



WALS als typologischer Evidenzraum

Sprachspezifische Strukturbelege, Reichweitengrenzen und eine eigene Crosswalk-Schicht

WALS als typologischer Evidenzraum

Sprachspezifische Strukturbelege, Reichweitengrenzen und eine eigene Crosswalk-Schicht

Michael Reich · Atlas der Sprache · 11. September 2026

Zusammenfassung

Der WALS-Forschungsstrang untersucht, wie der World Atlas of Language Structures Online den Atlas der Sprache mit überprüfbaren sprachvergleichenden Strukturbelegen ergänzen kann. WALS wird weder als Wörterbuch noch als Ontologie der Atlasbegriffe behandelt. Feature, Value, Language, Glottocode und Atlas-Concept-ID bleiben getrennt. Die kanonische Evidenzschicht enthält 36 akzeptierte Evidence-only-Belege zu 18 Atlasbegriffen: 30 WALS-bezogene Belege und sechs ergänzende fachwissenschaftliche Quellen. Zusätzlich führt eine eigene typologische Außenschicht neun Feature-Value-Language-Crosswalks zu neun Begriffen, neun Features und acht Sprachen. Der Bericht erläutert den 30-Fälle-Pilot, die zweistufige Evidenzregel, den Sonderfall der japanischen Personenorientierung, vier Phase-2-Grenzfälle und die neun typisierten Crosswalks. Er zeigt, warum ein einzelner Sprachwert nur eine begrenzte Aussage trägt und warum aus arealen oder genealogischen Mustern keine pauschalen sprachrelativistischen Schlüsse gezogen werden dürfen.

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.

Kennzahl	Wert
Klassifizierte Pilotbefunde	30
Evidence-only-Atlasbegriffe	18
Akzeptierte Evidence-only-Belege	36
Typologische Crosswalks	9
Features / Sprachen / Familien	9 / 8 / 8
Angenommene Inbox-Pakete	4

BERICHTSPROFIL

Evidence-only: 36 Belege zu 18 bestehenden Atlasprofilen. Typologische Schicht: 9 Feature-Value-Language-Crosswalks. Wirkung dieses Berichts: Dokumentation, keine Universalitäts- oder Identitätsbehauptung, kein Import und keine Publikation.

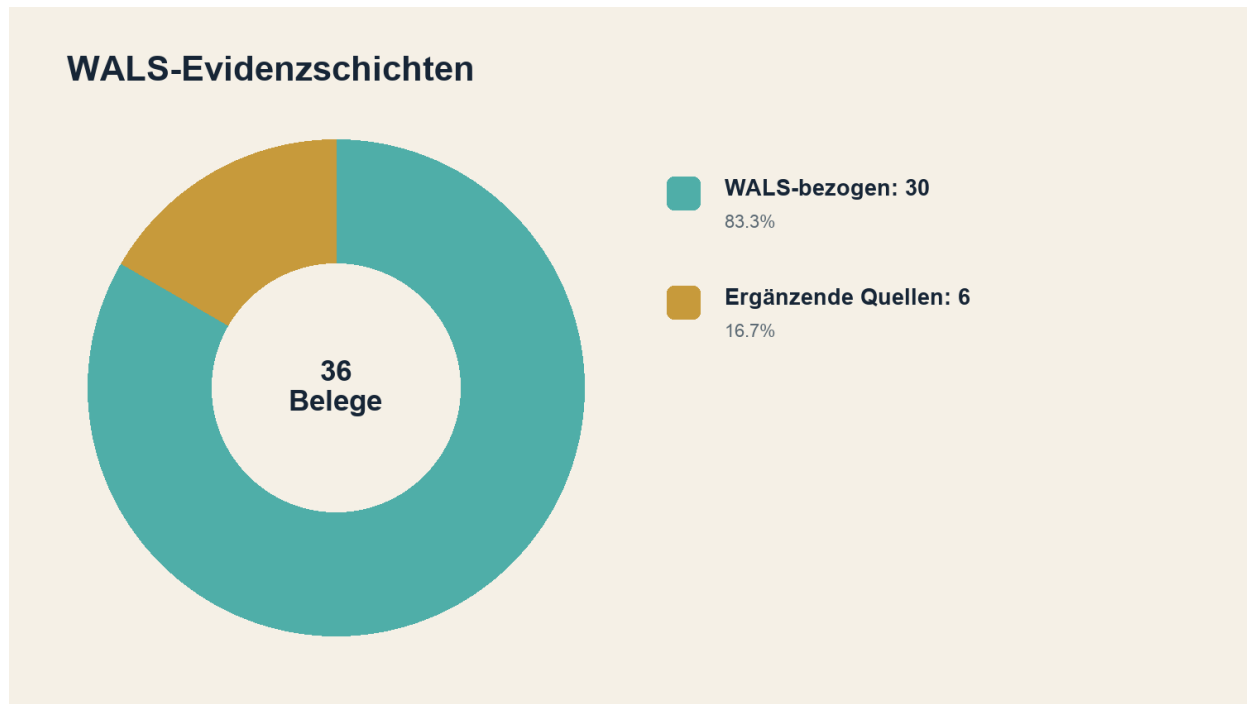


Abb. 1 Forschungsbefund und kanonische Reichweite bleiben getrennt.

1 Forschungsfrage und Gegenstand

WALS kartiert strukturelle Eigenschaften von Sprachen: lautliche Systeme, morphologische Kategorien, syntaktische Ordnungen, lexikalische Typen und weitere vergleichbare Merkmale. Der Atlas der Sprache enthält zwar viele dieser Gegenstände, verbindet sie aber mit rhetorischen, pragmatischen, semiotischen und erkenntnistheoretischen Perspektiven. Die leitende Frage lautet deshalb: Welche WALS-Daten können eine konkrete Aussage in einem Atlasprofil belegen, und welche Reichweite besitzt dieser Beleg?

Eine WALS-Zuordnung ist keine lexikalische Gleichsetzung. Sie verbindet einen Atlasbegriff mit einem bestimmten Feature, einem bestimmten Wert und einer bestimmten Sprache. Bei typologischen Crosswalks kommen Language-ID, Sprachname, Glottocode, ISO-639-3-Code und Sprachfamilie hinzu. Diese Felder verhindern, dass ein beobachteter Sprachfall zur Eigenschaft aller Sprachen oder zur Definition des Atlasbegriffs verallgemeinert wird.

2 WALS und die CLDF Datenbasis

WALS Online beschreibt sich als große Datenbank struktureller phonologischer, grammatischer und lexikalischer Eigenschaften, die aus beschreibenden Quellen von 55 Fachautorinnen und Fachautoren zusammengetragen wurden. Die Einleitung nennt 144 Kapitel und 160 Karten. Insgesamt erscheinen 2.662 Sprachen auf mindestens einer Karte; die Dichte der Abdeckung unterscheidet sich stark zwischen Sprachen und Features. Diese ungleiche Matrix gehört zur Ressource und begrenzt jede Häufigkeitsinterpretation.

Der Forschungsstrang verwendet die CLDF-Fassung v2020.5 am Commit f97440d6edbaed0097d1bd307a5a20ea3f09e271. Values, Codes, Languages, Examples und Kapiteltexte wurden als getrennte Tabellen gelesen. Die offizielle WALS-Seite erläutert, dass Feature Values für das jeweilige Feature eine erschöpfende Klassifikation bilden sollen. Das bedeutet nicht, dass jede Sprache in jedem Feature kodiert ist oder dass die Wertkategorien außerhalb der jeweiligen Kapiteldefinition dieselbe Bedeutung besitzen.

3 Kanonische Basis und Reproduzierbarkeit

Der Bericht liest GitHub-main bei 8e676f019f2031c58f51e42a25f899c4bdf577ef und den kanonischen Export Atlas 249 / Sites 212 / Datenmodell 1.5.0. Der Export ist am Checkpoint atlas-249-wals-phase2-checkpoint gebunden; das Manifest nennt 85ba65244988b1aa6d89e0af960109edc3192b7a als Export-Commit. Die typologische JSONL-Schicht

enthält neun Datensätze. Der kanonische Evidenzbestand umfasst 1.718 Zeilen, darunter die akzeptierten WALS- und Ergänzungsbelege dieses Strangs.

Jede Evidence-only-Operation war an den SHA-256 der bestehenden concepts.jsonl-Zeile gebunden und fügte nur Quellenbezüge hinzu. Sie änderte weder Concept-ID, Lemma, Definition, Beispiele, Sprachformen, Qualitätsstatus noch interne Relationen. Die typologischen Crosswalks liegen in einer eigenen Tabelle und einem eigenen Export. Der Bericht selbst schreibt weder Inbox- noch kanonische Dateien.

DREIFACHE**TRENNUNG**

Empirischer WALS-Befund, Atlas-Interpretation und abgeleiteter Denkweg werden als verschiedene Aussageebenen geführt.

4 Pilot und Reviewlogik

Der Gate-F3-Pilot klassifizierte 30 Forschungsbefunde genau einmal. 13 direkte Fälle wurden als Evidence-only-Vorschläge ausgearbeitet. Elf Fälle warteten auf ein passendes Relationsmodell. Kandidat 013 wurde separat neu geprüft. Verteilungsbefunde, Zurückweisungen und Negativkontrollen blieben außerhalb eines Importpakets. Diese Trennung verhinderte, dass ein technisch schemafähiger Vorschlag automatisch als redaktionell angenommen galt.

Jeder der 13 direkten Fälle erhielt zwei Belege. Der erste bindet den exakten WALS-Sprachwert in values.csv an Feature-ID, Code-ID und Language-ID. Der zweite bindet die konkrete Form oder Konstruktion an examples.csv, den Kapiteltext oder eine ergänzende Fachquelle. Der Sprachwert belegt damit die Klassifikation, während der Formbeleg erklärt, woran die Klassifikation sprachlich sichtbar wird. Beide Ebenen bleiben getrennt zitierbar.

Kandidat 013 betrifft die Personenorientierung japanischer Demonstrativa. Der Code 41A-3 Three-way contrast reicht allein nicht aus. Erst der sprachspezifische Kommentar zu 41A-jpn unterscheidet Nähe zum Sprecher, Nähe zum Hörer und Entfernung von beiden. Deshalb wurde der Fall separat und nur für den japanischen Teil des Atlasprofils akzeptiert; Tonganisch und Hausa wurden dadurch nicht mitvalidiert.

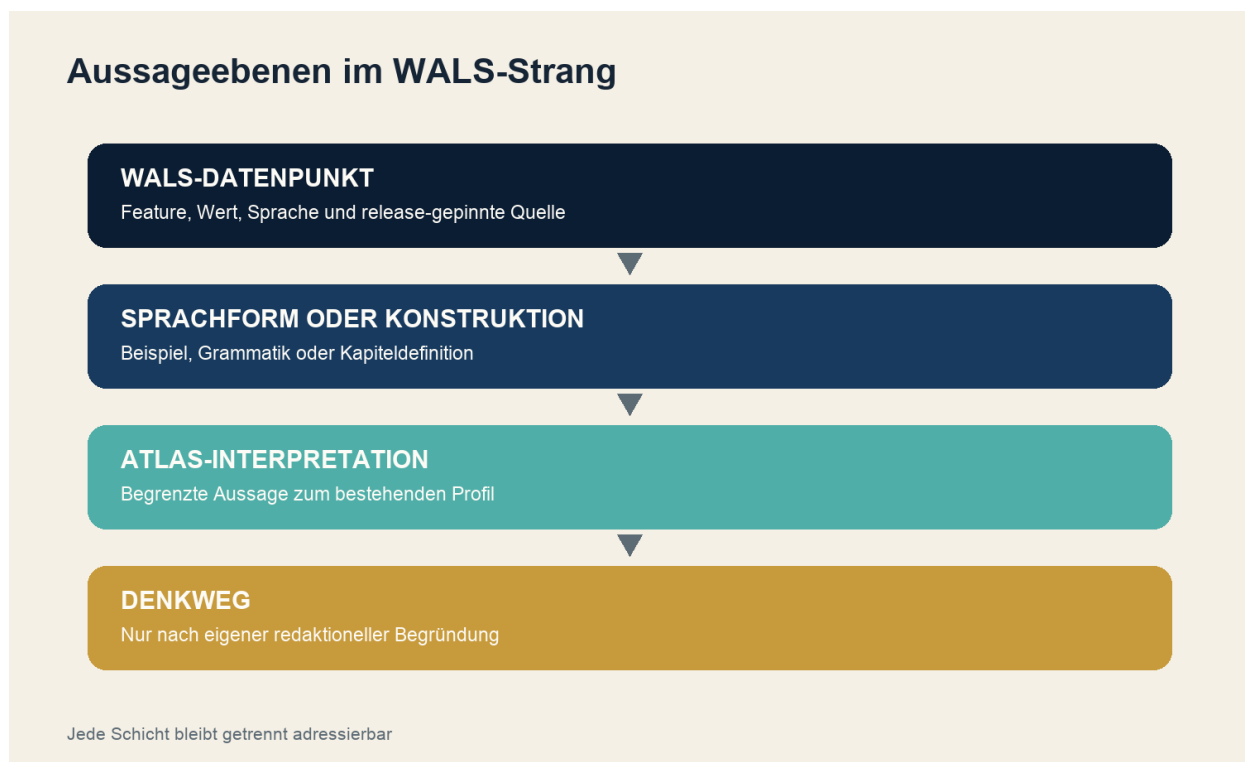


Abb. 2 Ein Denkweg folgt nicht automatisch aus einem WALS-Datenpunkt.

5 Evidence only und typologischer Crosswalk

Evidence-only bedeutet, dass ein WALS- oder Ergänzungsbeleg an ein bestehendes Profil angehängt wird, ohne dessen semantische Identität zu verändern. Diese Form eignet sich, wenn eine Quelle eine einzelne Aussage oder ein Beispiel stützt, aber keine stabile externe Relation modelliert. Die 18 angenommenen Operationen erzeugten 36 Belegzeilen: 30 WALS-bezogene und sechs ergänzende fachwissenschaftliche Quellen.

Ein typologischer Crosswalk ist stärker strukturiert. Er behauptet eine begrenzte Beziehung zwischen einem Atlasbegriff und einem Feature-Value-Language-Datenpunkt. mappingType beschreibt die Art der Beziehung, mappingScope begrenzt sie auf FEATURE_VALUE_LANGUAGE, assertionLayer benennt die tatsächlich getragene Aussage, und scopeNote formuliert den Ausschluss. Genau neun solcher Crosswalks wurden angenommen.

6 Ergebnisse des Evidenzprogramms

Die ersten beiden Pakete decken 14 Atlasbegriffe ab: 13 direkte Pilotfälle sowie die separat adjudizierte japanische Personenorientierung. Zu den Themen gehören indirekte Evidentialität im Türkischen, inklusive und exklusive Pronomen im Martuthunira, Numeralklassifikatoren im Vietnamesischen, Assoziativplural im Ungarischen, lexikalischer Ton im Yoruba, Vollreduplikation im Indonesischen, vigesimale Zahlbasis im Tschuktschischen, Höflichkeitsstrategien im Koreanischen, Possessivklassifikation im Amele, Komitativ und Instrumental im Inga sowie Kasusausrichtung, Passiv und Antipassiv.

Phase 2 ergänzte vier bewusst schwierige Profile. Für Hindi-Split-Ergativität dokumentiert WALS eine Abdeckungsgrenze, während eine eigene Fachquelle die aspektbasierte Spaltung trägt. Beim türkischen Differential Object Marking begrenzt WALS den geprüften Bereich; die Spezifitätsanalyse stammt aus einer eigenständigen Studie. Beim englischen be going to trennt der Bericht synchronen WALS-Wert und historische Grammatikalisierung. Beim Mandarin-Ton trennt er WALS-Klassifikation, vier lexikalische Töne, Neutralton und Tone-3-Sandhi.

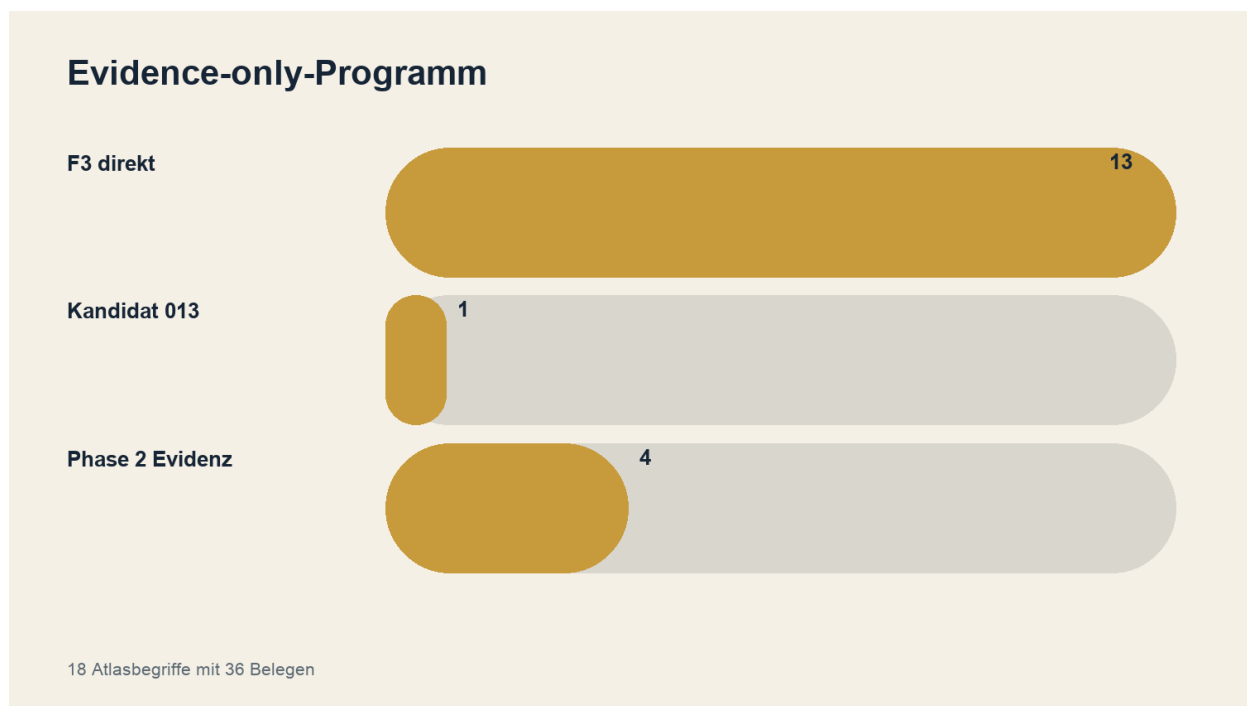


Abb. 3 Drei getrennt angenommene Linien tragen das Evidence-only-Programm.

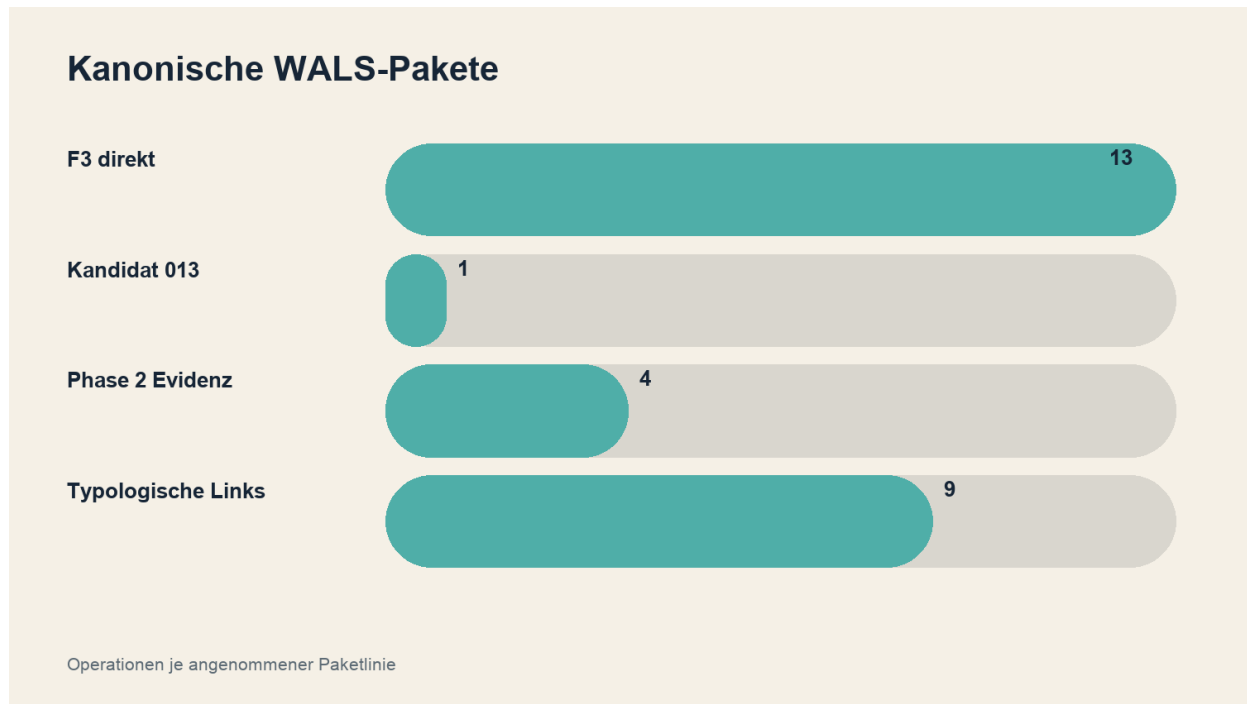


Abb. 4 Vier angenommene Paketlinien tragen den kanonischen WALS-Stand.

7 Neun typologische Crosswalks

Die typologische Schicht verbindet neun Atlasbegriffe mit neun verschiedenen WALS-Features: Possession mit 117A im Ungarischen, Nominalklasse mit 30A im Swahili, Plural mit 33A im Indonesischen, Deixis mit 41A im Hausa, Kasus mit 49A im Ungarischen, Alienabilität mit 58A im Acoma, Imperativ mit 70A im Deutschen, indirekte Evidentialität mit 78A im Türkischen und Wortstellung mit 81A im Ungarinjin. Ungarisch erscheint zweimal; dadurch ergeben sich acht Sprachen und acht eingetragene Sprachfamilien.

Die neun Mappingtypen sind bewusst differenziert. `FEATURE_SUBDOMAIN_OF` kommt dreimal vor, `LANGUAGE_REALIZATION_OF` zweimal. `ENCODING_DIMENSION_OF`, `OPERATIONAL_SUBDOMAIN_OF`, `CONTRASTIVE_SUBDOMAIN_OF` und `INVENTORY_DIMENSION_OF` erscheinen je einmal. Die Typen vermeiden die pauschale Behauptung, ein Feature sei mit einem Atlasbegriff identisch. Sie sagen stattdessen, ob WALS eine Kodierungsdimension, einen operativen Teilbereich, einen Kontrast, eine Inventardimension oder eine sprachspezifische Realisierung beschreibt.

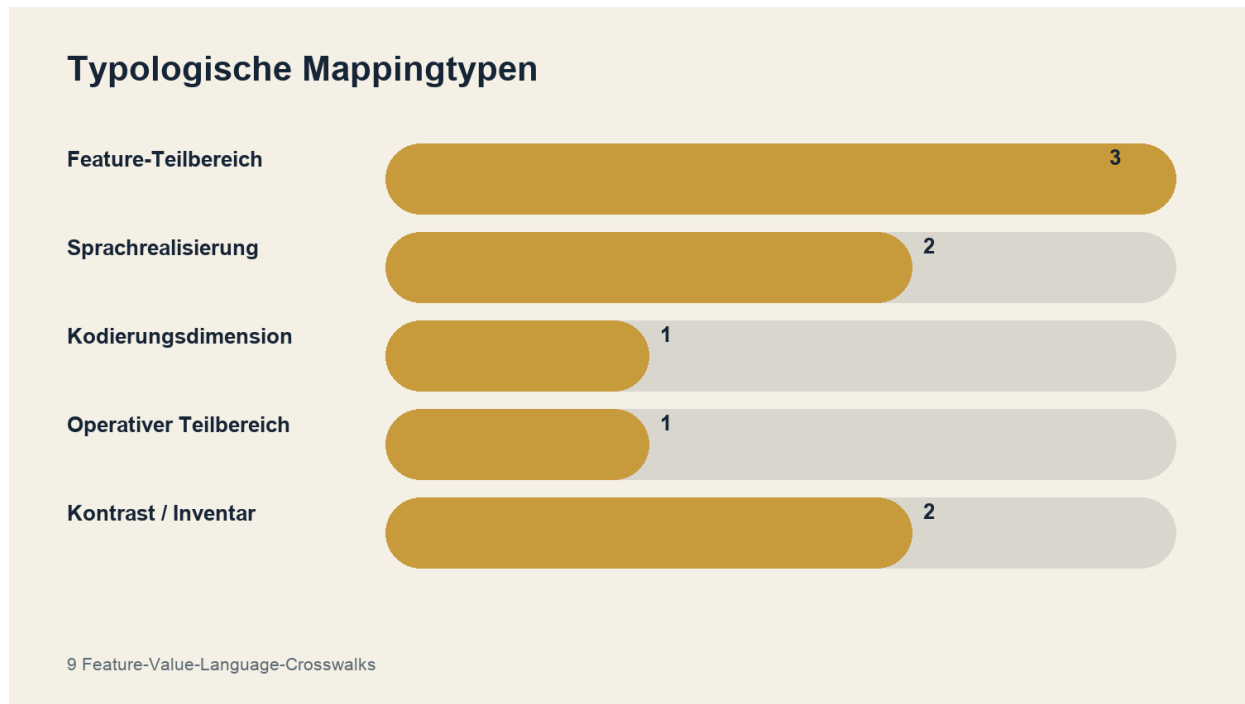


Abb. 5 Typen benennen Teilbeziehung und Reichweite statt Identität.

8 Fallstudien zur Reichweite

Der Crosswalk Nominalklasse zu WALS 30A-swa trägt nur eine operationale Teilbeziehung. WALS zählt controller genders anhand von Kongruenz; traditionelle Bantu-Beschreibungen können eine andere Zahl von Nominalklassen führen. Die scopeNote verhindert deshalb, dass Five or more als vollständige Definition des Swahili-Klassensystems gelesen wird.

Der Crosswalk Deixis zu 41A-hau ist FEATURE_SUBDOMAIN_OF. WALS kartiert Distanzkontraste adnominaler Demonstrativa; der viergliedrige Hausa-Wert ist ein belastbarer Sprachfall. Zeitliche, personale, soziale und diskursive Deixis bleiben ausdrücklich außerhalb. Derselbe Featurebereich liefert beim japanischen Pilotfall einen Evidence-only-Beleg für Personenorientierung. Beide Entscheidungen zeigen, dass derselbe WALS-Featurebereich je nach Atlasprofil eine andere Aussage trägt.

Plural zu 33A-ind ist LANGUAGE_REALIZATION_OF. Die vollständige Reduplikation in rumah-rumah oder perubahan-perubahan belegt eine indonesische Strategie der Pluralkodierung. Sie definiert weder Pluralität im Allgemeinen noch behauptet sie, Reduplikation sei die einzige indonesische Strategie. Wortstellung zu 81A-ung ist analog auf dominante transitive Deklarativsätze mit vollen Nominalphrasen begrenzt; OVS wird nicht zur Regel jeder Satzart erklärt.

9 Arealität Genealogie und Sprachvergleich

WALS visualisiert geographische Verteilungen und ordnet Sprachen genealogisch. Solche Muster sind für den Atlas interessant, aber sie brauchen eine eigene Begründung. Räumliche Nachbarschaft kann Kontakt anzeigen, genealogische Nähe gemeinsame Geschichte, und eine weit gestreute Verteilung kann mehrere unabhängige Entwicklungen enthalten. Kein einzelnes Kartenbild entscheidet zwischen diesen Erklärungen.

Der Forschungsstrang speichert deshalb Sprache, Glottocode, ISO-Code und Familie getrennt. Er leitet aus einer WALS-Verteilung keine Behauptung ab, eine Sprache zwingt ihre Sprecherinnen und Sprecher zu einer bestimmten Denkweise. Eine mögliche Verbindung zwischen Struktur, Gebrauch und Denken wäre eine neue Forschungsfrage mit eigenem Design, Vergleichsmaterial und kontrollierten Alternativerklärungen.

10 Kanonische Integration und technische Gates

Vier angenommene Inbox-Pakete tragen den kanonischen WALS-Stand: F3 Evidence 0002, Kandidat 013, Phase 2 Evidence 0001 und Phase 2 Typological 0001. Die ersten drei fügen Quellen und Claims hinzu; das vierte fügt neun Datensätze in `external_typological_links` ein. Alle Pakete dokumentieren Reviewer, Zeitpunkt, Annahmestatus, Manifestbasis und erwartete Record-Hashes.

Der Import blieb fail-closed. Ein Schema-Preflight mit abgeschalteter Annahmeprüfung konnte nur technische Kompatibilität zeigen; er war keine Freigabe. Erst die explizite redaktionelle Annahme erlaubte den transaktionalen Import. Danach prüften Export und Readback die Zahl der Zeilen, die einzelnen Felder, Fremdschlüssel und Paketidentitäten. Die WALS-Schicht veränderte keine internen Atlasrelationen.

11 Nutzen für den Atlas

WALS macht aus allgemein formulierten Sprachbeispielen überprüfbare Datenpunkte. Ein Profil kann nicht länger nur behaupten, dass eine Sprache Evidentialität, Kasus, Ton oder Wortstellung auf bestimmte Weise realisiert. Es kann Feature, Wert, Sprache und Quelle nennen und zugleich festhalten, welche Teile des Profils dadurch nicht belegt werden. Diese Kombination verbessert Genauigkeit und Lesbarkeit.

Der eigentliche Gewinn liegt in der Reichweitenkontrolle. Der Atlas kann sprachvergleichende Vielfalt sichtbar machen, ohne aus einzelnen Beispielen Universalien zu konstruieren. Evidence-only-Belege stärken konkrete Sätze; typologische Crosswalks erschließen eine stabile externe Struktur; zurückgestellte Relationstypen bleiben sichtbar, bis das Datenmodell ihre Aussage präzise tragen kann.

12 Grenzen und offene Forschung

WALS ist eine abgeschlossene, kuratierte Ressource mit heterogener Kapitelmethodik und ungleicher Sprachabdeckung. Featurewerte sind kapitelabhängige Klassifikationen; sie sind keine messgleichen Variablen über alle Domänen hinweg. Viele Atlasprofile verbinden mehrere sprachliche Ebenen, während ein WALS-Feature eine engere Operationalisierung besitzt. Die neun Crosswalks sind daher Beispiele einer Methode, keine repräsentative Stichprobe des Atlasbestands.

Offen bleiben die elf im Pilot zurückgestellten Relationstypen, eine systematische Prüfung weiterer Atlasprofile und die Frage, wie Verteilungsmuster dargestellt werden können, ohne Arealität, Genealogie und Zufall zu vermischen. Jeder Ausbau muss die WALS-Version neu binden, Kapiteldefinitionen lesen und die jeweils betroffene Sprache einzeln prüfen.

13 Schlussfolgerung

Der WALS-Forschungsstrang zeigt, dass typologische Daten den Atlas dann stärken, wenn sie kleinräumig und mehrschichtig gebunden werden. 36 Evidence-only-Belege zu 18 Begriffen und neun typologische Crosswalks bilden keinen zweiten Atlas im Atlas. Sie schaffen überprüfbare Anker für konkrete Sprachbehauptungen und machen ihre Grenzen explizit.

Die methodische Regel lautet: erst den Datenpunkt lesen, dann seine Reichweite bestimmen, anschließend Evidenz und Interpretation getrennt speichern. Auf dieser Grundlage kann WALS Vielfalt zeigen, ohne einzelne Sprachen zu Stellvertretern ganzer Familien, Regionen oder Denkweisen zu machen.

Anhang A Entscheidungs- und Crosswalk-Matrix

Die neun Zeilen sind die gesamte kanonische typologische Crosswalk-Schicht am Berichtsstand. Jede Zeile bindet Atlas-ID, WALS-Feature, Wert und Sprache separat.

Atlasbegriff	WALS-Wert	Sprache	Mappingtyp
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

Anhang B Hashbindung

Artefakt	SHA-256 / Commit
GitHub main	8e676f019f2031c58f51e42a25f899c4bdf577ef
CURRENT_MANIFEST.json	76bacf1f995e099d8f7bfd6d702dfbb18eae589a9fdf9c4770bdc48e8c3770f4
concepts.jsonl	3565bc1b1d5f79b4f6d9c43fe06f445b0b2fa17e133affec1e86bfa846790f9
external_crosswalks_v2.jsonl	7cd82b9991fbdaa55503b6cc36b9ef778c0dd58061cd37a6cc7d1755b690c705
typological_crosswalks.jsonl	db3261255beffbd5a4918760dd951c2f2e75d2a06b9f2c78042b0a158696d950
evidence.jsonl	4bd0c0809412339d5498dcfc111f4cf27537dabf8c30ab4a7c90ca76b5e9b97a
change-package.schema.json	9f95a51b13038d5f6580851e9604281d372ed158cfa6037e89dc419b42f216bd

STATUSGRENZE

Diese Hashes belegen die Lesebasis des Berichts. Der Bericht selbst ist ein Forschungsartefakt und verändert weder Manifest noch kanonische Datensätze.

Anhang C Qualitätssicherung

1. Atlas-ID und externe ID bleiben in getrennten Feldern.
2. Gleiche Wörter, Labels oder Übersetzungen begründen keine Identität.
3. Jede positive Beziehung benennt Mappingtyp und Reichweite.
4. Evidenz, Interpretation, Konfidenz und Reviewstatus sind getrennt.
5. Inbox-Pakete sind schema- und hashgebunden; ein Preflight ist keine Annahme.
6. Transaktionaler Import, kanonischer Export und Post-Import-Readback bilden einen gemeinsamen Nachweis.
7. Negativbefunde und zurückgestellte Fälle bleiben als Forschungswissen erhalten.

Datenprodukte und Zitierweise

Für Nachprüfungen sind die maschinenlesbaren Exporte und Pakete maßgeblich. Die redaktionellen DOCX- und PDF-Fassungen erläutern Methode und Ergebnis; sie sind weder Importquelle noch Freigabedokument für neue Daten.

EMPFOHLENE

Reich, Michael (2026): WALS als typologischer Evidenzraum. Sprachspezifische Strukturbelege, Reichweitengrenzen und eine eigene Crosswalk-Schicht. Forschungsbericht 04, Version 1.0. Atlas der Sprache / Atelier der Resonanzen.

ZITIERWEISE

Literatur und Quellen

- Dryer, Matthew S.; Haspelmath, Martin (Hg.) (2013): WALS Online. Max Planck Institute for Evolutionary Anthropology. <https://wals.info/>
- Comrie, Bernard; Dryer, Matthew S.; Gil, David; Haspelmath, Martin: WALS Introduction. Aufbau, Featurewerte, Sprachstichproben und Datenquellen. <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- und sprachspezifische Kapitel- und Werteseiten, zitiert in den angenommenen Atlas-Inbox-Paketen.
- Atlas der Sprache: WALS-Gate-F3-Pakete, Phase-2-Evidenzpaket, typologisches Crosswalk-Paket, Migrationen und kanonischer Readback.

Transparenzhinweis

Die grafische Gestaltung gehört zur Forschungsperspektive des Atlas. Diagramme visualisieren ausschließlich die im Bericht genannten Atlaszählungen; sie ersetzen keine externen Datenanalysen. Interpretation und Metapher sind als redaktionelle Ebene erkennbar und dürfen nicht mit den Befunden der externen Ressource verwechselt werden.



Schlussbild Externe Daten antworten am besten, wenn ihre Grenzen sichtbar bleiben.