



Historical Thesaurus of English and the Atlas of Language

Final report of the HTF research track and canonical evidence integration

Historical Thesaurus of English and the Atlas of Language

Final report of the HTF research track and canonical evidence integration

Michael Reich · Atlas of Language · 11 September 2026

Abstract

The HTF research track investigates how the *Historical Thesaurus of English* (HTE) can serve as historical evidence for the Atlas of Language without equating the HTE semantic hierarchy with the Atlas internal ontology. A purposive pilot set of 28 Atlas concepts was researched against HTE version 5.0 and reviewed for scope, polysemy, recorded chronology and level of abstraction. A multi-stage AI review distinguished 21 accepted evidence links, four context-only findings and three documented negative findings. Among the 21 accepted cases, ten are direct mappings, four point to an HTE category broader than the Atlas concept and seven to a narrower HTE category. The controlled accepted import added only 21 evidence records. It changed no Atlas concepts, definitions, internal relations, translations or crosswalk layers. Evidence IDs 1719 through 1739 are bound to the unchanged package hash. Following the later WordNet consolidation, the complete HTE contribution persists in the current Atlas 9.0.0 readback: 21 HTE records within 1,739 evidence records overall. The research demonstrates that a historical thesaurus is particularly valuable as source-specific provenance and semantic-field evidence, provided direct correspondence, broader scope, narrower scope, contextual proximity and non-match remain separate decision classes.

Keywords: Historical Thesaurus of English; HTE; historical semantics; semantic change; evidence; scope; polysemy; controlled import; provenance; Atlas of Language.

Metric	Value
Selected pilot concepts	28
Accepted HTE evidence records	21
Context-only / negative findings	4 / 3
Direct / HTE broader / HTE narrower	10 / 4 / 7
Current evidence inventory	1,739

REPORT

Research: purposive pilot covering 28 Atlas concepts. Canonical read layer: 21 scope-typed HTE evidence records. Effect of this report: research closeout and documentation; no external publication.

PACKAGE

ATLAS-INBOX-HTE-PILOT-V1-EVIDENCE-0001
 96d2a546a0fb13700ec3f63d8dcca9657394c32ca665be252347f488d1f36bb0

PROFILE

BINDING

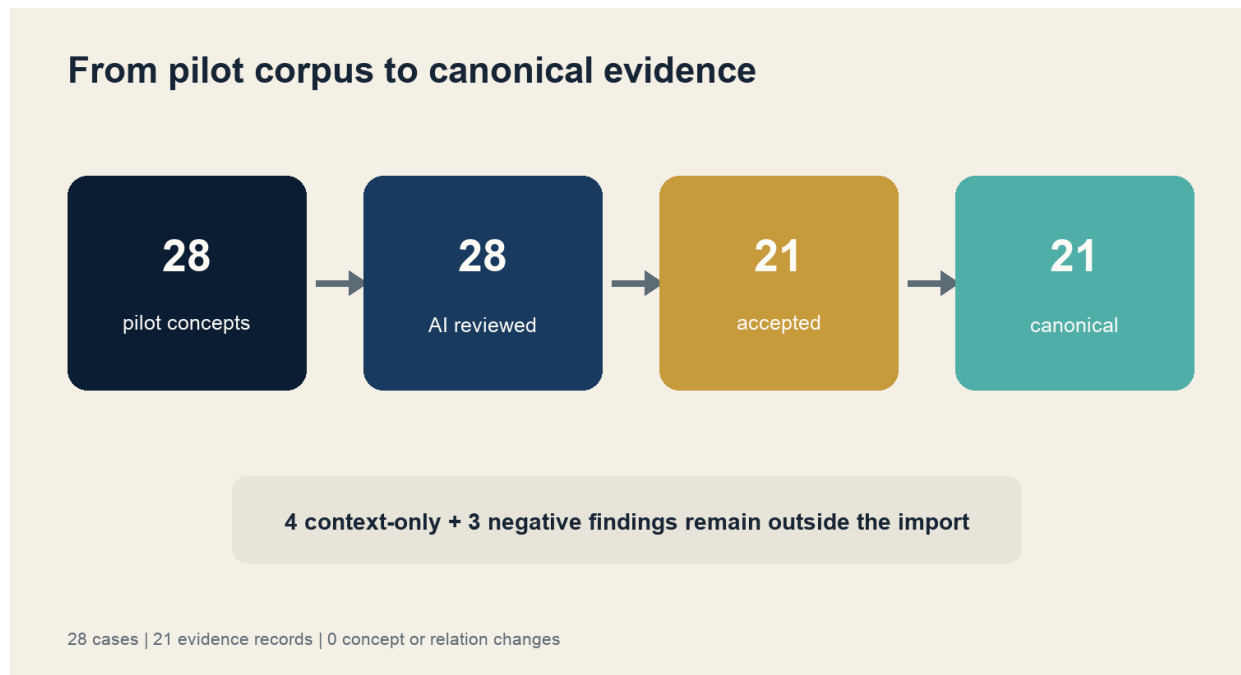


Fig. 1 From pilot corpus to canonical HTE evidence.

1 Research question and scope

The guiding question is: **How can a historically organized English semantic resource strengthen the Atlas empirically without turning external categories into presumed Atlas identities?** The pilot is neither a complete HTE survey nor a statistically representative sample. It is a methodological stress test using 28 deliberately diverse concepts from rhetoric, pragmatics, epistemology, narrative forms, literacy, morphology, syntax and translation.

Three layers remain separate throughout the project:

- **Source layer:** What does HTE version 5.0 actually document as a historical category or lexicalization?
- **Interpretive layer:** Does the external semantic field match the Atlas concept directly, is it broader or narrower, or does it provide context only?
- **Canonical layer:** Which reviewed claim may be added as evidence without changing Atlas identity or internal relations?

The project did not attempt to import the HTE classification, perform a bulk download, redefine Atlas concepts automatically, publish HTE raw data or claim that English lexical history represents cross-linguistic universals.

2 The external resource

HTE organizes the historical English lexicon onomasiologically, moving from semantic domains to linguistic expressions. Its current second edition, version 5.0, combines evidence from the *Oxford English Dictionary* and the *Thesaurus of Old English*. The official description lists approximately 235,249 categories and 804,830 entries under three major divisions: the external world, the mental world and the social world. Its hierarchy is semantic and prototypical; neighboring entries are not necessarily synonyms.

Feature	Relevance to the Atlas	Methodological consequence
Historical dates	Document recorded English usage	Do not treat dates as the origin of a universal concept
Polysemous lemmas	A word may occur in several categories	String equality is not identity evidence
Hierarchical categories	Broader and narrower regions are visible	Record scope direction explicitly
Versioned numbering	Categories and dates may be revised	Cite version 5.0 and access date

Feature	Relevance to the Atlas	Methodological consequence
Persistent category URLs	Individual findings remain inspectable	Preserve URL and category in the evidence record
Terms of use	Non-commercial research and limited extracts	Do not replicate raw data; document small source-bound findings only

Following the official guidance, the resource is cited as the *Historical Thesaurus of English*, second edition, version 5.0, University of Glasgow, with current access date and version. Its persistent research data record is available at [DOI 10.5525/gla.researchdata.2163](https://doi.org/10.5525/gla.researchdata.2163).

3 Data and method

3 1 Pilot design

The 28 concepts were chosen purposively. The set combines likely historical anchors, abstract boundary cases and negative controls. This makes the procedure capable of demonstrating not just matches but also the conditions under which an import must not occur.

Review stage	Input	Decision criterion	Durable output
Selection	28 Atlas concepts	thematic spread and risk cases	pilot-selection.jsonl
Source research	HTE v5.0	category, path, date, URL	pilot-crosswalk.jsonl
Initial typing	28 findings	direct, broader, narrower, context, negative	mapping proposal v0
AI review	all 28 cases	scope, polysemy, evidence text, false match	ai-expert-review-v1.jsonl
Owner gate	21 positive cases	explicit evidence-only acceptance	accepted package
Dry-run and import	unchanged package	21 additions, 0 duplicates, 0 conflicts	migration 0133
Readback	canonical export	package hash, IDs, counts, unresolved relations	verified final state

3 2 Multi-stage AI expert review

Because external human expert review is not economically feasible for this private project, a documented AI review was used as proportionate quality assurance. It simulated four separate perspectives: historical lexicography, semantic scope, data provenance and canonical integration safety. The AI had no authority to accept data. It revised three mapping types and downgraded four cases to context-only. Final acceptance of the 21 evidence records remained an explicit owner decision.

This procedure is not anonymous human peer review and is not represented as such. Its strength is reproducible review of every case, visible reasons for dissenting judgments and a clear boundary between recommendation and approval.

4 Results

Twenty-one of the 28 investigated Atlas concepts were accepted as canonical HTE evidence. Four findings provide useful research context but are not precise enough for an evidence link. Three concepts produced no defensible HTE anchor. The 75 percent acceptance rate describes this deliberately constructed pilot only and must not be extrapolated to the full Atlas.

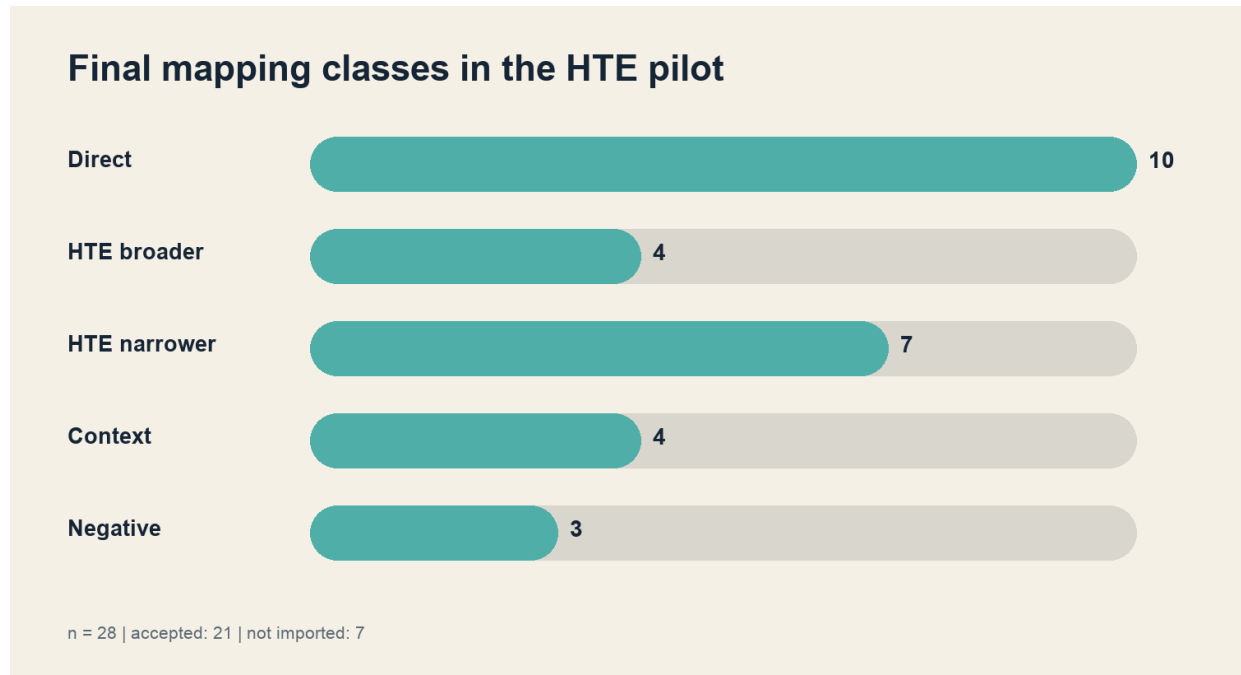


Fig. 2 Final distribution of mapping types.

Final class	Count	Share of pilot	Canonical effect
Direct	10	35.7%	additive HTE evidence
HTE broader	4	14.3%	additive evidence with scope warning
HTE narrower	7	25.0%	additive evidence with scope warning
Context	4	14.3%	research record, no import
Negative	3	10.7%	documented non-match, no import
Total	28	100%	21 imported, 7 not imported

4 1 Accepted evidence cases

Atlas concept	Domain	Final type	Claim boundary
Metaphor	Rhetoric/stylistics	HTE narrower	HTE path supports a specific historical subregion
Irony	Rhetoric/stylistics	direct	defensible historical semantic anchor
Euphemism	Rhetoric/stylistics	direct	defensible historical semantic anchor
Hyperbole	Rhetoric/stylistics	direct	defensible historical semantic anchor
Metonymy	Pragmatics	HTE broader	external category exceeds the Atlas concept
Pun	Word play/language puzzles	direct	historical lexicalization as evidence
Question	Pragmatics/conversation	HTE broader	communicative field exceeds Atlas profile
Request	Pragmatics/conversation	direct	defensible action and semantic anchor
Gesture	Pragmatics/conversation	HTE broader	HTE groups a wider field of expression
Attention	Philosophy/language/thought	HTE narrower	external region covers part of the Atlas concept
Memory	Philosophy/language/thought	HTE narrower	historical categories provide partial perspectives
Body	Philosophy/language/thought	HTE broader	HTE domain is ontologically more inclusive

Atlas concept	Domain	Final type	Claim boundary
Voice	Sound/phonetics/phonology	direct	historical expression field matches evidence target
Proverb	Poetics/narrative	direct	stable historical genre category
Fable	Poetics/narrative	direct	stable historical genre category
Quotation	Literacy/typography	direct	defensible historical category
Testimony	Argumentation/logic	HTE narrower	researched sense is narrower than Atlas profile
Evidence	Philosophy/language/thought	HTE narrower	source-bound subsense, not full identity
Interjection	Morphology/word form	direct	defensible grammatical-lexical anchor
Plural	Syntax/grammar	direct	historically documented grammatical category
Transliteration	Translation/untranslatability	HTE narrower	external operation is narrower than Atlas domain

4.2 Context-only and negative findings

Concept	Final status	Why no import?	Research value
Silence	context	semantic proximity without a precise binding	distinguishes neighborhood from evidence
Empathy	context	historical path covers related perspectives only	prevents modern back-projection
Cultural memory	context	complex compound without stable HTE identity	exposes composite Atlas concepts
Literacy/written discourse	context	lexical proximity without concept identity	separates word form from concept
Language death	negative	no adequate HTE anchor	documents an external coverage gap
Distributed cognition	negative	modern technical concept lacks historical coverage	prevents artificial backdating
Linguistic landscape	negative	disciplinary term lacks secure historical path	protects specialist Atlas scope

5 Mapping typology

The central methodological contribution is not the raw match count but the direction of semantic scope. A direct anchor can support narrowly formulated evidence. Broader or narrower HTE categories require an evidence statement that preserves the difference.

Type	Formal reading	Permitted claim	Invalid shortcut
Direct	Atlas and HTE scopes overlap defensibly	HTE documents a matching historical semantic anchor	the two ontologies are identical
HTE broader	HTE scope contains Atlas scope	Atlas concept lies within a wider historical field	the whole HTE field equals the Atlas concept
HTE narrower	HTE scope is part of Atlas scope	HTE documents a historical partial perspective	the partial perspective defines the whole concept
Context	proximity without robust binding	finding informs interpretation and later research	context is canonical evidence
Negative	no adequate anchor	reviewed non-match documents an external limit	absence means the concept did not exist

The three revised cases - Metaphor, Memory and Testimony - demonstrate the value of this typology. All remained positive research findings, but each had to move from an assumed direct correspondence to **HTE narrower**. The correction does not reduce the value of the source; it specifies what the source can support.

6 Controlled canonical integration

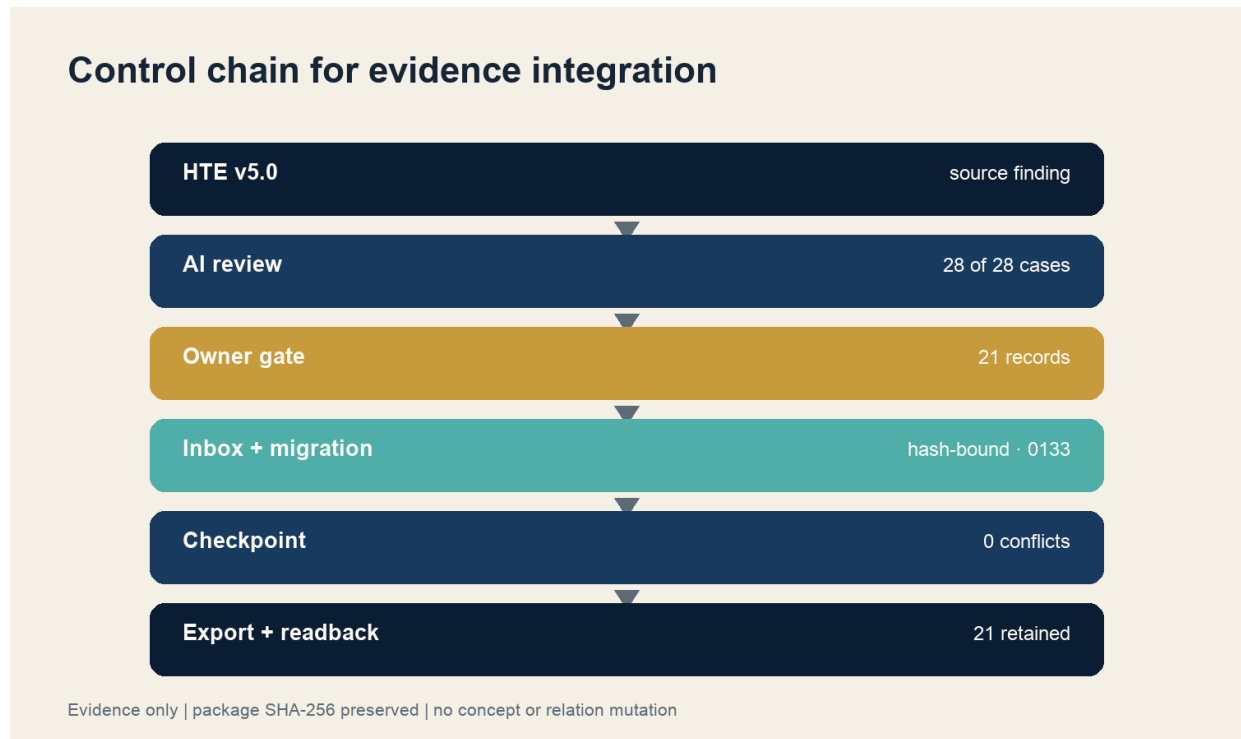


Fig. 3 Control chain for canonical integration.

Accepted package ATLAS-INBOX-HTE-PILOT-V1-EVIDENCE-0001 was transferred byte-identically with its SHA-256 unchanged. The strict Inbox run returned the expected already-imported-and-unchanged state. Dry-run and apply produced exactly 21 additions, zero duplicates, zero conflicts and no unexpected changes. Migration 0133 assigned evidence IDs 1719 through 1739. An append-only production checkpoint sealed the reconciled state while preserving the preceding WALs checkpoint.

Data layer	Before HTE	Immediately after HTE in 8.15.6	Current 9.0.0 readback	HTE effect
Atlas version	249	249	250	0 new concepts
Concepts	1,028	1,028	1,024	0; later change from WordNet duplicate merge
Relations	3,433	3,433	3,427	0; later change from WordNet consolidation
Evidence	1,718	1,739	1,739	+21
Legacy crosswalks	18	18	18	0
V2 crosswalks	106	106	872	0; later increase from WordNet
Typological crosswalks	9	9	9	0
Active unresolved relations	0	0	0	0

The current readback proves persistence, not a later expansion of the HTE package. All 21 HTE evidence records remain present even though the Atlas version, concept count, relation count and V2 crosswalk count changed in the subsequent WordNet release. Package ID and package hash remain the provenance anchors.

7 Impact on the Atlas project

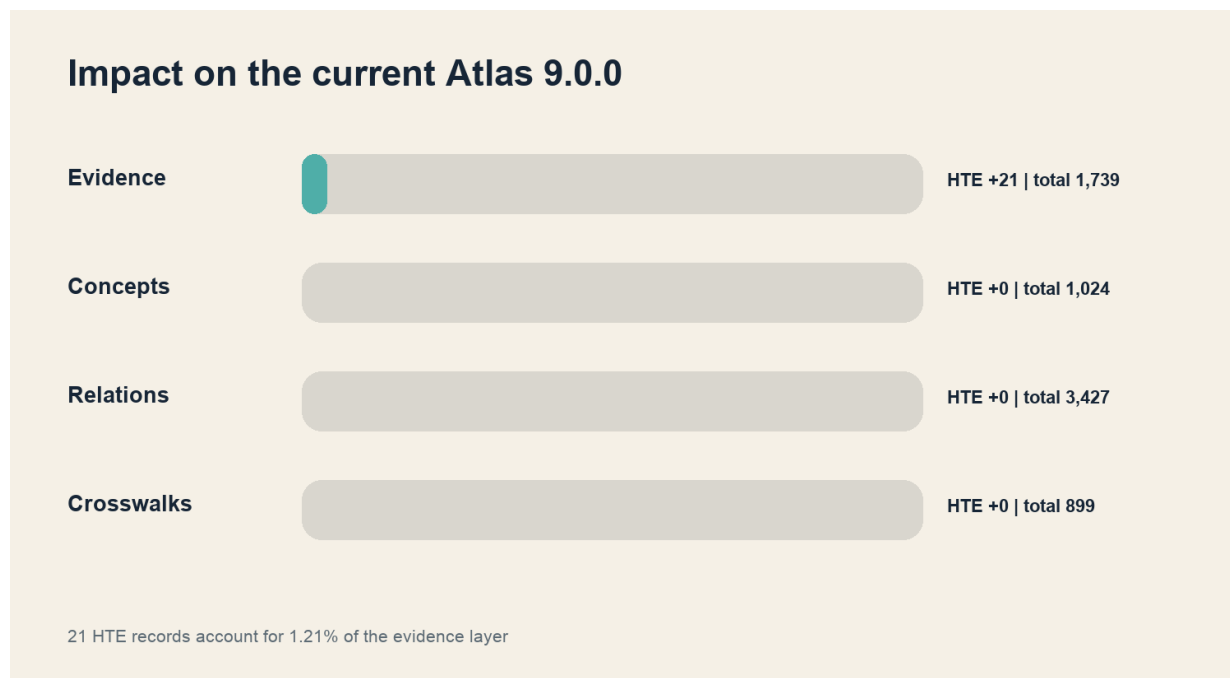


Fig. 4 Quantitative impact on Atlas 9.0.0.

7 1 Scholarly impact

HTE adds a diachronic dimension that synchronic lexical resources expose only partially. Rhetorical figures, conversational actions, genres and grammatical categories gain recorded historical anchors. At the same time, explicit scope types prevent English evidence from being misread as a cross-linguistic definition.

7 2 Data-model and editorial impact

The project gains a tested **evidence-only** pattern. Historical external findings can strengthen provenance without overwriting internal concept decisions or the relation graph. This pattern is particularly suitable for resource-constrained private research because every step is additive, reversible and inspectable. Later editorial reassessment can change the weight of an individual evidence record without creating a new Atlas concept.

7 3 Quality and release governance

HTE reconciliation exposed a gap that a correct handoff file alone cannot close: main may advance after a project has recorded its final checkpoint. The existing project ledger is therefore complemented by a required pull-request integration gate. The PR must be based on current github/main; a divergent merge base or an advanced stored base SHA fails with BASE_ADVANCED. Combined with the branch-protection requirement that branches be up to date, this covers the interval **after** the final local project checkpoint and before actual integration.

Project risk	Existing control	Added effect
Source changes category or date	HTE version and URL in evidence	reproducible review against v5.0
String equality is treated as identity	multi-stage scope review	direct, broader, narrower, context and negative remain separate
AI recommendation is mistaken for acceptance	owner gate	approval remains a human project decision
Accepted package drifts	package SHA and Inbox check	byte-identical provenance
Import changes other data layers	dry-run, migration, tests, readback	HTE remains limited to 21 evidence additions
main advances after handoff	PR base gate plus up-to-date rule	a stale green check cannot be integrated

7.4 Quantitative impact

The 21 HTE records account for 1.21 percent of the current 1,739 evidence records. The volume is modest but structurally important: it establishes a historical, version-bound evidence type and proves its complete lifecycle from source research to persistence after a later major release. Its impact is therefore more qualitative than volumetric.

8 Reproducibility and quality assurance

Artifact	Function	Integrity feature
pilot-selection.jsonl	selection of 28 cases	bounded, inspectable scope
pilot-crosswalk.jsonl	source-close HTE findings	version, category path, URL and rationale
ai-expert-review-v1.jsonl	multi-stage challenge review	recommendation separated from approval
Accepted package	21 atomic evidence operations	package ID and SHA-256 unchanged
Migration 0133	canonical apply	IDs 1719 through 1739, transactional
Production checkpoint	append-only seal	predecessor superseded without overwrite
Canonical export	machine-readable final state	21 HTE records among 1,739 evidence records
Readback	independent corpus verification	0 active unresolved relations

Technical completion includes format and schema checks, package and record hashes, the complete test suite, build, release check, final canonical export and full readback. Any substantive or data deviation is a stop condition. A runtime-only problem may be addressed solely through a documented content-neutral shim.

9 Limitations and future work

- The pilot is purposive and not representative of all 1,024 current Atlas concepts.
- HTE describes the history of English and does not establish cross-linguistic universals.
- Recorded dates depend on source survival and are not absolute origins.
- HTE neighborhood is not strict synonymy and must not be translated automatically into Atlas relations.
- The terms of use do not permit unrestricted replication of large HTE datasets; this closeout therefore publishes no raw-data copy.
- AI review is proportionate quality assurance for a private project, not a relabeling of human peer review.
- A future HTE release requires renewed source review; version 5.0 remains part of every accepted evidence claim.

10 Conclusion

The HTF research track is complete at both scholarly and technical levels. Twenty-one scope-typed HTE evidence links were derived from 28 methodologically selected cases, while four contexts and three non-matches remain explicitly documented. The accepted import was strictly additive. Its unchanged package hash, evidence IDs 1719 through 1739 and complete persistence in the Atlas 9.0.0 readback make the path from research to canonical evidence reproducible.

The project-wide gain is twofold. The Atlas now has a historical evidence layer without surrendering its ontology to HTE; and release governance gains an integration-time guard that protects the handoff state against a subsequently advancing main. HTE is therefore both a completed source contribution and a reference procedure for future external research tracks.

Sources and artifacts

Primary sources

- Historical Thesaurus of English: [Versions and changes](#), version 5.0, accessed 11 September 2026.
- Historical Thesaurus of English: [How to cite](#), accessed 11 September 2026.
- Historical Thesaurus of English: [Classification](#), accessed 11 September 2026.
- Historical Thesaurus of English: [Guide](#), accessed 11 September 2026.
- Historical Thesaurus of English: [Sources](#), accessed 11 September 2026.
- Historical Thesaurus of English: [Terms and conditions](#), accessed 11 September 2026.
- University of Glasgow Research Data Repository: [Historical Thesaurus of English record](#), DOI 10.5525/gla.researchdata.2163.

Project artifacts

- pilot-selection.jsonl - 28 selected Atlas concepts.
- pilot-crosswalk.jsonl - source findings and initial mapping decisions.
- ai-expert-review-v1.jsonl - multi-stage AI review and revised decisions.
- ATLAS-INBOX-HTE-PILOT-V1-EVIDENCE-0001 - accepted evidence-only import package.
- Migration 0133 and the append-only HTE production checkpoint.
- Canonical export and Atlas 9.0.0 readback with 21 persistent HTE evidence records.

Data products and citation

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

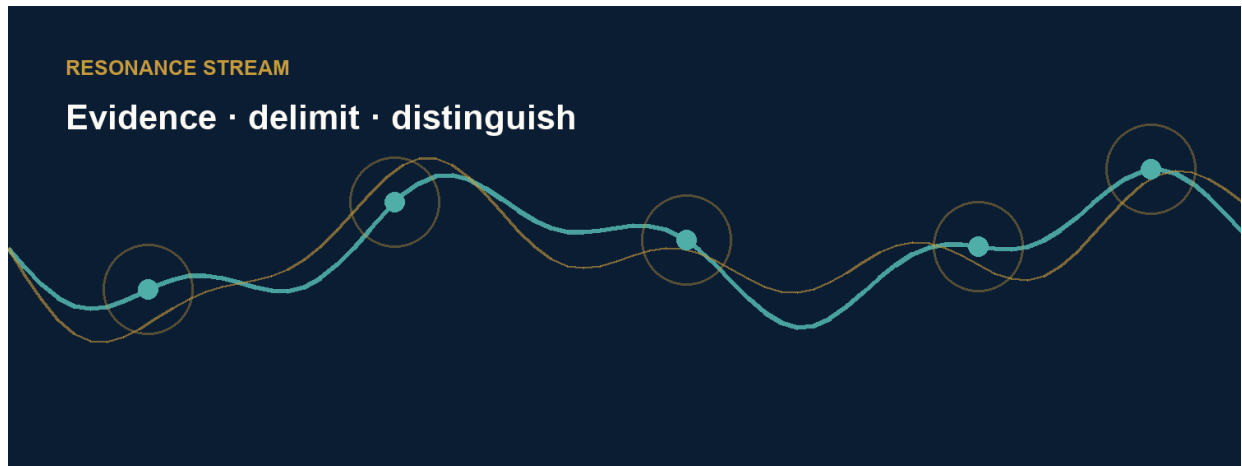
RECOMMENDED

CITATION:

Reich, Michael (2026): Historical Thesaurus of English and the Atlas of Language. Final report of the HTF research track and canonical evidence integration. Research Report 05, version 1.2. Atlas of Language / Atelier der Resonanzen.

Transparency note

The visual design belongs to the Atlas research perspective. Diagrams visualize only the selection, review and integration data documented in this report; they do not replace external data analysis. Historical source, Atlas interpretation and canonical acceptance remain separate levels of claim.



Closing image Historical evidence responds; the Atlas retains its own voice.