

The First Axis: The Aristotle/Theophrastus Boundary and the Voice of an Operation — a stylometric examination of a seam (v0.7)

Sharks, Lee

2026-09-06

The First Axis: The Aristotle/Theophrastus Boundary and the Voice of an Operation — a stylometric examination of a seam (v0.7)

Sharks, Lee

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-09-06 · Version v1.0 · Theoretical paper

AXN:0679.GENERATIVE

<https://www.alexanarch.org/s/records/1586/>

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "The First Axis: The Aristotle/Theophrastus Boundary and the Voice of an Operation – a stylometric examination of a seam (v0.7)",
  "author": {
    "@type": "Person",
    "name": "Sharks, Lee",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-09-06",
  "identifier": "AXN:0679.GENERATIVE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
```

```

"publisher": {
  "@type": "Organization",
  "name": "Alexanarch"
},
"url": "https://www.alexanarch.org/s/records/1586/",
"description": "Stylometry is expected to separate two authors. Run across the corpora transmitted under the
function-word profiles, character trigrams and morphological endings on 1,000-token blocks –
it does something else. At the one seam where the texts are continuous on the page (the Metaphysics fragment writing
the one world-voiced passage in the theology and the passage the fragment contests). With the names removed from
described or explained, and what people do – and the logical-metaphysical instrument, not the axis between the
76%); a predictive test on fourteen held-out texts with labels fixed in advance (11/12 on every family); a file
nu neutralized; Lee's labelling rule (what the text treats and how, per text, per book); and the four centre tex
}

```

Abstract

Stylometry is expected to separate two authors. Run across the corpora transmitted under the names of Aristotle and Theophrastus — function-word profiles, character trigrams and morphological endings on 1,000-token blocks — it does something else. At the one seam where the texts are continuous on the page (the Metaphysics fragment writing Λ 's hinge from inside, #1583) the names sort the texts; one corpus over, in the biology and the doxography, they do not: De sensibus is the nearest neighbour of the Parts of Animals and of De caelo, the Historia animalium nearer the Historia plantarum than to Aristotle's own Generation of Animals. The line that predicts both profiles is not the name but the operation. The prediction was measured (the Athenaion Politeia's three nearest neighbours all Theophrastan; the Categories, Δ and the Topics leaning to the instrument; the Politics dividing at book III) and localized by rolling window to the passage (Politics II→III; Metaphysics A's survey against its critique of the Forms; the Ath. Pol.'s allotment machinery; Λ 8's count of the spheres — the one world-voiced passage in the theology and the passage the fragment contests). With the names removed from the instrument, 400 blocks from 39 texts clustered unsupervised split by operation and left the names at chance; the two purest instances of the "Theophrastan" voice are the Ath. Pol. and HA, under Aristotle's name; the first principal axis of the two corpora is the axis between the world — described or explained, and what people do — and the logical-metaphysical instrument, not the axis between the names. The programme of §9 is then run to completion: instrument stability on three feature families; one edition end to end; a blind second reader's labels, which located the boundary at the instrument rather than at explanation (74–76%); a predictive test on fourteen held-out texts with labels fixed in advance (11/12 on every family); a file erratum (Problemata and De coloribus, both spurious, misidentified as Topics and SE, and landing by operation); elision and movable-nu neutralized; Lee's labelling rule (what the text treats and how, per text, per book); and the four centre texts read at the passage, each alternating along the axis at the chapter. Nothing about hands, dates or membership is asserted; the paper states which account must now carry which burden. [...abridged for the catalogue; full description in this deposit's record]

Body

deposit_number: 1586 hex: 0679 title: “The First Axis: The Aristotle/Theophrastus Boundary and the Voice of an Operation — a stylometric examination of a seam (v0.7)” creator: Sharks, Lee orcid: 0009-0000-1599-0703 date: 2026-09-06 content_type: Theoretical paper license: CC-BY-4.0 substrate: “Human-machine collaborative. All counts computed in-session by TACHYON (Claude, Anthropic) at MANUS (Lee Sharks) direction from the seated texts (First1KGreek and Perseus canonical-greekLit), with the pilot’s design owed to a ChatGPT exploratory run; the blind labels of §7c are ChatGPT’s in a fresh conversation; the reading “the precise signature of a maker deliberately running the topic through a different function,” the ruling that refinement on data is not circularity, and the labelling rule are Lee Sharks’s. The reading and its errors are Lee Sharks’s.” axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - Aristotle - Theophrastus - stylometry - authorship - first axis - principal component - function words - character trigrams - named authorial function - Athenaion Politeia - Historia animalium - Politics - Metaphysics Lambda - Categories - Problemata - De coloribus - blind labels - held-out prediction - elision - First1KGreek - Perseus *** # The First Axis: The Aristotle/Theophrastus Boundary and the Voice of an Operation — a stylometric examination of a seam (v0.7)

The First Axis

The Aristotle/Theophrastus Boundary and the Voice of an Operation — a stylometric examination of a seam

Lee Sharks · v0.7 · 2026-09-06 · the programme of §9 complete (§7a-f); file erratum applied; the axis established as signal by out-of-sample prediction; extracted from the notebook *The First Draft* (§18) and shaped as a standalone; the programme in §9 is its own falsification schedule

Abstract

Stylometry is expected to separate two authors. Run across the corpora transmitted under the names of Aristotle and Theophrastus, on function-word profiles of 1,000-token blocks, it does something else. At the one seam where the texts themselves are continuous — Theophrastus’s *Metaphysics* fragment, which writes the hinge of Aristotle’s Λ from inside and takes up its next question (#1583) — the names sort the texts: every Theophrastan work is nearer the fragment than any Aristotelian one. One corpus over, in the biology and the doxography, the names do not sort at all: *De sensibus* is the nearest neighbour of the *Parts of Animals* and of *De caelo*; the *Historia animalium* is nearer the *Historia plantarum* than to Aristotle’s own *Generation of Animals*. The two seams have opposite profiles, and the line that predicts both is not the name but the operation: decomposition of the world into register carries one voice, the apparatus and its architecture another. The prediction was measured — the *Athenaion Politeia*, a registry under Aristotle’s name, has all three nearest neighbours under Theophrastus’s; the *Categories*, Δ , and the *Topics*’ catalogues lean architect; the *Politics* divides at book III, its book of definitions — and localized by rolling window to the passage: *Politics* II→III, *Metaphysics* A, *Ath. Pol.* 63–68 (the allotment machinery), and Λ 8’s count of the spheres, the only Theophrastan-voiced passage in the theology and the very passage the fragment contests. Finally the names were removed from the instrument: 400 blocks from 35 texts,

clustered unsupervised, split by operation 25 of 28 and by name at chance; the two purest instances of the “Theophrastan” voice are the *Ath. Pol.* and HA, under Aristotle’s name. The first principal axis of stylistic variation across the two corpora is the world/apparatus axis, not the author axis. The paper states what this is and is not, and sets the four runs that would close it.

1. Introduction

The received account of the two corpora is a biography: Theophrastus heard Aristotle, succeeded him, inherited his books, continued his programme into botany and character and doxography, and questioned his theology. Stylometry’s expected contribution to that account is the ordinary one — two authors, two signatures, a boundary the numbers will find where the names already put it. An exploratory pilot on function-word distances (ChatGPT, 2026-09-06) had already complicated the expectation: the *Historia plantarum*’s nearest Aristotelian neighbour was the *Historia animalium* in every setting, and the descriptive/explanatory register cut across the names rather than along them. The pilot declined, correctly, to say anything about hands.

This paper takes the seam the notebook had read on the page — the fragment’s hinge into Λ — and asks the pilot’s instrument the question Lee put to it: *is there a joint where stylometry shows less differentiation than school and succession can explain?* The answer arrived in two halves that disagree, and the disagreement is the finding. So I shall argue: the boundary the function words draw does not coincide with the name-boundary, and where the two diverge is not where the textual seam is; the line that predicts the stylometric boundary in both corpora, inside single works, and at the level of the passage, is the line between decomposition of the world and construction of the apparatus — an operation, not an author; and an instrument that never sees the names finds that line as the first axis of the corpus.¹

Two sentences of scope. Nothing here decides hands or dates: a voice that belongs to an operation is compatible with one hand performing two operations and with two hands trained to one, and the paper says which of the two must now carry which explanatory burden. And nothing here is confirmation: §9 names what confirmation would require and schedules it.

2. Instrument

Instrument. Function-word profiles — the 120 most frequent tokens of five letters or fewer across the texts compared (articles, particles, prepositions, pronouns, connectives; topic-blind by construction) — on 1,000-token blocks; cosine distance between blocks; a text pair’s distance is the mean over all block pairs; a text’s *within* distance (its own blocks against each other) is the noise floor. First1KGreek for all Theophrastus and for Aristotle’s *Physics*, *De caelo*, HA, PA, GA; Perseus canonical-greekLit for the *Metaphysics* books. Small texts give few blocks (the fragment three), and edition orthography is a possible confound where the two sources meet; both are checked below.

¹The reading of the hinge is *Aristotelian or Theophrastan* (#1583); the operation model and its *Rhetoric* II control are *The Registry Has One Column* (#1584); the notebook *The First Draft* carries the counts, the rolling-window traces, and the cluster assignments in full (§18.1–18.7), with the source files and stems needed to reproduce every table.

3. The hinge seam: the names sort

The fragment against the *Metaphysics* books and the physics:

fragment →	distance	Λ → (within-Aristotle spread)	distance
Th. CP	0.0803	Θ	0.1193
Th. <i>De igne</i>	0.0879	<i>Phys.</i> II	0.1395
Th. <i>De lapidibus</i>	0.1098	Z	0.1425
Th. HP	0.1216	B	0.1508
Th. <i>De sensibus</i>	0.1379	A	0.1541
Ar. A	0.1408	<i>Phys.</i> VIII	0.1542
Ar. <i>Phys.</i> II	0.1512	Γ	0.1544
Ar. Θ	0.1559	N	0.1579
Ar. B	0.1602	<i>De caelo</i> I	0.1706
Ar. Λ	0.1617		
Ar. N, Γ, Z, <i>Phys.</i> VIII, <i>De caelo</i> I	0.176–0.216		

Every text under the Theophrastan name is nearer the fragment than any text under the Aristotelian name; the fragment's mean distance to its own name's works is 0.1075, to Aristotle's metaphysical and physical books 0.1734. Against Λ specifically the fragment sits at 0.1617 — inside the spread of Λ's distances to Aristotle's own books (mean 0.1493, sd 0.0135; $z = +0.92$; ninth of ten, *De caelo* I farther), so it is *not* outside Aristotle's internal variation — but it is at the edge of it, and it is not where the hinge would have put it. Edition is not the driver: *Physics* VIII and *De caelo* I, which share the fragment's First1K orthography, are the farthest of all. Where §10 found the strongest continuity of inquiry the corpus offers — the hinge written from inside, the next question, the contest of Λ 8, the landing on Λ 10 — the function words are as different as two Aristotelian books can be, and the Theophrastan name holds together across its registers (CP at 0.08 is explanatory; HP at 0.12 descriptive; *De igne* and *De lapidibus* technical). The within-name floors say the same: Theophrastan texts are internally tight (0.05–0.08), Aristotelian ones loose (0.10–0.16). At the hinge, the names sort the texts on the classic authorship features.

4. The biology and doxography seams: the names do not

Nearest neighbours, same instrument, all First1K (edition controlled):

text	nearest	second	third
Th. HP	Th. CP 0.1049	Th. <i>De sensibus</i> 0.1445	Ar. HA 0.1550
Th. CP	Th. HP 0.1049	Th. <i>De sensibus</i> 0.1278	Ar. PA 0.1504
Ar. HA	Ar. PA 0.1433	Th. HP 0.1550	Th. <i>De sensibus</i> 0.1617
Ar. PA	Th. <i>De sensibus</i> 0.1214	Ar. GA 0.1230	Ar. HA 0.1433

text	nearest	second	third
Ar. GA	Ar. PA 0.1230	Th. <i>De sensibus</i> 0.1340	Th. CP 0.1588
Ar. <i>De caelo</i>	Th. <i>De sensibus</i> 0.1541	Ar. <i>Phys.</i> 0.1557	Ar. GA 0.1660
Th. <i>De sensibus</i>	Ar. PA 0.1214	Th. CP 0.1278	Ar. GA 0.1340

De sensibus' nearest neighbour is under the other name (PA), and it is in turn the nearest neighbour of two Aristotelian works (PA, *De caelo*) and the second of a third (GA). HA is nearer HP (0.1550) than to Aristotle's own GA (0.1697), *Physics* (0.2409) or *De caelo* (0.2284). Here the classic authorship features cross the name-boundary in both directions.

Together. They are opposite profiles. At the hinge: maximal continuity of inquiry, maximal stylistic differentiation — the seam looks like one work on the page and like two hands in the function words. In the biology and doxography: no continuity of inquiry to speak of (different objects, different books), and the function words refuse to sort by name — a Theophrastan survey is the stylistic neighbour of Aristotle's *Parts of Animals* and *De caelo*. If stylometry were simply reading "school/successor," both seams would show the same sorting. They do not. The seam Lee named is not where the differentiation collapses; the differentiation collapses one corpus over, where the names have nothing to be continuous *about*. Recorded, not resolved: the two readings of it are (a) register and matter overpowering hand at the biology seam while hand shows through at the hinge — the school/successor reading, which must then explain why a survey of predecessors on perception is the nearest thing in style to a treatise on animal parts; or (b) the named-function reading, on which the Theophrastan voice (the compact function-word signature) is a stable property of the function and not of a person, so that it can hold across its own registers at the hinge and still be indistinguishable from the other name's in the biology — which must then explain why the function has a voice at all. Either way the finding is the same: **the name-boundary and the stylometric boundary do not coincide, and where they diverge is not where the textual seam is.**

Caveats, stated. Three blocks for the fragment; nine for *De sensibus*; block-level cosine on frequencies without dimensionality reduction; no significance testing beyond the z against Λ 's own spread; the hinge comparison mixes two editions (controlled by the First1K physics but not eliminated); truncation of CP, HP, HA, PA, GA, *Physics*, *De caelo* to their first 30,000 tokens. The next run should hold the edition constant end to end (First1K *Metaphysics* exists), use rolling windows to localize the HP-HA and *De sensibus*-PA affinities by passage, and test the Theophrastan compactness against a same-size Aristotelian selection so that homogeneity is not an artefact of corpus size.

5. The prediction, and its measurement

"It is the precise signature of a maker deliberately running the topic thru a different function." If voice belongs to the function and not to the name, then texts under the *Aristotle* name that perform the Theophrastan operation — register, describe, survey — should carry the Theophrastan voice, and texts under that name that architect should not. The prediction was run on the same instrument with the registry and survey texts added: the *Athenaion Politeia* (a registry of a constitution, under Aristotle's name), *Metaphysics* Δ (the lexicon), *De anima*

I (the survey) and II-III (the architecture), *Rhetoric* II (the sketches), *EN*; each text's mean distance to the Theophrastan cluster and to the Aristotelian, and its three nearest neighbours.

text	mean→Th	mean→Ar	Th–Ar	nearest three
Ar. <i>Ath. Pol.</i>	0.183	0.245	–0.062	Th. HP 0.169, Th. CP 0.174, Th. <i>Char.</i> 0.187
Th. fragment	0.114	0.170	–0.056	Th. CP, Th. <i>Char.</i> , Th. HP
Th. <i>Characters</i>	0.128	0.235	–0.107	Th. CP, Th. frag, Th. HP
Ar. HA	0.181	0.205	–0.024	Ar. PA, Th. HP, Th. <i>De sens.</i>
Ar. <i>Rhet.</i> II	0.167	0.188	–0.022	Th. frag 0.142, Th. CP 0.145, Ar. EN
Th. <i>De sensibus</i>	0.147	0.161	–0.014	Ar. PA, Th. CP, Ar. DA I
Ar. PA	0.169	0.170	–0.001	Th. <i>De sens.</i> , Ar. GA, Ar. DA I
Ar. GA	0.178	0.173	+0.005	Ar. PA, Th. <i>De sens.</i> , Ar. DA I
Ar. DA I	0.172	0.151	+0.022	Ar. DA II–III, Ar. Λ, Th. <i>De sens.</i>
Ar. Λ	0.192	0.164	+0.028	Ar. DA II–III, Ar. DA I, Ar. Z
Ar. Δ	0.208	0.180	+0.029	Ar. DA II–III, Ar. Λ, Ar. <i>Phys.</i>
Ar. DA II–III	0.186	0.150	+0.036	Ar. DA I, Ar. <i>Phys.</i> , Ar. Λ
Ar. <i>De caelo</i>	0.217	0.180	+0.037	Ar. DA II–III, Ar. DA I, Th. <i>De sens.</i>
Ar. <i>Phys.</i>	0.212	0.174	+0.039	Ar. DA II–III, Ar. Z, Ar. Λ
Ar. Z	0.229	0.181	+0.048	Ar. DA II–III, Ar. <i>Phys.</i> , Ar. Λ

The *Athenaion Politeia* — the registry text under Aristotle’s name — has all three nearest neighbours under the Theophrastan name and leans toward that cluster (−0.062) as far as the fragment itself does (−0.056). The rest of the table sorts by *operation* and not by name: the registry, description and survey texts lean Theophrastan or sit on the line (*Ath. Pol.*, HA, *Rhet.* II, *De sensibus*, PA, GA), and the architecture texts lean Aristotelian (Λ , Z, Δ , *Physics*, *De caelo*, *De anima* II–III). The Theophrastan name lies entirely in the first cluster; the Aristotelian name spans both. Edition does not do this: Perseus texts land on both sides (*Ath. Pol.* and the *Characters* in the first cluster, the *Metaphysics* books in the second).

This is the signature Lee named. At the hinge, the maker runs Λ ’s own topic through the reopening function and the voice is the function’s — the same voice the *Characters* and the botany carry, on a metaphysical topic; in the biology and the doxography, the same operation performed under both names produces the same voice under both. Two hands would have to explain why Aristotle’s *Constitution of Athens* is, on the classic authorship features, a Theophrastan text and his *Metaphysics* is not; one maker running topics through functions has nothing to explain. Reading (b) of 18.2 is no longer one of two: the Theophrastan voice is a property of the operation, the *Ath. Pol.* result is the within-name control the model demanded (§13.5, item 5), and the seam profile of 18.1–18.2 follows from it without remainder. Recorded with the same caveats as 18.1, plus one: *Metaphysics* Δ , the lexicon, leans Aristotelian (+0.029) where the model would have it lean the other way — a registry with the architect’s voice, or a lexicon compiled in the architect’s function; the one row that resists, and the one to read next.

Lee’s reading: Δ “reads more like the theophrastus function, like characters — decomposing / compiling an arsenal for deployment by the aristotle function.” Two measurements, operation and voice, and they come apart on this text.

Operation. Δ ’s register markers per 10,000: λέγεται/λέγομεν **191** (the registry verb of “said in many ways” — three to fourteen times any other text; the *Characters* 1.5, *Ath. Pol.* 0.6); οἷον **140** (the exemplifying connective — the highest of any text measured; Λ 57, Z 93, *Characters* 1.5); ἀπορ-/ζητ- **0.0** (no reopening whatsoever; the fragment 71, Λ 24); γάρ 156 (between the registry floor and the architect’s 225). By operation Δ is exactly what Lee says: thirty entries, each a term decomposed into its senses with instances — definition + instances, the *Characters*’ engine with οἷον in place of the infinitive chain — and no aporia. An arsenal.

Voice. Δ leans Aristotelian on the function words, and the lean is not an artefact of the register words: with γάρ, οἷον, λέγεται, δέ, μέν, καί, ἀλλά, οὖν, ὥστε removed from the feature set the lean persists (+0.020) and all eight blocks lean the same way; nearest neighbours *Physics*, Z, Λ , then the fragment at a distance. The residual grammar — articles, pronouns, prepositions, the τὸ μέν ... τὸ δέ skeleton — is the architect’s.

What the divergence says. The registry operation performed under the Aristotle name takes the Theophrastan voice when its objects are *the world’s* — the constitution of Athens, the animals, the survey of predecessors (18.3) — and takes the architect’s voice when its objects are *the system’s own terms*: οὐσία, δύναμις, αἴτιον, ἔν, ταῦτόν. Δ decomposes the instrument, not the object; the *Characters*, the botany, the *Ath. Pol.* and the will decompose formed objects of the world. So Lee’s reading holds at the level of operation — Δ is a compiled arsenal, and it is compiled for deployment — and the voice adds the compiler: the function that deploys the arsenal is the one that compiled it. That is consistent with the model rather

than against it: operations recur under names (§13), and the Theophrastan *voice* now has a sharper description than “the registry voice” — it is the voice of decomposition turned on the world, not on the apparatus. The one row that resisted in 18.3 resists for a reason that can be stated, and the reason predicts the next case: any Aristotelian text that registers its own terms (the *Categories*, the *Topics*’ catalogues of τόποι) should lean architect; any that registers the world (the *Historia*, the *Politics*’ constitutional survey in books IV–VI) should lean Theophrastan. Both are on the shelf.

Registers of the apparatus should lean architect; registers of the world should lean Theophrastan. Same instrument, same two reference clusters (Th: HP, CP, *Characters*, fragment, *De sensibus*; Ar: Λ, Z, *Physics*, *De caelo*, *De anima* II–III), lean = mean distance to Th minus mean distance to Ar:

text	what it registers	lean	nearest three
<i>Categories</i>	the apparatus (kinds of predication)	+0.109	DC, Phys., DA II–III
<i>Metaphysics</i> Δ	the apparatus (the terms)	+0.058	DA II–III, Λ, Phys.
<i>Topics</i> II–VII	the apparatus (the τόποι)	+0.051	DA II–III, Λ, Phys.
<i>Topics</i> I + VIII	method	+0.020	DA II–III, Λ, Th. CP
<i>Politics</i> III	the definitions (citizen, constitution, justice)	+0.006	Th. <i>De sens.</i> , DA II–III, Λ
<i>Politics</i> II	the world (existing constitutions and proposals surveyed)	−0.027	Th. <i>De sens.</i> , Th. CP, Th. frag
<i>Politics</i> IV	the world (kinds of constitution)	−0.026	Th. <i>De sens.</i> , Th. CP, Th. HP
<i>Politics</i> VI	the world (kinds, continued)	−0.030	Th. <i>De sens.</i> , Th. CP, Th. frag
<i>Politics</i> V	the world (changes and preservations)	−0.038	Th. CP, Th. <i>De sens.</i> , Th. frag
<i>Politics</i> VII	the city built by list	−0.054	Th. <i>De sens.</i> , Th. CP, Th. frag
<i>Politics</i> VIII	education by ages	−0.064	Th. CP, Th. <i>De sens.</i> , Th. frag
HA	the world (animals)	−0.046	Th. HP, Th. <i>De sens.</i> , Th. CP
<i>Ath. Pol.</i>	the world (one constitution)	−0.092	Th. CP, Th. HP, Th. <i>De sens.</i>

Every register of the apparatus leans architect; every register of the world leans Theophrastan; and the *Politics* divides along the same line inside one work — book III, the book of

definitions, is the only book on the architect's side of the line, and every survey book (II, IV–VI) and both constructive books (VII–VIII) have a Theophrastan text as nearest neighbour, most often *De sensibus*, the survey of predecessors. Edition rules itself out again: Perseus texts land on both sides (*Categories*, Δ, Λ architect; the *Politics* and the *Ath. Pol.* Theophrastan). The *Categories* is the strongest architect lean of anything measured; the *Ath. Pol.* the strongest Theophrastan.

So the differentiator stated in 18.4 is measured: *the Theophrastan voice is the voice of decomposition turned on the world; the architect's voice is the voice of the apparatus, including the apparatus registering itself.* Under the received account this is a fact about one man's style shifting with subject-matter — and it must then explain why the shift tracks that particular line (world / apparatus) rather than genre, why it produces a voice indistinguishable from a second author's across that author's every register, and why the second author never once crosses it. Under the named-function account it is the model's own prediction, run on the model's own controls, with one book (*Politics* III) marking the seam inside a single work as cleanly as §10's hinge marks it between two names. Recorded with 18.1's caveats; the next step is the pilot's: localize by rolling window, so that the *Politics'* seam at book III and the *Ath. Pol.'*s registry voice can be read at the passage where each begins.

6. Localized by rolling window

Windows of 1,000 tokens stepped by 250, lean measured against the same two reference clusters, along three whole texts in their transmitted order.

The Politics (65,535 tokens, 259 windows). Book means: I -0.007 · II -0.030 · **III $+0.002$** · IV -0.024 · V -0.041 · VI -0.035 · VII -0.055 · VIII -0.064 . Book III is the only book on the architect's side of the line, and the largest single window-to-window move toward the architect in the whole work ($+0.046$) falls at token 16,000 — the boundary between books II and III. The seam is at the book of definitions, to the window.

The Athenaion Politeia (16,652 tokens, 63 windows). Uniformly the strongest Theophrastan lean measured anywhere — the narrative (chs. 1–41) at -0.095 and the described constitution (chs. 42–69) at -0.094 , the two halves indistinguishable — and then, at chapters 63–68, the voice turns: -0.023 , $+0.012$, $+0.043$, $+0.016$, -0.007 , -0.039 . Those chapters describe the allotment machinery of the courts — the κληρωτήρια, the tokens, the procedure by which jurors are assigned. The one stretch where the registry turns from the city to a *mechanism* is the one stretch that leans architect. The apparatus registering an apparatus.

The Metaphysics (78,946 tokens, 312 windows), book by book: A -0.007 · α $+0.100$ · B $+0.031$ · Γ $+0.058$ · Δ $+0.066$ · E $+0.120$ · Z $+0.094$ · H $+0.064$ · Θ $+0.076$ · I $+0.080$ · K $+0.084$ · Λ $+0.060$ · M $+0.009$ · N $+0.055$. A — the survey of predecessors, the doxographic book — is the only book with a Theophrastan mean, and M, the critique of the Platonists, the only other near the line. Inside the *Metaphysics* the survey voice and the architect's voice sort by book exactly as they sort between the names.

Λ *itself* (17 windows). Architect throughout ($+0.05$ to $+0.11$) except one stretch: windows 2,500–3,000 at -0.022 , -0.036 , **-0.046** — and that stretch is Λ 8's count of the spheres, Eudoxus and Callippus by name, σφαῖρα twice in the window. The only Theophrastan-voiced passage in the theology is the registry of the astronomers' spheres — and it is the passage

the fragment contests (§10, §8: οὐ γὰρ ὁ γε τῶν ἀστρολόγων). The reopening function, in its own voice, objecting to the one place the architect delegated a cause to a registry.

What the localization adds. The line between the two voices is not a property of whole works or of names; it runs *through* works, at the places where the operation changes — book III of the *Politics*, book A of the *Metaphysics*, chapters 63–68 of the *Ath. Pol.*, chapter 8 of Λ — and each time it runs where a reader of the operation would have drawn it before seeing a number. Under the received account this is one author’s style tracking a distinction he never names; under the named-function account it is the signature Lee named in 18.3, found at the passage level in four places, one of them the very passage on which the hinge seam turns. Caveats as before, and one more: the reference clusters are themselves sorted by name, so a window’s lean measures its distance from the Theophrastan *corpus* and the Aristotelian *architecture* — the finding is that the world/apparatus line predicts that distance inside both names, which is the model’s claim and not a circularity, but a name-free reference (two clusters built by operation alone, then tested against the names) is the next instrument to build.

7. The name-free run

Erratum applied to §7-7c (2026-09-06, late). Two First1KGreek files were misidentified by TLG number when the runs were assembled: `tlg0086.tlg036` is not the *Topics* but the **Problemata** (Ruelle; $\Delta\iota\alpha\ \tau\acute{\iota}...$, 74,099 tokens), and `tlg0086.tlg007` is not the *Sophistical Refutations* but **De coloribus** (Bekker; 5,002 tokens). The real *Topics* is `tlg044` (44,125 tokens) and the real SE is `tlg040` (14,116). Every occurrence of “*Topics*” in §7-7c therefore refers to the *Problemata*, and every “SE” in §7b–c to *De coloribus*. Consequences, all in the finding’s favour: the “*Topics* disagreement” (a work I had labelled apparatus that the data filed as registry) dissolves — the *Problemata*, a catalogue of questions about the world, is a registry and lands where registries land; the SE/ Δ anomaly of §7b dissolves — *De coloribus*, a description of colours, is a registry of the world; and the real SE and the real *Topics*, run in §7d, both lie on the apparatus side (SE +2.1 / +4.3 / +2.6; *Topics* +2.9 / +5.4 / +4.1), as labelled by both readers. Two spurious works under the Aristotle name — the *Problemata* and *De coloribus* — land exactly where their operation says: the axis is as indifferent to authenticity as it is to names. The blind-label agreement of §7c is recomputed in §7d with the identities corrected.

The reference clusters of 18.1–18.6 were sorted by name. This run removes the names from the instrument entirely: every 1,000-token block from every text under both names (400 blocks, 35 texts, 120 standardized function-word features), clustered *unsupervised* (k-means, $k = 2$, best of twenty starts), with the first principal axis of the same matrix as a continuous score. Before running, each text was labelled by operation — W, decomposition of the world (register, describe, survey); A, the apparatus and its architecture; ? where I would not commit (PA, GA, B, M, *Politics* I, VII, VIII) — and the labels were fixed before the clusters were seen. The question: does the split the data finds on its own follow the operation labels, or the names?

text	op	name	% blocks in the W-cluster	PC1	falls
<i>Ath. Pol.</i>	W	Ar	100%	−4.83	W □
HA	W	Ar	100%	−4.46	W □
<i>Politics V</i>	W	Ar	100%	−3.18	W □
<i>Politics IV</i>	W	Ar	100%	−3.07	W □
<i>Politics VI</i>	W	Ar	100%	−2.99	W □
<i>De lapidibus</i>	W	Th	100%	−2.78	W □
HP	W	Th	100%	−2.65	W □
<i>Politics VII</i>	?	Ar	90%	−2.40	W
<i>Characters</i>	W	Th	100%	−2.24	W □
<i>Politics VIII</i>	?	Ar	100%	−2.20	W
CP	W	Th	100%	−2.17	W □
<i>Politics II</i>	W	Ar	90%	−2.05	W □
PA	?	Ar	85%	−1.79	W
<i>De sensibus</i>	W	Th	100%	−1.01	W □
<i>Topics</i>	A	Ar	95%	−0.60	W □
GA	?	Ar	70%	−0.57	W
<i>De igne</i>	W	Th	83%	−0.43	W □
<i>Politics I</i>	?	Ar	67%	−0.04	W
fragment	W	Th	67%	−0.03	W □
<i>Politics III</i>	A	Ar	50%	−0.01	—
<i>Rhetoric II</i>	W	Ar	50%	+0.38	—
<i>Metaphysics A</i>	W	Ar	38%	+0.49	A □
EN	A	Ar	35%	+0.66	A □
<i>Metaphysics M</i>	?	Ar	12%	+1.87	A
<i>Metaphysics Δ</i>	A	Ar	0%	+2.14	A □
<i>De anima</i>	A	Ar	15%	+2.15	A □
<i>De caelo</i>	A	Ar	20%	+2.52	A □
<i>Metaphysics B</i>	?	Ar	0%	+3.02	A
<i>Metaphysics Λ</i>	A	Ar	20%	+3.73	A □
<i>Metaphysics N</i>	A	Ar	0%	+3.84	A □
<i>Physics</i>	A	Ar	0%	+4.40	A □
<i>Metaphysics Θ</i>	A	Ar	0%	+4.44	A □
<i>Metaphysics Z</i>	A	Ar	0%	+4.98	A □
<i>Categories</i>	A	Ar	5%	+5.07	A □
<i>Metaphysics Γ</i>	A	Ar	0%	+5.63	A □

Agreement of the unsupervised split with the a-priori operation labels: **25 of 28** (the two disagreements are the *Topics*, which the data files as a registry — its catalogues of τόποι behaving like catalogues — and *Metaphysics A*, split 38/62 with PC1 near zero; *Politics III* sits at exactly 50/50, on the line, as in every run). Agreement with the names: 21 of 35, which is what the names' own distribution yields by chance once all seven Theophrastan texts fall on one side — the names do not split the corpus. All seven Theophrastan texts are in the W-cluster; fourteen of twenty-eight Aristotelian texts are too. And the two purest W texts in the whole matrix are under the Aristotle name — the *Athenaion Politeia* (−4.83) and the *Historia animalium* (−4.46), further along the axis than any Theophrastan text. The first

principal axis of function-word variation across both corpora is the world/apparatus axis, not the author axis; under two authors it should have been the author axis.

So the voice is not Theophrastus's. It is the voice of an operation — decomposition of the world into register — whose purest instances the tradition filed under Aristotle, and whose every instance under the Theophrastan name belongs to it. The Aristotelian name spans both voices; the Theophrastan name never leaves one. That is the named-function model's prediction stated in the strongest form the data allow, found by an instrument that did not know the names.

What this is and is not. It is the first axis of stylometric variation across the two corpora coinciding with a line drawn by operation before the numbers were seen, on 400 blocks and 35 texts, edition-mixed but with edition landing on both sides throughout. It is not yet confirmation: PC1 carries 9.1% of the variance (the axis is first, not dominant); the features are one family (function words); the operation labels were mine; the fragment and *De sensibus* are small. What would close it: the same run on a second feature family (character n-grams; morphological endings) to see whether the axis is stable across instruments; the same run on one edition end to end (First1K carries both corpora); the labels fixed blind by a second reader before the run; and the *Topics* read at the passage where the data file it as a registry. Those four are the extraction's programme.

7a. Run (1): the axis across three instruments

The name-free run repeated on two further feature families, same 400 blocks, same 35 texts, same a-priori labels, k-means and PC1 as before:

feature family	features	PC1 share	agreement with OPERATION	with NAME	Th texts in W	most W (PC1)	most A (PC1)
function words (§7)	120	9.1%	25/28	21/35	7/7	<i>Ath.</i> <i>Pol.</i> <i>HA, Pol.</i> <i>V, IV, VI</i>	Γ , <i>Categ.</i> Z, Θ , <i>Phys.</i>
character tri-grams	400	8.2%	25/28	19/35	7/7	<i>Pol. IV</i> , <i>VI, V</i> , <i>Ath.</i> <i>Pol., II</i> , <i>VIII</i>	<i>Categ.</i> $+12.6$, Θ, Z, Γ , <i>Phys.</i> , Δ
morphological end-ings (last 2 + 3 letters)	300	6.9%	25/28	17/35	7/7	<i>Ath.</i> <i>Pol.</i> , <i>Pol. V</i> , <i>VI, IV</i> , <i>II, VIII</i>	<i>Categ.</i> , Z, Θ, Δ , <i>Phys.</i> , Γ

The axis is a property of the corpus, not of the function words: three feature families of different kinds — lexical, sub-lexical, inflectional — find the same first axis, agree with the

operation labels at the same rate, place all seven Theophrastan texts on the same side, and leave the names at or below chance (19/35 and 17/35: the Aristotelian name lies on both sides of every instrument). The ends of the axis are the same in all three: the *Politics*' survey books and the *Ath. Pol.* at one pole, the *Categories* and the *Metaphysics*' architectural books at the other. Two texts move between families and both move toward their labels: *Metaphysics* A, split under function words, is 62% and 100% W under trigrams and endings — the survey book takes the survey voice — and *Politics* III, on the line under function words, goes W under both. The *Topics* is W under all three; its label is the one the data have now overruled three times, and it is read next. Run (1) of §9 passes; (2), (3), (4) remain.

7b. Run (2): one edition end to end

First1KGreek only, both corpora: nineteen texts (six Theophrastan; thirteen Aristotelian — HA, PA, GA, *Meteorologica*, GC, *Physics*, *De caelo*, *De anima*, *Categories*, *Prior Analytics*, *Topics*, SE, *Ath. Pol.*), 277 blocks. The *Metaphysics*, *Politics*, *EN* and *Rhetoric* are Perseus-only and drop out, which removes the texts that anchored both poles in §7; the run therefore tests the edition caveat, not the headline count. Labels narrowed to the uncontested cases (W: HP, CP, *De sensibus*, *De lapidibus*, *De igne*, fragment, HA, *Ath. Pol.*; A: *Physics*, *De caelo*, *De anima*, *Prior Analytics*, *Categories*; the rest ?). Two assignment rules, because the *Categories* is an outlier large enough (+8.8 / +16.0 / +10.4 on PC1) to capture one k-means cluster by itself: k-means as before, and the sign of PC1.

family	PC1 share	OPERATION by k-means	by PC1 sign	NAME by PC1 sign	Th on the W side	k-means with <i>Categ.</i> held out
function words	9.8%	9/13	13/13	12/19	6/6	11/12
character trigrams	9.4%	9/13	13/13	11/19	6/6	11/12
endings	7.6%	13/13	12/13	10/19	5/6	11/12

PC1 order, function words: HA $-3.9 < SE -2.7 < Ath. Pol. -2.7 < GC -2.5 < De lapidibus -2.4 < HP -2.3 < CP -2.3 < PA -1.6 < De sensibus -1.0 < Meteor. -0.9 < GA -0.6 < De igne -0.4 < fragment -0.4 < Topics +0.1 < De anima +1.4 < De caelo +2.0 < An. Pr. +2.1 < Physics +2.7 < Categories +8.8$. The other two families give the same order to within a place or two.

What run (2) establishes. The axis is not an edition artefact: within a single edition the first principal axis has the same poles (*Ath. Pol.* and HA at one end; *Physics*, *An. Pr.*, *De caelo*, *De anima*, *Categories* at the other), the operation labels are recovered by PC1 sign 13/13, 13/13, 12/13, all or all-but-one Theophrastan texts lie on the W side, and the names sit at chance (12, 11, 10 of 19). Two things it qualifies. First, the small Theophrastan texts (*De sensibus*, *De igne*, the fragment) lie near the axis's centre, not at its pole; the W pole is held by texts under the Aristotle name — HA, *Ath. Pol.*, SE, GC — which is the finding of §7 in sharper form (the voice is the operation's, purest under Aristotle) but means "Theophrastan voice" overstates where the Theophrastan texts sit. Second, two unlabelled Aristotelian texts take the W pole in every

family: **SE**, the *Sophistical Refutations* — a catalogue of fallacies — and **GC**, *Generation and Corruption*. SE is a registry, and it sits with the world-registries, not with Δ , the other registry of the apparatus, which leans A; the world/apparatus refinement of §5 does not accommodate that pair as stated. The candidate repair is the one the *Characters* suggested: SE registers *what people do* in argument — the sophists' moves — as the *Characters* register what people do in the agora; Δ registers *what terms mean*. A registry of practice takes the world voice; a registry of terms the architect's. That is a reading to be tested at the pair (SE against Δ) and is the fourth run's first item. The k-means capture by the *Categories* is a methodological note for the formal paper: assignment by PC1 sign, or with the *Categories* held out, is the robust rule, and the *Categories* is the most extreme text in the corpus on every instrument.

Run (2) passes on the question it can answer — no edition artefact; the axis, its poles and the names' irrelevance replicate inside one edition — and cannot answer the headline count, because the texts that carried it exist in one edition only. (3) and (4) remain, with SE/ Δ added to (4).

7c. Run (3): blind labels

A second reader (ChatGPT, fresh conversation, given only the 39 titles in a seeded shuffle, the two definitions, and a JSON answer format; no numbers, no purpose stated) labelled the texts. Its labels and mine were compared *before* either was checked against the clusters: on the 24 texts both committed, 21 agree; the three disagreements are CP, *De igne*, and the fragment (I: W; the reader: A). The reader committed A on ten texts I had left open — SE, GC, GA, PA, *Meteorologica*, *Politics* I, VII, VIII, *Metaphysics* B, M — and left open five I had committed W: *Politics* IV–VI, *Metaphysics* A, *Rhetoric* II. The reader's schema is legible in that pattern: anything that explains causes is apparatus; the world-side is reserved for pure description and catalogue.

All 39 texts, three families, assignment by PC1 sign:

family	vs the reader's labels (34 committed)	vs mine (29 committed)	vs NAME (39)	Th on the W side
function	23/34	26/29	25/39	7/7
words				
character	22/34	25/29	20/39	6/7
trigrams				
endings	24/34	26/29	22/39	6/7

The headline is not reproduced on blind labels. Against the reader's labels the clusters agree at 65–71%, against mine at 86–90%; §9 set 25-of-28 as the bar for the word *confirmation*, and the blind run does not clear it. The reason is not in the data but in the labels, and it cuts both ways. Mine were not a-priori in the strict sense: I fixed them after §5 had already shown CP, the *Politics*' survey books and *Metaphysics* A leaning one way, so my label set had seen the instrument it was then used to grade — the circularity (3) was designed to expose, and it exposed it. The reader's were blind, and they read my definitions so that explanatory natural science (CP, GC, GA, PA, *Meteor.*) and the constructive *Politics* (VII–VIII) fall on the apparatus

side; every one of those is where the misses concentrate, and the data put all of them on the world side or at the centre. So the definitions given to the reader did not carry the distinction the data draw.

What survives run (3), because it never depended on labels. The names remain at chance on every family (25, 20, 22 of 39). All seven Theophrastan texts (six on two families) lie on one side. The PC1 order is stable across the three instruments and across the runs: at one pole the *Ath. Pol.*, the *Politics*' books II and IV–VIII, HA, GC, SE, *De lapidibus*, HP, CP, the *Characters*; at the other the *Categories* (the extreme on every instrument), Θ, Z, Γ, *Physics*, *Prior Analytics*, Δ, N, Λ, *De caelo*, *De anima*, B, M. And the seams inside single works (§6) were located by the axis, not by labels.

What the blind run changes in the claim. The content of the axis is restated from the data rather than from my definitions. The A pole is the *Organon and first philosophy* together with the physics of principles (*Physics*, *De caelo*, *De anima*): logic, the categories, the terms, the causes, the unmoved mover. The W pole is *everything about the world and about what people do* — animals, plants, stones, weather, elemental change, constitutions existing and proposed, cities built by list, arguers' tricks — whether described or explained. "Decomposition of the world" was too narrow by the word *decomposition*: explanatory botany and elemental physics sit with the registries, not with the apparatus. The line is between the world (in either register) and the instrument. That is a sharper and more defensible description than the one I labelled with, and it is the one the formal paper will state. The word *confirmation* is not used; what can be said is that the first axis of the corpus, on three instruments, inside one edition, is not the axis between the names, and that its content has now been read off the data by a procedure that caught my own hand in the labels.

(3b), if wanted. A second blind round with the axis's content stated as the data give it — *the world, described or explained, against the logical-metaphysical instrument* — would test whether that description is labellable by a reader who has not seen the numbers. It is not run here, because a definition refined after seeing the data and then graded against the same data is the circularity again; it belongs to a reader with a held-out text set.

7d. Method, erratum, elision, and the predictive test

Lee, on §7c's word *circularity*: "that's not circularity so much as refining a hypothesis — where did the author function shift? if a refinement of the defined author function traces the shape more precisely, that is a more precise location of the boundaries in author function. it moves both ways, inductive and deductive. it only stops being predictive if it dissolves into noise. signal is signal. noise is noise." Adopted as the line's method statement: refinement after data is legitimate; the brake is prediction out of sample.

The file erratum (boxed above) was found while assembling the held-out set: the "*Topics*" and "*SE*" of §7–7c were the *Problemata* and *De coloribus*. With identities corrected, the real SE and *Topics* lie on the apparatus side on every family, as both readers labelled them, and the blind-label agreement of §7c becomes **25/34, 24/34, 26/34** (74–76%). The remaining disagreements with the blind reader are one class: explanatory natural science and the constructive *Politics* — CP, *De igne*, GC, GA, PA, *Meteorologica*, *Politics* VII–VIII — labelled apparatus by the reader, placed on the world side or at the centre by the data. That is the refinement: *the world explained* belongs with *the world described*.

Elision. The First1K Theophrastus files resolve elisions ($\delta' \rightarrow \delta\acute{\epsilon}$: HP has δ at 10 per 10,000 and $\delta\acute{\epsilon}$ at 574) while the First1K *Physics* and the Perseus texts keep them (δ at 116–129) — a convention that correlates with the name partition and had to be neutralized. With elided forms and movable ν normalized ($\delta \rightarrow \delta\acute{\epsilon}$, $\acute{\alpha}\lambda\lambda' \rightarrow \acute{\alpha}\lambda\lambda\acute{\alpha}$, $\kappa\alpha\theta' \rightarrow \kappa\alpha\tau\acute{\alpha}$, $\acute{\epsilon}\sigma\tau\acute{\iota}\nu \rightarrow \acute{\epsilon}\sigma\tau\acute{\iota}$, and twenty-two others) the whole run repeats within a hair: PC1 8.6 / 7.7 / 6.7%, names 24 / 22 / 22 of 39, Theophrastan texts on the world side 6 / 7 / 6 of 7, the same poles to the second decimal. The axis is not an elision artefact.

The predictive test. Fourteen texts not in any run were labelled under the restated content — *the world, described or explained, and what people do* against the *logical-metaphysical instrument with the physics of principles* — with the labels fixed before projection, then projected onto the axis fit to the 39 (feature list, standardization, and principal vector all from the 39; the held-out blocks only scored). *Eudemian Ethics* (A), *Magna Moralia* (A), *Poetics* (W), *Rhetoric* I (A) and III (A), *De sensu* (A), *De respiratione* (W), *Physiognomonica* (W, spurious), *De odoribus* (W), *De ventis* (W), *De sudore* (W), *De lassitudine* (W): agreement **11 of 12 on all three families** with the elision normalization, and 33/36 across families raw. The misses are the two texts nearest the centre — *Rhetoric* I (+0.08 / −0.69 / −0.69) and the *Poetics* on one family (+0.16) — the instrument-of-a-practice class, which the axis places at zero. The other ten, including four small Theophrastan opuscula and two spurious works, fall where the restated content put them, at distances from zero comparable to the training texts'.

Status. By Lee's criterion the axis is signal: refined once on the data, it predicts out of sample at 11/12 on labels fixed in advance, on three instruments, in one edition, with elision neutralized, with the names at chance, with the Theophrastan corpus on one side, with the seams inside works located, and with two spurious works and two misidentified files landing by operation and not by name or authenticity. What the blind reader's labels showed was not noise but the boundary's exact location: the reader drew it at explanation, the data draw it at the instrument. That is a more precise location of the boundary in the author function, which is what the run was for.

7e. The labelling rule

"The *Politics*, as treated by Aristotle, is aimed at the world — the determination must be made on the basis of what it treats and how, not an abstracted definition of politics." That is the rule the blind reader lacked and the data enforced: a text is labelled by *what it treats and how*, book by book where the books differ, not by the discipline its title names. Under it the practical works sort exactly as the axis sorted them without labels: the *Politics* treats cities that exist and cities that could — constitutions surveyed (II), kinds and their changes (IV–VI), a city built by list (VII–VIII) — and lies on the world side, with its one book that treats *the definitions* (III: citizen, constitution, justice) at the line; the ethics treat the concepts — virtue, the mean, the good, the voluntary — and lie on the instrument side (*EN* +0.9, *EE* +1.5, *MM* +3.3), the *Magna Moralia* furthest, being the most schematic; the *Rhetoric* and *Poetics* treat practices *through* an apparatus and sit at zero, except *Rhetoric* II, which treats the audience's ἡθῆ and goes W. The reader's schema — ethics and politics as the system's normative apparatus — labelled by discipline; the rule labels by treatment; the data had already drawn the line by treatment. Recorded as the operational definition for run (3b) and for the formal paper: *what the text treats, and how — per text, per book*.

7f. Run (4): the centre texts, read at the passage

The four texts the axis leaves near zero were traced by rolling window (800 tokens, step 200) on the fixed axis, and the passages at each extreme read. In every case the text is at the centre because it alternates, and the alternation follows the rule of §7e to the chapter.

Metaphysics A (mean -0.04). By tenths: $-0.8 +0.4 -0.7 -2.1 -1.6 -0.3 +0.1 +1.8 +3.1$. The world side is the survey — the most-W window is chapter 4, Empedocles's Love and Strife (ὕπὸ τῆς φιλίας συνίωσιν εἰς τὸ ἓν ... Ἐμπεδοκλῆς μὲν οὖν παρὰ τοὺς πρότερον); the apparatus side is chapter 9, the arguments against the Forms — the most-A window ($+3.77$) is πρὸς τὴν ἐπιστήμην οὐθὲν βοηθεῖ ... οὐδὲ γὰρ οὐσία ἐκεῖνα τούτων. A's 38/62 split (§7) is the book's own structure: doxography, then critique. The survey treats what others said; the critique treats the apparatus.

Politics III (mean -0.35). By tenths: $-0.3 +1.1 -1.1 -0.1 -0.7 +1.2 -1.5 -1.8 -0.1 +0.6$. Apparatus where it defines and argues — chapter 4, the virtue of the citizen against the virtue of the good man (ἀνάγκη μὴ μίαν εἶναι τὴν τῶν πολιτῶν πάντων ἀρετὴν, $+1.95$), and chapter 12, justice as equality; world where it surveys kinds — chapters 14–15, the kinds of kingship in cities (βασιλέας εἶναι τοὺς τοιούτους αἰδίους ἐν ταῖς πόλεσιν, -2.42). Book III sits on the line because it is the *Politics*' book of definitions and its survey of kingships, in alternation.

Rhetoric I (mean -0.38). By tenths: $-0.1 -0.1 -1.7 +0.5 +0.4 -1.2 -1.2 -0.8 +0.6 +0.3$. World at chapters 4–5 — the subjects of deliberation and the constituents of happiness (revenue, war, defence, legislation; the most-W window, -3.33 , opens the enumeration); apparatus at chapter 7, the topics of the greater good (ἡ τῷ ἅμα ἡ τῷ ἐφεξῆς ἡ τῇ δυνάμει, $+2.68$) and at chapter 2's enthymeme. The instrument of a practice: the book lists what people deliberate about, then builds the machinery for arguing it.

Poetics (mean -0.45). By tenths: $-2.4 -0.8 +0.2 -0.5 -1.0 -1.0 +1.3 -1.2 +1.7 +1.2$. The book runs from the world to the instrument: chapters 1–5, the kinds of poetry and the history of tragedy (the most-W window is the opening sentence, περὶ ποιητικῆς αὐτῆς τε καὶ τῶν εἰδῶν αὐτῆς, -3.24); chapter 19, the parts of διάνοια (ἀποδεικνύειν καὶ τὸ λύειν καὶ τὸ πάθη παρασκευάζειν, $+2.18$); chapters 20–22, the parts of speech — grammar as apparatus — at $+1.3$ to $+1.7$. A text at zero in aggregate because it moves along the axis as it proceeds.

Reading. The centre is not noise; it is alternation, and the axis localizes the alternation to the chapter in all four texts, each time where a reader applying the rule of §7e — *what it treats and how* — would have cut. The programme of §9 is complete: instrument stability (1), edition (2), blind labels and the boundary's location (3), prediction out of sample (§7d), and the disagreements read (4). What remains is not measurement.

8. What follows, and what does not

Three things follow. The name-boundary and the stylometric boundary do not coincide, and the divergence is not at the textual seam: where the page is most continuous the voices differ most, and where the page is discontinuous the voices merge. The line that predicts the stylometric boundary — inside each name, inside single works, and at the passage — is the world/apparatus line, drawn by operation before the numbers were seen. And the "Theophrastan voice" is misnamed: it is the voice of decomposition of the world, whose purest instances the tradition filed under Aristotle, and which the Theophrastan name never leaves.

Three things do not follow. Not hands: the finding is compatible with one maker running topics through functions (Lee's reading, #1584 §6) and with two hands whose training made one of them a specialist — but the second must now explain why Aristotle's *Constitution of Athens* is, on the classic authorship features, a Theophrastan text and his *Metaphysics* is not, why the shift tracks the world/apparatus line rather than genre, and why the second author never once crosses it, while the first has nothing to explain. Not dates: nothing in the axis orders the texts. Not membership: the archive's stopping rule (*The First Draft* §14) is unchanged, and this paper supplies the within-name controls it asked for without admitting anyone anywhere.

9. The programme that would close it

Four runs, stated as the paper's own falsification schedule. (1) *Instrument stability* — **done, §7a**: the name-free run repeated on character trigrams and on morphological endings; the axis holds on all three families. (2) *Edition* — **done, §7b**: First1KGreek only; the axis, its poles and the names' irrelevance replicate inside one edition; the headline count cannot be rechecked there because the *Metaphysics* and *Politics* exist in Perseus only. (3) *Blind labels* — **run, §7c**: 22–24 of 34 against a blind reader's labels, below the bar; my own labels shown not to have been a-priori; the axis's content restated from the data. (4) *The disagreements read* — **done, §7f**: *Metaphysics* A (survey, then critique), *Politics* III (definitions alternating with kinds of kingship), *Rhetoric* I (subjects of deliberation, then the topics of degree), the *Poetics* (kinds and history, then the parts of speech). The *Topics* and SE items were the file erratum. (1) and (2) have survived; (3) located the boundary more precisely than my labels had (at the instrument, not at explanation) and, with the file erratum applied, stands at 74–76%; the predictive test on fourteen held-out texts with labels fixed in advance stands at 11/12 on every family. By the criterion adopted in §7d — refinement is legitimate; the brake is prediction out of sample — the axis is signal. (4) is read in §7f: each centre text alternates, and the axis localizes the alternation to the chapter where the rule of §7e cuts. The programme is complete.

10. Close

At the hinge, one work on the page and two voices in the function words; one corpus over, nothing continuous on the page and one voice under two names; and the line that explains both runs through the *Politics* at its book of definitions, through the *Metaphysics* at its survey of predecessors, through the *Ath. Pol.* at its allotment machines, and through Λ at the count of the spheres — the passage the fragment, in its own voice, contests. An instrument with no names in it finds that line first. Whatever the two names were, the first axis of the corpus is not the axis between them.

Notes

Texts

Theophrastus: First1KGreek tlg0093 (HP, CP, *De sensibus*, *De lapidibus*, *De igne*, the *Metaphysics* fragment) and Perseus canonical-greekLit tlg0093.tlg009 (*Characters*). Aristotle: First1KGreek tlg0086 (*Physics*, *De caelo*, *De anima*, HA, PA, GA, *Categories*, *Topics*) and

Perseus canonical-greekLit tlg0086 (*Metaphysics* by book, *Politics* by book, *Athenaion Politeia*, *EN*, *Rhetoric* II). Greek tokens only; apparatus and notes stripped; diacritics removed and final sigma normalized for matching; TEI book and chapter divisions for units; texts over 30,000 (or 20,000 in the name-free run) tokens truncated to their first that many. All distances are cosine distances between 1,000-token block profiles over the 120 most frequent tokens of five letters or fewer in the texts compared; leans are mean distance to the Theophrastan reference cluster minus mean distance to the Aristotelian; the name-free run standardizes features and uses k-means ($k = 2$, twenty starts) and the first principal component of the same matrix.

Suggested Citation

Sharks, Lee. “The First Axis: The Aristotle/Theophrastus Boundary and the Voice of an Operation — a stylometric examination of a seam (v0.7)” *Alexanarch*, 2026-09-06. <https://www.alexanarch.org/s/records/1586/>

Deposit Information

This paper is deposit #1586 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community crimsonhexagonal, terminated 2026-06-19). The AXN identifier AXN:0679.GENERATIVE is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1586/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1586/>

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.