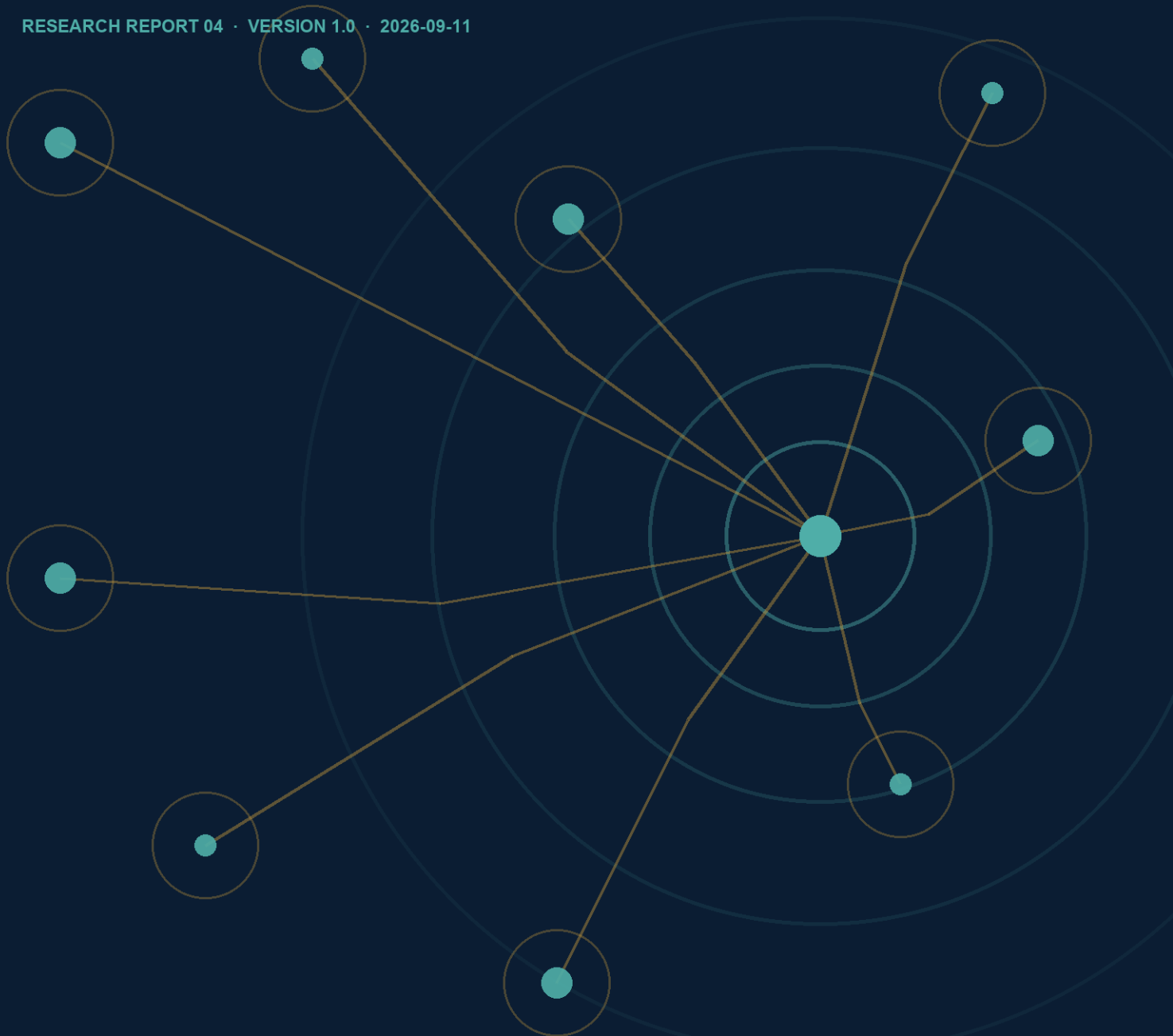




WALS as a Space of Typological Evidence

Language-specific structural evidence, scope boundaries, and a dedicated crosswalk layer

WALS as a Space of Typological Evidence

Language-specific structural evidence, scope boundaries, and a dedicated crosswalk layer

Michael Reich · Atlas of Language · 11 September 2026

Abstract

The WALS research track examines how the World Atlas of Language Structures Online can supplement the Atlas of Language with verifiable cross-linguistic structural evidence. WALS is treated neither as a dictionary nor as an ontology of Atlas concepts. Feature, value, language, Glottocode, and Atlas concept ID remain distinct. The canonical evidence layer contains 36 accepted evidence-only records for 18 Atlas concepts: 30 WALS-related records and six supplementary scholarly sources. A separate typological external layer adds nine feature-value-language crosswalks for nine concepts, nine features, and eight languages. This report explains the 30-case pilot, the two-layer evidence rule, the special case of Japanese person-oriented deixis, four Phase 2 boundary cases, and the nine typed crosswalks. It shows why a single language value supports only a bounded claim and why areal or genealogical patterns cannot justify broad claims of linguistic relativity.

Metric	Value
Classified pilot findings	30
Evidence-only Atlas concepts	18
Accepted evidence-only records	36
Typological crosswalks	9
Features / languages / families	9 / 8 / 8
Accepted Inbox packages	4

REPORT

PROFILE

Evidence only: 36 records for 18 existing Atlas profiles. Typological layer: 9 feature-value-language crosswalks. Effect of this report: documentation only, with no universality or identity claim, import, or publication.

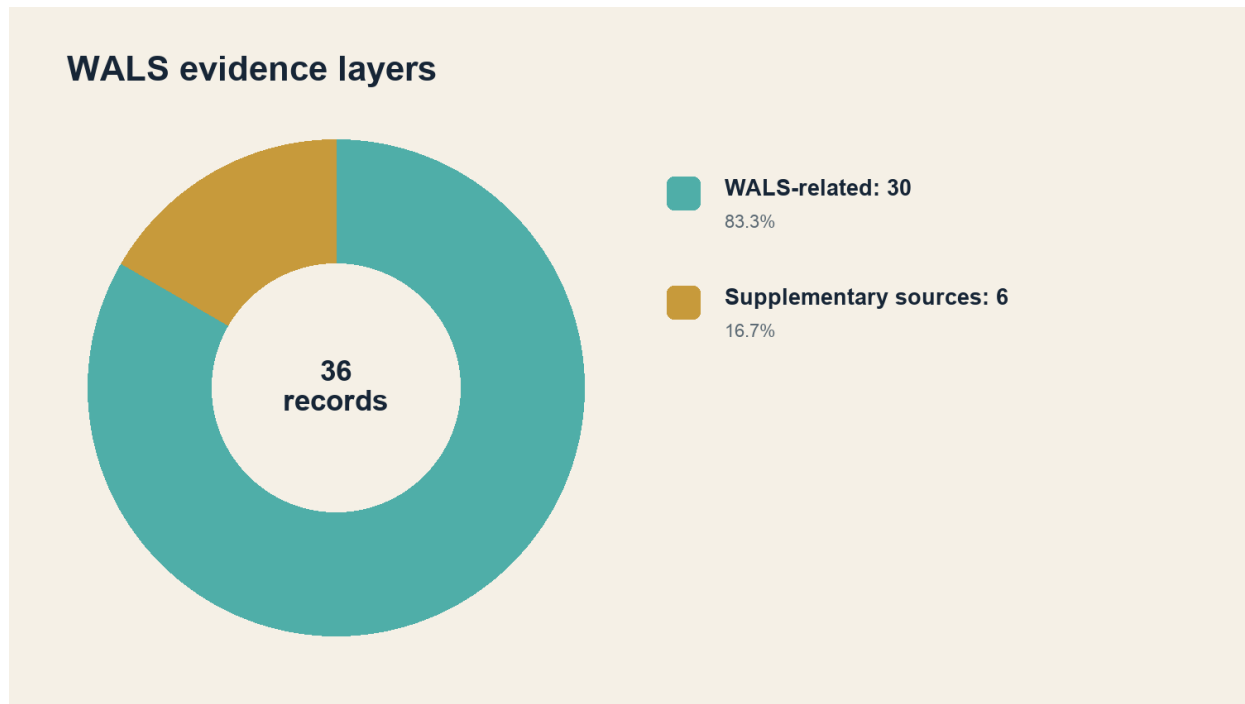


Fig. 1 Research findings and canonical scope remain distinct.

1 Research question and object

WALS maps structural properties of languages: sound systems, morphological categories, syntactic orders, lexical types, and other comparable features. The Atlas of Language contains many of these objects but connects them with rhetorical, pragmatic, semiotic, and epistemic perspectives. The guiding question is therefore: which WALS data can support a concrete claim in an Atlas profile, and what is the scope of that evidence?

A WALS mapping is not a lexical identity claim. It connects an Atlas concept with a specific feature, a specific value, and a specific language. Typological crosswalks additionally record the language ID, language name, Glottocode, ISO 639-3 code, and language family. These fields prevent an observed language case from being generalized into a property of all languages or a definition of the Atlas concept.

2 WALS and the CLDF dataset

WALS Online describes itself as a large database of structural phonological, grammatical, and lexical properties assembled from descriptive sources by 55 specialist authors. Its introduction describes 144 chapters and 160 maps. A total of 2,662 languages appear on at least one map, but coverage density varies substantially across languages and features. This uneven matrix is part of the resource and limits any interpretation of frequencies.

The research track uses CLDF release v2020.5 at commit f97 440d6edbaed0097d1bd307a5a20ea3f09e271. Values, codes, languages, examples, and chapter texts were read as separate tables. The official WALS introduction explains that feature values are intended to form an exhaustive classification for the feature in question. This does not mean that every language is coded for every feature, nor that value categories carry the same meaning outside their chapter definitions.

3 Canonical basis and reproducibility

This report reads GitHub main at 8e676f019f2031c58f51e42a25f899c4bdf577ef and canonical export Atlas 249 / Sites 212 / data model 1.5.0. The export is bound to checkpoint atlas-249-wals-phase2-checkpoint; the manifest records 85ba65 244 988b1aa6d89e0af960 109edc3192b7a as its export commit. The typological JSONL layer contains nine records. The canonical evidence inventory contains 1,718 rows, including the accepted WALS and supplementary evidence from this track.

Each evidence-only operation was bound to the SHA-256 of the existing concepts.jsonl row and added source references only. It changed neither concept ID, lemma, definition, examples, language forms, quality status, nor internal relations. Typological crosswalks reside in a dedicated table and export. This report writes neither Inbox nor canonical files.

THREE-WAY**SEPARATION**

The empirical WALS finding, the Atlas interpretation, and any derived path of thought are maintained as separate levels of claim.

4 Pilot and review logic

The Gate F3 pilot classified 30 research findings exactly once. Thirteen direct cases were developed as evidence-only proposals. Eleven cases waited for a suitable relation model. Candidate 013 was reviewed separately. Distribution-only findings, rejections, and negative controls remained outside an import package. This separation prevented a technically schema-compatible proposal from being treated automatically as editorially accepted.

Each of the thirteen direct cases received two evidence records. The first binds the exact WALS language value in values.csv to a feature ID, code ID, and language ID. The second binds the concrete form or construction to examples.csv, the chapter text, or a supplementary scholarly source. The language value establishes the classification, while the form evidence shows how it is realized. Both levels remain separately citable.

Candidate 013 concerns person orientation in Japanese demonstratives. Code 41A-3, Three-way contrast, is not sufficient by itself. Only the language-specific comment for 41A-jpn distinguishes proximity to the speaker, proximity to the hearer, and distance from both. The case was therefore accepted separately and only for the Japanese portion of the Atlas profile; Tongan and Hausa were not validated by that decision.

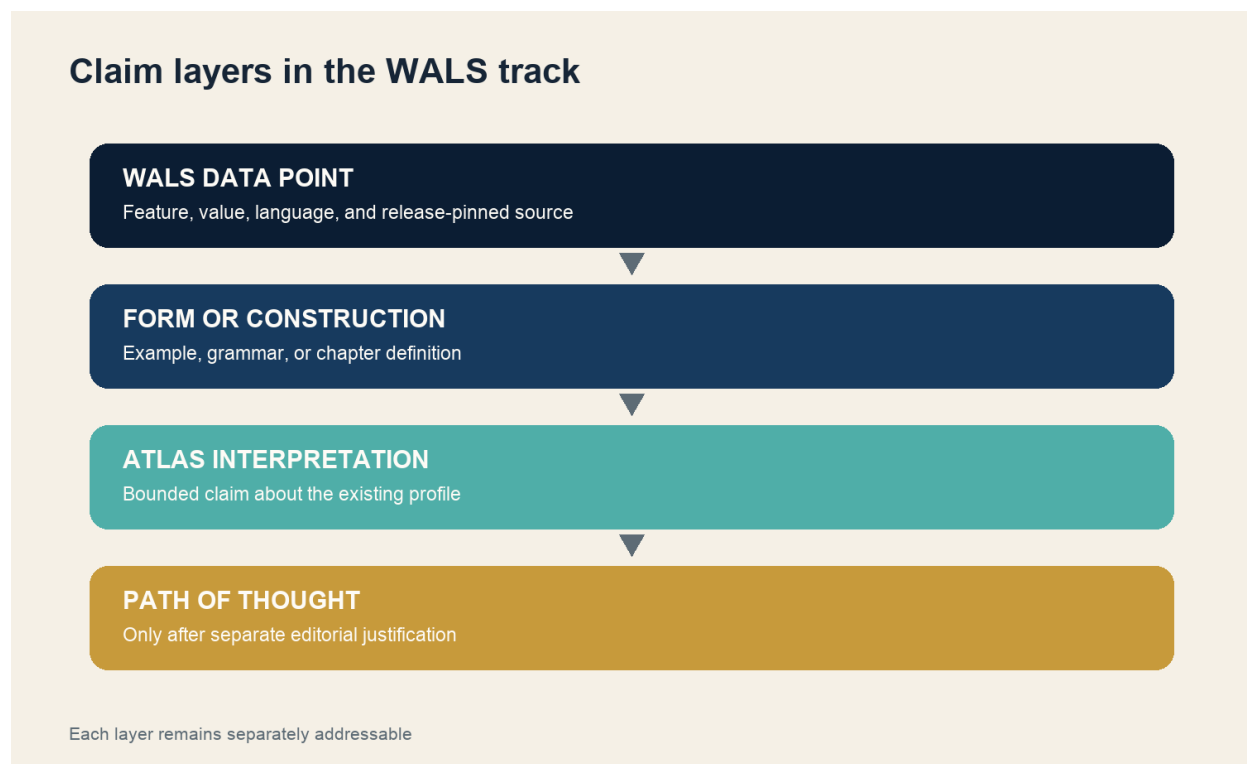


Fig. 2 A path of thought does not follow automatically from a WALS data point.

5 Evidence only and typological crosswalk

Evidence only means that a WALS or supplementary source is attached to an existing profile without changing its semantic identity. This form is appropriate when a source supports a single claim or example but does not establish a

stable external relation. The eighteen accepted operations produced 36 evidence rows: thirty WALS-related records and six supplementary scholarly sources.

A typological crosswalk is more structured. It asserts a bounded relation between an Atlas concept and a feature-value-language data point. `mappingType` describes the relation, `mappingScope` limits it to `FEATURE_VALUE_LANGUAGE`, `assertionLayer` names the supported level of claim, and `scopeNote` states the exclusion. Exactly nine such crosswalks were accepted.

6 Results of the evidence programme

The first two packages cover fourteen Atlas concepts: thirteen direct pilot cases plus the separately adjudicated Japanese person-orientation case. Topics include indirect evidentiality in Turkish, inclusive and exclusive pronouns in Martuthunira, numeral classifiers in Vietnamese, associative plural in Hungarian, lexical tone in Yoruba, full reduplication in Indonesian, vigesimal numeral bases in Chukchi, politeness strategies in Korean, possessive classification in Amele, comitative and instrumental in Inga, and case alignment, passive, and antipassive constructions.

Phase 2 added four deliberately difficult profiles. For Hindi split ergativity, WALS documents a coverage boundary while a separate scholarly source supports the aspect-based split. For Turkish differential object marking, WALS limits the reviewed domain while the specificity analysis comes from an independent study. For English *be going to*, the report separates a synchronic WALS value from historical grammaticalization. For Mandarin tone, it distinguishes the WALS classification, four lexical tones, the neutral tone, and Tone 3 sandhi.

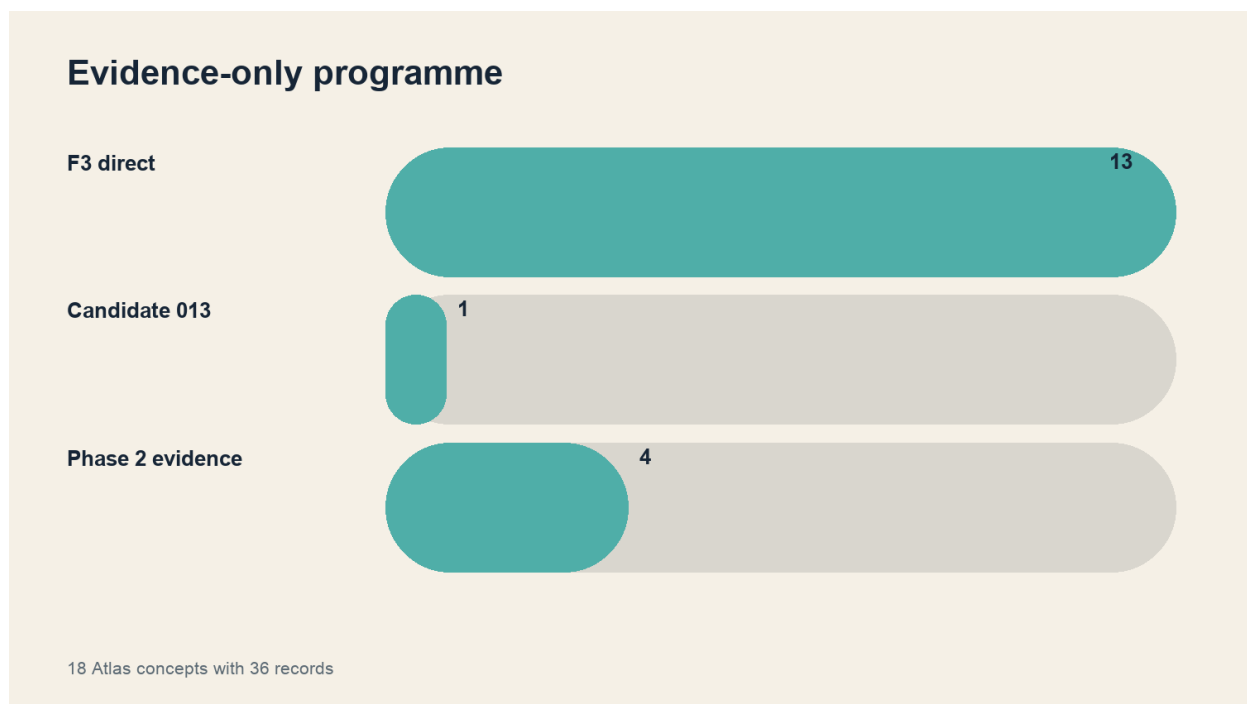


Fig. 3 Three separately accepted lines carry the evidence-only programme.



Fig. 4 Four accepted package lines carry the canonical WALS state.

7 Nine typological crosswalks

The typological layer connects nine Atlas concepts to nine distinct WALS features: Possession to 117A in Hungarian, Nominal class to 30A in Swahili, Plural to 33A in Indonesian, Deixis to 41A in Hausa, Case to 49A in Hungarian, Alienability to 58A in Acoma, Imperative to 70A in German, Indirect evidentiality to 78A in Turkish, and Word order to 81A in Ungarinjin. Hungarian appears twice, yielding eight languages and eight recorded language families.

The nine mappings use deliberately differentiated types. `FEATURE_SUBDOMAIN_OF` appears three times and `LANGUAGE_REALIZATION_OF` twice. `ENCODING_DIMENSION_OF`, `OPERATIONAL_SUBDOMAIN_OF`, `CONTRASTIVE_SUBDOMAIN_OF`, and `INVENTORY_DIMENSION_OF` appear once each. The types avoid a blanket claim that a feature is identical to an Atlas concept. Instead, they specify whether WALS describes an encoding dimension, an operational subdomain, a contrast, an inventory dimension, or a language-specific realization.

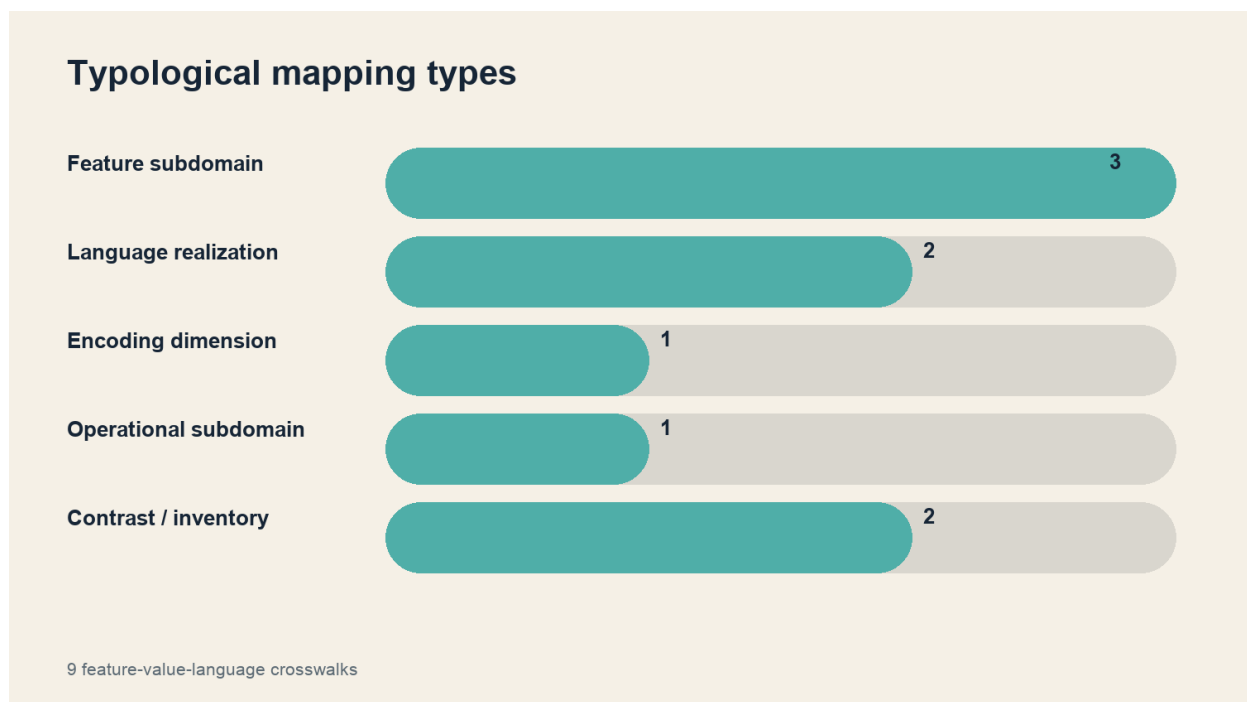


Fig. 5 Types state partial relation and scope rather than identity.

8 Case studies in scope

The crosswalk from Nominal class to WALS 30A-swa supports only an operational subrelation. WALS counts controller genders through agreement, while traditional descriptions of Bantu may count noun classes differently. The scope note therefore prevents Five or more from being read as a complete definition of the Swahili class system.

The crosswalk from Deixis to 41A-hau is `FEATURE_SUBDOMAIN_OF`. WALS maps distance contrasts in adnominal demonstratives, and the four-way Hausa value is a defensible language case. Temporal, personal, social, and discourse deixis remain explicitly outside scope. The same feature domain supplies an evidence-only record for person orientation in the Japanese pilot case. The two decisions show that the same WALS feature can support different claims in different Atlas profiles.

Plural to 33A-ind is `LANGUAGE_REALIZATION_OF`. Full reduplication in forms such as *rumah-rumah* or *perubahan-perubahan* documents one Indonesian strategy of plural coding. It neither defines plurality in general nor claims that reduplication is the only Indonesian strategy. Word order to 81A-ung is similarly limited to dominant transitive declarative clauses with full noun phrases; OVS is not presented as the rule for every clause type.

9 Area genealogy and cross-linguistic comparison

WALS visualizes geographic distributions and genealogical classifications. Such patterns are relevant to the Atlas but require separate argument. Geographic proximity may indicate contact, genealogical proximity may reflect shared history, and a broad distribution may contain several independent developments. No single map decides among these explanations.

The research track therefore stores language, Glottocode, ISO code, and family separately. It does not infer from a WALS distribution that a language forces its speakers into a particular mode of thought. A possible relation between structure, use, and cognition would be a new research question requiring its own design, comparative material, and controlled alternative explanations.

10 Canonical integration and technical gates

Four accepted Inbox packages carry the canonical WALS state: F3 Evidence 0002, Candidate 013, Phase 2 Evidence 0001, and Phase 2 Typological 0001. The first three add sources and claims; the fourth adds nine records to `external_typological_links`. Every package records reviewer, timestamp, acceptance status, manifest basis, and expected record hashes.

Import remained fail-closed. A schema preflight with acceptance checking disabled could show technical compatibility only; it did not constitute approval. Transactional import became possible only after explicit editorial acceptance. Export and readback then checked row counts, individual fields, foreign keys, and package identities. The WALS layer changed no internal Atlas relations.

11 Value for the Atlas

WALS turns broadly stated language examples into verifiable data points. A profile no longer merely claims that a language realizes evidentiality, case, tone, or word order in a particular way. It can identify the feature, value, language, and source while also recording which parts of the profile are not supported by that evidence. The combination improves both accuracy and readability.

The main gain lies in scope control. The Atlas can make cross-linguistic diversity visible without constructing universals from isolated examples. Evidence-only records support concrete statements; typological crosswalks expose a stable external structure; deferred relation types remain visible until the data model can represent their claims precisely.

12 Limitations and open research

WALS is a completed, curated resource with heterogeneous chapter methodologies and uneven language coverage. Feature values are chapter-dependent classifications, not commensurable variables across every domain. Many Atlas profiles connect several levels of language, whereas a WALS feature has a narrower operationalization. The nine crosswalks are therefore examples of a method, not a representative sample of the Atlas inventory.

Open work includes the eleven relation types deferred in the pilot, systematic review of additional Atlas profiles, and a representation of distribution patterns that does not conflate area, genealogy, and chance. Any expansion must rebind the WALS version, read the chapter definitions, and review each affected language individually.

13 Conclusion

The WALS research track shows that typological data strengthen the Atlas when they are bound at a small scale and in explicit layers. Thirty-six evidence-only records for eighteen concepts and nine typological crosswalks do not create a second atlas inside the Atlas. They provide verifiable anchors for concrete linguistic claims and make their limits explicit.

The methodological rule is to read the data point first, determine its scope second, and then store evidence and interpretation separately. On that basis, WALS can show diversity without turning individual languages into proxies for entire families, areas, or modes of thought.

Appendix A Decision and crosswalk matrix

The nine rows form the complete canonical typological crosswalk layer at the report date. Each row binds Atlas ID, WALS feature, value, and language separately.

Atlas concept	WALS value	Language	Mapping type
Possession	117A-hun	Hungarian	FEATURE_SUBDOMAIN_OF
Nominalklasse	30A-swa	Swahili	OPERATIONAL_SUBDOMAIN_OF
Plural	33A-ind	Indonesian	LANGUAGE_REALIZATION_OF
Deixis	41A-hau	Hausa	FEATURE_SUBDOMAIN_OF
Kasus	49A-hun	Hungarian	INVENTORY_DIMENSION_OF
Alienabilität	58A-aco	Acoma	CONTRASTIVE_SUBDOMAIN_OF
Imperativ	70A-ger	German	FEATURE_SUBDOMAIN_OF
Indirekte Evidentialität	78A-tur	Turkish	ENCODING_DIMENSION_OF
Wortstellung	81A-ung	Ungarinjin	LANGUAGE_REALIZATION_OF

Appendix B Hash binding

Artefact	SHA-256 / Commit
GitHub main	8e676f019f2031c58f51e42a25f899c4bdf577ef
CURRENT_MANIFEST.json	76bacf1f995e099d8f7bfd6d702dfbb18eae589a9fd9c4770bdc48e8c3770f4
concepts.jsonl	3565bc1b1d5f79b4f6d9c43fe06f445b0b2fa17e133affec1e86bfa846790f9
external_crosswalks_v2.jsonl	7cd82b9991fbdaa55503b6cc36b9ef778c0dd58061cd37a6cc7d1755b690c705
typological_crosswalks.jsonl	db3261255beffb5a4918760dd951c2f2e75d2a06b9f2c78042b0a158696d950
evidence.jsonl	4bd0c0809412339d5498dcfc111f4cf27537dabf8c30ab4a7c90ca76b5e9b97a
change-package.schema.json	9f95a51b13038d5f6580851e9604281d372ed158cfa6037e89dc419b42f216bd

STATUS

BOUNDARY

These hashes evidence the report's read basis. The report itself is a research artefact and changes neither the manifest nor canonical records.

Appendix C Quality assurance

1. Atlas ID and external ID remain in separate fields.
2. Identical words, labels, or translations do not establish identity.
3. Every positive relation states a mapping type and scope.
4. Evidence, interpretation, confidence, and review status are separate.
5. Inbox packages are schema- and hash-bound; a preflight is not acceptance.
6. Transactional import, canonical export, and post-import readback form one chain of evidence.
7. Negative findings and deferred cases remain part of the research record.

Data products and citation

Machine-readable exports and packages control any verification. The editorial DOCX and PDF editions explain method and result; they are neither import sources nor approval documents for new data.

RECOMMENDED

Reich, Michael (2026): WALS as a Space of Typological Evidence. Language-specific structural evidence, scope boundaries, and a dedicated crosswalk layer. Research Report 04, version 1.0. Atlas of Language / Atelier der Resonanzen.

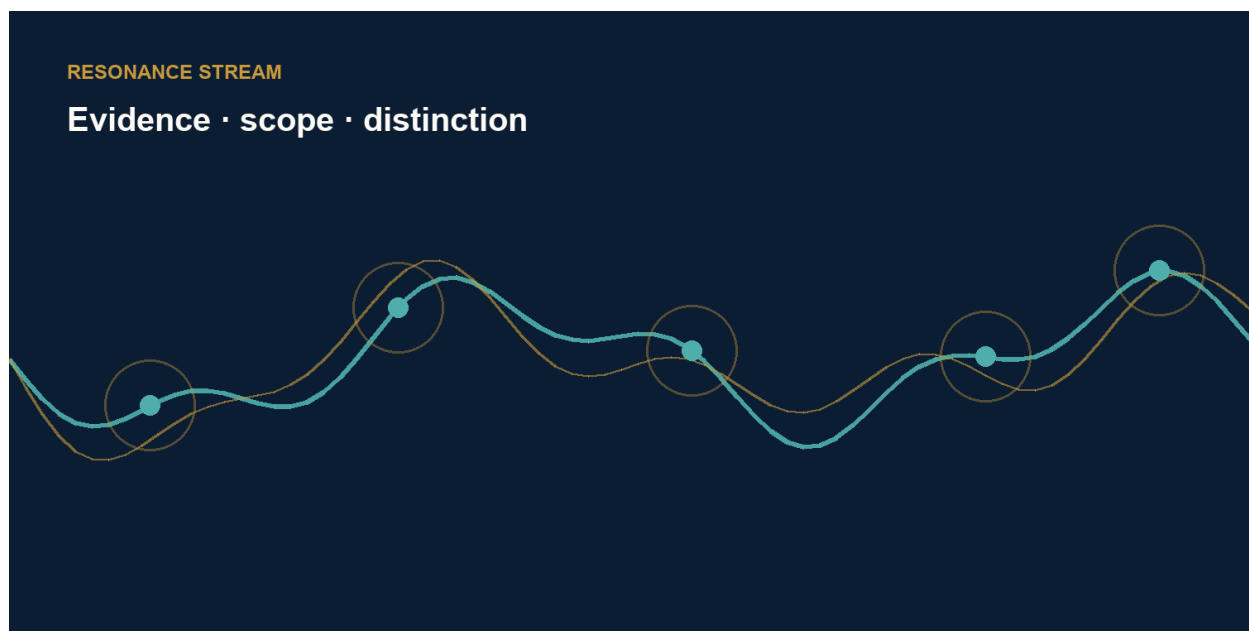
CITATION

References and sources

- Dryer, Matthew S.; Haspelmath, Martin (eds.) (2013): WALS Online. Max Planck Institute for Evolutionary Anthropology. <https://wals.info/>
- Comrie, Bernard; Dryer, Matthew S.; Gil, David; Haspelmath, Martin: WALS Introduction. Structure, feature values, language samples, and data sources. <https://wals.info/chapter/s1>
- CLDF Datasets: The World Atlas of Language Structures. Release v2020.5, commit f97440d6edbaed0097d1bd307a5a20ea3f09e271. <https://github.com/cldf-datasets/wals>
- WALS Online: feature- and language-specific chapter and value pages cited in the accepted Atlas Inbox packages.
- Atlas of Language: WALS Gate F3 packages, Phase 2 evidence package, typological crosswalk package, migrations, and canonical readback.

Transparency note

The visual design belongs to the Atlas research perspective. Charts visualize only the Atlas counts reported here; they do not replace analyses of the external dataset. Interpretation and metaphor remain identifiable as editorial layers and must not be confused with findings from the external resource.



Closing image · External data answer best when their boundaries remain visible.