

# Computergestützte Beobachtungen zur Schriftentwicklung in Papsturkunden des Hochmittelalters – neue Instrumente in praxi angewandt

Von **Benedikt Hotz** und **Benjamin Schönfeld**

## Vorwort

Die Forschungsarbeit an mittelalterlichen Urkunden wird in zunehmendem Maße digital<sup>1</sup>. Auf die Frage aber, welchen Nutzen digitale Instrumente für die Paläographie im 21. Jahrhundert bieten, erhält man seitens paläographisch arbeitender Forscher zumeist recht unterschiedliche Antworten. Versucht man diese auf einen gemeinsamen Nenner zu bringen, lässt sich konstatieren, dass die Erwartungen an „den Computer“ zuweilen recht hoch sind<sup>2</sup>. Nicht anders verhielt es sich, als der Teilbereich Paläographie des BMBF-Projektes „Schrift und Zeichen. Computergestützte Analyse hochmittelalterlicher Papsturkunden. Ein Schlüssel zur Kulturgeschichte Europas“<sup>3</sup> im August 2012 die Arbeit aufnahm. Gegen die immer wieder kritisierte Subjektivität und mangelhafte Nachvollziehbarkeit der Paläographie erschien der Computer in den Augen vieler Beobachter als Allheilmittel, ebenso gegen das oftmals mühsame „manuelle“ Lesen handschriftlicher Texte mittels vollautomatisch arbeitender OCR-Software oder als Instrument zur Visualisierung abstrakter Entwicklungen innerhalb der Schrift, die andernfalls schwer vermittelbar wären.

Dieser Beitrag<sup>4</sup> möchte in Form eines Werkstattberichts über die bisherige Arbeit des Teilprojekts zu diesen Erwartungen Stellung nehmen. Er soll Erwartungen, die sich als Irrwege erwiesen haben, ebenso dämpfen, wie neue Optionen der digitalen Paläographie aufzeigen, die sich stattdessen im Zuge der Arbeiten aufgezeigt haben und die Frage beleuchten, wie der Computer die Arbeit des Paläographen verändert, an welchen Stellen er sie bereichern kann und wo Erwartungen korrigiert werden müssen.

Eine der Hauptaufgaben des BMBF-Projektes „Schrift und Zeichen“ besteht darin, Antworten auf zukunftsweisende Fragen zu geben, welche Strukturen und Methodiken der klassischen Paläographie<sup>5</sup> effektiv und auf sinnvolle Art und Weise in technische, digitale

---

<sup>1</sup> Irmgard FEES, Urkundendigitalisierung und Mittelalterforschung. Expertengespräch / Workshop zur Begleitung des DFG-Projekts „VdU – Virtuelles deutsches Urkundennetzwerk“, 2012 Marburg, in: Blätter für deutsche Landesgeschichte 149 (2013), S. 127–185; online unter <http://documents.icar-us.eu/documents/2013/12/urkundendigitalisierung-und-mittelalterforschung.pdf> (letzter Zugriff: 4. August 2014).

<sup>2</sup> Vgl. hierzu die Zusammenfassung der Thesen und Standpunkte hinsichtlich der Nutzung digitaler Instrumente in der diplomatischen Forschung bei Georg VOGELER (Hg.), Digitale Diplomatie. Neue Technologien in der historischen Arbeit mit Urkunden (AfD Beiheft 12), Köln u. a. 2009.

<sup>3</sup> Das BMBF-Projekt wird von Juni 2012 bis Juli 2015 an der Friedrich-Alexander-Universität Nürnberg/Erlangen und an der Ludwig-Maximilians-Universität in München durchgeführt. Siehe hierzu die Homepage des Projekts: <http://www5.cs.fau.de/de/papsturkunden-des-hohen-mittelalters/> (letzter Zugriff: 4. August 2014). Vgl. auch den Beitrag von Viktoria TRENKLE in diesem Band.

<sup>4</sup> Bei diesem Beitrag handelt es sich um die geringfügig veränderte Fassung eines Vortrages, der im Rahmen der BMBF-Tagung „Papsturkunden zwischen internationaler Vernetzung und Digitalisierung. Neue Zugangsweisen zur europäischen Schriftgeschichte“ am 24. März 2014 im Historischen Kolleg, München gehalten wurde.

<sup>5</sup> Gudrun BROMM, Neue Vorschläge zur paläographischen Schriftbeschreibung, in: Peter RÜCK (Hg.), Methoden der Schriftbeschreibung (Historische Hilfswissenschaften 4), Stuttgart 1999, S. 21–42.

Analysevorgänge<sup>6</sup> übersetzt werden können<sup>7</sup>. Es ist klar, dass in einem solchen Projekt, in welchem in verschiedener Hinsicht sicherlich Pionierarbeit verrichtet wird, durchaus Irrwege beschritten werden können und müssen. Dies stellt aus unserer Sicht vielmehr einen produktiven Prozess dar. Dieser Punkt soll nachfolgend anhand der Überlegungen, die zur Schaffung eines OCR-Programms für Handschriften im Projekt angestellt wurden, erläutert werden.

### **Grenzen moderner Informationstechnologie im Hinblick auf Paläographie: Das Beispiel der OCR-Programme**

Wie bereits mehrfach in einschlägigen Fachpublikationen zur digitalen Paläographie für mittelalterliche Quellenkorpora konstatiert wurde, ist der oft geäußerte Wunsch nach einem solchen Programm zwar nachvollziehbar, im vorliegenden Fall jedoch technisch nicht auf sinnvolle Weise realisierbar<sup>8</sup>. Die Menge an zeitaufwendig aufzubereitendem Trainingsmaterial für den zugrundeliegenden Algorithmus ist sehr hoch, was das Verhältnis von Kosten und Nutzen äußerst ungünstig gestaltet. Auch wenn Optimierungsverfahren in Form von Word- oder Glyphspotting hier vordergründig Effektivierung versprechen, müssten auch solche Instrumente für die Vielzahl an Charakter- und Ausführungsvarianzen in den vom Projekt „Schrift und Zeichen“ untersuchten Schriften zunächst ebenso zeitraubend erarbeitet werden. Funktionstüchtigkeit in adäquater Form ist dabei zwar durchaus möglich, der erforderliche Zeitaufwand relativiert jedoch den Nutzen nachhaltig. Ändert sich die Schriftgrundlage – in mittelalterlichen Handschriften naturgemäß häufig der Fall –, so wird das Programm wiederum fehleranfällig und neues Training anhand des veränderten Materials wäre notwendig. Eine im Wandel befindliche Schrift, als die sich die kuriale Minuskel im 11. und 12. Jahrhundert offenbart, erhöht den Arbeitsaufwand für die Generierung eines OCR-Tools um ein Vielfaches.

Somit ist ein OCR-Programm für Papsturkunden des Hochmittelalters technisch durchaus nicht unmöglich, aufgrund des äußerst hohen Zeitaufwandes jedoch ineffektiv, da jeder paläographisch geübte Forscher entsprechende Urkunden mit weit geringerer Fehleranfälligkeit und in wesentlich geringerer Zeit transkribieren könnte, als ein computergestütztes Werkzeug dies ohne zusätzliche Optimierungen könnte. Aus diesem Grunde wurde im Verlauf der weiteren paläographischen Arbeit im Projekt „Schrift und Zeichen“ ein Paradigmenwechsel unternommen, von einer Automatisierung des Lese- und Analyseprozesses hin zu einer Präzisierung und Beschleunigung des „klassischen“ paläographischen Ansatzes. Dieser Ansatz sollte dabei schwerpunktmäßig in zwei nicht

---

<sup>6</sup> Peter A. STOKES, Describing Handwriting. (2011): Digital Ressource and Database of Palaeography, Manuscripts and Diplomatic (DIGIPAL): <http://www.digipal.eu/blogs/blog/describing-handwriting-part-i/> (letzter Zugriff: 4. August 2014); Wernfried HOFMEISTER u. a., Forschung am Rande des paläographischen Zweifels. Die EDV-basierte Erfassung individueller Schriftzüge im Projekt DAMaLS, in: Malte REHBEIN u. a. (Hg.), Kodikologie und Paläographie im digitalen Zeitalter (Schriften des Instituts für Dokumentologie und Editorik 2), Norderstedt 2009, S. 261–292.

<sup>7</sup> Zur Quellenbasis und methodischen Vorgehensweise der Teilprojekte vgl. den Beitrag von Viktoria TRENKLE in diesem Band.

<sup>8</sup> Arianna CIULA, The palaeographical method under the light of a digital approach, in: REHBEIN, Kodikologie (wie Anm. 6), S. 219–236, und Mark STANSBURY, The computer and the classification of script, ebd. S. 237–250.

minder relevanten Bereichen mittelalterlicher Papsturkundenforschung angewandt werden: zum einen auf dem Feld der Identifizierung von Schreiberhänden, zum anderen hinsichtlich der Nachzeichnung der paläographischen Transformationsprozesse, welche die Urkundenschrift im Untersuchungszeitraum durchlief – beides auf möglichst objektiver Grundlage<sup>9</sup>. Dieser Ansatz ist nicht vollständig neu, denn vergleichbare Projekte aus dem kodikologischen Bereich erzeugten hier bereits wichtige Vorkenntnisse<sup>10</sup>. Trotzdem wurde schon bald deutlich, dass die Arbeitsweisen und Erkenntnisse aus dem buchschriftlichen Bereich nur sehr begrenzt auf das Corpus der Urkundenpaläographie übertragbar sind. Die automatische Generierung von Musterbuchstaben als „Normwert“, wie sie etwa das SPI-Tool vornimmt, lässt sich in der wesentlich dynamischeren, von einem stetigen Wandlungsprozess geprägten Urkundenschrift des 12. Jahrhunderts nicht sinnvoll anwenden. Auch tendieren auf Buchschriften basierende paläographische Arbeiten im Allgemeinen zu einer starken Kategorisierung, die sich für den geschäftsschriftlichen Bereich nur bedingt anwenden lassen dürfte<sup>11</sup>. Damit lautet im hiesigen Kontext die Frage, wie das Auge des Paläographen durch den Rechner geschärft und ergänzt werden kann.

### **Exkurs zur Schriftentwicklung an der päpstlichen Kurie im 11. und 12. Jahrhundert**

Zum besseren Verständnis sei an dieser Stelle kurz der konkrete Hintergrund dieser Aufgabenstellung umrissen: Die Ablösung der Kuriale durch die Diplomatische Minuskel sowie deren allmähliche Gotisierung im 11. und 12. Jahrhundert gilt als ein Entwicklungsprozess, der im Zuge der Reformen Leos IX. um das Jahr 1050 einsetzte und in den ersten Jahrzehnten des 12. Jahrhunderts zum Abschluss kam<sup>12</sup>. Im Gegensatz zur zweiten Entwicklungslinie der päpstlichen Schrift, der allmählichen Gotisierung der kurialen Urkundenminuskel, war die Einführung der Diplomatischen Minuskel weniger ein allmählicher, aus den Schreibgewohnheiten der *scriptores* gewissermaßen „natürlich“ vor sich gehender Prozess, sondern ein mit entsprechenden Wandlungen in der Kanzlei einhergehender Einschnitt „von oben“. Bemerkenswert ist dabei, dass noch vor der endgültigen Verbannung

<sup>9</sup> Vgl. hierzu die Anmerkungen von Imran SIDDIQI u. a., Writing property descriptors: a proposal for typological groupings, in: *Gazette du livre médiéval* 56–57 (2011), S. 42–57, hier S. 42f., der in einer überschaubaren, konstruktiv gewählten Auswahl an objektiven Termini zur Schriftbeschreibung ein zentrales Forschungsobjekt sah. Die Notwendigkeit, paläographische Argumentationsstrukturen objektiver wie quantitativer zu gestalten, betont auch Peter A. STOKES, Computer-aided palaeography, present and future, in: REHBEIN, *Kodikologie* (wie Anm. 6), S. 309–340, hier exemplarisch S. 309f.

<sup>10</sup> So etwa das Projekt DAMaS: HOFMEISTER, Paläographischer Zweifel (wie Anm. 6) oder das (J)SPI Tool: CIULA: Palaeographical method (wie Anm. 8).

<sup>11</sup> Vgl. den Beitrag von Petros SAMARA in diesem Band.

<sup>12</sup> Vgl. hierzu auch die Einzelstudien zur gotischen Schrift von Walter HEINEMEYER, *Studien zur Geschichte der gotischen Schrift*, Köln u. a. 1982, und zur römischen Kuriale von Paul RABIKASKAS, *Die römische Kuriale in der päpstlichen Kanzlei* (*Miscellanea Historiae Pontificiae* 20 = *Collectionis* 59), Rom 1958; Paolo RADICIOTTI, *La curiale romana nuova. Parabola discendente di una scrittura*, in: *Archivio della Società Romana di Storia Patria* 112 (1989), S. 39–113; DERS., *Attorno alla storia della curiale romana*, in: *Archivio della Società Romana di Storia Patria* 122 (1999), S. 105–123; Walter MÜLLER-SCHÖLL, *Die curiale Minuskel des ausgehenden 11. und 12. Jahrhunderts*, Göttingen 1924/1925. Zur Schrift in hochmittelalterlichen Papsturkunden vgl. darüber hinaus beispielsweise Thomas FRENZ, *Papsturkunden des Mittelalters und der Neuzeit*, 2. aktual. Aufl. Stuttgart 2000, hier v. a. den Literaturüberblick auf S. 30–32.

der Kurialis aus den Urkunden erste Gotisierungstendenzen in der päpstlichen Kanzlei festzustellen sind. Der Zeitraum, in dem die Diplomatische Minuskel in ihrer „Reinform“ genutzt wurde, erweist sich somit als verhältnismäßig kurz.

Das Einfließen gotischer Elemente der Buchstabengestaltung und des Schriftbildes ist ein deutlich längerer Prozess als die Ablösung der Kuriale, bedingt durch den fließenden Charakter dieser Entwicklung. Aus paläographischer Sicht befand sich die römische Kurie somit am Puls der Zeit. Die benutzte Schrift war nicht mehr, wie noch zu Beginn des 11. Jahrhunderts, ein regional begrenztes, traditionsreiches Relikt, sondern folgte aktuellen Entwicklungen. Hierin lässt sich die Handschrift des neuen, nicht-römischen, oft aus den schriftgeschichtlich für die Ausprägung der Gotisierung so bedeutenden Regionen Frankreichs und des römisch-deutschen Reiches stammenden Kanzleipersonals erkennen.

### **Die paläographische Arbeit am Digitalisat: verfügbare Werkzeuge, Vorteile und Auswirkungen auf die Perspektiven der digitalen Paläographie**

Doch zurück zur Ausgangsfrage: Wie gestaltet sich die Übersetzung klassisch-paläographischer Methoden in die Arbeit mit digitalen Werkzeugen? Auch in diesem Fall bestehen Schwierigkeiten, da zunächst anspruchsvolle technische Hürden in Erscheinung treten.

Da für die Bearbeitung der Urkunden nicht mehr einzelne Originale in unterschiedlichen Archiven konsultiert werden müssen, sondern nun eine Vielzahl digital auf der Festplatte oder im Internet verfügbare Bilder zur Verfügung stehen<sup>13</sup>, galt es zunächst ein Instrument zu schaffen, mit dem eine umfassende Aufnahme, Systematisierung, Einteilung, Ordnung und Speicherung des relevanten Quellenmaterials durchgeführt werden kann. Zu diesem Zweck wurde vom Teilbereich Mustererkennung in Kooperation mit den anderen Projektteilen zunächst ein Programm entwickelt, mit dessen Hilfe bestimmte Bereiche einer Urkunde markiert und gewissermaßen kommentiert werden können. Dieses wird im Folgenden als Annotierungstool bezeichnet werden<sup>14</sup>. Vergleichbare Programme existieren zwar bereits, jedoch bietet das hier präsentierte Tool neben der Speicherung der Daten in XML zudem die Möglichkeit, mithilfe von polygonalen Annotationen die Markierungen wesentlich präziser zu gestalten und an das gegenüber Buchtexten oftmals wesentlich unregelmäßigere Schriftbild von Urkunden anzupassen.

Darüber hinaus wurden in einem zweiten Schritt jeweils ein System zur Visualisierung eigenständiger Bildausschnitte aus diesen Annotationen, sowie zur Messung bestimmter Abstände und Größen innerhalb einer Urkunde entwickelt. Alle genannten Programme basieren also weniger auf Algorithmen, welche seitens der Mustererkennung für die automatische Analyse der Schrift entwickelt wurden, sondern stellen eher Hilfsinstrumente für die klassische Arbeit der Paläographen dar. Der hier erkennbare Unterschied in Zielsetzungen der unterschiedlichen Disziplinen wird, bei grundsätzlich ertragreicher Zusammenarbeit, vermutlich zunächst bestehen bleiben. Als eher theoretische Konsequenz dieser Erkenntnis

---

<sup>13</sup> Theo KÖLZER, *Diplomatik, Edition, Computer*, in: Georg VOGELER (Hg.), *Digitale Diplomatie. Neue Technologien in der historischen Arbeit mit Urkunden*, Köln u. a. 2009, S. 13–27.

<sup>14</sup> Vgl. hierzu auch den Beitrag von Vincent CHRISTLEIN in diesem Band.

zeichnet sich nämlich ab, dass die Arbeit weniger automatisiert, als vielmehr methodisch durch verbesserte Möglichkeiten vertieft und teilweise beschleunigt wird. Im Anschluss an Jinna Smit, die bereits im Jahre 2011 nach dem „Death of the Palaeographer“<sup>15</sup> fragte und diesen bereits verneinte, ließe sich also formulieren: Die Arbeit des Paläographen wird durch den Einsatz des Computers weder unwichtiger noch weniger.

Bisher basieren paläographische Untersuchungen, von wenigen Ausnahmen abgesehen, auf individuellen Aussagen des Forschers, der sich dabei auf seine Expertise und das Augenmaß verlässt. Dies führt zu dem Phänomen, dass entsprechende Publikationen oftmals einen gewissen Hang zur Metaphorik aufweisen. Ohne Beigabe von Abbildungen der untersuchten Schrift sind sie deshalb jedoch oftmals kaum nachvollziehbar. Wie definiert sich, um ein willkürliches Beispiel zu konstruieren, „eine sanft geschwungene, jedoch sichere Kanzleihand“, wenn kein Schriftbeispiel vorliegt? Die Qualität der jeweiligen Arbeit ist dabei ausdrücklich unberührt – problematisch ist vielmehr die Nachvollziehbarkeit der Erkenntnisse<sup>16</sup>.

Diesem Mangel soll Abhilfe geschaffen werden, indem subjektiv geschätzte Werte soweit wie möglich durch exakte Messungen ersetzt werden. Hier ließe sich einwenden, dass bloße Messwerte an sich gleichfalls kaum aussagekräftiger sind als wortreiche Umschreibungen. Es ist richtig, dass der Bezug zum beschriebenen Objekt, das bedeutet: die Beigabe entsprechenden Bildmaterials, nach wie vor sehr wichtig bleibt. Zudem bedürfen auch Messwerte der Interpretation und sind nicht für sich allein gültig. Sie müssen also systematisch in die Schriftuntersuchung einbezogen werden, anstatt um ihrer selbst willen mitgeliefert zu werden. Dennoch bieten Messwerte im Vergleich zu verbalen Beschreibungen eine bessere Grundlage für Einschätzungen, zumal sie sich, etwa durch Diagramme oder als Grundlage für Statistiken, besser für weitergehende Visualisierungen eignen.

Neben den Messungen ist es vor allem die Möglichkeit, durch Bildausschnitte die paläographische Untersuchung systematisch auf bestimmte Bereiche eines Textes zu lenken, welche die Wahrnehmung einer Schrift durch den Forscher beeinflusst. Dieser „Schabloneneffekt“ wirkt sich zunächst positiv auf die Arbeit aus, denn es wird auf diese Weise möglich, umfangreiches Datenmaterial punktgenau und direkt zu vergleichen, ohne Bildkopien in großen Mengen auf dem Schreibtisch ausbreiten zu müssen. Vor diesem Hintergrund ist also eine Präzisierung der paläographischen Befunde denkbar, da das Auge des Forschers nicht durch den übrigen Kontext abgelenkt wird, sondern sich konzentriert auf ein Detail einer Schrift richten kann. Aufgrund fehlender längerfristiger Praxis ist diese These gleichwohl noch nicht erwiesen.

---

<sup>15</sup> Jinna SMIT, The death of the palaeographer? Experiences with the Groningen Intelligent Writer Identification System (GIWIS), in: AfD 57 (2011), S. 413–246.

<sup>16</sup> Auf das Problem einer „gewisse[n] Poesie der Handschriftencharakterisierung“ wies bereits Werner MALECZEK, Die eigenhändigen Unterschriften der Kardinäle - ein Spiegelbild ihrer Persönlichkeit? Mit einem Überblick über eigenhändige Unterschriften auf Urkunden vom Frühmittelalter bis ins 13. Jahrhundert, in: Stefan WEINFURTER (Hg.), Päpstliche Herrschaft im Mittelalter: Funktionsweisen – Strategien – Darstellungsformen, Ostfildern 2012, S. 239–300, hier v. a. S. 269–273 hin.

## **Digitale Bearbeitung paläographischer Forschungsfragen in der Praxis**

Wie gestaltet sich nun die praktische Arbeit mit den zuvor erwähnten Programmen? Am Anfang steht grundsätzlich das Annotierungstool, mit dem die Voraussetzungen für die spätere Arbeit geschaffen werden. Dessen effektive Funktionsweise, die eine zügige Bedienung ermöglichen soll, galt bereits während des Entwicklungsprozesses als wichtigste Prämisse.

([Abbildung 1](#))

Zunächst wird eine „Datenbank“ erstellt, indem Bilder mittels einer entsprechenden Option in das Programm geladen werden. Alle Fenster im Programm sind frei verschiebbar, was die Übersichtlichkeit gerade im Bereich der Urkunde im Bedarfsfall erhöht. Die Urkunde selbst wird im zentralen Fenster angezeigt, wo später auch die Annotationen vorgenommen werden. Zunächst jedoch müssen Kategorien erstellt werden, nach denen die Markierungen klassifiziert werden können. Mithilfe des Fensters im unteren Bereich des Bildschirms werden diese Klassen definiert, etwa „Wort“, „Textbereich“, „Graphisches Symbol“. Es handelt sich also um Überbegriffe derjenigen Objekte, die annotiert werden sollen, weshalb der Paläograph bereits vor der Benutzung des Programms die Schrift sowie das zugrundeliegende Medium (Urkunden, Codices, Akten, Individualschriftgut oder ähnliches) in einem theoretischen Konzept analysiert haben sollte. Die jeweilige Ausprägung der Klasse, also beispielsweise „Rota“ für „Graphisches Symbol“ wird in die Tabellenstruktur direkt oder mittels einer Eingabemaske eingefügt. ([Abbildung 2](#))

Zuvor allerdings müssen diejenigen Bereiche der Urkunde, die der entsprechenden Kategorie angehören sollen, durch ein Quadrat oder ein Polygon im Bild optisch markiert werden. Hierfür stehen grundsätzlich die Optionen „Rechteck“ und „Polygon“ zur Verfügung. Somit besteht der Annotationsprozess zusammenfassend aus folgenden Stufen: ([Abbildung 3](#))

1. Vorarbeit: Erstellung eines theoretischen Konzepts, das der Annotation zugrunde liegt. Hierbei ist insbesondere die Art der behandelten Quelle mit allen Implikationen zu berücksichtigen. Handelt es sich zum Beispiel um kalligraphisch anspruchsvolles, mit künstlerischer Intention zur Repräsentation gestaltetes Schriftgut oder um „Geschäftsschriftgut“, also eine ästhetisch hierzu meist konträre, vielmehr schwerpunktmäßig der Information dienende Schrift? Liegt ein größeres, homogenes Corpus (etwa ein Codex) vor, das in einem oder wenigen Arbeitsschritten von einer oder zumindest wenigen Personen erstellt wurde oder handelt es sich um ein Konvolut an Schriftstücken, das wie im Fall von Urkunden zumeist von einer größeren Anzahl teilweise verschiedenster Personen gefertigt wurde? Jede dieser hier zu stellenden Fragen birgt paläographische Implikationen, denen bei der Erstellung der Klassenstruktur Rechnung getragen werden sollte. Zwar kann diese Struktur theoretisch auch nachträglich angepasst werden, sobald jedoch Daten erhoben wurden, kann jeder Eingriff in die bestehende Systematik gravierende Konsequenzen für die Verknüpfung und damit die Logik der betreffenden Daten nach sich ziehen.

2. Umsetzung dieses Konzepts durch Anlage von Klassen im Programm, die gleichzeitig als Überbegriffe für spätere Annotationen dienen können. Dieser Schritt kann, eingedenk der genannten Einschränkungen, bei Bedarf jederzeit wiederholt oder modifiziert werden,

sodass sich die Struktur der Annotationen veränderten Forschungsfragen anpassen kann.

3. Optische Auszeichnung der gewünschten Bereiche einer Urkunde, wobei hier auch mehrere Markierungen übereinander möglich sind. Dies kann später mithilfe des Snippettools genutzt werden, um nur bestimmte Segmente ausgeben zu lassen. ([Abbildung 4](#))

4. Bezeichnung der markierten Stelle, womit das Annotierungstool somit auch zur Erstellung digitaler Editionen genutzt werden könnte, wenn Auffälligkeiten im Text direkt markiert und mit Bezug zur entsprechenden Stelle im Bild im XML-Format gespeichert werden. XML selbst bildet, zumeist unter Anwendung des Sprachstandards TEI, die Grundlage zahlreicher Ansätze zur digitalen Edition<sup>17</sup>. Das Abspeichern der Informationen im XML-Standard erhöht zudem deren Nutzbarkeit an anderer Stelle, etwa in Online-Datenbanken, und erleichtert das spätere Abrufen der Daten auch abseits der Programme, die dazu im Projekt „Schrift und Zeichen“ entstanden sind.

Das erste dieser Programme ist ein Skript, welches anhand der zuvor annotierten Bereiche neue Bildausschnitte, die hier als Snippets bezeichnet werden, generiert. Obwohl zum gegenwärtigen Zeitpunkt noch keine graphische Benutzeroberfläche existiert, ist das Tool jedoch bereits voll funktionsfähig. Für die Verfeinerung der Auswahl stehen einige Parameter bereit, die dem Programm übergeben werden können. Sie ermöglichen es beispielsweise, die Generierung der Snippets auf einen bestimmten, ebenfalls annotierten Bereich einer Urkunde einzugrenzen. Anstatt also grundsätzlich alle Annotationen einer Urkunde in Snippets umzusetzen, ist es beispielsweise möglich, Wörter eines bestimmten Wertes (beispielsweise „Innocentius“) innerhalb eines bestimmten Textabschnittes (etwa „Kontext“) ausgeben zu lassen, um auf diese Art die Schreibweise des Papstnamens im Fall der Bestätigung einer Vorurkunde zu untersuchen, ob dieser Name beispielsweise in irgendeiner Weise systematisch hervorgehoben wurde. Dies kann auf Wunsch auch auf bestimmte Urkunden eingegrenzt werden, womit gezielt Bereiche für die paläographische Untersuchung extrahiert werden können. Vornehmlich der Übersichtlichkeit dient weiterhin die Option, alle erzeugten Ausschnitte in ein Sammelbild zu integrieren. Eine Visualisierungsoption in Form eines geeigneten Sammelbildes befindet sich derzeit zwar noch im Entwicklungsprozess, die erarbeiteten Vorversionen versprechen aber bereits schon jetzt großen Nutzen. Weiterhin existieren dank der Speicherung der Annotationen in XML zusätzliche Möglichkeiten, die Bildausschnitte mit den zugrundeliegenden Daten zu verknüpfen und somit schnell eine gute Arbeitsgrundlage zu schaffen. ([Abbildung 5](#))

Mit Hilfe der Formatierungssprache XSLT können Metadaten der Annotationen mit den Einzelbildern verknüpft und in übersichtlicher Form ausgegeben werden. Der Vorteil besteht in diesem konkreten Beispiel in den vielfältigen und sehr übersichtlichen Möglichkeiten zur Formatierung der gespeicherten Daten. Zweifellos von Nachteil ist es jedoch, dass für Änderungen an der Formatierung bislang Kenntnisse in XSLT notwendig sind. Erweiterte Benutzerfreundlichkeit zu erreichen ist somit nach wie vor das Ziel weiterer Anstrengungen. Eine Option für künftige Projekte wäre folglich, in ähnlicher Weise auf die erzeugten Daten

---

<sup>17</sup> Vgl. hierzu den Beitrag von Georg VOGELER in diesem Band.

zuzugreifen und diese, etwa im Rahmen einer webbasierten Suchmaske, für jedermann leicht durchsuchbar zu machen. Eine Umsetzung einer solchen Benutzeroberfläche innerhalb des Projektes „Schrift und Zeichen“ kann aufgrund des im konkreten Fall geringen Mehrwertes bei gleichzeitig hohem Zeitaufwand und limitierten Ressourcen nicht geleistet werden.

Die Funktionalität der oben beschriebenen Programme sei im Folgenden an einem kurzen Anwendungsbeispiel demonstriert. Als Untersuchungsgegenstand dienen dabei die Datumzeilen päpstlicher Privilegien<sup>18</sup>. Bekanntermaßen wurden diese bis etwa zur Mitte des 12. Jahrhunderts autograph durch den Kanzleileiter eingetragen. Ihr Vokabular ist aufgrund des naturgemäß formelhaften Charakters der Zeile stark begrenzt. Konkret boten sich die von Johannes von Gaeta, dem späteren Papst Gelasius II., datierten Diplome an, da sie sich aufgrund von dessen langer Amtszeit über mehrere Pontifikate erstrecken. Zudem dürfte sich die Schrift des späteren Papstes getrost als markant bezeichnen lassen, wodurch sich umso effektiver einige, unter seinem Namen ausgestellte Urkunden erschließen lassen, die sich zwar in den eigentümlichen Buchstabenformen an seine Gewohnheiten anschließen, sich im Duktus dagegen gravierend unterscheiden. Der Algorithmus sollte nun die Meinung des „klassischen“ Paläographen gegenprüfen, ob die entsprechenden Fälle von Johannes' eigener Hand stammen oder nicht. ([Abbildung 6](#))

Dazu wurden der EDV Daten als Trainingsmaterial übergeben, die aus zweifelsfrei zuzuordnenden Urkunden stammten. Hierfür wurden in allen Stücken, die eine Datierung unter dem Namen Johannes' aufweisen, die Datumzeilen zunächst als Ganze annotiert. Im Anschluss daran wurde die Auszeichnung auf Wortbasis wiederholt beziehungsweise verfeinert. Dem Algorithmus war also bekannt, wie der untersuchte Datar schrieb und wie er sicher nicht schrieb. Über die Zweifelsfälle „wusste“ der PC hingegen nichts, wodurch sich die klassische paläographische Analyse und die digitalen Messungen folglich nicht gegenseitig in ihrem Ergebnis beeinflussen konnten. Das Resultat der entsprechenden Berechnung war überzeugend, denn in fast allen dem Computer vorgelegten Fällen stimmte die paläographische Einschätzung mit den algorithmisch erzeugten Berechnungen überein. Hinzuzufügen ist jedoch, dass diesem Ergebnis die wie erwähnt markante Handschrift des Johannes entgegenkam, sodass tatsächliche paläographische „Härfälle“ nicht ohne weiteres der EDV zur endgültigen Lösung vorgelegt werden können. Wie sich gezeigt hat, können die so produzierten Bildausschnitte bereits als solche für weiterführende paläographische Untersuchungen herangezogen werden.

Technisch fußt auf dem Tool zur Erstellung von Bildausschnitten noch ein weiteres Programm, das der Erstellung von Statistiken auf der Basis von Annotationen beziehungsweise Snippets dient. Es ist das paläographisch vielleicht interessanteste der bisher vorgestellten Werkzeuge, da es konkreten Messungen an der Urkunde dient. Bevor wir die Funktion im Detail erläutern werden, lässt es sich jedoch nicht umgehen, auf die technischen Schwierigkeiten einzugehen, welche die statistischen Auswertungen des vorliegenden Quellenmaterials behindern. Es handelt sich dabei namentlich um die unbekannte Auflösung der retrodigitalisierten Urkunden. Messungen innerhalb einzelner Urkunden sind unter dieser

---

<sup>18</sup> Die Auswahl erfolgte in Absprache mit dem Teilbereich Mustererkennung, da die Datumzeilen gleichzeitig den idealen Prüfstein für einen dort entwickelten Algorithmus zur automatischen Schreiberverifikation darstellten.

Voraussetzung zwar möglich, die eigentlich relevanten urkundenübergreifenden Vergleiche jedoch sind ohne Umwege nicht zu bewerkstelligen. Der größte Teil der digitalen Abbildungen setzt sich aus Photographien zusammen, die im Zuge des BMBF-Projektes „Schrift und Zeichen“ retrodigitalisiert wurden<sup>19</sup>. Der Zielsetzung der damaligen Fotografen kann es kaum entsprochen haben, detailgetreue Wiedergaben im Sinne moderner Archivdigitalisierungen anzufertigen. Vielmehr ging es um die Sicherung des Originals für die praktischen Zwecke der späteren Edition beziehungsweise Regestierung, wofür die vorhandenen Fotos völlig ausreichend waren, ein beigefügter Maßstab deshalb als nicht nötig erschien. Leider geht jedoch hierdurch die Fähigkeit verloren, die tatsächlichen, exakten Urkundenmaße zu ermitteln. Selbst wenn also die Fotografien in einheitlicher Auflösung gescannt wurden, ist es nicht mehr möglich, eine gemessene Pixelanzahl direkt in den tatsächlichen metrischen Wert der Urkunde umzurechnen. Anders ausgedrückt kann eine Erhöhung der Distanz von der Kamera des Fotografen gegenüber der Urkunde von wenigen Zentimetern das Ergebnis des Statistik-Tools um Dutzende Pixel verzerren. Die Übertragbarkeit der Messwerte zwischen den einzelnen Exemplaren geht damit verloren. Sollte das Tool allerdings auf ein Corpus angewandt werden, welches nach einheitlichen Kriterien direkt am Original digitalisiert wurde, entfällt diese Problematik und die erzeugten statistischen Werte können direkt verglichen werden. Im Fall des im Projekt „Schrift und Zeichen“ verwendeten Materials ist dies jedoch nicht durchgängig möglich. Ein möglicher Ausweg wäre es, sämtliche Messungen in Relation zu einem bekannten, in allen Urkunden gleichen Wert durchzuführen. Spätestens seit den Forschungen Frank M. Bischoffs ist jedoch bekannt, dass solch einheitliche Werte in den Urkunden des 11. und 12. Jahrhunderts völlig fehlen<sup>20</sup>.

Derzeit funktioniert das Statistikscript wie das Snippettool bislang noch ohne eine systematisch konzipierte Benutzeroberfläche. Das Ergebnis ist eine Textdatei, in der bestimmte Messungen und Relationen notiert sind. Dazu zählen neben der Zahl der Zeilen, deren durchschnittliche Höhe, die entsprechende Standardabweichung vom Durchschnittswert, das Verhältnis von Wortbreite zu Worthöhe, und einige weitere Messungen wie etwa zu den graphischen Symbolen. Messbar sind hier beispielsweise das Seiten-Höhen-Verhältnis der Rota, der Einzug der graphischen Symbole im Verhältnis zum Textkörper sowie Seitenränder und Abstände. ([Abbildung 7](#))

Viele dieser Werte mögen auf den ersten Blick wenig hilfreich erscheinen, können jedoch in einem größeren Zusammenhang paläographische Entwicklungen nachvollziehbarer gestalten. Als Beispiel kann die Relation der Worthöhe zur Wortbreite dienen. Bei einem grundsätzlich gleichen Wortpool kann die Entwicklung des Duktus im Laufe der Jahrzehnte gut nachvollzogen werden, wobei die Formelhaftigkeit der päpstlichen Urkundensprache in Bezug auf das zu untersuchende Quellenmaterial sich hier als vorteilhaft erwies. Die Streckung der Schäfte und die Betonung der Vertikale als grundsätzliche Merkmale der gotischen Schriften führen im Schriftduktus zu entsprechenden Veränderungen. So darf man schließen, dass mit stärker gestreckten, dafür jedoch weniger breiten Buchstaben der genannte

<sup>19</sup> Vgl. hierzu auch den Beitrag von Viktoria TRENKLE in diesem Band.

<sup>20</sup> Frank M. BISCHOFF, *Urkundenformate im Mittelalter. Größe, Format und Proportionen in Zeiten expandierender Schriftlichkeit (11.–13. Jahrhundert)* (Elementa diplomatica 5), Marburg an der Lahn 1996, S. 53–84.

Koeffizient von Wortbreite und -höhe sich signifikant ändert. In ein Diagramm eingetragen, ließe sich der soeben skizzierte Prozess graphisch darstellen, anstatt ihn nur mit Worten abstrakt zu beschreiben.

### **Zusammenfassung**

Digitale Analysemethoden verändern die paläographische Wissenschaft nachhaltig, wenn auch auf andere Weise als vielfach prognostiziert. Obwohl auf das Auge des menschlichen Forschers sowie dessen Korrektur- und Einfühlungsvermögen noch lange nicht verzichtet werden kann, ist die Paläographie mehr und mehr im Begriff, eine „Kunst des Messens“ zu werden, wie es der vielzitierte Paläograph Bernhard Bischoff bereits vor über 30 Jahren pointiert prophezeit<sup>21</sup>. Um auf diesem Felde zukünftig erfolgreich zu sein, bedarf es weitreichender und fortdauernder Kommunikation zwischen den einschlägigen Forschungsdisziplinen, Archiven und weiteren involvierten Institutionen, um den Digitalisierungsprozess vor allem hinsichtlich Geschwindigkeit, Bestandsauswahl und Qualitätsgarantie in sinnvoller Weise umzusetzen. Seitens der paläographischen wie der diplomatischen Forschung bedarf es zudem umfassender Überlegungen, wie, wann, an welchen Stellen und in welcher Form eigene Forschungsmethoden nützlich in digitale Analysestrukturen umgesetzt werden sollen. Nur so kann ein fruchtbarer Dialog über wissenschaftliche Disziplingrenzen hinaus, allem voran zwischen Geisteswissenschaften und der Informatik entstehen und Ergebnisse hervorbringen. Technische Hürden oder Irrwege stellen dabei keine Ausschlusskriterien dar, sondern müssen vielmehr als wichtige Etappen einer notwendigen Grundlagenforschung gesehen werden. Schließlich aber, trotz aller Zuversicht, eine eindringliche Warnung: Die Skepsis, die den digitalen Analysemethoden in der Paläographie oftmals noch entgegengebracht wird, beruht zum Teil auf zu hoch gesteckten oder auf Missverständnissen beruhenden Anforderungen und Akzeptanzhürden gegenüber dem Computer. Noch immer befindet sich die digitale Paläographie vielmehr in einer Konzeptionsphase, in der es effektive, sinnvolle und umsetzbare Forschungs- und Anwendungsfelder erst zu eruieren gilt.

---

<sup>21</sup> Bernhard BISCHOFF, Paläographie des römischen Altertums und des abendländischen Mittelalters (Grundlagen der Germanistik 24), Berlin <sup>4</sup>2009, S. 19.

**Abstract**

Vor dem Hintergrund der zunehmenden Digitalisierung geschichtswissenschaftlicher Forschungsarbeit skizziert der vorliegende Beitrag am Beispiel der schriftgeschichtlichen Untersuchung hochmittelalterlicher Papsturkunden neue und innovative Möglichkeiten, Strukturen und Methodiken der klassischen Paläographie effektiv und auf sinnvolle Art und Weise in technische Analysevorgänge zu übersetzen. Dabei soll jedoch deutlich vor Augen geführt werden, dass durch den wachsenden Einsatz EDV-gestützter Werkzeuge die Arbeit des „menschlichen“ Paläographen keineswegs ersetzt, zukünftig wohl aber modifiziert werden wird. Darüber hinaus sollen mit dem Aufzeigen der sich deutlich offenbarenden technischen Grenzen die sich zwischen euphorischem Fortschrittsstreben und kategorischer Ablehnung bewegenden Erwartungen an den Computer in der geisteswissenschaftlichen Forschungswelt ein Stück weit „humanisiert“ werden.

Abbildung 1

Dateiliste. Zeigt alle Bilder an, die in die Datenbank geladen wurden

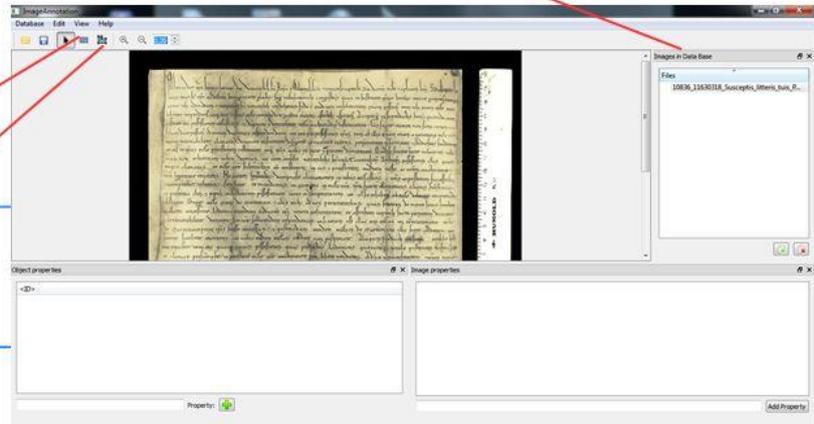
Rechteck erstellen/annotieren

Polygon erstellen /annotieren

Hauptfenster.  
Zeigt das Bild der Urkunde und Annotationen an.

„Object properties“: Kategorien bzw. Informationen für Annotationen

„Image properties“: Kategorien bzw. Informationen, die das gesamte Bild/Urkunde betreffen



[Zurück zum Text](#)

Abbildung 2

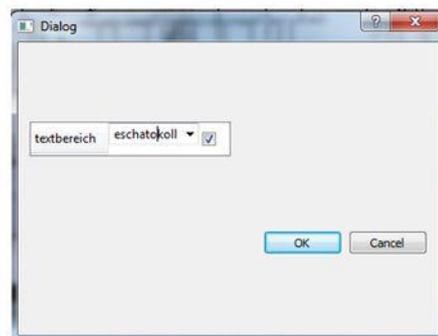
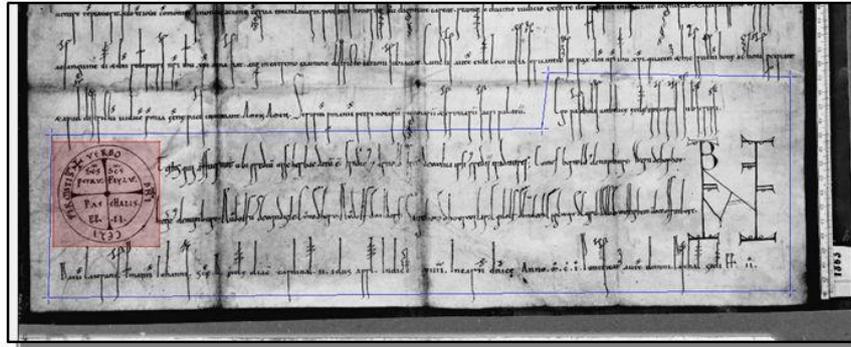
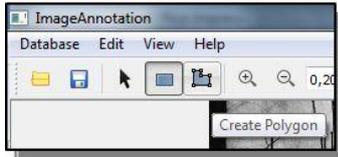


Abbildung 3



Legende:

blau=bestehende  
Annotation

rot=aktuelle  
Annotation

[Zurück zum Text](#)

Abbildung 4

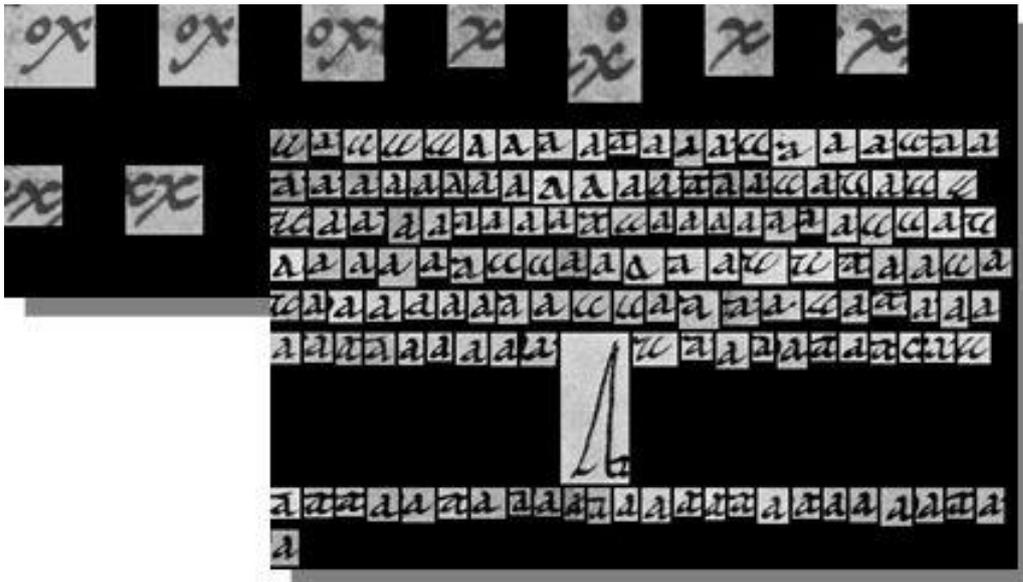
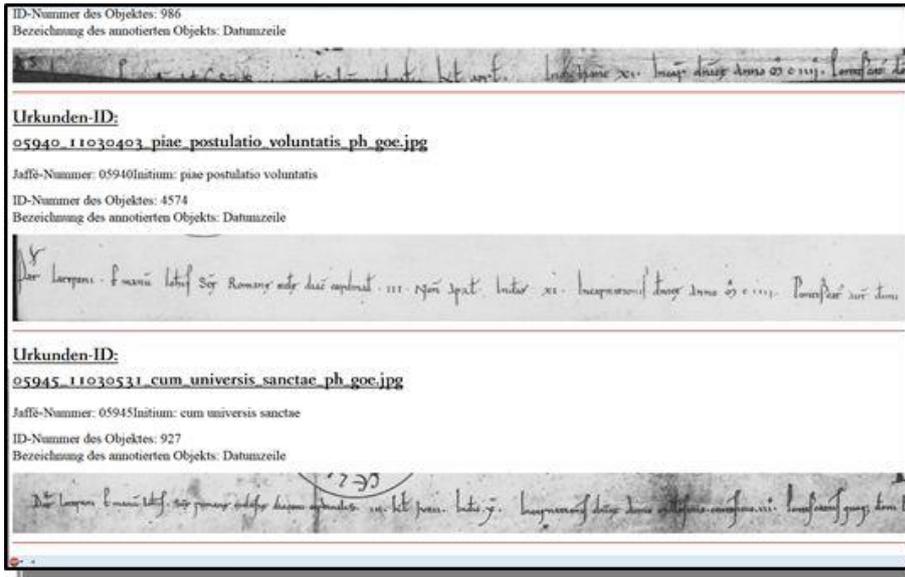
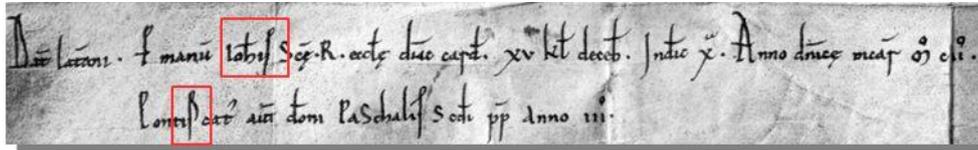


Abbildung 5

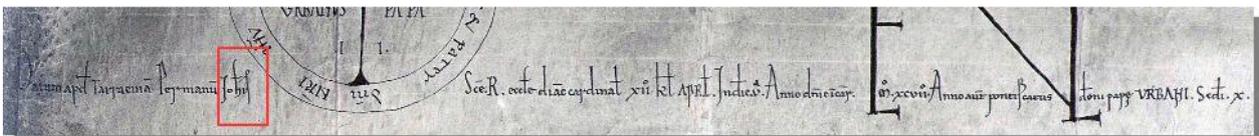


[Zurück zum Text](#)

Abbildung 6



Typisches Schriftbeispiel für Johannes v. Gaeta



Schriftbeispiel mit ähnlichen Formen (z.B. fi-Ligatur), aber gänzlich anderem Duktus – selber Schreiber?

Abbildung 7

```

statistics - Editor
Datei Bearbeiten Format Ansicht ?
06232_11090410_pfae_postulatio_voluntatis_ph_goe.jpg
- kontext
  lines: 18
  mean height: 62.1333333333
  standard deviation: 4.54410485012
- erste zeile
  height (relative to mean line height): 3.15450643777
- eschatokoll
  height (relative to mean line height): 10.0912017167
- benevalete
  height (relative to mean line height): 5.21459227468
  width-height ratio: 0.916666666667
  distance to right kontext edge (relative to mean line height): 3.49248927039
- rota
  height (relative to mean line height): 7.00107296137
  distance to left kontext edge (relative to mean line height): 0.289699570815
- komma
  not annotated / part of charter
- wort
  mean width-height ratio: 1.19610644005
- datumzeile
  height (relative to mean line height): 0.708154506438
- other
  distance rota - benevalete (relative to mean line height): 6.32510729614
  minimum symbol distance to kontext (relative to mean line height): 0.321888412017
05888_11010111_sicut_iniusta_poscentibus_ph_goe.jpg
- kontext
  lines: 27
  mean height: 79.375
  standard deviation: 3.26359336724
- erste zeile
  height (relative to mean line height): 2.7968503937
- eschatokoll
  height (relative to mean line height): 15.105511811
- benevalete
  height (relative to mean line height): 5.2031496063

```

[Zurück zum Text](#)