence; a code without order is a set; a code with a meaning is folk etymology; a code without a scale discipline is apophenia; a code without a qualic ground is a formalism describing an object that does not exist.

2. The ontological ladder

Endolinguistics distinguishes four levels. They are levels of a *model*, not stages of a process, and the single most common analytic error is to collapse one into the next.



2.1 The cualo: an originating sensation, not a meaning

Before any combinatorics, an isolated consonant produces a qualitative sensation within a system. In the author's formulation: *"Antes de toda combinatoria, la consonante aislada produce una sensación cualitativa por sistema... núcleos fenomenológicos irreductibles."*

These are *cualos* — irreducible units of lived appearing. They are continuous, non-discrete, and pre-symbolic. Crucially, a cualo is a sensation, not a meaning and not a concept. S is felt as friction, sliding, continuity, a going-out; N as internal resonance, interiority; M as closure, containment, sonorous mass, belonging.

Two guardrails apply from the start. First, the cualo is indexed — and the indexation runs at more than one scale. At the collective scale, as the OAS protocol already states, the mapping between vocal configuration and affective sensation is hypothesized as stable within a macrosystem and may differ across macrosystems. To this the present paper adds the individual scale: the sensation is also more or less stable within an individual, inside the boundaries of their endolanguage. The originating sensation is therefore not only macrosystemic; it is also individual, and it corresponds to that particular system — the speaker’s own endolanguage, the most intimate scale of system the discipline recognizes (§7.1). The two indexations do not compete: the macrosystem gives the sensation its shared horizon, the individual endolanguage gives it its lived instance, and neither can be read off the other. And between those two poles lie all the intermediate scales — groups, dialects, languages — each a psychic system that indexes the sensation in its own way (§2.2). Second, the cualo is not accessible as an object. It is articulable only through traces. This is why the discipline formalizes *conditions of passage* rather than the cualo itself.

2.2 The OAS class: discrete, articulatorily anchored, meaningless

The OAS system — *Originating Affective Sound* — is Toledo’s own contribution to a discipline he did not found. The discipline itself (endolinguistics, the four levels of language, the exo/endolanguage distinction) was created by Meulemans and Elias; the class inventory, the qualic mathematics, the falsifiability protocol and the bonding formalism are Toledo’s additions to it. The reference inventory, declared for the Indo-Iranian-European macrosystem, discretizes the consonantal space into seven classes — though, as the end of this section states, the inventory’s true anchor is not the macrosystem but the psychic system under study.

class	symbol	articulatory region	affective profile (a direction, not a meaning)	realizations
P	Φ	labial	expansion / outward projection	p b f v w
T	Θ	dental–alveolar	boundary / delimitation	t d θ ð
K	X	velar / guttural	generative depth	k g x h q c
S	Σ	sibilant	externalization	s z ʃ ʒ
L	Λ	liquid (l + r)	flow / continuity	l r ʎ ʟ ʀ ʁ
M	M	labial nasal	material mass / belonging	m ŋ
N	Ξ	nasal (non-labial)	interiority	n ŋ ɲ ɳ

The protocol states the class’s properties negatively and precisely. A class is *discrete at the modeling level; has no internal scalar gradiance; has no semantic meaning; is not a code; cannot function as an attractor on its own*. It is an atomic structural unit, articulatorily grounded and therefore not arbitrary, and its consonants are its *realizations* — class identifiers, not literal phonemes.

The affective profile in the table is the most misread line in the whole theory. It is a felt affective direction, pre-symbolic, at the level of the class region as a whole — never a property of any individual phoneme and never a semantic atom. The protocol’s own formula: the class “*encodes an affective direction of structural emphasis, not an invariant symbolic meaning*.”

From this follows the axiom that governs everything above it:

Zero class valence. For any class $x \in \Sigma$, the compositional valence of x considered in isolation is $\nu_{\text{comp}}(x) = 0$. Classes contribute tendencies, not relations.

One refinement, added here, generalizes the protocol's indexation. The reference inventory above is declared for the Indo-Iranian-European macrosystem (OAS-IIE-M v1.0), and the protocol states that the mapping between vocal configuration and sensation is stable within a macrosystem. But the macrosystem is not the ground of the class system; it is one scale of it. The OAS system may be redefined by system. This does not mean more symbols for the consonantal system: the seven-class skeleton — the discrete partition, the symbols, the zero-valence axiom — is the invariant of the model. (Where the symbolic inventory *could* legitimately grow is elsewhere: tone, vowels, suprasegmentals, accents — the qualic layer of §5 — may eventually receive symbols of their own. That extension is the subject of another paper; here it is only marked as open.) What may drift from system to system is the *content* of each class: the originating sensation, the pre-symbolic affective profile, and even the realizations, which may be nuanced toward other sounds. And this drift can occur at every scale at which a psychic system exists — an individual, a group, a dialect, a language, a family.

The anchor of the inventory is therefore not the macrosystem but the psychic system. OAS, like the codes, is linked to a psychic system — whether that system is a single person or a group. A declared inventory is always indexed to one: OAS-IIE-M v1.0 is the declaration for one macrosystem, not the definition of the classes as such. The falsifiability discipline survives intact, and in fact gains precision: what C0 demands is that whichever psychic system is under study, its class content be declared and versioned *for that system* before any code is read, and revised only formally. Stability is required per system; identity across systems is never assumed.

2.3 The code: the first relational object

A code is *a bounded structural combination of two or three OAS classes*. Three properties make it what it is.

It is relational. Code formation is the constitutive act that turns atomic, non-relational class-units into a structural object in which relation exists. Below this level there are no relations, only tendencies.

It is ordered. *"El código no es un conjunto, sino una estructura orientada."* The unordered set of its classes is the nucleus — the invariant conserved under transformation — but the code itself is the nucleus under a specific order. The unit of study is the concrete order, not the nucleus.

It is bounded to two or three classes. There is no quaternary primitive. A form exhibiting four or more distinct classes is not a longer code: it is a saturated composite, a molecule, and must be separated into morphemes each carrying its own code — never truncated to three.

And it has one semantic property, stated as a restriction rather than a capacity: *a code does not determine meaning; it organizes a field of semantic possibility*. Section 6 develops what this means and what it forbids.

2.4 The molecule: bonding, valence, saturation

The class is atomic, but the code is the smallest bond-capable unit. Bonding begins only after a code nucleus has been formed. Two structurally distinct operations must never be conflated: the internal integration of classes into a code (internal cohesion) and the external bonding of codes into molecules (inter-code bonds).

Valence is not assigned by counting classes; it is *derived* from what remains exposed after internal organization. For a code c , the interface set $\mathcal{I}(c)$ collects its externally available relational surfaces — determined by arity, internal cohesion, class-specific tendencies, positional ordering, and macrosystem constraints — and $\nu(c) = |\mathcal{I}(c)|$. The saturation index

$$\sigma(W) = \frac{\text{occupied bond sites}}{\text{available bond sites}}$$

marks a closed system at $\sigma(W) = 1$. A ternary code does not automatically have higher valence than every binary; triadic organization merely *permits* higher maximal valence, because it can leave more structurally distinct interfaces free.

3. Ten things a code is not

Each item below is a failure mode observed in practice. Together they are more diagnostic than any positive formula.

1. Not a consonant string. The skeleton ($b \cdot r$) is the visible realization; the code $\{P, L\}$ is the object. Several skeletons realize one code, and the mapping runs from pronunciation to class, never from spelling to letter.
2. Not read off spelling. The code comes from IPA and is dialect-aware. Latin-American Spanish *luz*, with $/s/$, gives $\{L, S\}$; Castilian *luz*, with $/\theta/$, gives $\{L, T\}$. Same written word, different code, because the code tracks pronunciation.
3. Not a morpheme, root, or lexeme. Codes are not morphological units. Affixes and terminatives are stripped before the code is read: *shatter* yields SH-T, not $\{L, S, T\}$, because *-er* is terminative.
4. Not an etymology. Binary and ternary codes are not etymologies. Historical linguistics is an input to the reading, not the thing being read.
5. Not a meaning. Writing $\{K, T\} = \text{bounded matter}$ is a category error. A code does not denote.
6. Not the sum of its classes' affective profiles. This is the most seductive error and the most explicitly forbidden. "X = depth plus Σ = coming-forth, therefore the code means *emerges and covers*" mixes the category of *sensation* with the category of *meaning* and reifies both. The profiles are pre-symbolic sensations; they are not semantic atoms to be added. What makes a code readable is bonding, not composition.
7. Not a set. Order carries quality. $\{T, L\}$ under the order $\Theta \cdot \Lambda$ and under $\Lambda \cdot \Theta$ share a nucleus and differ in orientation (§4.2).
8. Not a unit of a single language. Single-language isolation inflates false positives and invites narrative rescue. The reading is comparative by construction: the shared value of a code appears in the comparison of languages, not inside one.
9. Not a rival of historical linguistics. "*Endolinguistics does not replace historical linguistics; it presupposes it.*" IPA, etymology, sound laws and attested usage are inputs, read through a different ontology. And the relation is positive, not merely non-competitive: endolinguistics *works with* genealogy. Within a coderivative field, the codes serve as instruments of greater traceability — where the surface forms diverge across millennia of sound change, the code remains the most followable thread through the descent (§7).
10. Not an entity of the language. This is the ontological claim, and §9 is devoted to it. Codes do not exist in languages the way phonemes do. Whether something else does — something the codes detect — is the deepest question in the theory, and the paper ends on it.

4. The code as a formal object

4.1 Nucleus and order

Let $\Sigma = \{P, T, K, S, L, M, N\}$ be the class inventory — seven elements, finite, versioned, declared before any hypothesis is examined. (Formulae use the Latin class letters; the presentation symbols Φ Θ X Σ Λ M Ξ belong to the prose, where a code is displayed rather than computed.) A binary code is an ordered pair and a ternary an ordered triple over Σ . The nucleus is the underlying unordered set:

$$\mathcal{N}(C) = \{A, B\} \quad \text{for } C = (A, B), \quad \mathcal{N}(C) = \{A, B, C\} \quad \text{for a ternary.}$$

The nucleus is the deep structural binding — what remains constant through transformation. It is what lets the theory distinguish *structural change* (the code is destroyed) from *qualitative change* (the same code, reoriented).

The counting follows immediately: 21 binary nuclei and 35 ternary nuclei, hence 56 nuclei; imposing order gives $21 \times 2 + 35 \times 6 = 252$ ordered codes. This is the whole primitive inventory of the discipline. Its smallness is a feature: no code can be invented outside it without a formal revision of the protocol.

4.2 Inversion, and the axiom that order carries quality

Define the inversion operator ι by $\iota(A, B) = (B, A)$ and $\iota(A, B, C) = (C, B, A)$. It is an involution, $\iota^2 = \text{id}$, and it conserves the nucleus: $\mathcal{N}(\iota C) = \mathcal{N}(C)$.

Now let q be a *qualic function* mapping a code into a space of psycho-semantic qualities. The central axiom of the qualic mathematics is:

$$\mathcal{N}(\iota C) = \mathcal{N}(C) \quad \text{but} \quad q(\iota C) \neq q(C).$$

Inversion is structurally conservative and qualitatively non-commutative. The result is not another object; it is another quality of the same nucleus.

The canonical illustration is Θ – Λ against Λ – Θ . The first orients toward settling, ground, terrain, stability — *terra, Erde, terrain*. The second orients toward rotation, wheel, rhythm, turning — *rotate, rueda, rota*. The author’s own gloss is exact and worth preserving: “*no oposición semántica directa, sino reorientación cualitativa del mismo núcleo.*” Not a semantic opposition; a qualitative reorientation of one nucleus.

If the qualic space Q admits a metric d_Q , the qualic difference $\delta(C) = d_Q(q(C), q(\iota C))$ becomes a comparable magnitude — a way to speak of degrees of reorientation without ever assigning a meaning to either orientation.

4.3 Ternaries: resolution, not addition

A ternary is not a binary with one more class. Two formulations converge on this.

Qualically, a ternary often arises as the qualic resolution of a binary tension: the third element is not additive but a stabilizer that integrates the tension generated by inversion. Structurally, a ternary carries an irreducibly triadic surplus that its pairwise projections cannot recover: writing A_{ij} for the pairwise compatibility of classes i, j and H_{ijk} for the full stabilization of the triple,

$$\Delta H_{ijk} = H_{ijk} - (A_{ij} + A_{jk} + A_{ik}),$$

and $\Delta H_{ijk} \neq 0$ marks exactly the configurations in which the triple is not the sum of its pairs. The difference between binary and ternary is geometric, not arithmetic.

Insertion is the operation that makes this visible in real words. A binary $\{P, S\}$ ($\Phi \cdot \Sigma$, Latin *pes-*) takes Ξ internally and yields $\Phi \cdot \Xi \cdot \Sigma$ — *pensar*. The operation is positional: which class enters, and where (initial, internal, final). And it is found, not computed: the same insertion appears realized across the macrosystem, sometimes as a segmental nasal (Spanish *pensar*), sometimes as a nasal vowel (French *pensé*).

4.4 The bounded transformation calculus

Five operations are admissible: within-class substitution (S), permutation (P), inversion (I), reduction (R, ternary→binary) and expansion (E, binary→ternary). All must be declared in advance, historically plausible, and bounded by an explicit maximum $k_{\max}$.

The bound is not bureaucratic. Without it, any form can be rescued by unlimited rewriting, and the theory explains everything, which is to say nothing.

One rule outranks the calculus, and it is easy to state and easy to violate: the operations are already realized in existing words. The analyst's task is to *find* the realized inverse, reduction or expansion — *terra* as the realized inverse of *Erde* — not to generate transformations abstractly and then hunt for words that fit. Computing a transformation and finding one are different epistemic acts producing different objects.

5. The qualic dimension

A purely structural reading of the code is epistemologically insufficient. The discrete formalism describes conservation; it cannot explain difference. This section names what it presupposes.

5.1 Vowels are qualic operators, not code

Vowels do not function as structural units. They are operators of quality: they modulate a nucleus without altering it.

The canonical pair is *macro* and *micro*. The nucleus $M \cdot X \cdot A$ is constant across both. What changes is the vowel: A opens — amplitude, exteriority; I contracts — precision, interiorization. The result is a qualic displacement of orientation, not a structural change. No discrete combinatorics accounts for it, and no reading of the code is complete without it.

5.2 Prosody, suprasegmentals, endorhythm

Endolinguistics cannot restrict itself to the segment. Intonation, stress and duration organize sense non-locally, and one consonantal structure can acquire radically different orientations under different prosodic patterns.

Endorhythm is the deepest term here and the most easily misunderstood. It is not poetic metre and not the audible rhythm of speech. It is the internal organization of the linguistic flow — repetition and variation, tension and resolution, continuity and cut — and, at the personal level, a speaker's capacity and orientation *for manifestation*. The same person speaking English, French or Spanish manifests different audible rhythms while the underlying endorhythm remains constant. It is not counted; it is perceived. And it organizes the emergence of sense systematically.

5.3 The unit and the field

The general form of the relation is the same across domains of knowledge. For a discrete unit u_X and its qualic field F_X , the realized phenomenon is $\Phi_X(u_X, F_X)$ — never u_X alone. In physics, event = $\Phi_Q(\text{state}, \text{manifestation-field})$; in biology, expression = $\Phi_G(\text{sequence}, \text{regulatory-field})$; in information, meaning = $\Phi_B(\text{distinction}, \text{semantic-pragmatic field})$. And in language:

$$\text{living language} = \Phi_E(\text{code}, \text{prosodic-psychic field}).$$

The code is structurally necessary and ontologically insufficient. The field is not a residue left over after formalization; it is the condition under which the code operates at all.

One notational remark closes the section. The consonantal system has its seven symbols and needs no more (§2.2). This layer — vowels, tone, suprasegmentals, accents — has none yet, and it is the one place where the symbolic inventory of the model could legitimately grow: a symbology of the qualic operators, parallel to the symbology of the classes. That extension is deliberately left to another paper; here it is only marked as open.

6. What a code does to meaning

6.1 Orientation, not denotation

The disciplined form of the semantic claim is stated in the negative, and the negation is the content:

Not “code X means Y”, but: within a concrete linguistic system, the words instantiating a code tend to organize themselves in a limited set of related semantic and psycho-symbolic regions — possibly with structured polarity inversions across languages or strata.

Three consequences follow. The relation is tendential, so counterexamples do not automatically refute but must be weighed. The relation is system-indexed, so the same code may orient differently in Germanic and in Iranian without contradiction. And the relation is *not stored*: a database of the discipline should never contain a table Code | Meaning. It should contain System | Code | Coderivative network | Observed semantic field | Evidence. The field is a result of analysis, not an attribute of the code.

6.2 The multi-depth stack: the *canis* principle

This is the single most important thing to understand about endolinguistic meaning, and it is where endolinguistics parts company with linguistics proper.

In linguistics, meaning is relatively stable and singular with its signifier. *Canis* means “dog” in Latin; context may modulate, but the pairing holds.

In endolinguistics, a code carries meaning by depths — a resonant stack held together by psychic resonance. The author’s own example: *canis* → root *can* → skeleton (k·n) → code {K, N}. Its meaning runs by levels: dog at the surface/social level; a totemic animal at the archetypal level; a numeral depth, *cien* / hundred; and a generative power, *can*, *gen*.

These depths are connected psychically, not etymologically. To characterize a code one surfaces its depth-stack across domains — an animal, a numeral, a verb of power, a totem — and exposes the relations between the depths. A concept list of basic vocabulary shows one thin slice of that stack and should never be mistaken for the whole.

This is also why translation is not synonymy. There are no deep synonyms: the same code is the same word pronounced with a different rhythm; a different code is a different word. English *break*, German *brechen* and Spanish *quebrar* share {K,L,P}; Spanish *romper*, with {L,M,P}, does not. Translation preserves the code.

6.3 The reference lies in the relations among the words

This is the operational heart of the method, and it is bottom-up.

The reference of a code is never a top-down definition and never the class sensations. It is the logical relation exposed among the real words. The author's worked example never writes " $X + \Sigma = X$ ". It exposes the thread: "*La causa de que habitemos una casa, es porque caemos ahí; se da el caso de que sea el lugar donde caemos. 'Caer' para descansar es la casa.*" And the governing remark: "*el pensamiento no equipara términos, funciona lógicamente.*" The code does not equate *casa = causa = caso*. It exposes a shared logical orientation among them.

So the procedure to characterize a code is: gather the family that shares the nucleus, across the macrosystem, *found and not invented*; expose the relations between specific words as pairs and logical threads — *man* $\leftrightarrow$ *mind* ("he who minds"), *name* $\leftrightarrow$ *mind* ("to name is to fix in mind"); and let the orientation emerge from those relations without reducing it to a label.

6.4 Polarity inversion, disciplined

The same code can organize opposite poles of one axis: K-L-D appears in *calidus* (hot) and in *kalt* (cold). Treated loosely, this is an escape hatch that makes the theory irrefutable. Treated strictly, it is a prediction.

The constraint is that a reading may invoke inversion only if it declares in advance that the field admits a polar structure and specifies *which axis* is relevant, together with the poles and the scope conditions. Inventing an axis after the fact to save a match is disallowed, and doing so is itself grounds for rejecting the reading.

7. The scales of observation

This section states the methodological layer of the definition. Without it, everything above can still be applied in a way that produces nothing but confirmation.

7.1 System, system of languages, macrosystem

The terminology is fixed and the levels do not substitute for one another.

term	referent	examples
System	a particular language	Spanish, German, Persian, Russian, Nahuatl
System of languages	a genealogically related group: branch or subbranch	West Germanic, Romance, Iranian
Macrosystem	a complete language family	Indo-Iranian-European, Semitic, Turkic, Mayan

The methodological consequence is a strict order of work. One characterizes a code's behaviour *within a system* first, without importing data from its neighbours. One then repeats the exercise in a second system. Only then does one compare the two, and only because they belong to the same system of languages. Ascending to the macrosystem is the last step, not the first.

Below the table there is one further scale, which the ladder of §2 already implied: the individual endolanguage — the deep linguistic organization of one speaker. It is the limiting case of a system, and it is where the *cualo* has its lived instance (§2.1). The scales of the table are collective; this one is personal; the discipline needs both, because a sensation that is stable in a macrosystem is still *had* by individuals, each within the boundaries of their own endolanguage.

And a language is not a dictionary. It is a dynamic, psychological, partly unconscious system of thinking and of human linking — a living psychic entity, not an inventory of forms.

7.2 The coderivative: a different object, from a different discipline

The discipline needs a term that historical linguistics does not supply, and this is it.

Definition. A coderivative is a cognate in which the endolinguistic code remains observable.

The essential point is not that the coderivative is a narrower cognate. It is that cognate and coderivative are two distinct objects belonging to two distinct disciplines. The cognate is the object of historical linguistics: a form whose descent from a common ancestor is demonstrated by regular correspondences. The coderivative is the object of endolinguistics: the unit in which a code is still readable, and therefore the unit on which any endolinguistic reading can rest. Neither displaces the other, and neither is a version of the other.

They are related asymmetrically. Endolinguistics presupposes the evidence of cognation — kinship is not invented — and then selects its own object on top of it. Given a protoform with descendants A, B and C, where A and B preserve the code and C has undergone change severe enough that the code can no longer be identified under the declared rules, all three remain cognates for historical linguistics; only A and B are coderivatives. Every coderivative is a cognate; not every cognate is a coderivative.

Distinct, however, does not mean separate — and this must be said clearly: the cognate and the coderivative work together. The cognate is attested history and linked genealogy — the documented forms, the demonstrated descent, the chain of correspondences that no one may invent. The coderivative is linked *to it*: it inherits that attested anchoring, and adds the code's observability on top. Without the cognate, the coderivative would float — a resemblance with no history behind it, which is the definition of apophenia. Without the coderivative, the cognate's history would stay closed to endolinguistic reading. The two objects are complementary by construction: one supplies the evidence of descent, the other supplies the traceable thread through it (§3, item 9).

The definition therefore does two things at once. It keeps the rigour of historical linguistics, because the genealogical evidence is not the endolinguist's to invent; and it adds the criterion proper to endolinguistics, which is the *observable persistence of the code*.

7.3 The coderivative network is the unit of observation

This is the correction that reorganizes the whole method. The basic unit of endolinguistics is neither the word nor the code. It is the coderivative network.

The order of work follows:

1. Network → describe its semantic field.

2. Code → identify which networks reuse it.
3. Comparison between networks → look for strong relations.
4. Hypothesis → explain why those relations appear, if they do.

The code is therefore not the starting point. It is the criterion of grouping that brings several networks into dialogue. And both phases are endolinguistic: describing a single network within a system is not a preliminary to the discipline, it is the discipline answering its first question. The second phase answers a different question.

Treating the network rather than the word as the unit also imposes a natural honesty on the evidence. *Break / broke / broken / breach* count as one piece of evidence, not four, so a highly productive family cannot dominate a reading by sheer morphological fertility.

7.4 Genealogical field and systemic field

A network does not live inside one system; it crosses them. The descendants of one protoform are distributed across branches, and each language realizes a different part of the network.

Two distinct objects follow, and conflating them produces most of the confusion in cross-linguistic readings:

- The genealogical field belongs to the whole network, from the protoform through every descendant. It is what genealogy conserves.
- The systemic field is what that network acquires inside one particular language. It is what the system specializes.

English *bear* participates simultaneously in both: in the great historical field of the network, and in the semantic system of present-day English. They answer different questions and require different evidence. Genealogy conserves the identity of the network; the system modulates its realization. They do not do the same work.

7.5 The addition fallacy

Now the prohibition that gives the section its point.

Within a network, historical linguistics already explains the semantic continuity. *Bear* keeps the field of carrying; *bore* keeps the field of perforating; *break* keeps the field of rupture. Those facts need no code to explain them; they have an etymological explanation already.

It follows that the goal of endolinguistics is not to fuse the meanings of the networks in order to produce a supposed meaning for the code. Saying “*bear* is carrying and *bore* is perforating, therefore the code means *to carry by perforating*” is not a bold hypothesis; it is a methodological error of the same family as composing the class profiles (§3, item 6). The code is the same; it is *used* differently depending on the network and the context.

The real research question is different, and much harder:

Are there systematic relations between independent coderivative networks that reuse the same code within the same linguistic system?

Not “what meaning do they share”, but “what relation holds between their fields”. They may share a type of transformation, or a narrative structure, or nothing relevant at all. Which of these obtains is what patient observation decides — and “nothing relevant” is a genuine result, not a failure.

7.6 Strong relations and weak relations

The distinction between endolinguistics and pattern-hunting reduces to one operational question: is the relation strong or weak?

Weak relations are easy to produce by free association and analogy. Any two semantic fields can be connected by a sufficiently ingenious paraphrase. This is the terrain of apophenia, and ingenuity is precisely the wrong instrument.

Strong relations recur between independent networks, survive the widening of the material, persist across registers, and remain when set against networks that were not chosen to please. A connection between two networks is more solid to the extent that it appears in many words of both, holds across registers of the language, and does not depend on the analyst's selection of favourable examples.

The evidence, then, is never an elegant analogy. It is the persistence of a relation between independent networks under comparison and restraint.

7.7 The depth hypothesis

One consequence of the network-level view deserves separate statement, because it is easy to over-read.

Suppose two reconstructed protoforms in the same macrosystem are treated by comparative linguistics as distinct roots, because the correspondences do not license reconstructing a single ancestor. Suppose further that both preserve the same code and that their networks show a stable, non-anecdotal semantic relation. Endolinguistics may then formulate a depth hypothesis: that the two roots descend from a common ancestor prior to the reconstructed stage — while conceding that the comparative method cannot demonstrate it.

This does not contest the reconstruction. It adds a heuristic layer above it. And it earns its place only under strict conditions: same macrosystem; same observable code; a stable relation between the networks; no simpler explanation available (borrowing, secondary analogy, chance convergence); and — decisively — the hypothesis must generate new predictions, such as which under-studied root sharing the code should occupy a nearby semantic region.

Endolinguistics here does what other sciences do: it does not reconstruct the unobservable ancestor directly; it generates hypotheses about unobservable structure from observable regularity.

7.8 What the identity thesis does and does not say

A formulation in circulation needs to be stated precisely, because loosely read it looks like a contradiction of everything in this section. The formulation is that within a macrosystem there are no cognates: *three*, *tri*, *tres*, *drei* are the same word pronounced with a different rhythm, the macrosystem being one language and its "languages" rhythmic variants.

Read as a denial of cognation, this would put endolinguistics at war with comparative linguistics, and it would make the genealogical discipline of §7.1–§7.7 incoherent. That is not what it says. Cognates exist; they are simply not endolinguistic objects. The identity thesis is a claim in the endolinguistic register, about coderivatives: one code realized under different rhythms is one word. The distinction of §7.2 is what makes the two statements compatible — they are about different objects, established by different disciplines, and neither is answerable to the other's criteria.

Two consequences follow for practice.

First, the fact that two forms of one macrosystem share a code is not an inheritance "confound" to be controlled away. In the endolinguistic register it is the phenomenon, not the noise. Genealogy is

what tells us *which* forms belong to one coderivative network, so that networks are not pooled and fields are not summed; it is not a competitor to be neutralized.

Second, the two registers govern different scales. Genealogy is the regime of evidence *within* a macrosystem: it separates a finding from a coincidence, and the coderivative network is its unit. Resonance is what governs comparison *between* macrosystems, where no genealogy is available and history cannot arbitrate at all.

Insight. The rule is not "ignore genealogy" and not "reduce codes to inheritance". It is: within a macrosystem, genealogy delimits the object and resonance reads it; between macrosystems, only resonance is available, and the claim must be stated as such.

8. A worked illustration: B–R in English

The point of this section is to show the discipline operating by hand — including the places where it declines to conclude. Nothing here is a test; it is the method walking.

System: English. Code: {P, L}, realized as the skeleton (b·r). No meaning assumed.

Step 1 — collect without filtering. Not by prefix. Searching "br-" is a search for a spelling, and it silently excludes *bear, bore, bar, burden, berry, borrow, bury*. The code is not a word-initial cluster; it is a class pair, wherever it is realized.

Step 2 — separate the networks. The collected candidates fall into genealogically independent networks. Two suffice here.

Network 1 — PIE *b ^h er-	Network 2 — PGmc <i>brinnanq</i> / <i>brennaną</i>
bear, bearer, bearing, borne, born birth, childbirth, childbearing burden, burdensome, bier	burn, burning, burnt, burner

Step 3 — describe each field internally, without comparing. Network 1 organizes itself by internal relations rather than by a single gloss: *bear* is the fundamental action; *bearer* the agent; *bearing* the process, capacity or mode; *burden* the object undergone; *borne* the result; *born* a specialized result of one modality of the action; *birth* the derived event; *bier* the instrument. A first descriptive approximation of the field — and the hedging is part of the claim — is something like: *an entity sustains, carries or maintains another through a process or interval to a resulting state*. Note that plain "to carry" fails to cover *birth* and *bearing*, while this formulation covers the network without forcing any member.

Network 2 organizes around burning, consuming, transformation by fire.

Step 4 — stop. Here the discipline requires restraint. We do not look for a word that unifies "carrying" and "burning". We do not conclude that {P, L} means transformation, energy, or process. What we may state is exactly four things: the networks belong to distinct roots; both preserve the code; each has a coherent field of its own; both fields are recorded for comparison.

Only with ten or twenty independent {P, L} networks in hand does the second question become askable: are there groupings, proximities, oppositions, polarities — or no relation at all? The connection is not presupposed. It is discovered, or it is not.

Step 5 — what the second phase would require. The comparison of many networks at once is not an exercise for a dictionary and a good eye; it needs a corpus with declared properties — pronunciation-level forms, genealogical annotation, semantic context — that no existing resource yet provides in adequate shape. Designing and building that corpus is a separate undertaking. Until it exists, phase

two remains what it should remain: designed, declared, and not yet run. Nothing in the definition of the code depends on its outcome.

9. The ontological status of a code

Everything so far has treated the code as an instrument. That position now has to face its strongest objection — and the resolution of that objection is the deepest result this paper has to state. It emerged, in exactly the order reproduced here, from the working dialogue on which this paper is based.

9.1 The dilemma

Two positions seem to exhaust the options.

Option A — instrumentalist. Codes are only a model of observation. They do not exist in languages; they exist in the theory. On this reading the analogy is with coordinates: Cartesian coordinates do not form space, taxonomy does not create species, a coordinate system does not generate a surface. If someone objects "codes do not exist in languages", the answer is: correct, and neither does the equator.

Option B — structural realist. Codes describe an objective regularity that languages reuse historically. They would not be an invention of the observer but a structural property of the linguistic phenomenon itself.

9.2 The fact that forces the question

Consider the root *b^her-*. *Across five or six thousand years it produces* bear, burden, birth, gebären, Geburt, bharati* — and through all of it, the code remains observable. Genealogy has conserved it.

If the code were purely an arbitrary tool of the investigator, it would be an extraordinary coincidence that it remained readable through millennia of sound change. So the instrumentalist position, taken alone, leaves the central fact unexplained. But the realist position, taken alone, proves too much: if the code is a real structure of language, one expects it to *do* something semantic — and the whole of §6 has shown that it does not denote and must not be read as if it did.

The first move of the resolution is to separate two questions that the dilemma runs together:

- Structural existence. The code may exist as an objective regularity.
- Semantic content. The code possesses no semantic content of its own. Content appears in the networks and in the systems.

These are independent. Physics offers the parallel: coordinates are an instrument, but the symmetries of spacetime are not an instrument — they are properties of the universe discovered *through* mathematical models. Perhaps codes occupy an analogous position.

9.3 Why symbols and not letters

This is the moment at which a notational choice reveals itself as ontology.

The OAS classes are written with symbols — Φ Θ X Σ Λ M Ξ — and not with letters, because letters belong to a language and symbols belong to the model. A class never claims to be a phoneme; it is an abstract entity. And the code should be understood the same way. Not as

B–R

but as

an abstract structure which in English we observe as BR, which in another language could be observed through other, phonetically equivalent realizations — and which would still be the same structure.

The letters are how a particular language shows us the code; the symbols are how the model names it. Confusing the two is confusing the observation with the observed.

9.4 Three terms, kept apart

The theory therefore gains an explicit ontological distinction, and every sentence in the discipline should respect it:

1. Representation of the code. The concrete consonants used to identify it in a language or a reconstruction (B, R, ...). Conventional, and dependent on the linguistic system.
2. Structural code. The abstract relation that remains recognizable through genealogy. It is not a word, not a meaning, and not a phoneme.
3. Semantic field. The historical manifestation that a coderivative network acquires within a linguistic system.

In this scheme the code is no longer a merely arbitrary tool, and no hidden meaning either. It would be an abstract structure discovered through the analysis of languages, while the symbols used to represent it — like the OAS classes themselves — are the formal language of the model. The formulation avoids two extremes at once: the essentialism of "the code means X", and the absolute instrumentalism of "the code exists only in the theory". A third possibility opens: *the code as a structural regularity whose existence is itself an empirical hypothesis of endolinguistics*.

9.5 The resolution

The final step is sharper still, and it is the author's: keep the codes as instruments, and relocate what persists.

Codes are methodological instruments. What persists through genealogy is not the codes, but a historical structure that the codes allow us to observe.

The astronomical analogy is exact. Celestial coordinates are not the stars. Without the coordinates we could not describe the motion of the stars. The coordinates do not persist; the celestial objects do. Here, the codes are the coordinates — and what persists is something else.

Three levels follow, and they should be kept apart in every sentence one writes:

1. The historical reality. Networks of related words evolve over millennia. The continuity is real.
2. An inherited structure that travels with those networks. Not yet defined ontologically. This is precisely what the codes detect.
3. The codes. Not that structure — the instrument that makes it visible.

Even the vocabulary should shift. Rather than "the code B–R *appears* in...", the accurate formulation is "the code B–R detects a structural continuity". The difference is enormous, because it stops the instrument from being identified with the phenomenon.

9.6 The entity left open

What is that continuity? The honest answer is that we do not yet know. Perhaps the right word is a pattern. Or a structural continuity. Or an inherited configuration. The best term is not yet settled — and the theory gains by leaving it open. What we do know is negative and firm:

It is not a word. It is not a meaning. It is not a phoneme. And it is not the code.

The position, stated in full:

Endolinguistics does not claim that codes exist as entities of languages. It claims that structural continuities exist across linguistic genealogy. The codes constitute the methodological instrument by which those continuities can be identified, described and compared.

This formulation is stronger than either extreme. It does not reify the codes, but neither does it reduce everything to a methodological artifact. It keeps the ontology of the phenomenon separate from the epistemology of the method — which strengthens a theory rather than weakening it. And it leaves open a scientific question of the first order:

What exactly is the structural continuity that runs through linguistic genealogies, and that the codes allow us to trace?

If, in time, the evidence suffices, that continuity can be characterized further. If not, the codes remain exactly as useful as instruments of observation. The question does not belong to endolinguistics alone; it has become an ontological and scientific problem in its own right.

9.7 The same principle, one level down

The resolution for codes is the resolution for classes, because the two stand at the same epistemological level.

	OAS class	Endolinguistic code
nature	abstract pre-verbal model	abstract logico-mathematical model
has meaning of its own?	no	no
function	organizes the pre-verbal affective space	organizes the observation of words
where content comes from	the psychophysical system produces the observed phenomenon	the linguistic system produces the observed semantic field

A class no more is an emotion than a code is a meaning. The classes are called pre-verbal not because they precede language chronologically, but because the model does not start from words or from their meanings: it starts from an abstract layer that organizes the subsequent analysis. And both are anchored the same way: OAS and the codes alike are linked to a psychic system — be it a single person or a group — and neither sits, by definition, on top of macrosystems (§2.2). The two theories were never independent. They are two applications of one epistemological principle:

Abstract models do not contain the phenomenon they study; they supply a coordinate system for observing how that phenomenon organizes itself in a concrete system.

10. Falsifiability as internal discipline

A definition that cannot fail is not a definition. The constraints below are part of the object — the shape of the discipline the theory imposes on itself — not a report of anything measured.

Declaration before reading. The class inventory is declared before any code is examined, and revised only by explicit versioning (C0). The root is normalized first — affixes and terminatives stripped (C0★) — and every extracted variant remains traceable alongside its projected class-code (C0★★).

Bounds on transformation. The transformation set and its maximum $k_{\max}$ are declared in advance (C1), with a bounded bridge-path between forms — no teleportation (C1†) — and structural conservation up to the declared equivalence (C2).

Plausibility and signature. Every proposed instantiation must be consistent with the historical phonology of its language (C3), and a reading must predict a *qualic signature* — prosody, endorhythm, vowel-operator direction — treating its systematic absence as counterevidence (C3★).

Inversion under declaration. Polarity inversion is admissible only with a pre-declared axis, poles and scope (C4/C5). A post-hoc axis invented to save a match is not a rescue; it is a rejection.

How a reading fails. A code reading is rejected or revised when its instantiations do not recur beyond what the language's phonotactics would produce anyway; when most of them exceed the declared bounds; when the classes must be redefined to capture them; when the phonological paths are implausible; when the gathered words show no semantic concentration; when inversion is invoked after the fact; or when a rival code under identical bounds accounts for the same words as well or better.

"Falsifiability does not demand total success; it demands that failure is meaningful." A code may be active in some strata of a language and inactive in others; that is a finding, not an embarrassment.

Two remarks close the section. First, the empirical program that these constraints frame — the comparison of many coderivative networks under declared bounds and controls — requires a corpus with properties (pronunciation-level forms, genealogical annotation, stratified semantics) that no existing resource yet provides adequately; the program is therefore stated here as design, not as result. Second, one malformed test must be named so that it is not repeated: asking judges whether a *composed* field-meaning "fits" a set of concepts is not a test of the theory, because it presupposes that the code denotes — the very thing the theory denies. A test that assumes the negation of the hypothesis can neither refute it nor confirm it.

11. The definition, assembled

Everything above compresses into the following. The long form first, since the short forms are what get misquoted.

An endolinguistic code is a bounded, ordered combination of two or three OAS classes — themselves discrete, articulatorily anchored, meaningless partitions of the consonantal space, indexed to the psychic system under study, be that an individual, a group, a dialect, a language or a macrosystem — extracted from the normalized root of a word by way of its pronunciation, not its spelling. Its unordered class-set is its nucleus, the invariant conserved under a bounded calculus of admissible transformations; its order carries quality, so that inversion preserves the nucleus while reorienting the quality. A code has no meaning of its own. Within a concrete linguistic system it orients a field of semantic possibility, observable only through the coderivative networks that instantiate it, and legible only through the relations among the real words of those networks, across the several depths at which meaning is held by psychic resonance. It is the smallest bond-capable unit of the model, combining with other codes into molecules. It presupposes a continuous qualic field — vowels as operators of quality, prosody, endorhythm — which it cannot itself contain. And it is an instrument of observation: what persists through linguistic genealogy is not the code, but a structural continuity that the code allows us to detect.

For working use, the definition can be carried as a checklist. Anything presented as a code reading should be able to answer all ten.

#	requirement
1	classes declared and versioned before the analysis (C0)
2	code extracted from IPA, dialect-aware, root-normalized (C0*, C0**)
3	nucleus and order both stated; the unit of study is the concrete order
4	transformation set and $k_{\max}$ declared in advance; operations <i>found</i> in real words, not computed
5	scale declared: system, system of languages, or macrosystem
6	coderivative networks separated; no pooling of independent networks
7	each network's field described before any comparison
8	no meaning assigned to the code; no composition of class profiles; no summing of network fields
9	qualic signature considered (prosody, endorhythm, vowel direction), and its absence weighed (C3*)
10	the language of detection used throughout: the code <i>detects</i> a continuity; it does not <i>appear</i> in words

And the one-line version, for when only one line is available: *a code is a bounded, ordered pair or triple of affective sound classes that orients — but never denotes — a field of meaning, observable only through networks of coderivatives inside a declared system; what persists through history is not the code, but what the code detects.*

The paper ends where the theory ends: on a question rather than a thesis. The instrument is defined. The object it detects is not yet named. The difference between the two is where the science lives.

References

- Meulemans, C. S., and J. Á. Elias Fernández y Cuevas. *Introducción a la endolingüística Decaglot.* 1993. (*Foundation of the discipline; prior to OAS.*)
- Toledo Martínez, Alejandro. *Los códigos de la endolingüística.* Mexico: Escuela de Lingüística Aplicada Didier Elias Meulemans, 2024.
- Toledo Martínez, Alejandro. *Teoría de la Endolingüística: Fundamentos y Aplicaciones.* Editorial ELADEM, 2025. ISBN 9798291155158.
- Toledo Martínez, Alejandro. "OAS Inventory Protocol: Versioned Originating Affective Sound Classes, IIE-M v1.0." 2025.
- Toledo Martínez, Alejandro. "La matemática cuántica de los códigos endolingüísticos: inversión, núcleo e integración binario-ternaria." ELADEM, 2025.
- Toledo Martínez, Alejandro. "La dimensión cuántica de la endolingüística: prosodia, endorritmo y cualidad en la organización profunda del lenguaje." ELADEM, 2025.
- Toledo Martínez, Alejandro. "On the Falsifiability and Structural Limits of Endolinguistics." 2025.
- Toledo Martínez, Alejandro. "Endolinguistic Molecules: Inter-Code Bonding, Valence Saturation, and Compositional Structures." 2025.
- Toledo Martínez, Alejandro. "From Discrete Unit to Qualic Field: Quantum Particle, Gene, Bit, and Endolinguistic Code in a Qualic Ontology." 2025.
- Toledo Martínez, Alejandro. *Metafísica Cuántica: hacia una ontología del aparecer y una lógica de la articulación.* Qualic Metaphysics, 2025.

Toledo Martínez, Alejandro. *Elementos y cualidades pre-simbólicos y pre-verbales del endolenguaje (epistemología del endorritmo)*. ELADEM, 2025.

Basave Fernández del Valle, Agustín. *Tratado de metafísica: teoría de la habencia*. Mexico City: Editorial Limusa, 1982.

Jakobson, Roman. *Selected Writings I: Phonological Studies*. The Hague: Mouton, 1962.

Klein, Felix. *A Comparative Review of Recent Researches in Geometry*. 1872/1893. (*Invariance under transformation.*)

Merleau-Ponty, Maurice. *Phenomenology of Perception*. London: Routledge, 1962.

Saussure, Ferdinand de. *Course in General Linguistics*. 1916.

Weyl, Hermann. *Symmetry*. Princeton: Princeton University Press, 1952.