

Transkribus in den Altertumswissenschaften: Potenziale, Herausforderungen und Perspektiven

Doris Vickers BA MA

Empfohlene Zitierweise: Vickers, Doris (2026). Transkribus in den Altertumswissenschaften: Potenziale, Herausforderungen und Perspektiven. UR: Das Journal, 9(1), S. 48-56. DOI: <https://doi.org/10.48646/ur.20230102/ur.20260906>

Lizenziert unter der CC-BY-ND 4.0 International Lizenz.

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung - Keine Bearbeitungen 4.0 International Lizenz zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie <http://creativecommons.org/licenses/by-nd/4.0/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.

1. Einleitung

Die Altertumswissenschaften (verstanden als alle Disziplinen, die sich mit den Kulturen der Antike und ihrer Nachwirkung befassen) stehen traditionell vor besonderen Herausforderungen im Umgang mit ihren Quellen. Ein Großteil der überlieferten Dokumente liegt nicht in modernen Editionen vor, sondern in schwer zugänglichen, häufig beschädigten Originalen, die in komplexen Schriftsystemen verfasst sind. Dazu zählen Inschriften, Papyri, Pergamenthandschriften oder mittelalterliche Abschriften antiker Texte. Diese Quellen sind nicht nur sprachlich anspruchsvoll, sondern erfordern auch paläographische Expertise, da die Schriften oft nur schwer zu entziffern sind.

Die Erschließung solcher Materialien beruhte bislang auf der akribischen Arbeit von Spezialist*innen, die ihre Ergebnisse in kritischen Editionen oder Katalogen veröffentlichten. Dieser Prozess war zeit- und ressourcenintensiv, sodass meist nur ein Bruchteil des überlieferten Materials wissenschaftlich bearbeitet werden konnte. Mit der Digitalisierung großer Bestände und dem Einsatz KI-gestützter Technologien hat sich jedoch ein Paradigmenwechsel vollzogen. Arbeitsprozesse lassen sich nicht nur beschleunigen, sondern auch neue Forschungsfragen eröffnen. Ein zentrales Werkzeug in diesem Zusammenhang ist *Transkribus* (<https://www.transkribus.org/>), eine Plattform zur automatisierten Texterkennung historischer Dokumente.

Der folgende Beitrag untersucht die Funktionsweise von *Transkribus*, seine bisherigen Anwendungen in den Altertumswissenschaften, die Chancen und Grenzen dieser Technologie sowie Perspektiven für die künftige Forschung.

2. Transkribus: Technologische Grundlagen

Transkribus entstand im Rahmen des EU-geförderten Projekts *READ* (Recognition and Enrichment of Archival Documents) und wird heute von der Read-Coop SCE, einer genossenschaftlich organisierten Forschungsinfrastruktur, weiterentwickelt. Die Plattform bietet eine Vielzahl von Funktionen, die speziell auf die Bedürfnisse der historischen Forschung zugeschnitten sind.

Im Zentrum steht die *Handwritten Text Recognition* (HTR), also die automatisierte Erkennung von Handschriften mittels maschinellen Lernens. Hinzu kommt eine *Optical Character Recognition* (OCR) für gedruckte Texte auch aus älteren Drucktraditionen sowie eine Layout-Analyse, die Textbereiche, Spalten, Überschriften oder Randnotizen automatisch identifiziert. Ergänzt wird dieses Spektrum durch Editionswerkzeuge, mit denen Transkriptionen annotiert und in gängige wissenschaftliche Formate (beispielsweise XML/TEI) exportiert werden können.

Der HTR-Prozess erfolgt mehrstufig: Zunächst wird eine sogenannte *Ground Truth* erstellt. Die beinhaltet manuell angefertigte Referenztranskriptionen, die dem System als Trainingsmaterial dienen. In der Regel umfasst dieser Schritt 50–200 Seiten, abhängig von Vielfalt und Schwierigkeit des Materials. Anschließend wird ein neuronales Netz mit diesen Daten trainiert, sodass es Schriftarten, Sprache und Textstruktur der Quelle

erkennt. Das resultierende Modell kann dann auf größere Textmengen angewendet werden und maschinelle Transkriptionen erzeugen, die den Arbeitsaufwand erheblich reduzieren.

Ein entscheidendes Merkmal von *Transkribus* ist die Möglichkeit der fortlaufenden Optimierung: Nutzerinnen und Nutzer überprüfen die automatisch erzeugten Transkriptionen, korrigieren Fehler und speisen die Ergebnisse zurück in den Trainingsprozess. Dieses Active Learning führt dazu, dass die Modelle mit zunehmender Nutzung präziser werden und sich besser an unterschiedliche Schriftarten und historische Kontexte anpassen. Gerade in den Altertumswissenschaften mit ihren unregelmäßigen und schwer lesbaren Schriften ist dieser Mechanismus von besonderer Bedeutung.

3. Anwendungsfelder in den Altertumswissenschaften

3.1 Epigraphik

In der Epigraphik eröffnet *Transkribus* neue Perspektiven. Traditionell werden Inschriften durch Abreibungen, Zeichnungen oder Fotografien dokumentiert. Doch bei verwitterten, beschädigten oder fragmentarischen Texten bleibt die Lesbarkeit eingeschränkt. Digitale Abbildungen können mit *Transkribus* automatisiert transkribiert werden, wodurch nicht nur der Arbeitsaufwand sinkt, sondern auch ganze Inschriftencorpora systematisch und vergleichbar erschlossen werden können. Erste Pilotprojekte zeigen, dass eigens trainierte Modelle antike Buchstabenformen zuverlässig erkennen und so zur verbesserten Lesbarkeit beitragen.

3.2 Paläographie und Handschriftenkunde

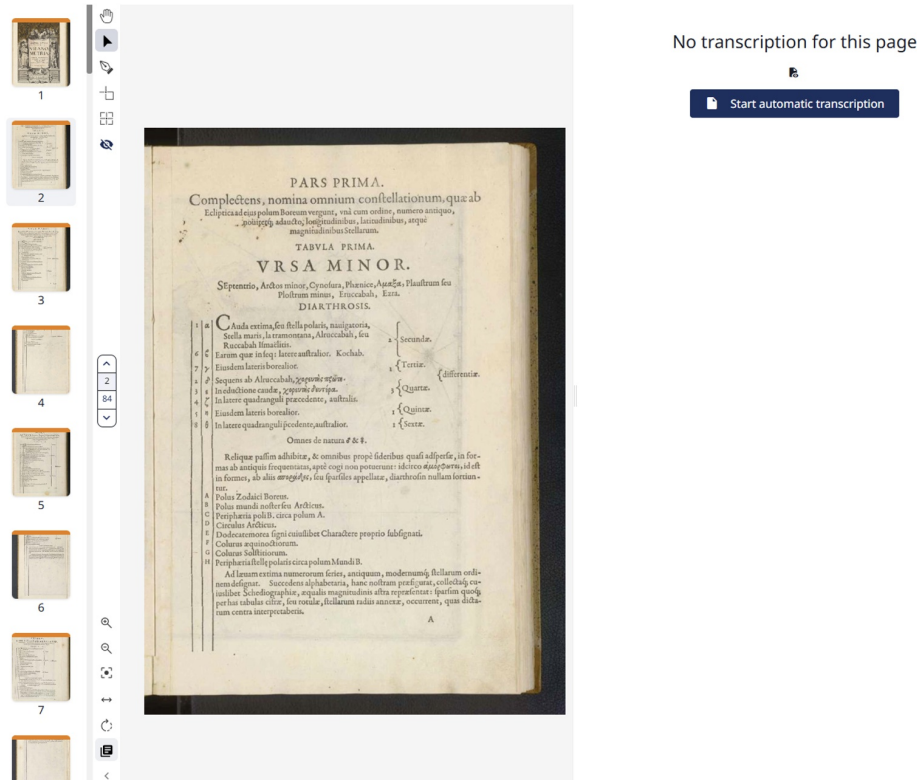
Auch auf dem Gebiet der Paläographie entstehen neue Ansätze. HTR-gestützte Analysen erlauben es, große Korpora quantitativ auszuwerten. Dadurch werden neue Erkenntnisse über Schreibtraditionen, Abkürzungsformen oder die Entwicklung von Schriftstilen möglich. Forschende können nicht nur einzelne Handschriften, sondern ganze Überlieferungszusammenhänge vergleichen und digital visualisieren.

3.3 Archiv- und Bibliotheksforschung

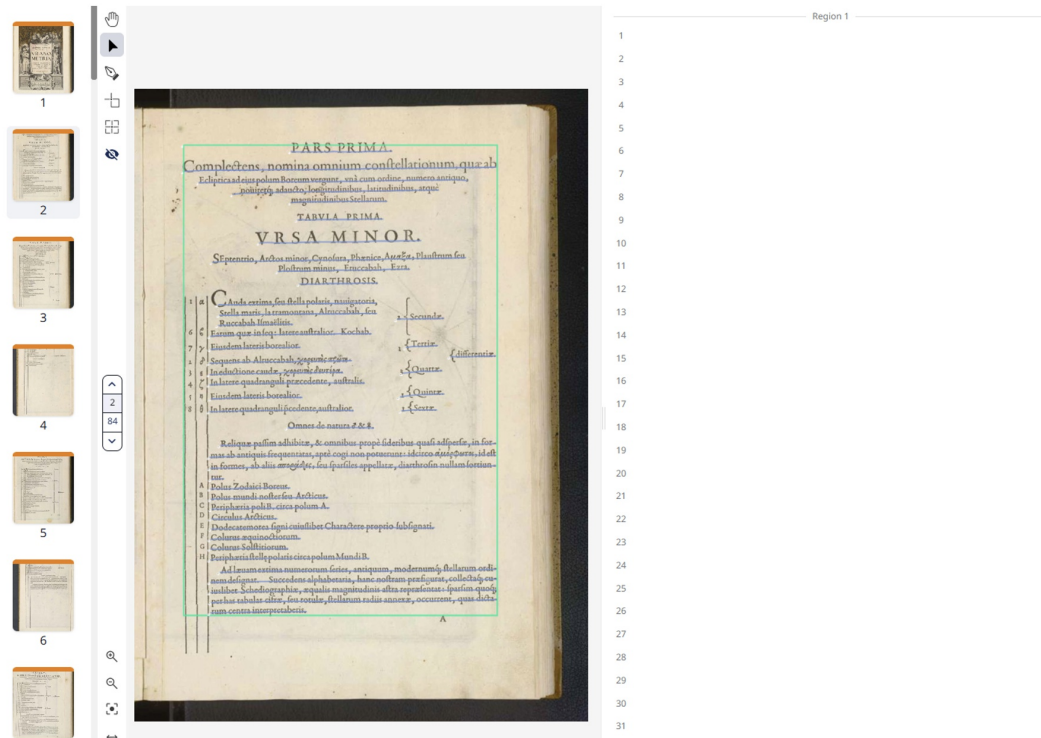
Schließlich bietet *Transkribus* auch in der Archiv- und Bibliotheksforschung neue Chancen. Viele für die Altertumswissenschaften relevante Dokumente sind bislang nur indirekt über Kataloge oder Exzerpte zugänglich. Mit *Transkribus* lassen sich solche Materialien digital erschließen und dadurch die Überlieferungsgeschichte antiker Texte neu beleuchten. Besonders die automatische Auswertung handschriftlicher Findmittel und Kataloge eröffnet bislang unzugängliche Forschungszugänge.

4. Beispielworkflow: Erste Schritte mit Transkribus

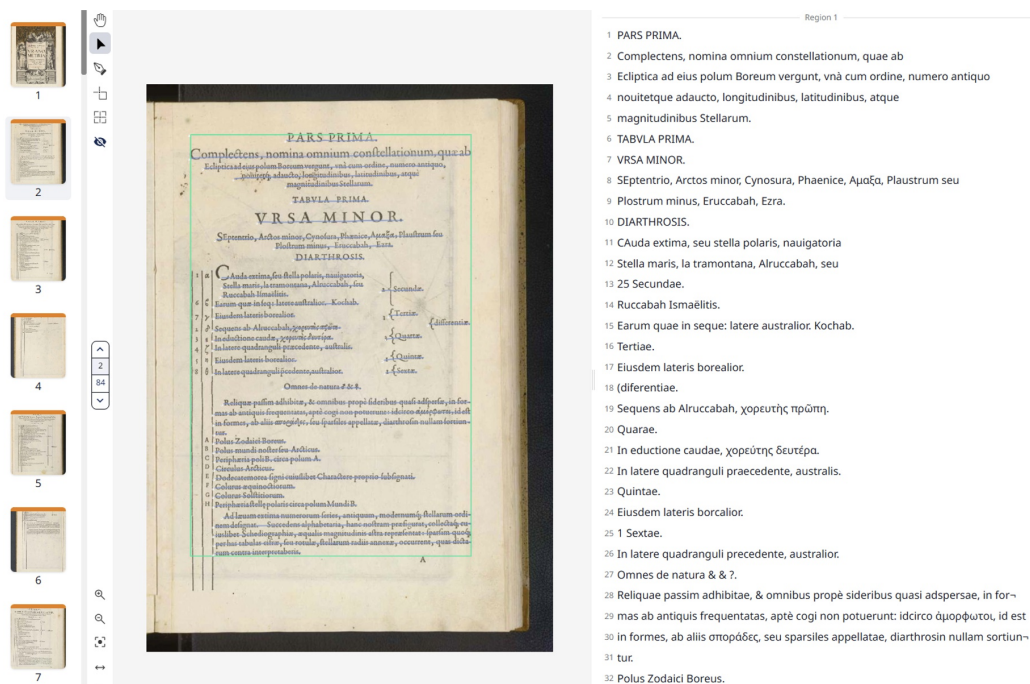
Ein einfaches Szenario beginnt damit, dass Forschende vorhandene Digitalisate, etwa Papyrusfragmente, Handschriftenseiten oder Inschriftenfotografien, in die *Transkribus*-Plattform hochladen.



Schon ohne vorherige Trainingsphase kann die Software auf sogenannte *Baseline-Modelle* zurückgreifen, die für verschiedene Schriftarten und historische Kontexte entwickelt wurden. Diese vortrainierten Modelle liefern erste automatische Transkriptionen, die sich anschließend im Editor von Transkribus anzeigen, durchsuchen und exportieren lassen.



Auf diese Weise können Forschende relativ schnell einen Überblick über größere Textmengen gewinnen, ohne zuvor umfangreiche Vorarbeiten leisten zu müssen. Die automatisch erzeugten Transkriptionen werden in einem nächsten Schritt überprüft und, wo nötig, korrigiert.



Schon diese einfache Kombination von automatischer Texterkennung und manueller Nachbearbeitung spart Zeit und erleichtert den Zugang zu Quellen, die bislang schwer lesbar oder nur eingeschränkt nutzbar waren. Die korrigierten Texte lassen sich dann mit Annotationen versehen und in gängige Austauschformate (etwa XML/TEI) exportieren, sodass sie in digitale Editionsumgebungen integriert oder für weitere Analysen aufbereitet werden können.

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2 <Page xmlns="http://schema.primeresearch.org/PAGE/gts/pagecontent/2013-07-15">
3   <Metadata>
4     <Creator>prov=READ-COOP:name=PyLaia@TranskribusPlatform:version=2.31.0:model_id=37855:lm=none:date=08_09_2025_10:54</Creator>
5     <Created>2021-11-21T12:03:29.463+01:00</Created>
6     <LastChange>2025-09-08T10:56:03.051+02:00</LastChange>
7     <TranskribusMetadata docId="839308" pageId="3108522" pageNr="2" status="IN_PROGRESS" userId="95194" imgPrin="https://files.transkribus.eu/Get?id=OGMEKNDKCHJODBMZFWBQF&amp;fileType=view" ima
8   </Metadata>
9   <Page imageFilename="3108522.jpg" imageWidth="3127" imageHeight="4361">
10     <TextRegion orientation="0.0" id="tr_1" custom="readingOrder (index:0)">
11       <Coords points="140,398 340,361 2504,3631 2504,388"/>
12       <TextLine id="tr_1_tl_1" custom="readingOrder (index:0)">
13         <Coords points="1065,451 1200,486 1246,454 1316,478 1422,457 1476,486 1551,457 1773,457 1773,341 1727,351 1684,313 1616,311 1511,357 1403,305 1338,330 1230,313 1154,359 1065,335"/>
14         <Baseline points="1080,432 1116,435 1153,435 1190,435 1227,435 1263,435 1300,435 1337,435 1374,432 1411,434 1447,435 1484,435 1521,435 1558,435 1595,432 1631,432 1668,432 1705,432 1742
15       <TextEquiv>
16         <Unicode>PARS PRIMA.</Unicode>
17       </TextEquiv>
18     </TextLine>
19     <TextLine id="tr_1_tl_2" custom="readingOrder (index:1)">
20       <Coords points="225,592 2497,581 2497,432 2443,427 2324,484 2216,432 2078,492 1862,411 1770,462 1481,486 1090,439 916,481 716,438 505,481 224,427"/>
21       <Baseline points="240,570 448,564 556,561 664,561 772,561 880,559 988,558 1096,558 1204,558 1312,555 1420,555 1528,554 1636,549 1744,549 1852,549 1960,546 2068,546 2176,545 2284,543 23
22       <TextEquiv>
23         <Unicode>Complectens, nomina omnium constellationum, quae ab</Unicode>
24       </TextEquiv>
25     </TextLine>
26     <TextLine id="tr_1_tl_3" custom="readingOrder (index:2)">
27       <Coords points="430,668 508,705 732,654 2103,640 2200,697 2292,668 2294,589 2162,559 554,611 432,578"/>
28       <Baseline points="444,647 536,645 629,647 722,644 815,641 908,641 1001,641 1093,641 1186,641 1279,638 1372,640 1465,638 1558,637 1651,634 1743,632 1836,632 1929,629 2022,629 2115,629 2
29       <TextEquiv>
30         <Unicode>Ecliptica ad eius polum Boreum vergunt, vñā cum ordine, numero antiquo</Unicode>
31       </TextEquiv>
32     </TextLine>
33     <TextLine id="tr_1_tl_4" custom="readingOrder (index:3)">
34       <Coords points="662,781 2735,716 1862,781 1965,732 2092,784 2094,643 665,662"/>
35       <Baseline points="680,728 751,728 822,719 893,716 964,715 1035,713 1106,711 1177,710 1248,707 1319,708 1390,710 1461,707 1532,707 1603,704 1674,704 1745,704 1816,701 1887,701 1958,699
36       <TextEquiv>
37         <Unicode>nouitque adaucto, longitudinibus, latitudinibus, atque</Unicode>

```

Optional kann dieser Workflow durch die Erstellung einer eigenen *Ground Truth* erweitert werden. Wenn Forschende feststellen, dass die Erkennungsqualität für ihr spezielles Material nicht ausreicht, können sie eine Auswahl an Seiten manuell transkribieren und damit ein eigenes Modell trainieren. Dieses Modell wird Schritt für Schritt verfeinert, indem Korrekturen in den Trainingsprozess zurückfließen. So entsteht eine maßgeschneiderte Lösung, die sich langfristig für größere Textcorpora eignet und die Genauigkeit der Transkriptionen deutlich erhöht.

5. Chancen und Vorteile

Die Potenziale von *Transkribus* lassen sich auf mehreren Ebenen beschreiben. Zunächst ermöglicht die Plattform eine erhebliche Effizienzsteigerung. Die Transkription großer Dokumentenmengen, die bisher Jahre in Anspruch nahm, kann nun in deutlich kürzerer Zeit durchgeführt werden. Darüber hinaus eröffnet die Arbeit mit umfangreichen Korpora die Möglichkeit zu Massendatenanalysen, die neue Fragestellungen etwa zur Verbreitung von Schriften oder zur Vernetzung antiker Schreiber*innen erlauben. Ebenso bedeutsam ist die verbesserte Zugänglichkeit: digitalisierte und automatisiert transkribierte Texte lassen sich weltweit teilen, was internationale Kooperationen erleichtert und es auch kleineren Institutionen ermöglicht, auf bedeutende Quellenbestände zuzugreifen. Nicht zuletzt trägt *Transkribus* zur Demokratisierung der Forschung bei, da die Plattform so gestaltet ist, dass nicht nur große Forschungszentren, sondern auch kleinere Projekte und sogar Einzelpersonen mit ihr arbeiten können. Besonders hervorzuheben ist schließlich

die Rolle von *Transkribus* als Brücke zwischen den Altertumswissenschaften und den *Digital Humanities*: indem philologische und historische Fragestellungen mit Methoden der Informatik und der künstlichen Intelligenz verbunden werden, entsteht ein interdisziplinärer Raum, in dem traditionelle Kompetenzen und neue Technologien produktiv aufeinandertreffen.

6. Herausforderungen und Chancen

Den genannten Chancen stehen jedoch auch erhebliche Herausforderungen gegenüber. Eine zentrale Hürde liegt im hohen Initialaufwand: damit ein Modell zuverlässig arbeiten kann, muss zunächst eine ausreichende Menge an *Ground Truth* erstellt werden, was umfangreiche manuelle Arbeit und fachliche Expertise erfordert. Hinzu kommt, dass die Qualität der Ergebnisse stark vom Zustand des Ausgangsmaterials abhängt. Beschädigte oder ungewöhnlich gestaltete Texte bereiten dem System weiterhin große Schwierigkeiten. Auch technische Anforderungen spielen eine Rolle, denn Forschende benötigen nicht nur hochauflösende Digitalisate, sondern auch die notwendigen digitalen Kompetenzen, um die Plattform effizient einsetzen zu können. Schließlich sind rechtliche Fragen zu berücksichtigen: da nicht alle digitalisierten Handschriften und Inschriften frei zugänglich sind, können urheberrechtliche Bestimmungen oder institutionelle Zugangsregelungen die Nutzung von *Transkribus* erheblich einschränken.

7. Zukunftsperspektiven und Fazit

Die Entwicklung von *Transkribus* verspricht eine kontinuierliche Verbesserung der Texterkennung. Fortschritte im Bereich des *Deep Learning* und multimodaler Modelle werden künftig eine noch präzisere Analyse ermöglichen, die nicht nur Buchstaben, sondern auch Kontexte berücksichtigt – besonders relevant für fragmentarische oder beschädigte Texte. Zudem ist mit der Etablierung offener, gemeinschaftlich entwickelter Modelle zu rechnen, etwa für griechische Minuskelschriften.

Langfristig wird die Integration von *Transkribus* in bestehende Forschungsinfrastrukturen entscheidend sein, um nachhaltige digitale Editionen zu schaffen.

Der Einsatz von *Transkribus* markiert einen Meilenstein in der digitalen Transformation der Altertumswissenschaften. Zwar bleibt philologische Expertise unersetzlich, doch eröffnet die Plattform die Möglichkeit, Material in bislang unerreichtem Umfang zu erschließen und neue Forschungsfragen zu entwickeln. Damit trägt *Transkribus* wesentlich dazu bei, den Zugang zu zentralen Quellen der Antike zu erweitern und die Disziplin nachhaltig zu verändern.