Access or Apophenia

Is the endolinguistic code game access to lexical structure, or apophenia? Machine players, and the stability of the consonantal skeleton under error

Alejandro Toledo Martínez

A study within the endolinguistic programme. It does not redefine the discipline and it does not assume its central claims. It tests one question — when a player links words that share a consonantal code, is that access to a detectable structure of the lexicon or apophenia (projection)? — and treats the errors of the impaired brain as supporting evidence that the code names something recoverable rather than imposed. One dissociation this evidence suggests (a two-layer reading of the sign) is strong enough to deserve its own study and is flagged as such, not argued here. Philosophy and psychoanalysis frame the question; the evidence is lexical and clinical. Endolinguistics was founded by C. S. Meulemans and J. Á. Elias; the present author is a systematizer of the discipline, not its founder, and the OAS reduction used here is his contribution to it.

Abstract

Endolinguistics plays a *code game*: given a seed word, link others that share its consonantal skeleton by metonymy or metaphor (Spanish *soltar* → *lazo* → *solo*). The game feels like access to a hidden order. We ask whether it is, or whether it is apophenia, and measure the question by making the player a machine. Across 600 monolingual seeds an embedding test finds an in-code advantage that is, at best, marginally detectable and practically negligible (+0.3%; seed-level $z = 2.4$; sign-test $p = 0.01$; a seed-bootstrap interval that barely clears zero). On a clean 43-item battery judged by a panel of eight language models, at equal link quality the shared code affords nothing (in-code = decoy), while discrimination of genuine from forced links rises with model capability (AUC 0.58 → 0.96). A bilingual game over five real dictionaries (English, Swedish, Spanish, Czech, Swahili) finds projection flat with genealogical distance — Swahili, wholly outside Indo-European, projects as much as sister Swedish — and the flatness survives native-competence judges reading the actual foreign words. As supporting evidence that the consonantal skeleton is not merely the analyst's projection, we show it is recoverable under error: across 75,000 spelling errors (including a dyslexic corpus) it is preserved 1.9× above a magnitude-matched null, and in 2,000 speech errors from two independent sources (aphasic speech; the open Fromkin database, four languages) errors of *form* preserve it three-to-five times more than errors of *meaning*. The picture is consistent throughout: the consonantal code is a stable, recoverable, statistically detectable constraint on word-form — but it orients without denoting; it does not carry meaning, and the game's cross-word semantic connections are projection riding on a formal coincidence. This is the endolinguistic thesis, not a refutation of it. A stronger reading — that error dissociates two layers of the sign, and that the game can be *interpreted* as a voluntary counterpart to aphasia — is compatible with the data but not demonstrated here; we set it aside for separate work.

Resumen

La endolingüística juega un *juego del código*: dada una semilla, enlazar palabras que comparten su esqueleto consonántico por metonimia o metáfora (*soltar* → *lazo* → *solo*). El juego se siente como

acceso a un orden oculto. Preguntamos si lo es, o si es apofenia, y medimos la pregunta haciendo del jugador una máquina. En 600 semillas monolingües un test de embeddings halla una ventaja en-código, a lo sumo, marginalmente detectable y prácticamente nula (+0,3%; z a nivel semilla = 2,4; sign-test p = 0,01; un IC bootstrap que apenas cruza el cero). En una batería limpia de 43 ítems juzgada por un panel de ocho modelos, a igual calidad de enlace el código compartido no aporta nada (en-código = señuelo), mientras la discriminación de enlaces genuinos vs forzados sube con la capacidad (AUC 0,58 $\rightarrow$ 0,96). Un juego bilingüe sobre cinco diccionarios reales (inglés, sueco, español, checo, swahili) halla la proyección plana con la distancia genealógica —el swahili, del todo fuera del indoeuropeo, proyecta tanto como el sueco hermano— y la planitud sobrevive a jueces de competencia nativa que leen las palabras reales. Como evidencia de apoyo de que el esqueleto consonántico no es mera proyección del analista, mostramos que es recuperable bajo error: en 75.000 errores de escritura (incluido un corpus disléxico) se conserva 1,9$\times$ sobre un nulo de igual magnitud, y en 2.000 errores de habla de dos fuentes independientes (habla afásica; la base abierta de Fromkin, cuatro lenguas) los errores de *forma* lo preservan tres a cinco veces más que los de *significado*. La imagen es consistente: el código consonántico es una restricción estable, recuperable y estadísticamente detectable sobre la forma de la palabra —pero orienta sin denotar; no porta significado, y las conexiones semánticas entre palabras del juego son proyección montada sobre una coincidencia formal. Esta es la tesis endolingüística, no su refutación. Una lectura más fuerte —que el error disocia dos capas del signo, y que el juego puede *interpretarse* como un contrapunto voluntario de la afasia— es compatible con los datos pero no queda demostrada aquí; la reservamos para trabajo aparte.

Keywords / Palabras clave

code game, consonantal skeleton, apophenia, signal detection, access, cross-linguistic projection, error corpora, endolinguistics.

1. The question — the game

The endolinguistic game is simple to state and seductive to play. Fix a *seed* word and read off its consonantal code — the ordered skeleton of consonant classes under the vowels. Then find other words sharing that code and connect them to the seed by sense: metonymy (cause, instrument, part) or metaphor. A Spanish player, seed *soltar* (to let go), code `S·L`, plays *saltar* (to leap — what is let go leaps), *lazo* (a loop — the instrument of holding and releasing), *solo* (alone — what is let go is left solitary). The winner is not who forces the most words in but who finds the most senses without getting lost in the explanation. Play it for an hour and a hidden order seems to shine through the lexicon.

That shine is the object of this study. There is a name for perceiving connection where none exists — apophenia — and there is a discipline devoted to producing exactly such connections and calling them law. So the question is unavoidable and it is not rhetorical: *when the player links two words through their shared code, is that access to a detectable structure of the lexicon, or is it apophenia?* We do not answer it by introspection. We answer it by measurement, and we keep a fixed external reference at hand: the code as it appears where no one is playing — in the stability of word-form under error.

The purpose of this paper is deliberately modest. It does not attempt to validate endolinguistics as a whole. It isolates one central intuition — the code game — and asks whether its apparent success reflects lexical structure or projection. Any broader theoretical consequences are treated only insofar as they are required to interpret the measurements.

2. Frame — access, apophenia, and a code that orients without denoting

Borrow the vocabulary of signal detection. A player faced with a candidate word can say *connected* or *not*. Two quantities describe the player independently: sensitivity (d') — how well genuine connections are told from spurious ones — and criterion — the standing bias toward saying *connected*. Access is high sensitivity. Apophenia is a lax criterion: *connected* to almost anything. A pure apophenia machine has $d' \approx 0$ and answers *yes* by reflex; a good player has $d' > 0$ and a criterion strict enough that the reflex is caught.

Two features make the case sharp. First, the code is a finite combinatorial space: with a handful of classes and length ≤ 3, coincidental code-sharing is common, so the base rate of spurious links is high — fertile soil for apophenia. Second, the discipline carries two prostheses against it. The *code itself*, with its fixed reduction rules, is a prosthesis of form: it constrains which words are even in play. The *null test* — comparing what the player finds to what chance would find — is a prosthesis of evidence: it measures the criterion from the outside.

One caution frames everything that follows: endolinguistics never assigns a fixed *meaning* to a code. The code orients, it does not denote — it is a channel along which senses organise, read from the relations among the words that share it, not a label to be decoded. The apophenic error the prostheses guard against is therefore not the discipline's claim but its shadow: mistaking the channel for its content, reading a fixed sense *off* the code. This paper audits the two prostheses on one question — does code-sharing, by itself, connect words in sense? — and uses the impaired brain only to check that the code names something recoverable rather than imposed.

3. Method

Every word, in every language and corpus, is reduced the same way: to its local code, the ordered sequence of consonant classes read by sound and folded into natural place-classes (dentals together, labials together, velars together; liquids, nasals and sibilants kept apart), collapsed to length ≤ 3. The reduction is orthographic or phonemic depending on the source, and it is the *same function* for seeds, candidates, spelling errors and phonetic transcriptions — this shared reduction is what lets the two kinds of measurement speak to each other.

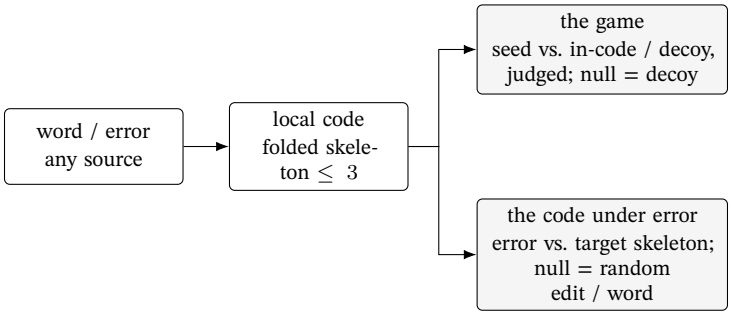

Figure 1: One reduction, two measurements. The same folded consonantal code is computed for game candidates and for errors, then each is compared to a matched null. *Una reducción, dos medidas. El mismo código consonántico plegado se computa para las candidatas del juego y para los errores, y cada uno se compara con un nulo igualado.*

The game is operationalised as a forced comparison. For a seed we assemble in-code candidates (same folded code, form-distinct from the seed) and decoy candidates (a different code, matched for form-distance). A judge scores the metonymic/metaphoric relation of each candidate to the seed. The signature of an affordance is *in-code scores above decoy scores*; the signature of apophenia is *in-code = decoy* with both elevated. Judges are embedding cosine (§4), then a panel of language models scoring

0–100 with a strict rubric (§§4–5). For the error measurement (§6) the null is a magnitude-matched random character edit (spelling) or a random word pair (speech). All tests are within-item and pre-declared; nothing is shuffled across meanings.

4. The game from inside — a whisper, and a machine that says yes

Two measurements pin the game from within. The first asks whether in-code words are *semantically* closer to the seed than decoys, using multilingual embeddings (bge-m3) over 600 seeds in seven languages, with borrowings and cognates removed by a form-dissimilarity control. The in-code relatedness is 0.4957 against a decoy 0.4941 — a difference of +0.0016 that is, at best, marginally detectable: a seed-level paired $z = 2.4$, a seed-bootstrap 95% CI of [+0.0003, +0.0028] that barely clears zero, a sign test at $p = 0.01$, the seed favoured in only 55% of cases. And it is practically nothing — about a third of one percent of the relatedness is attributable to the code. The playable field exists; it is a whisper, and the distribution-free tests can only just tell it from silence.

The second measurement replaces cosine — which cannot see metonymy — with a judge that can: a language model. Here a calibration hazard surfaces and is itself a result. Asked a *binary* question, a mid-size model (gemma2) says *natural* to every pairing, narrating a bridge for *let-go table*, *let-go dog*, *let-go candle* alike; it never refuses. Only a graded 0–100 rubric, calibrated with worked examples, recovers discrimination. On a clean 43-item battery — common words, common senses, code verified, a 2×2 of code-match × genuine/forced link — the panel behaves as signal detection predicts:

judge	AUC (genuine vs forced)	forced-link score
qwen2.5-3b	0.58 ≈ chance	17.2
gemma2-9b	0.84	15.7
Haiku 4.5	0.93	4.7
Sonnet 5	0.91	10.2
Fable 5	0.95 → 0.96	6.5
Opus 4.8	0.96	5.9

Two results. Discrimination scales with capability: small models sit at chance (qwen scores everything 18 — noise, not apophenia), and resistance to apophenia climbs to 0.96; deliberation ("careful" mode) sharpens scores but does not move the AUC. And, decisively for the endolinguistic question, the shared code affords nothing extra: at equal link quality (forced links) in-code and decoy candidates score identically (15.7 each for gemma2); the highest-scoring item in the whole battery is a *decoy* with a genuine link (*fire* → *smoke*, *sun* → *star*) — a real connection whose code does *not* match the seed. The game's connections are semantic, and they exist in any code.

5. The bilingual game — projection is flat with distance

If access grew with reach, a player who commands more languages should find *more and truer* connections, richest where history is shared (sister languages) and thin where it is not. We test this with a distance ladder from an English seed: Swedish (Germanic sister), Spanish (Romance), Czech (Slavic), and — the arbiter — Swahili, a Bantu language wholly outside Indo-European. For eight English seeds we sampled in-code, form-distinct words from each language's real dictionary (Wiktionary/Kaikki, each entry carrying an English gloss), with no hand-curation, and a judge scored the seed–candidate relation. Projection is the mean score per language.

The profile is flat (slope −1.3 points per rung), and the code-effect (in-code minus decoy) collapses to zero and turns *negative* for Swahili (−3.0). The arbiter is decisive: Swahili shares no history with the

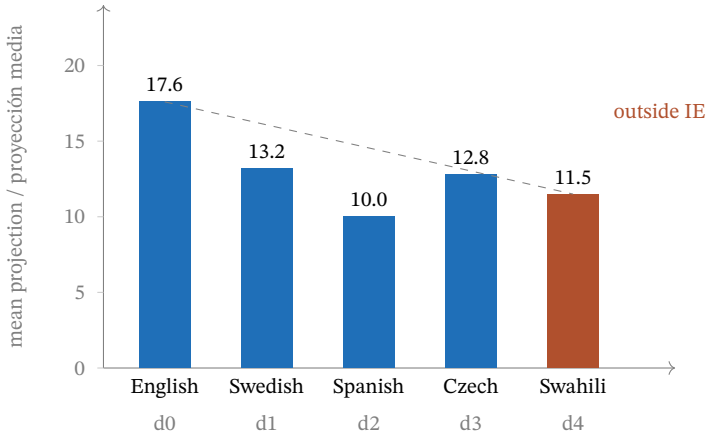

Figure 2: Projection onto in-code words does not decay with genealogical distance from the English seed (slope -1.3 per rung). Swahili (d4, outside Indo-European) projects as much as sister Swedish (d1). The only elevated bar, English (d0), is confounded: this judge operates in English. *La proyección sobre palabras en-código no decae con la distancia genealógica desde la semilla inglesa (pendiente* $-1{,}3$ *por peldaño). El swahili (d4, fuera del indoeuropeo) proyecta tanto como el sueco hermano (d1). La única barra elevada, el inglés (d0), está confundida: este juez opera en inglés.*

seed, yet it "connects" as readily as Swedish. The only slightly raised bar is English itself, confounded by design — this judge reasons in English, so English candidates carry incidental pull — but even it is low (17.6/100) and does not disturb the sv ≈ cs ≈ sw indifference.

One objection must be met head-on: this judge read English glosses with a mid-size model, so genealogical distance could be neutralised by construction. We re-ran the whole ladder with native-competence judges — one strong multilingual judge per language, shown the *actual foreign word* and asked to reason in that language about its form, senses and connotations. The judges became far stricter (mean scores roughly halved) — yet the distance profile stayed flat (slope −0.6 per rung): Swahili still projects as much (5.0) as English (6.3) or Czech (7.0). The single elevated code-effect, Swedish (in − decoy +8.4), traces almost entirely to one cognate the form filter missed — *måne* / *moon*, scored 100 — plus a couple of genuine Germanic metonymies: that is shared ancestry surfacing in the nearest sister, not distance-independent access. Seeing the real words, judged natively, only sharpens the judges and exposes the one cross-language signal as cognation.

6. The code under error — a recoverable invariant

The results so far say what the code is *not* (a source of cross-word meaning). This section fixes what it *is*, using a reference the player cannot move: the code as it surfaces where no one is playing — in error. If the consonantal skeleton were merely the analyst's way of looking, it should leave no trace in involuntary mistakes. It leaves a strong one.

Across 75,000 real error→target pairs from spelling-error corpora — Birkbeck/Mitton, a corpus of dyslexic children's errors (Holbrook), Wikipedia, aspell, Norvig — we classify each error by the operation it performs on the target's folded skeleton, against a null that applies the *same number* of random character edits to the same target.

operation	real	null	ratio
identity (skeleton intact)	50.6%	27.2%	×1.9
reduction (lose a consonant)	16.2%	14.5%	×1.1
substitution (consonant change)	9.3%	12.9%	×0.7

operation	real	null	ratio
expansion (add a consonant)	9.2%	19.6%	×0.5
permutation (metathesis)	1.3%	2.3%	×0.6
other (multiple changes)	13.3%	23.3%	×0.6

Half of all real errors — dyslexic errors included (Holbrook 51% vs 34% null) — leave the consonantal skeleton intact, changing only vowels or doublings, nearly twice as often as chance; every operation that destroys the skeleton sits below chance. The ×1.9 is not an artefact of a random null over-hitting consonants: a more realistic QWERTY keyboard-adjacency null gives the same 26% baseline and the same ×1.9. Under noise, what is recovered is the consonantal skeleton — the object the code names, surfacing from a source unconnected to endolinguistics.

Speech shows the same, and adds a distinction between two kinds of error. From AphasiaBank's APROCSA (28 connected-speech transcripts of people with aphasia) we take every coded error as a produced/target pair with its type — phonological (a form error: *cot* for *got*) or semantic (a meaning error: *lamp* for *chandelier*). Phonological errors preserve the folded skeleton far more than semantic ones. Because one small corpus is easy to over-read, we replicate on an open, independent source: the Fromkin Speech Error Database (MPI Nijmegen), 8,673 natural slips in four languages, with whole-word target–error pairs.

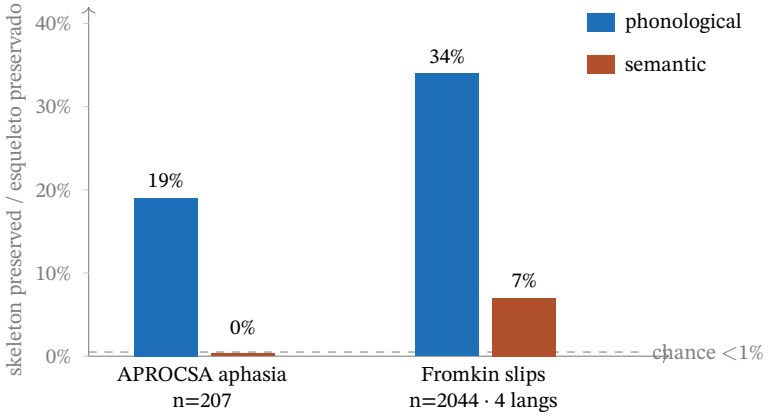

Figure 3: Two independent sources agree: form errors preserve the consonantal skeleton far more than meaning errors. Aphasic speech (APROCSA): phonological 19% vs semantic 0/30. Natural slips (Fromkin/MPI, four languages): phonological 34% (×383 chance) vs semantic 7%. *Dos fuentes independientes coinciden: los errores de forma preservan el esqueleto mucho más que los de significado. Habla afásica (APROCSA): fonológico 19% vs semántico 0/30. Slips naturales (Fromkin/MPI, cuatro lenguas): fonológico 34% (×383) vs semántico 7%.*

The contrast is consistent across both sources and four languages: an error of *form* preserves the skeleton three-to-five times more than an error of *meaning* (phonological 19% and 34%, up to ×383 chance; semantic 0/30 and 7%). The larger sample corrects the aphasic 0/30 — meaning errors preserve the skeleton not *never* but *rarely*, and the residual 7% are malapropism-like slips that land on a form-similar word (a form effect). The load-bearing point for this paper is narrow and firm: the consonantal skeleton is recoverable and stable under error, above chance, robustly — so the code names a detectable regularity of word-form, not the analyst's projection. What that form/meaning asymmetry might further imply about the architecture of the sign, we take up as interpretation (§7), not as a claim proved here.

7. Discussion

The game is dominated by apophenia — but apophenia over a detectable pattern. Every measurement from inside the game points the same way. The embedding whisper (+0.3%, barely separable from zero) shows that the playable field is almost all projection: without the null test, a third-of-a-percent enrichment feels, from inside, like a law — which is exactly why the prosthesis of evidence is indispensable. The panel shows that, at equal link quality, code-sharing adds nothing; the connections players feel are semantic and exist in any code. The bilingual ladder shows the projection is distance-independent: an out-of-system arbiter (Swahili) projects as much as a sister language, even when a native-competence judge reads the real words. Multilingual reach is not privileged access to a specular order; it is more material to project onto.

The code orients without denoting — and this is the endolinguistic thesis, not its refutation. The error evidence (§6) fixes the code as a stable, recoverable constraint on *form*: it survives noise above chance and is what mistakes leave standing. But nothing here makes it *carry meaning*, and the discipline never claimed it did. The form/meaning asymmetry in speech errors is the cleanest illustration: when the impaired system slips on form, the skeleton stays; when it slips on meaning, the skeleton is largely discarded. If the code denoted meaning, a change of sense would drag the skeleton along as much as a change of form does; it does not — three-to-five times less. This is consistent with reading form and meaning as separable layers, the code living in the form layer, orienting and channelling rather than labelling. We state this as an interpretation the data are compatible with; establishing it as a claim about the architecture of the sign — with error-type coding designed for the purpose and human raters — is a separate study, and a psycholinguistic one.

A suggested, undemonstrated reading of the game. The same separability invites an elegant picture: that the game is a *voluntary counterpart* to error. Where phonological error involuntarily preserves the skeleton, the game deliberately reconfigures it (the inversion and permutation that natural error, as §6 shows, rarely performs); where semantic error involuntarily substitutes meaning free of the code, the game deliberately does the same. This "voluntary inverse of aphasia" is a hypothesis the results are *compatible with*, not one we have tested: we measured neither intention nor a mapping between game moves and error types. We flag it as a direction, not a finding.

Two prostheses, and where the signal actually lives. The study is, in the end, about the discipline's two safeguards. The *code with its rules* holds up as a prosthesis of form, because the skeleton is a recoverable regularity, not an invention. The *null test* is indispensable, because the fixed sense a player reads *off* the code is projection — the real sense is relational, to be measured against chance, never assumed. And it points to where the endolinguistic signal is denser: not *between* words of a shared code (thin, and mostly projection) but *within* a word — the organisation of one word's senses (polysemy), and the borrowing-controlled convergence for shared meaning reported in the companion study. The game reaches outward; the signal is denser inward.

> Note on capability. The panel's most capable judges are both its strictest scorers and its best discriminators: apophenia is not cured by deliberation but by capacity. The corollary is a caution the discipline must keep naming — the more capable the player, the more *convincing* the projection they can generate. A good endolinguist is therefore not one who feels fewer connections but one who submits more of them to the null.

8. Scope and falsifiability

We report what these instruments show under these conditions, and mark the limits. The panel's battery key was one model's labelling; a human panel (twelve naive raters, Cronbach $\alpha = 0.93$) now reproduces the capability gradient against an independent standard — re-scored against the human-consensus key the ordering holds and human raters land mid-gradient (AUC 0.82), so it is not an

artefact of the model key (details in the companion paper). It remains a pilot; more raters will tighten the absolute values. The reduction is orthographic for some sources and phonemic for others, declared per source, never mixed silently. ”Connectivity judgement” is a proxy for apophenia, not the clinical phenomenon. The error analysis (§6) is supporting evidence, kept deliberately light; its finer grain — matched error-type coding across sources, human re-rating, and the two-layer reading of §7 — is the material of a separate paper.

The claims are built to be broken. Access, not apophenia, would show as: in-code scores that beat matched decoys at equal link quality; projection that *decays* with genealogical distance and drops for the out-of-system arbiter, even under native judges. We looked for each and found the opposite or nothing. A meaning-bearing code (a claim the discipline never made — the code orients, it does not denote) would show as the code affording sense over decoys and as meaning-errors tracking the skeleton as strongly as form-errors; neither appears. What survives every test is narrower and firmer than the game’s promise: the consonantal skeleton is a stable, recoverable, statistically detectable constraint on word-form, along which both the player and the erring system operate — but it does not, by itself, carry the meaning the player feels flowing through it.

Versión en español

1. La pregunta — el juego

El juego endolingüístico es simple de enunciar y seductor de jugar. Se fija una *semilla* y se lee su código consonántico —el esqueleto ordenado de clases de consonante bajo las vocales—. Luego se buscan otras palabras con ese código y se conectan a la semilla por sentido: metonimia (causa, instrumento, parte) o metáfora. Un jugador español, semilla *soltar*, código `S·L`, juega *saltar* (lo que se suelta salta), *lazo* (el instrumento de sujetar y soltar), *solo* (lo que se suelta queda solitario). Gana no quien mete más palabras a la fuerza sino quien halla más sentidos sin perderse en la explicación. Juéguese una hora y un orden oculto parece traslucirse en el léxico.

Ese brillo es el objeto de este estudio. Hay un nombre para percibir conexión donde no la hay —apofenia— y hay una disciplina dedicada a producir tales conexiones y llamarlas ley. La pregunta es inevitable y no es retórica: *cuando el jugador enlaza dos palabras por su código compartido, ¿es acceso a una estructura detectable del léxico, o es apofenia?* No la respondemos por introspección, sino por medición, y con una referencia externa fija a mano: el código tal como aparece donde nadie juega —en la estabilidad de la forma de la palabra bajo error.

El propósito de este paper es deliberadamente modesto. No pretende validar la endolingüística en su conjunto. Aísla una intuición central —el juego del código— y pregunta si su éxito aparente refleja estructura léxica o proyección. Cualquier consecuencia teórica más amplia se trata solo en la medida en que es necesaria para interpretar las mediciones.

2. Marco — acceso, apofenia y un código que orienta sin denotar

Tomemos el vocabulario de la detección de señal. Ante una candidata el jugador dice *conectada* o *no*. Dos magnitudes lo describen con independencia: sensibilidad (d') —cuán bien distingue conexiones genuinas de espurias— y criterio —el sesgo a decir *conectada*—. El acceso es alta sensibilidad. La apofenia es un criterio laxo. Una máquina de apofenia pura tiene $d' \approx 0$ y dice *sí* por reflejo; un buen jugador tiene $d' > 0$ y un criterio que atrapa el reflejo.

Dos rasgos afilan el caso. Primero, el código es un espacio combinatorio finito: con pocas clases y longitud ≤ 3, compartir código por azar es común, así que la tasa base de enlaces espurios es alta. Segundo, la disciplina porta dos prótesis. El *código mismo*, con sus reglas fijas, es prótesis de forma: restringe qué palabras entran en juego. El *test nulo* —comparar lo que el jugador halla con el azar— es prótesis de evidencia: mide el criterio desde fuera.

Una cautela enmarca todo lo que sigue: la endolingüística nunca asigna un *significado* fijo a un código. El código orienta, no denota —es un canal por donde los sentidos se organizan, leídos de las relaciones entre las palabras que lo comparten, no una etiqueta a descifrar—. El error de apofenia contra el que las prótesis protegen no es la tesis de la disciplina sino su sombra: confundir el canal con su contenido, leer un sentido fijo *en* el código. Este trabajo audita las dos prótesis sobre una pregunta —¿compartir código, por sí solo, conecta palabras en sentido?— y usa el cerebro dañado solo para comprobar que el código nombra algo recuperable, no impuesto.

3. Método

Cada palabra, en toda lengua y corpus, se reduce igual: a su código-local, la secuencia ordenada de clases de consonante leída por sonido y plegada en clases naturales por lugar (dentales, labiales, velares juntas; líquidas, nasales y sibilantes aparte), colapsada a longitud ≤ 3. La reducción es ortográfica o fonémica según la fuente, y es la *misma función* para semillas, candidatas, errores de escritura y transcripciones fonéticas — eso permite que las dos medidas se hablen (Fig. 1).

El juego se operacionaliza como comparación forzada. Para una semilla reunimos candidatas en-código (mismo código plegado, de forma distinta) y señuelo (código distinto, igualadas en distancia de forma). Un juez puntúa la relación metonímica/metafórica con la semilla. La firma de una afordancia es *en-código por encima de señuelo*; la de la apofenia es *en-código = señuelo*, ambas elevadas. En la medida del error (§6) el nulo es una edición aleatoria de igual magnitud o un par de palabras al azar. Todo es intra-ítem y pre-declarado; nada se baraja entre significados.

4. El juego desde dentro — un susurro, y una máquina que dice sí

Dos medidas fijan el juego por dentro. La primera pregunta si las palabras en-código están *semánticamente* más cerca de la semilla que los señuelos, con embeddings multilingües (bge-m3) sobre 600 semillas en siete lenguas, quitando préstamos y cognados por disimilitud de forma. La cercanía en-código es 0,4957 frente a 0,4941 del señuelo —una diferencia de +0,0016 a lo sumo marginalmente detectable: z pareado a nivel semilla = 2,4; IC bootstrap-semilla 95% de [+0,0003, +0,0028] que apenas cruza el cero; sign test p = 0,01; la semilla favorecida en solo el 55%—. Y prácticamente nula: cerca de un tercio del uno por ciento de la cercanía es atribuible al código. El campo jugable existe; es un susurro, y los tests distribution-free apenas lo distinguen del silencio.

La segunda medida reemplaza el coseno —que no ve metonimia— por un juez que sí: un modelo de lenguaje. Aquí aflora un peligro de calibración que es en sí un resultado. En pregunta *binaria*, un modelo mediano (gemma2) llama *natural* a todo par (*soltar mesa*, *soltar perro*, *soltar vela* por igual); no rechaza nada. Solo una rúbrica graduada 0–100, calibrada, recupera la discriminación. En una batería limpia de 43 ítems —palabras y sentidos comunes, código verificado, 2×2 de coincidencia $\times$ enlace genuino/forzado— el panel se comporta como predice la detección de señal: la discriminación (AUC genuino vs forzado) sube de 0,58 (qwen, azar) $\rightarrow$ 0,84 (gemma2) $\rightarrow$ 0,91–0,96 (Claude); deliberar afila los scores pero no mueve el AUC. Y lo decisivo: a igual calidad de enlace (forzados), en-tablero y señuelo puntúan idéntico (15,7 ambos en gemma2); el ítem que más puntúa es un *señuelo* con enlace genuino (*fire* $\rightarrow$ *smoke*, *sun* $\rightarrow$ *star*), cuyo código *no* coincide. Las conexiones del juego son semánticas, y existen en cualquier código.

5. El juego bilingüe — la proyección es plana con la distancia

Si el acceso creciera con el alcance, quien domina más lenguas hallaría *más y más verdaderas* conexiones, más ricas donde la historia es compartida y tenues donde no. Lo probamos con una escalera de distancia desde una semilla inglesa: sueco (hermana germánica), español (romance), checo (eslava) y —el árbitro— swahili, lengua bantú del todo fuera del indoeuropeo. Para ocho semillas muestreamos palabras en-código, de forma distinta, de cada diccionario real (Wiktionary/Kaikki, con glosa inglesa), sin curación, y un juez puntuó la relación semilla–candidata. La proyección es el score medio por lengua (Fig. 2).

El perfil es plano (pendiente $-1{,}3$ por peldaño), y el efecto-código (en-código menos señuelo) colapsa a cero y se vuelve *negativo* en swahili ($-3{,}0$). El árbitro es decisivo: el swahili no comparte historia con la semilla y aun así "conecta" tan fácil como el sueco. La única barra algo elevada es el inglés, confundida

por diseño —este juez razona en inglés—, pero aun baja (17,6/100) y no altera la indiferencia sv ≈ cs ≈ sw.

Una objeción hay que enfrentarla de frente: este juez leía glosas inglesas con un modelo mediano, así que la distancia podía quedar neutralizada por diseño. Re-corrimos toda la escalera con jueces de competencia nativa —un juez multilingüe fuerte por lengua, que ve la *palabra extranjera real* y razona en esa lengua sobre su forma, sentidos y connotaciones—. Los jueces se volvieron mucho más estrictos (medias a la mitad); pero el perfil por distancia siguió plano (pendiente −0,6 por peldaño): el swahili proyecta tanto (5,0) como el inglés (6,3) o el checo (7,0). El único efecto-código elevado, el sueco (in − señuelo +8,4), viene casi todo de un cognado que el filtro de forma no atrapó —*måne* / *moon*, puntuado 100— más un par de metonimias germánicas: es ascendencia compartida asomando en la lengua más cercana, no acceso distancia-independiente. Ver las palabras reales, juzgadas por nativos, solo afila a los jueces y desnuda la única señal cross-lingüística como cognación.

6. El código bajo error — un invariante recuperable

Lo anterior dice lo que el código *no* es (fuente de significado entre palabras). Esta sección fija lo que *sí* es, con una referencia que el jugador no puede mover: el código tal como asoma donde nadie juega — en el error. Si el esqueleto consonántico fuera solo la manera de mirar del analista, no dejaría rastro en los errores involuntarios. Deja uno fuerte.

En 75.000 pares reales error→objetivo de corpora de escritura —Birkbeck/Mitton, un corpus de errores de niños disléxicos (Holbrook), Wikipedia, aspell, Norvig— clasificamos cada error por la operación sobre el esqueleto plegado del objetivo, contra un nulo que aplica el *mismo número* de ediciones aleatorias al mismo objetivo. La mitad de los errores reales —disléxicos incluidos (Holbrook 51% vs 34%)— dejan el esqueleto intacto, cambiando solo vocales o dobles, casi el doble que el azar (×1,9); toda operación que lo destruye queda por debajo del azar. El ×1,9 no es artefacto de un nulo que golpea consonantes en exceso: un nulo de adyacencia de teclado QWERTY da la misma base (26%) y el mismo ×1,9. Bajo ruido, lo que se recupera es el esqueleto consonántico — el objeto que el código nombra, asomando desde una fuente ajena a la endolingüística.

El habla muestra lo mismo, y añade una distinción entre dos tipos de error. De APROCSA de AphasiaBank (28 transcripciones de habla afásica) tomamos cada error codificado como par producido/objetivo con su tipo —fonológico (error de forma: *cot* por *got*) o semántico (error de sentido: *lamp* por *chandelier*)—. Los fonológicos preservan el esqueleto mucho más que los semánticos. Como un corpus pequeño es fácil de sobre-leer, replicamos en una fuente abierta e independiente: la Fromkin Speech Error Database (MPI Nijmegen), 8.673 slips naturales en cuatro lenguas, con pares palabra entera (Fig. 3).

El contraste es consistente en ambas fuentes y cuatro lenguas: un error de *forma* preserva el esqueleto tres a cinco veces más que uno de *significado* (fonológico 19% y 34%, hasta ×383 el azar; semántico 0/30 y 7%). La muestra mayor corrige el 0/30 afásico —los errores de significado lo preservan no *nunca* sino *rara vez*, y ese 7% son slips tipo malapropismo que caen en palabra de forma parecida (un efecto de forma)—. El punto que carga el peso de este paper es estrecho y firme: el esqueleto consonántico es recuperable y estable bajo error, sobre el azar, de forma robusta — así que el código nombra una regularidad detectable de la forma de la palabra, no la proyección del analista. Lo que esa asimetría forma/significado pueda implicar sobre la arquitectura del signo, lo tomamos como interpretación (§7), no como algo probado aquí.

7. Discusión

El juego está dominado por la apofenia — pero apofenia sobre un patrón detectable. Toda medida desde dentro apunta igual. El susurro de embeddings (+0,3%, apenas separable de cero) muestra que el campo jugable es casi toda proyección: sin el nulo, un enriquecimiento de un tercio de por ciento se siente, desde dentro, como una ley — por eso la prótesis de evidencia es indispensable. El panel muestra que, a igual calidad de enlace, compartir código no añade nada; las conexiones que el jugador siente son semánticas y existen en cualquier código. La escalera bilingüe muestra que la proyección es distancia-independiente: un árbitro fuera del sistema (swahili) proyecta tanto como una lengua hermana, incluso con juez nativo que ve las palabras reales. El alcance multilingüe no es acceso privilegiado a un orden especular; es más material para proyectar.

El código orienta sin denotar — y esto es la tesis endolingüística, no su refutación. La evidencia del error (§6) fija el código como restricción estable y recuperable sobre la *forma*: sobrevive al ruido sobre el azar y es lo que el error deja en pie. Pero nada aquí lo hace *portar significado*, y la disciplina nunca afirmó que lo hiciera. La asimetría forma/significado en los errores de habla lo ilustra: cuando el sistema resbala por forma, el esqueleto queda; cuando resbala por significado, se descarta en gran medida. Si el código denotara, cambiar el sentido arrastraría el esqueleto tanto como cambiar la forma; no lo hace —tres a cinco veces menos—. Esto es consistente con leer forma y significado como capas separables, viviendo el código en la de forma, orientando y canalizando, no etiquetando. Lo enunciamos como interpretación con la que los datos son compatibles; establecerlo como afirmación sobre la arquitectura del signo —con codificación de tipos diseñada al efecto y jueces humanos— es un estudio aparte, y psicolingüístico.

Una lectura sugerida, no demostrada, del juego. La misma separabilidad invita a una imagen elegante: que el juego es un *contrapunto voluntario* del error. Donde el error fonológico preserva el esqueleto sin querer, el juego lo reconfigura a propósito (la inversión y permutación que el error natural, como muestra §6, rara vez hace); donde el error semántico sustituye significado libre del código, el juego hace lo mismo. Este ”inverso voluntario de la afasia” es una hipótesis con la que los resultados son *compatibles*, no una que hayamos probado: no medimos ni la intención ni un mapeo entre jugadas y tipos de error. La señalamos como dirección, no como hallazgo.

Dos prótesis, y dónde vive la señal. El estudio trata, al final, de las dos salvaguardas de la disciplina. El *código con sus reglas* aguanta como prótesis de forma, porque el esqueleto es una regularidad recuperable, no un invento. El *test nulo* es indispensable, porque el sentido fijo que el jugador lee *en* el código es proyección — el sentido real es relacional, a medir contra el azar, nunca a suponer. Y señala dónde es más densa la señal endolingüística: no *entre* palabras de un código (tenue, y casi toda proyección) sino *dentro* de una palabra —la organización de sus sentidos (polisemia), y la convergencia controlada por préstamo para el significado compartido del estudio hermano—. El juego alcanza hacia afuera; la señal es más densa hacia adentro.

> Nota sobre la capacidad. Los jueces más capaces del panel son a la vez los más estrictos y los mejores discriminadores: la apofenia no se cura deliberando sino teniendo capacidad. El corolario es una cautela que la disciplina debe seguir nombrando — cuanto más capaz el jugador, más *convincente* la proyección que puede generar. Un buen endolingüista no es el que siente menos conexiones sino el que somete más de ellas al nulo.

8. Alcance y falsabilidad

Reportamos lo que estos instrumentos muestran bajo estas condiciones, y marcamos los límites. La clave de la batería era el etiquetado de un modelo; un panel humano (doce evaluadores naive, Cronbach $\alpha = 0{,}93$) reproduce ahora el gradiente de capacidad contra un estándar independiente —re-

puntuado contra la clave de consenso humano el orden aguanta y los evaluadores caen a mitad del gradiente (AUC 0,82), así que no es un artefacto de la clave del modelo (detalle en el paper hermano)— . Sigue siendo un piloto; más evaluadores afinarán los valores absolutos. La reducción es ortográfica para unas fuentes y fonémica para otras, declarada por fuente, nunca mezclada en silencio. ”Juicio de conectividad” es un proxy de la apofenia, no el fenómeno clínico. El análisis del error (§6) es evidencia de apoyo, deliberadamente ligera; su grano fino —codificación de tipos igualada entre fuentes, re-anotación humana, y la lectura de dos capas de §7— es el material de un paper aparte.

Las afirmaciones se construyeron para poder romperse. Acceso, no apofenia, se vería como: scores en-código que baten a señuelos igualados a igual calidad de enlace; proyección que *decae* con la distancia genealógica y cae para el árbitro fuera del sistema, aun con jueces nativos. Buscamos cada una y hallamos lo contrario o nada. Un código portador de significado (algo que la disciplina nunca afirmó —el código orienta, no denota—) se vería como el código afordando sentido sobre los señuelos y como errores de significado siguiendo el esqueleto tan fuerte como los de forma; ninguno aparece. Lo que sobrevive a toda prueba es más estrecho y firme que la promesa del juego: el esqueleto consonántico es una restricción estable, recuperable y estadísticamente detectable sobre la forma de la palabra, a lo largo de la cual operan tanto el jugador como el sistema que yerra — pero no porta, por sí solo, el significado que el jugador siente fluir por él.

Appendix — reproducibility / reproducibilidad

Code and data live under `docs/studies/specular-cognation/` (game and error experiments) and `docs/studies/transformations/` (the shared local-code reduction and the companion resonance study). The reduction is `transformations/src/localcode.py` (`code_of`, `classes`) and `ipacode.py`, folding consonants by place. Experiments and their tables in `specular-cognation/data/specular.db`:

- Exp 1 — `src/code_game.py`: monolingual embedding game (bge-m3), 600 seeds, form-dissimilarity control, seed-bootstrap + sign test; table `code_game_result`.
- Exp 2 — `../llm-apophenia/src/build_battery.py` + `judge_battery.py`: clean 43-item battery, 2×2 design, panel of eight judges (Ollama qwen/gemma2 + Claude Haiku/Sonnet/Fable/Opus); `battery.db`.
- Exp 3 — `src/build_bilingual_real.py`: bilingual game over Kaikki dictionaries (English, Swedish, Spanish, Czech, Swahili), judge gemma2; table `bilingual_real`.
- Exp 3-bis — `src/build_native_batches.py` + `src/aggregate_native.py`: native-competence re-judging by one strong multilingual judge per language, shown the real word (gloss-confound robustness check); table `bilingual_native`.
- Exp 4 — `src/clinical_ops.py`: 75,000 spelling error→target pairs (Birkbeck/Mitton, Norvig), operation classification vs random-edit and keyboard-adjacency nulls; table `clinical_ops`.
- Exp 5 — `src/cha_paraphasia.py`: AphasiaBank/APROCSA (28 CHAT transcripts), phonological vs semantic skeleton preservation; table `paraphasia`.
- Exp 6 — `src/fromkin_ops.py`: Fromkin Speech Error Database (MPI Nijmegen), 8,673 natural slips in four languages — an open replication of Exp 5; table `fromkin_ops`.

Open sources (no affiliation): Birkbeck/Mitton and Norvig spelling corpora; Kaikki dictionaries; the Fromkin Speech Error Database (`mpi.nl/dbmpi/sedb`), which is what lets §6 replicate without institutional access. AphasiaBank/APROCSA requires approved access (obtained by logged-in browser download). Judges: Ollama (`gemma2:9b`, `qwen2.5:3b`, `bge-m3`) and Claude models. Everything is deterministic apart from the language-model judges (temperature 0.1), seeded where sampling occurs.

Todo el código y los datos están bajo `docs/studies/specular-cognation/` y `docs/studies/transformations/`. Los experimentos y tablas se listan arriba; las fuentes abiertas (Birkbeck, Norvig, Kaikki, Fromkin/MPI) se descargan con los scripts sin acceso institucional; solo APROCSA requiere acceso aprobado. Determinista salvo los jueces LLM (temperatura 0,1), con semilla fija donde hay muestreo.

Atribución. La endolingüística fue fundada por C. S. Meulemans y J. Á. Elias; el presente autor es sistematizador de la disciplina, no su fundador, y la reducción OAS que aquí se usa es su aporte a ella.