Two Kinds of Stable: Decomposing Consonant Stability into Identity and Regularity

Anonymous ACL submission

Abstract

Computational historical linguistics has long assumed that consonants are more diachronically stable than vowels, an assumption operationalized in the class-based methods that dominate cognate detection. "Stability" conflates two properties — identity (how often a segment matches its counterpart) and regularity (when it changes, whether the change stays within a phonological class) — that existing methods do not separate. We separate them and measure each directly over Needleman-Wunsch alignments of ASJP wordlists across ten language families. Regularity universally favors consonants: in all ten families, consonant changes stay within phonological class more often than vowel changes do. Identity does not pattern this way: six families show higher vowel identity, one is a tie, and three show higher consonant identity. Class-based methods such as DOLGO and SCA succeed because their class structures capture regularity rather than because consonants have higher identity. The assumption that consonants are more stable than vowels holds, but only for regularity.

1 Introduction

"Stable" is two things. A sound can be stable in the sense that it is preserved as itself in cognate words across related languages: Latin /p/ in *pater* and Sanskrit /p/ in *pitár*. It can also be stable in the sense that when it does change, the change stays within a phonological class: PIE *p becomes /f/ in Germanic, not /n/ or /l/. Identity and regularity are separate properties of a phoneme's behavior over historical time. A method that finds one does not thereby find the other.

Computational historical linguistics has not treated them separately. Since Dolgopolsky (1964), the working assumption has been that consonants are more diachronically stable than vowels. This assumption is operationalized in the consonant-class matching method (Dolgopolsky, 1986; Turchin et al., 2010; Kassian et al., 2021), in ASJP-derived distance metrics (Brown et al., 2008; Wichmann et al., 2025), in the sound-class systems behind LexStat and SCA (List, 2012b,a), and in PMI-based weighted alignment (Jäger, 2013, 2018). Vowels are discarded in some pipelines (Dolgopolsky, 1986; Turchin et al., 2010) and downweighted in others (Brown et al., 2008; Wichmann et al., 2025). No existing method decomposes "stability" into these two components.

We separate the two properties by measuring each directly over aligned cognate pairs from 10 language families in the ASJP database. Identity is the rate at which an aligned segment in one language matches the segment in a related language. Regularity, in our usage, is the rate at which a target segment that differs from the source nevertheless remains within the same phonological class. This differs from the Neogrammarian sense of regular sound correspondence — a change can be exceptionlessly regular while crossing class boundaries (/p/ > /h/) — and captures phonetic proximity of change rather than lawfulness.

Across 10 language families, regularity favors consonants in every family. Ten out of ten. Identity does not behave this way: six families favor vowels, one is a tie (Indo-European), and three favor consonants. The three on the consonant side of identity are Turkic, Dravidian, and Uralic, all in Dolgopolsky's (1964) Nostratic macrofamily proposal, which raises the possibility that the consonant-stability assumption reflects properties of that grouping rather than a cross-linguistic universal; we examine this in Section 4.4. Our principal evidence that the two measures diverge is this contrast between a pattern that holds in every family and one that does not.

If consonant stability in identity is sample-dependent rather than universal, then methods that discard or downweight vowels may be discarding signal that varies by family. The regularity advan-

tage, by contrast, appears robust enough to justify consonant-class approaches — but for reasons that have not been directly tested.

The framework decomposes a phoneme's diachronic behavior into identity and regularity and measures each directly. Existing class-based methods such as DOLGO and SCA succeed because their class structures already capture regularity rather than because consonants have higher identity. The decomposition is the contribution.

2 Related Work

Dolgopolsky (1964) introduced ten consonant equivalence classes in the course of comparative work on the language families of Northern Eurasia, defining cognacy as identity of the first two consonants of a root and discarding vowels entirely. The classes were grouped so that "phonetic correspondences inside a 'type' are more regular than those between different types" (as cited in List, 2012b). The criterion is empirical: correspondences should be more consistent within a class than across classes. The operationalization measured class identity and discarded vowels. Turchin et al. (2010) refined the system to nine classes for cross-family permutation testing; Kassian et al. (2021) expanded it to thirteen. Across the lineage, vowels remain outside the unit of measurement, and class-level identity on the first consonants of a root is the only diachronic property the system quantifies.

The ASJP database (Brown et al., 2008; Wichmann et al., 2025) supplies the aligned wordlists used here. ASJPcode transcribes 34 consonant symbols and 7 vowel symbols, an inventory asymmetry that gives consonants more discriminative weight before any stability claim is made. Brown et al. (2013) compile 692 recurrent sound correspondences over ASJP: 582 consonant-only, 64 vowel-only, 46 mixed. This roughly 9:1 consonant-to-vowel ratio is the most frequently cited empirical evidence for cross-linguistic C > V claims; our decomposition attributes the asymmetry to regularity rather than identity. The SCA sound-class system (List, 2012b) extends Dolgopolsky (1986), and LexStat (List, 2012a) builds it into a cognate-detection pipeline; both produce a single similarity score that combines identity and within-class regularity without separating them. ALINE (Kondrak, 2000, 2003) applies feature-weighted Needleman-Wunsch with a salience hyperparameter that assigns higher weight to consonant identities than vowel identities. This is the clearest case of consonant primacy treated as a design choice rather than an empirical finding. These methods score consonants and vowels differently, but the differential treatment is a setting rather than a result.

Hruschka et al. (2015) model regular versus sporadic sound change in Turkic via Bayesian MCMC, the closest prior treatment of regularity as a quantifiable property. Their measure is calibrated within one family against a sporadic-change null; ours measures within-class containment across 10 families and reads it against identity on the same alignments. Turkic is one of the three families on the consonant side of identity in our results. List (2019) infers sound correspondence patterns across multiple languages: the system learns which correspondence classes are present in the data, where we measure the rate at which segment changes stay within a given class. Jäger (2013, 2018) learn segment-to-segment PMI scores over ASJP by weighted alignment, the methodological ancestor of our transformation matrix. Hierarchical clustering of the Jäger (2018) learned PMI matrix separates vowels from consonants at the highest level, but this is a discriminability claim, not a stability claim. Each of these works quantifies a single property of sound change; our framework measures two properties on the same alignments and contrasts them.

Recent neural models for protoform reconstruction and reflex prediction (Meloni et al., 2021; Kim et al., 2023; Akavarapu and Bhattacharya, 2023, 2024; Lu et al., 2024a,b) treat consonants and vowels symmetrically as tokens and optimize for prediction of held-out segments. Their contribution is predictive where ours is diagnostic; the C > V assumption is not directly tested.

Edit distance (Levenshtein, 1966) remains a common baseline in cognate detection, and weighted alignment methods such as PMI scoring can improve on it for tree inference (Jäger, 2013). Our transformation-matrix distance does not (76.4% versus 79.4% quartet accuracy on Indo-European); it is a derived comparison, not the focus of this work. On family-specific phonological history, Blevins (2004) documents at least fourteen independent final-consonant-loss events across Austronesian, a mechanistic account for the V > C identity result we report on that family; on the Dravidian side, the high consonant identity rate is consistent with consonants having changed relatively little, a segment-level preservation pattern. These accounts

explain why specific families fall the way they do; they do not measure C versus V regularity or test the assumption directly.

No existing method — class-based, PMI-based, neural, or edit-distance — measures identity and regularity as separate quantities over the same aligned data. The framework in Section 3 is built to do so.

3 Methods

3.1 Data

The data source is the ASJP database, v21 (Wichmann et al., 2025), distributed as a CLDF dataset. We use ten language families: Indo-European, Austronesian, Sino-Tibetan, Pama-Nyungan, Afro-Asiatic, Atlantic-Congo, Uto-Aztecan, Uralic, Dravidian, and Turkic. Family membership and subgroup structure follow the Glottolog classification (Hammarström et al., 2024) as recorded in the distribution's classification fields. The ten were selected as the largest families by ASJP coverage with sufficient Swadesh-list data; the criterion is data availability, not theoretical grouping.

Identity and regularity are computed directly from Needleman-Wunsch alignments of word-forms across shared concepts. No cognate annotation is used. The two measures are read off the same alignment data, which avoids inheriting cognacy judgments from any specific cognate-detection pipeline. The languages meeting the coverage threshold yield over 3.4 million possible within-family pairs across the ten families. Four families (Indo-European, Sino-Tibetan, Afro-Asiatic, Atlantic-Congo) are analyzed on uniform random samples of 10,000 language pairs (`random.seed(42)`); the remaining six are analyzed exhaustively: Austronesian (1,119,007 pairs), Pama-Nyungan (17,004), Uto-Aztecan (3,003), Dravidian (2,415), Turkic (2,145), and Uralic (1,128). Bootstrap 95% confidence intervals on the per-family C minus V regularity advantage clear zero in every family, with the narrowest interval (Uralic) lower-bounded at +15.1 percentage points.

The 40-item ASJP Swadesh list (Holman et al., 2008) and ASJPcode transcription (34 consonant symbols and 7 vowel symbols) are used as-is, with no additional normalization.

3.2 Alignment

For each language pair we apply Needleman-Wunsch global alignment to the word-forms of every shared concept on the Swadesh list, with match = $+1$, mismatch = -1, and gap = -1. Swadesh items are short — typically three to six segments — so global alignment is appropriate. We use uniform-cost alignment rather than a class-aware scheme such as SCA (List, 2012b), because class-aware alignment bakes phonological similarity into the alignment itself and would conflate the identity and regularity signals we measure separately. The alignment is deliberately class-neutral; class sensitivity enters only at measurement.

Gap-containing alignment positions are excluded from identity and regularity computation. Both measures are computed over (segment, segment) pairs only. The output of this stage, pooled across the 40 Swadesh items for a given language pair, is the set of aligned segment pairs used by every subsequent step.

3.3 Phoneme Feature Encoding

Each phoneme is encoded as a 12-dimensional binary feature vector, grounded in the standard feature inventory of PanPhon (Mortensen et al., 2016). Dimensions 0–4 carry vowel features: three height dimensions (high, mid, low) and two frontness dimensions (front, back). Dimensions 5–11 carry consonant features: five place dimensions (labial, coronal, palatal, velar, uvular/glottal) and two manner dimensions (stop/nasal, fricative/approximant).

The active dimensions are {5, 6, 8, 10, 11} on the consonant side and {0, 1, 2, 3, 4} on the vowel side; palatal and uvular/glottal place are not active (Appendix E). This encoding is used only for the transformation matrix and the Frobenius-distance baseline (Section 4.6); the identity and regularity measures do not use it and operate on ASJPcode symbols directly, including all consonants.

3.4 Transformation Matrix

For each language pair we fit a 12×12 transformation matrix T by ridge regression, mapping source-segment feature vectors to aligned target-segment vectors; it is related to PMI-based weighted alignment (Jäger, 2013, 2018) but operates over feature vectors rather than discrete segments. Implementation detail is in Appendix E.

A scalar distance derived from T — the Frobenius distance $\|T - I\|_F$ between the fitted matrix and the identity matrix — is used as a transformation-matrix distance baseline in the quartet-accuracy comparison reported in Section 4.6. The identity and regularity measures de-

fined in Section 3.5 are computed directly from the same aligned segment pairs that T is fit on; the matrix and the two measures share the alignment data as input, but the measures do not read from T itself.

The transformation matrix serves only as the apparatus for the Frobenius-distance baseline; the identity and regularity decomposition is computed directly on the aligned pairs.

3.5 Operationalizing Identity and Regularity

Identity. For each aligned segment pair (s, t), identity equals 1 if $s = t$ and 0 otherwise. The C identity rate is the mean identity over all aligned pairs whose source segment is a consonant; the V identity rate is the mean over aligned pairs whose source segment is a vowel. We compute identity per language pair and then average within family. Per-family rates appear in Appendix F.

Regularity. For each aligned segment pair (s, t) where $s \neq t$ and s, t belong to the same phonological category (both consonants or both vowels), regularity equals 1 if s and t belong to the same phonological class and 0 otherwise. Cross-category changes (s a consonant and t a vowel, or vice versa) are excluded from regularity in both numerator and denominator. The class structure is three consonant classes by manner (stops and nasals; fricatives; approximants) and three vowel classes by height (high; mid; low). We compute regularity per language pair and average within family. The class-to-ASJPcode assignment is given in Appendix D.

Class structure and the artifact concern. Three consonant classes by manner and three vowel classes by height yield parallel class counts. Different partitions of feature space could yield different class granularities, and a coarser consonant partition would mechanically inflate the within-class rate on the consonant side. We test this directly: C $>$ V regularity holds for every balanced class definition we try, and breaks only when the vowel side is collapsed into a degenerate class (Appendix B). The universal C $>$ V regularity result is not an artifact of our choice of three-by-three classes.

Partition of the alignment data. Identity and regularity are computed on the same aligned pairs but partition them differently. A pair where $s = t$ contributes to identity but not to regularity, because no change occurred: /p/ aligned to /p/ is one such case. A pair where $s \neq t$ within the same major category contributes to regularity but not to identity, as in /p/ aligned to /b/, where both are stops. A pair that changes across classes, such as /p/ aligned to /l/, contributes to neither, since the segments differ in manner class. Cross-category pairs (a consonant aligned to a vowel) contribute to neither regularity measure. The two measures are disjoint in what they read from the alignment data, yet defined over the same aligned segment pairs.

The contribution of the framework is that these two numbers are produced separately and can be contrasted. Neither measure is novel in isolation.

3.6 Statistical Tests

For each of the ten families we record whether C $>$ V or V $>$ C on each measure. A unanimous 10/10 outcome under a one-sided sign test has $p = (1/2)^{10} = 0.000977$, since only one of 2^{10} outcomes is unanimous in the predicted direction. The Wilcoxon signed-rank statistic on the same data is $W = 55$, the maximum possible at $n = 10$ (sum of ranks $1 + 2 + \cdots + 10$), with the same one-sided exact $p = 0.000977$. The two tests give the same p-value at the unanimous extreme by construction: there is one extreme outcome and 2^{10} equally likely arrangements of signs. We report both because either test alone is open to method-specific objections at small n.

Bootstrap 95% confidence intervals on the mean within-class rate advantage are computed by 10,000 resamples at the language-pair level within each family, then aggregated. The family-level Pearson correlation between identity advantage $(C - V)$ and regularity advantage $(C - V)$ is computed with $n = 10$. At this sample size the correlation alone has limited power; the principal evidence for divergent patterning between the two measures is the contrast between a universal pattern in regularity and a non-universal one in identity, not the correlation coefficient.

Results are stable across five Needleman-Wunsch alignment parameter settings: consonants show higher regularity than vowels under every setting tested. The ridge regularization parameter λ affects the fitted T and hence the Frobenius-distance baseline of Section 4.6; the identity and regularity rates are computed directly from the alignments and are independent of λ. A full λ sweep on the Frobenius distance is not reported here and remains for future work.

4 Results

4.1 Regularity Universally Favors Consonants

Across all ten language families, the within-class regularity rate is higher for consonants than for vowels. The direction is unanimous. Table 1 reports per-family rates, the C minus V advantage, and bootstrap 95% confidence intervals on the advantage.

The largest C minus V advantage is in Turkic, with a C/V ratio of $2.64\times$; the smallest is in Uralic at $1.71\times$. The cross-family averages are 41.0% C regularity and 18.9% V regularity, a ratio of $2.17\times$. Effect sizes are large in every family. The narrowest bootstrap interval is Uralic, lower-bounded at $+15.1$ percentage points, which still places the family advantage well above zero.

The unanimous 10/10 outcome is significant under both a one-sided sign test and a Wilcoxon signed-rank test ($p = 0.000977$ for each; Section 3.6). The mean C minus V advantage across the ten families is $+22.1$ percentage points, with a bootstrap 95% confidence interval of $[+19.0, +25.4]$.

Within-family subgroups give a further consistency check. Across the 29 subgroups defined by Glottolog within eight of the ten families, all 29 show C $>$ V regularity. These subgroups are not independent observations, since subgroups within a family share genealogical history, but the consistency across subgroups indicates the pattern is not driven by outlier families. The full subgroup breakdown is in Appendix A.

One family-level dissociation is worth flagging before Section 4.3 examines the cross-measure pattern in full. Austronesian shows the widest V $>$ C identity gap of any family in our sample (57.7% V identity, 37.7% C identity, a gap of 20 percentage points), yet Austronesian's regularity rates still place consonants well above vowels with a C/V ratio of $2.18\times$. The two measures separate cleanly even within a single family.

4.2 Identity Shows No Universal Direction

The pattern for identity is different. Of the ten families, six show V $>$ C identity, one shows a tie (Indo-European, with a gap of 0.2 percentage points), and three show C $>$ V identity (full per-family rates in Appendix F).

The three families on the C $>$ V side of identity are Turkic ($+10.7$pp), Dravidian ($+3.2$pp), and Uralic ($+5.5$pp). The six on the V $>$ C side are

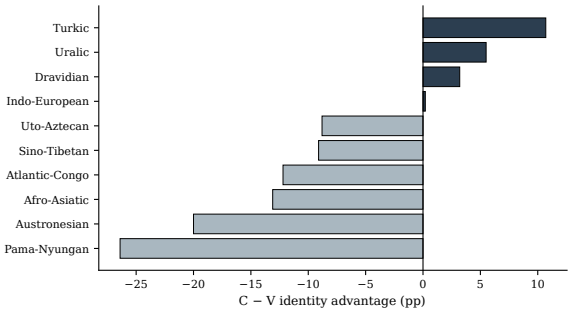

Figure 1: C $-$ V identity advantage by family, sorted. Dark bars mark the three families with C $>$ V identity (Turkic, Uralic, Dravidian); Indo-European is within rounding of a tie; light bars mark the six families with V $>$ C identity.

Sino-Tibetan (-9.1pp), Afro-Asiatic (-13.1pp), Atlantic-Congo (-12.2pp), Uto-Aztecan (-8.8pp), Austronesian (-20.0pp), and Pama-Nyungan (-26.4pp). The widest gap in either direction is Pama-Nyungan at -26.4pp; the widest C $>$ V gap is Turkic at $+10.7$pp.

The Indo-European tie is notable: IE has the most extensively documented comparative phonology and is the family most likely to anchor prior intuitions about C/V stability, yet shows a tie rather than the expected consonant advantage.

Treating Indo-European's $+0.2$pp gap as effectively zero, the sign test on identity runs across the nine families with a clear direction: six V $>$ C, three C $>$ V. We use a two-sided binomial test because we have no directional prior on identity, unlike regularity. The exact two-sided p-value is approximately 0.51, which does not reject the null of no consistent direction, though with $n = 9$ the test has limited power to detect a direction if one exists. Subgroup analysis is applied to the regularity result, where universality is the claim under test, and not to identity, where no universal direction is claimed.

We return to the convergence among these three families in Section 4.4.

4.3 The Two Measures Pattern Differently

The two findings point in different directions on the same alignment data: regularity favors consonants in 10 of 10 families, identity in only 3 of 10. The family-level tally is the principal evidence that the measures pattern differently. The Pearson correlation between the per-family advantages is near zero ($r = -0.015$, $n = 10$), but at this sample size the correlation is underpowered (95% CI approx-

Family	C reg.	V reg.	C − V	C/V ratio	95% CI
Turkic	51.7%	19.6%	+32.1	2.64×	[+30.9, +33.4]
Pama-Nyungan	47.0%	18.2%	+28.8	2.58×	[+27.9, +30.1]
Dravidian	43.0%	17.0%	+26.0	2.53×	[+24.9, +27.5]
Uto-Aztecan	39.6%	16.9%	+22.7	2.34×	[+20.8, +23.5]
Austronesian	40.1%	18.4%	+21.7	2.18×	[+20.8, +23.3]
Atlantic-Congo	36.9%	18.2%	+18.7	2.03×	[+17.4, +20.0]
Afro-Asiatic	35.2%	16.7%	+18.5	2.11×	[+17.1, +19.4]
Sino-Tibetan	42.4%	22.0%	+20.4	1.93×	[+19.1, +22.1]
Indo-European	36.4%	20.3%	+16.1	1.79×	[+13.8, +16.5]
Uralic	37.3%	21.8%	+15.5	1.71×	[+15.1, +17.3]
Mean	41.0%	18.9%	+22.1	2.17×	[+19.0, +25.4]

Table 1: Regularity by family. C reg. and V reg. are the proportion of within-category changes that stay within phonological class, computed separately for consonants and vowels. The C − V column is the consonant advantage in percentage points. The bootstrap 95% CI is on the C − V advantage, computed from 10,000 resamples of the analyzed pairs within each family. Every family shows C > V regularity, and every interval lies above zero.

imately $[-0.63, 0.62]$); the tally contrast, not the correlation, carries the claim.

4.4 The Nostratic Convergence

The three families showing C > V identity (Turkic, Dravidian, Uralic) are all affiliated with the Nostratic macrofamily as proposed by Dolgopolsky (1964); none of the five non-affiliated families in our sample shows C > V identity, though affiliation alone does not predict it (Indo-European is a tie, Afro-Asiatic shows V > C). We use Nostratic affiliation as a pre-defined grouping criterion enumerated independently of our data, not as an endorsement of the macrofamily's genealogical validity, which remains contested. We treat the convergence as a post-hoc observation: the sample is small, a vowel-harmony confound is a natural alternative, and a confirmatory test would require a pre-registered analysis on a larger, typologically balanced sample. Appendix G gives the full enumeration, the harmony-confound analysis, and the conditions a confirmatory test would require.

4.5 Robustness

The universal C > V regularity result is stable under the methodological choices we vary: five Needleman-Wunsch alignment settings on three sampled families (Appendix C), five alternative class definitions, of which the three balanced ones preserve C > V in all ten families while two with degenerate vowel classes do not (Appendix B), and all 29 Glottolog subgroups across eight families (Appendix A). Identity, a direct match count, does not depend on class structure and is not separately tested.

4.6 Method Comparison

The cognate-detection F1 evaluation, unlike the identity and regularity analyses, uses gold-standard cognate annotations from five Lexibank datasets: `mannburmish` (Sino-Tibetan), `sidwellbahnaric` (Austro-Asiatic), `hsiuhmongmien` (Hmong-Mien), `dravlex` (Dravidian), and `savelyevturkic` (Turkic).

Quartet accuracy. The Frobenius-distance transformation-matrix method underperforms Levenshtein on quartet accuracy in all four families tested, by 1.2 to 3.7 percentage points (for example, 76.4% versus 79.4% on Indo-European). Weighted alignment methods can outperform plain edit distance on tree inference (Jäger, 2013); ours does not, and we make no claim that it improves on edit distance. The contribution is the decomposition reported in Sections 4.1 through 4.3, not the distance.

Cognate detection F1. Across the five Lexibank datasets, DOLGO achieves a mean F1 of 0.846 and SCA achieves 0.847, a difference of less than 0.1%. TURCHIN, which removes vowels entirely, scores 7.4 percentage points lower at 0.772. DOLGO and SCA achieve near-identical F1 despite DOLGO's greater simplicity. We return to the interpretation in Section 5.

5 Discussion

5.1 Why Regularity Universally Favors Consonants

Dolgopolsky's own grouping criterion describes regularity rather than identity: he grouped sounds so that correspondences inside a type are more regular than those between types (as cited in List, 2012b). The universal C > V regularity we find confirms that this property holds in every family, and the DOLGO ≈ SCA result (Section 4.6) is its operational signature: finer scoring on top of the coarse classes adds nothing because the classes already capture the regularity. The mechanism behind the universality itself is an open question. Two informal accounts in the typological literature are consistent with the data without being tested by it.

The articulatory account holds that consonants are more discretely articulated than vowels. Consonant sound changes tend to be discrete substitutions that preserve manner or place (/p/ > /b/, /t/ > /d/) more often than they cross articulatory dimensions. Vowels drift along continuous formant axes, crossing height and frontness boundaries more readily. Discrete targets produce higher within-class rates.

The lexical-anchor account holds that consonants serve as perceptual anchors for lexical identity. Changes that preserve manner class preserve the distinctiveness of lexical items even when phonetic detail shifts. Vowels carry suprasegmental functions and are under different functional pressure, making cross-class drift less disruptive.

The universality across our ten families is consistent with both accounts but constitutes neither. A study comparing within-class change rates against articulatory-distance metrics directly would be the appropriate test.

5.2 Why Identity Varies Family by Family

Identity is a per-segment match rate, sensitive to the specific historical sound changes a family has undergone. Where regularity reflects a general property of how sounds change when they change, identity reflects which sounds have changed and how often in a given family. The first is type-level; the second is token-level and history-dependent.

Two cases illustrate. Austronesian shows V > C identity by 20.0 percentage points. Blevins (2004) documents at least fourteen independent final-consonant-loss events across the family. Systematic loss of final consonants in one branch but not another mechanically reduces consonant identity between cognates: when a consonant is deleted, it cannot match. Austronesian's V > C identity is a predictable consequence of this well-documented pattern. The family still shows C > V regularity, because consonant changes that do occur stay within class.

Dravidian shows C > V identity by 3.2 percentage points. The high consonant identity rate is consistent with consonants having changed relatively little in the family. The +3.2pp gap is consistent with a segment-level preservation pattern.

The Nostratic concentration and the harmony observation in Section 4.4 sit alongside these family-specific histories as a separate post-hoc pattern. Identity variation is what we should expect given that family-specific phonological histories differ.

5.3 The Brown Correspondence Asymmetry

Brown et al. (2013) report 692 recurrent sound correspondences in ASJP, 582 consonant-only against 64 vowel-only. We attribute this 9:1 asymmetry to regularity rather than identity, on two arguments.

The mechanistic argument: Brown et al. count recurrent correspondences, those that appear repeatedly across independent language pairs. Recurrence is what regularity produces. When a consonant changes predictably within a manner class, the same correspondence type (e.g., /p/ ∼ /b/) appears across many independent language pairs and gets counted as recurrent. When a vowel changes less predictably, changes scatter across many idiosyncratic correspondences rather than concentrating into recurrent ones. Recurrence is a threshold phenomenon: predictable changes concentrate into a small number of high-count correspondences that clear the recurrence bar, while unpredictable changes scatter across many low-count correspondences that do not. Small differences in predictability can therefore produce large differences in recurrent-correspondence counts. The 2.17× regularity gap in our sample is the kind of mechanism that would generate a recurrence asymmetry of the shape Brown et al. report.

The eliminative argument: if the asymmetry were driven by identity, C > V identity should be broad, but it holds in only three of our ten families. A measure favoring consonants in three of ten cannot account for a 9:1 ratio; regularity, which favors consonants in all ten, can.

We have not formally derived the 9:1 ratio from our 2.17× regularity gap; these are different measurements on different data. With that scope, the

asymmetry is consistent with the regularity finding.

5.4 Implications for Computational Methods

Three implications follow from the decomposition, framed as inferences and not as tested claims.

First, regularity is the defensible basis for consonant weighting. Methods that weight consonant correspondences more heavily have empirical support, and that support comes from regularity. Future methods could weight by within-class correspondence rate rather than by category membership alone.

Second, vowel discarding is not universally justified. Identity varies by family, so pipelines that discard vowels may lose family-specific signal, particularly in families with low consonant identity (Austronesian, Pama-Nyungan, Afro-Asiatic). The TURCHIN F1 penalty (7.4pp below DOLGO and SCA) is consistent with this.

Third, the framework is a diagnostic tool: for a new family, the decomposition asks whether consonant stability there is an identity effect, a regularity effect, or both.

We do not test whether explicit regularity-weighting improves downstream phylogenetic or cognate-detection performance. A pre-registered test on non-Nostratic families would be the appropriate confirmation.

Across these threads, the result is a refinement rather than a refutation. Consonant primacy in computational historical linguistics is grounded, but in regularity, not in identity. The distinction matters because the two components have different cross-linguistic distributions and carry different implications for method design.

6 Conclusion

Consonant stability, treated in computational historical linguistics as a single property, decomposes into two. Segment identity asks how often an aligned segment matches its counterpart. Regularity asks how often a change that does occur stays within a phonological class. The two are operationally distinct, and the assumption that consonants are diachronically more stable than vowels holds for one of them and not the other. Across our ten families, regularity universally favors consonants (Section 4.1); identity does not (Section 4.2). The assumption was not wrong, but it was underspecified. The contribution is to separate the two components and measure each directly.

The class-based methods that descend from Dolgopolsky (1986) succeed because they capture consonant regularity, which is what their class structures encode. The DOLGO $\approx$ SCA result in Section 4.6 is the operational signature of this: finer-grained scoring on top of the coarse classes does not improve cognate-detection F1, because the relevant signal is already present. The most frequently cited empirical anchor for cross-linguistic C $>$ V claims, Brown et al.'s (2013) roughly 9:1 recurrent-correspondence ratio, is best attributed to regularity rather than identity. Methods that treat consonant identity as universally diagnostic make a family-dependent assumption the data do not universally support.

Three questions remain open. First, why regularity universally favors consonants: the articulatory and lexical-anchor accounts are consistent with the data, and a direct comparison of within-class change rates against articulatory-distance metrics would be the appropriate test. Second, whether the concentration of C $>$ V identity in Nostratic-affiliated families (Section 4.4) generalizes: a pre-registered analysis on a larger and typologically balanced sample, with affiliation as a predictor variable, would be the appropriate confirmation. Third, whether explicit regularity-weighting improves cognate-detection or phylogenetic performance: the present results show that the existing class structures already capture regularity implicitly, and a test on non-Nostratic families would tell us whether the implicit capture is optimal.

The value of separating the two components is that we now know what makes the field's methods work: regularity.

Limitations

The 12-dimensional feature encoding excludes palatal place (dimension 7) and uvular/glottal place (dimension 9), so palatal and uvular/glottal consonants — including the common segment /h/ — do not contribute to the transformation-matrix Frobenius-distance baseline (Section 3). This is a scope condition on that baseline only; the identity and regularity measures operate on ASJPcode symbols directly and include all consonants. A second scope condition follows from the data source: the ASJPcode transcription uses 34 consonant symbols and 7 vowel symbols, an inventory asymmetry that gives consonants more discriminative resolution by transcription design, and the regularity measure op-

erates at this granularity. Finer-grained vowel transcription, under which vowels would carry as many distinctions as consonants, is not tested. Nor are results on lexicons larger than the 40-item Swadesh list.

Ten language families is a small sample, selected for ASJP coverage rather than typological balance; family-level statistics are correspondingly underpowered, and the divergent-patterning finding rests on the universality contrast rather than the correlation (Section 4.3). The Nostratic convergence reported in Section 4.4 is a post-hoc observation: a pre-registered analysis on a larger and typologically balanced sample, with Nostratic affiliation as a predictor variable, would be needed for a causal claim. The mechanism for universal C > V regularity is also untested here; the articulatory and lexical-anchor accounts in Section 5 are consistent with the data without being established by it. Several robustness and downstream questions remain unswept: the ridge parameter λ for the Frobenius baseline of Section 4.6, identity-rate stability under alignment perturbations, and whether explicit regularity-weighting improves downstream cognate-detection or phylogenetic performance. The Section 4.6 F1 evaluation is scoped to five Lexibank datasets and generalizes only insofar as comparable gold-standard cognate annotations are available for other families.

Ethical Considerations

This work uses the publicly available ASJP database, which comprises short standardized wordlists compiled from published linguistic sources. The analysis involves no human subjects, no personal or sensitive data, and no new data collection or annotation. The contribution is diagnostic rather than applied, and we foresee no risk of harm from the analysis or its dissemination.

The ten families analyzed include both well-documented language families and families containing endangered or under-resourced languages. We make no claim that ASJP's coverage of any specific language community is exhaustive; our analysis operates over the standardized wordlists as distributed.

References

V. S. D. S. Mahesh Akavarapu and Arnab Bhattacharya. 2023. Cognate transformer for automated phonological reconstruction and cognate reflex prediction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6852–6862, Singapore. Association for Computational Linguistics.

V. S. D. S. Mahesh Akavarapu and Arnab Bhattacharya. 2024. Automated cognate detection as a supervised link prediction task with cognate transformer. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 965–975, St. Julian's, Malta. Association for Computational Linguistics.

Juliette Blevins. 2004. The mystery of Austronesian final consonant loss. *Oceanic Linguistics*, 43(1):208–213.

Cecil H. Brown, Eric W. Holman, and Søren Wichmann. 2013. Sound correspondences in the world's languages. *Language*, 89(1):4–29.

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

Roderic F. Casali. 2008. ATR harmony in African languages. *Language and Linguistics Compass*, 2(3):496–549.

Bernard Comrie. 1981. *The Languages of the Soviet Union*. Cambridge University Press, Cambridge.

Aharon B. Dolgopolsky. 1964. Gipoteza drevnejšego rodstva jazykovyx semej Severnoj Evrazii s verojatnostnoj točki zrenija. *Voprosy Jazykoznanija*, (2):53–63. English translation: "A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia," in Shevoroshkin and Markey (eds.), *Typology, Relationship and Time*, Karoma, 1986, pp. 27–50.

Aharon B. Dolgopolsky. 1986. A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia. In Vitalij V. Shevoroshkin and Thomas L. Markey, editors, *Typology, Relationship and Time: A Collection of Papers on Language Change and Relationship by Soviet Linguists*, pages 27–50. Karoma, Ann Arbor.

Harald Hammarström, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2024. Glottolog 5.0. https://glottolog.org.

Eric W. Holman, Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, and Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica*, 42(3–4):331–354.

Daniel J. Hruschka, Simon Branford, Eric D. Smith, Jon Wilkins, Andrew Meade, Mark Pagel, and Tanmoy Bhattacharya. 2015. Detecting regular sound changes in linguistics as events of concerted evolution. *Current Biology*, 25(1):1–9.

Gerhard Jäger. 2013. Phylogenetic inference from word lists using weighted alignment with empirically determined weights. *Language Dynamics and Change*, 3(2):245–291.

Gerhard Jäger. 2018. Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Data*, 5:180189.

Mark Kaiser and Vitalij V. Shevoroshkin. 1988. Nostratic. *Annual Review of Anthropology*, 17:309–329.

Alexei S. Kassian, George Starostin, Ilya M. Egorov, Ekaterina S. Logunova, and Anna V. Dybo. 2021. Permutation test applied to lexical reconstructions partially supports the Altaic linguistic macrofamily. *Evolutionary Human Sciences*, 3:e32.

Young Min Kim, Kalvin Chang, Chenxuan Cui, and David R. Mortensen. 2023. Transformed protoform reconstruction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 24–38, Toronto, Canada. Association for Computational Linguistics.

Grzegorz Kondrak. 2000. A new algorithm for the alignment of phonetic sequences. In *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, pages 288–295, Seattle, WA. Association for Computational Linguistics.

Grzegorz Kondrak. 2003. Phonetic alignment and similarity. *Computers and the Humanities*, 37(3):273–291.

Bhadriraju Krishnamurti. 2003. *The Dravidian Languages*. Cambridge University Press, Cambridge.

Vladimir I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710.

Johann-Mattis List. 2012a. LexStat: Automatic detection of cognates in multilingual wordlists. In *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, pages 117–125, Avignon, France. Association for Computational Linguistics.

Johann-Mattis List. 2012b. SCA: Phonetic alignment based on sound classes. In *New Directions in Logic, Language, and Computation*, volume 7415 of *Lecture Notes in Computer Science*, pages 32–51, Berlin, Heidelberg. Springer.

Johann-Mattis List. 2019. Automatic inference of sound correspondence patterns across multiple languages. *Computational Linguistics*, 45(1):137–161.

Liang Lu, Jingzhi Wang, and David R. Mortensen. 2024a. Improved neural protoform reconstruction via reflex prediction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*.

Liang Lu, Peirong Xie, and David R. Mortensen. 2024b. Semisupervised neural proto-language reconstruction. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.

Carlo Meloni, Shauli Ravfogel, and Yoav Goldberg. 2021. Ab Antiquo: Neural proto-language reconstruction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4460–4473. Association for Computational Linguistics.

David R. Mortensen, Patrick Littell, Akash Bharadwaj, Kartik Goyal, Chris Dyer, and Lori Levin. 2016. PanPhon: A resource for mapping IPA segments to articulatory feature vectors. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3475–3484, Osaka, Japan.

Peter Turchin, Ilia Peiros, and Murray Gell-Mann. 2010. Analyzing genetic connections between languages by matching consonant classes. *Journal of Language Relationship*, 3:117–126.

Søren Wichmann, Eric W. Holman, Cecil H. Brown, Matthew S. Dryer, and Qibin Ran. 2025. The ASJP database (version 21). `https://asjp.clld.org`. CLLD distribution.

A Subgroup-Level Regularity

Table 2 reports the regularity rates for each Glottolog subgroup within the eight families for which subgroup-level data were computed. Each subgroup is analyzed on the language pairs internal to it. Consonant regularity exceeds vowel regularity in all 29 subgroups. The subgroups within a family share genealogical history and are not independent observations; the consistency indicates the family-level result is not driven by a single subgroup.

B Class Definition Robustness

We test the regularity computation under five alternative phonological class definitions to check that the result does not depend on the specific three-by-three partition used in the main analysis. Table 3 reports each definition and the number of families showing C > V regularity under it. The three definitions that assign balanced class structures to consonants and vowels preserve C > V regularity in all ten families. The two definitions that collapse vowels into a coarse or degenerate class show the pattern weaken or reverse, which indicates the result depends on a fair class structure rather than on finer-grained consonant classes.

Family	Subgroup	N	C reg.	V reg.
Sino-Tibetan	Sinitic	502	54.8%	25.7%
	Burmo-Qiangic	102	37.2%	20.8%
	Bodic	92	40.6%	17.2%
	Kuki-Chin-Naga	86	45.0%	16.3%
	Himalayish	39	41.9%	16.5%
	Brahmaputran	39	45.6%	18.1%
	Macro-Bai	24	47.3%	17.5%
	Karenic	16	34.7%	22.8%
	Macro-Tani	14	51.6%	18.8%
	Kho-Bwa	9	45.2%	14.3%
	Nungish	7	41.4%	14.2%
Afro-Asiatic	Chadic	240	36.9%	17.0%
	Semitic	116	36.1%	22.0%
	Cushitic	51	38.8%	16.8%
	Berber	33	35.2%	16.3%
Austronesian	Malayo-Polynesian	1447	39.5%	18.4%
	Austronesian (other)	21	33.5%	29.8%
	Atayalic	11	33.2%	26.5%
	East Formosan	10	42.7%	15.4%
Atlantic-Congo	Volta-Congo	1654	37.3%	18.2%
	North-Central Atlantic	80	37.0%	19.3%
	Mel	16	45.2%	22.3%
Dravidian	South Dravidian	58	42.3%	17.3%
	North Dravidian	6	34.3%	8.0%
	Central Dravidian	6	45.7%	19.8%
Uto-Aztecan	Southern Uto-Aztecan	60	41.1%	14.9%
	Northern Uto-Aztecan	18	42.0%	14.9%
Turkic	Common Turkic	65	52.3%	21.3%
Indo-European	Classical Indo-European	594	37.0%	21.2%

Table 2: Subgroup-level regularity within eight families. N is the number of languages in the subgroup. C reg. and V reg. are the within-class regularity rates. Consonant regularity exceeds vowel regularity in all 29 subgroups.

Class definition	Balance	C > V
Original (3 C, 3 V)	balanced	10/10
Place-based (4 C, 3 V)	balanced	10/10
Balanced binary (2 C, 2 V)	balanced	10/10
Binary both (2 C, 2 V)	uneven	9/10
Manner + binary V (3 C, 2 V)	skewed	1/10

Table 3: Regularity under five class definitions. The three balanced definitions preserve C > V regularity in all ten families. The two definitions with uneven or degenerate vowel classes do not.

Setting	C reg.	V reg.
SCA default (gap $= -2$)	57.1%	45.4%
SCA low gap (gap $= -1$)	61.7%	46.9%
SCA high gap (gap $= -4$)	51.6%	41.3%
DOLGO model	57.9%	46.9%
ASJP model	53.4%	42.9%

Table 4: Regularity under five Needleman-Wunsch settings (Turkic). Consonant regularity exceeds vowel regularity in every setting.

C Needleman-Wunsch Parameter Sensitivity

Table 4 reports consonant and vowel regularity rates under five Needleman-Wunsch alignment parameter settings, computed on Turkic. Consonant regularity exceeds vowel regularity under every setting. The same direction holds for the other two families tested (Austronesian, Indo-European).

D Class-to-ASJPcode Assignment

Table 5 gives the assignment of ASJPcode symbols to the three consonant classes and three vowel classes used in the regularity computation. All consonant and vowel symbols are assigned to a class; no symbols are excluded from the identity or regularity measures.

E Transformation Matrix Detail

The transformation matrix T is fit by ridge regression with $\lambda = 1.0$, taking the source-segment feature vector as input and the aligned target-segment feature vector as output. T captures the systematic feature-level relationship between source and target inventories over the aligned data. Ridge regularization controls instability when feature dimensions are correlated and prevents overfitting on language pairs with few aligned tokens.

The active consonant dimensions were deter-

Class	ASJPcode symbols
Stops and nasals	p, b, t, d, k, g, q, m, n, N
Fricatives	f, v, s, z, S, Z, c, x, X, G, h, T
Approximants	r, l, w, y, j, 5, 7
High vowels	i, u
Mid vowels	e, o, 3
Low vowels	a, E

Table 5: ASJPcode symbols assigned to each phonological class in the regularity computation.

mined empirically by running the encoding function on a reference set of common consonants (`p`, `t`, `k`, `s`, `m`, `n`), none of which activate dimension 7 (palatal place; ASJPcode `c`, `j`, `5`, `N`) or dimension 9 (uvular/glottal place; ASJPcode `q`, `G`, `h`, `7`). Palatal and uvular/glottal consonants are therefore not represented in the Frobenius-distance baseline, though they are included in the identity and regularity measures, which do not use this encoding.

F Per-Family Identity Rates

Table 6 reports per-family vowel and consonant identity rates and their difference. Three families show C > V identity, one is within rounding of a tie (Indo-European), and six show V > C.

Family	V id.	C id.	C − V
Turkic	51.6%	62.3%	+10.7
Uralic	43.7%	49.2%	+5.5
Dravidian	55.9%	59.1%	+3.2
Indo-European	43.6%	43.8%	+0.2
Uto-Aztecan	63.5%	54.7%	−8.8
Sino-Tibetan	52.1%	43.0%	−9.1
Atlantic-Congo	41.4%	29.2%	−12.2
Afro-Asiatic	44.4%	31.3%	−13.1
Austronesian	57.7%	37.7%	−20.0
Pama-Nyungan	58.2%	31.8%	−26.4

Table 6: Identity by family, sorted by C − V difference. V id. and C id. are the proportion of aligned segments where source equals target, computed separately for vowel-sourced and consonant-sourced pairs.

G The Nostratic Convergence

We did not predict this pattern in advance. We use Nostratic affiliation as a pre-defined grouping criterion — a specific list of families enumerated independently of our data, not an endorsement of the macrofamily's genealogical validity, which remains contested in mainstream historical linguistics.

Of the three families showing C > V identity in our sample (Turkic, Dravidian, Uralic), all three belong to the Nostratic macrofamily as proposed by Dolgopolsky (1964) and standardized in subsequent Nostraticist work (Kaiser and Shevoroshkin, 1988). The Nostratic proposal covers six families: Indo-European, Afro-Asiatic, Kartvelian, Uralic, Dravidian, and the Turkic, Mongolic, and Tungusic group (Altaic).

Five of our ten families overlap with the Nostratic proposal (Indo-European, Afro-Asiatic, Uralic, Dravidian, and Turkic); five do not (Austronesian, Sino-Tibetan, Pama-Nyungan, Uto-Aztecan, Atlantic-Congo). All three C > V identity families are Nostratic-affiliated, and the remaining two Nostratic-affiliated families (Indo-European and Afro-Asiatic) show a tie and V > C respectively. Every C > V family in our sample is Nostratic-affiliated, while affiliation alone does not predict C > V.

A natural alternative explanation invokes vowel harmony as a confound. The background phonological literature documents Turkic and Uralic as families with widespread vowel harmony (Comrie, 1981); Dravidian as having weaker patterns, with a reconstructed Proto-Dravidian alternation (**i, *u* → **e, *o* before low **a*) and synchronic harmony in a few languages including Telugu (Krishnamurti, 2003); Afro-Asiatic as lacking Turkic/Uralic-style vowel harmony; and a substantial subset of Atlantic-Congo as exhibiting ATR vowel harmony (Casali, 2008). Languages with vowel harmony can show depressed vowel identity rates because aligned vowels in cognate pairs are more likely to differ by harmonic context. If harmony were the mechanism, the C > V identity pattern should track these typological categorizations.

It does not. Afro-Asiatic lacks Turkic/Uralic-style harmony yet shows the widest V > C identity gap among Nostratic-affiliated families in our sample. Atlantic-Congo exhibits ATR harmony in a substantial subset of its languages yet also shows V > C identity. Across the ten families, every family showing C > V identity is Nostratic-affiliated (3 of 3; no non-overlap family shows C > V), whereas the harmony categorization does not track the pattern (Dravidian's harmony is weak yet shows C > V; Atlantic-Congo has harmony yet shows V > C). With $n = 10$ we cannot make a strong causal claim, but the two variables are not equally well-supported

by the data.

If the C > V identity pattern is concentrated in the families proposed for Nostratic affiliation, the generalization from this set to a universal claim about consonant identity may reflect properties of that specific grouping rather than a cross-linguistic law. The observation does not establish that the Nostratic hypothesis is correct, and it does not establish that the consonant-stability assumption is wrong. It establishes that the identity component of that assumption may be sample-dependent in ways the regularity component is not. Testing this properly would require a pre-registered analysis on a larger and more typologically balanced sample, with both Nostratic affiliation and presence of vowel harmony as predictor variables.