



Hanse.Quellen.Lesen! Die Spätzeit der Hanse gemeinsam entdecken

Transkriptionsrichtlinien

Grundlage jeder Transkription sind möglichst eindeutige Regeln, wie der Text in der Vorlage (dem historischen Dokument/der Archivalie) in moderne Schrift übersetzt werden soll. Hierbei gibt es verschiedene Herausforderungen, wie etwa Buchstaben, die auf einer modernen Computertastatur nicht vorkommen oder Korrekturen/Ergänzungen im Text, bei denen nicht ohne weiteres klar ist, wie man sie ‚abtippen‘ kann. Für viele dieser Probleme haben wir im Folgenden Regeln aufgeführt, um angesichts die vielen Transkribierenden möglichst einheitliche und für alle nachvollziehbare Ergebnisse zu gewährleisten.

Das übergeordnete Ziel der Transkriptionen ist, sie als Trainingsmaterial für eine machine-learning gestützte Handschriftenerkennung zu verwenden. Das bedeutet für die Transkriptionsrichtlinien, dass so nah wie möglich am tatsächlichen Buchstabenbestand der Vorlage, also der archivalischen Quelle transkribiert werden soll. Normalisierung und Anpassungen an modernes Deutsch sind daher die Ausnahme. Auch die Lesbarkeit für heutige Nutzer steht zunächst nicht im Vordergrund (entsprechende Überarbeitungen finden ggf. später statt).

Buchstabenbestand

- Grundsätzlich gilt: *Buchstabengetreu transkribieren, nicht nach Lautwert*. Ein ‚v‘ wird als ‚v‘ transkribiert, auch wenn es im Wort als ‚u‘ gesprochen würde. Umgekehrt gilt gleiches für ‚u‘ (Typische Beispiele: ‚vnd‘ statt ‚und‘ oder ‚zuuor‘ statt ‚zuvor‘)
- *Groß- und Kleinschreibung*: auch hier wird buchstabengetreu gemäß der Vorlage transkribiert, wobei grundsätzlich im Zweifel die Kleinschreibung verwendet wird. Großschreibung wird nur dann gewählt, wenn für den Buchstaben eine eindeutige, sich von dem Kleinbuchstaben unterscheidende Schreibweise vorliegt. Die relative Größe des Buchstabens im Vergleich zu seinem Umfeld spielt keine Rolle. Auch Namen etc. werden somit nur groß geschrieben, wenn dies der Vorlage entspricht.
- *Lateinische Schreibschrift*: wird nicht besonders markiert und regulär buchstabengetreu transkribiert. Wie bei der in den Dokumenten vorherrschenden Schrift, der Kurrentschrift, werden ‚u‘ und ‚v‘ getreu den Buchstaben in der Vorlage transkribiert, gleiches gilt für Groß- und Kleinschreibung.
- *Unklar oder Unleserlich*: unleserliche Stellen werden mittels des gleichnamigen Buttons in der Mittelleiste des Interfaces als ‚unclear‘ markiert. Unklare Zeichen



werden durch ‚\$‘ (Dollar-Zeichen) ersetzt. Im Idealfall sollte hierbei die ungefähre Anzahl der geschwärzten oder unleserlichen Zeichen der Zahl der eingefügten ‚\$‘ entsprechen. Auf diese Weise werden noch untranskribierte Stellen für zukünftige Nutzer*innen leicht auffindbar.

- *Doppelte Buchstaben* (insbesondere bei ‚t‘, ‚f‘ und ‚s‘, aber auch sonst): Werden ebenfalls vorlagengetreu transkribiert, auch wenn dies nicht der modernen Schreibweise oder dem Wortlaut entspricht bzw. wenn Buchstaben ausschließlich in doppelter Form auftreten und somit auch einfach geschrieben werden könnten.
- *Abkürzungen und Kürzel:*
 - Einfache Abkürzungen* (für einzelne Buchstaben oder Silben wie ‚en‘ am Wortende) werden stillschweigend aufgelöst, sofern es sich um einheitlich verwendete Kürzel mit eindeutiger Auflösung handelt. Dies gilt auch für das einem ‚q‘ oder ‚g‘ ähnelnde Kürzungszeichen für lateinisch ‚us‘. Faustregel: Eindeutige und einheitliche Abkürzungen für bis zu zwei Buchstaben können stillschweigend aufgelöst werden.
 - Komplexere Abkürzungen* wie Titel oder Anreden (bspw. ‚Erb.‘ für ‚Erbare‘, ‚Maytt:‘ für ‚Mayestät‘) werden nicht aufgelöst. Abkürzungszeichen wie ‚.‘ oder ‚:‘ hinter der Abkürzung werden übernommen, wie sie in der Vorlage vorkommen. Auch hier findet keine Normalisierung statt. Faustregel: Umfasst die aufgelöste Abkürzung (der Text in den eckigen Klammern) mehr als zwei Zeichen, wird die Abkürzung nicht aufgelöst.
- ‚y‘ / ‚ii‘ / ‚ij‘: Entgegen der üblichen Praxis, sich für die drei Zeichen(kombinationen) auf eine Schreibweise festzulegen, wird auch hier gemäß dem Buchstabenbestand der Vorlage transkribiert und keine Normalisierung vorgenommen (Bsp. ‚Maii‘, ‚bey‘). Zu beachten: Das ‚y‘ wird im 16. und 17. Jhd. im niederdeutschen Raum oftmals mit zwei Punkten versehen. Da dies nicht ohne weiteres über eine Standardtastatur mittels Transkribus Web Interface umsetzbar ist, kann hier der Einfachheit halber das reguläre ‚y‘ verwendet werden.
- ‚I‘ / ‚J‘ bzw. ‚i‘ / ‚j‘: Gerade für das moderne, großgeschriebene ‚I‘ bzw. ‚J‘ gibt es in der Regel keine zwei klar abgrenzbaren Entsprechungen, vielmehr findet derselbe Schrift-Buchstabe Verwendung. Da dies in bisherigen Texten so gehandhabt wurde, werden die ‚I‘ / ‚J‘ –Buchstaben als ‚I‘ transkribiert (etwa ‚Iohann‘ statt ‚Johann‘). Im Falle der Kleinbuchstaben lassen sich in einigen Fällen unterschiedliche Schreibweisen von ‚i‘ und ‚j‘ feststellen, in der Regel durch eine Unterlänge bei ‚j‘. Diese werden dann entsprechend unterschiedlich (als ‚i‘ oder ‚j‘) gemäß der Vorlage transkribiert.
- *Zahlen*: römische Zahlen werden wie sie in der Vorlage stehen transkribiert (etwa als ‚xvii‘), für gewöhnlich sind die Buchstaben hierbei klein geschrieben. Im Falle von Mischformen (14 als ‚i4‘ geschrieben) wird zur arabischen Ziffer normalisiert.
- *Währungen*: Grundsätzlich gilt auch hier, dass diese gemäß der Vorlage als Kürzel nicht aufgelöst werden. Gerade die für die lübische Mark verwendeten Kürzel sind allerdings oftmals nicht ohne weiteres in modernen Buchstabenbestand zu



überführen. Im Falle nicht buchstabengetreu zu transkribierender Kürzel soll daher wie folgt transkribiert werden: ‚ml‘ für Mark, ‚ßl‘ für Schilling und ‚dl‘ für Pfennig.

Zeichensetzung

- *Zeichensetzung allgemein*: Auch hier wird übernommen, was auf der Vorlage (eindeutig) erkennbar ist. Einheitlichkeit oder gar moderne Orthographie sowie das Einteilen in Sinnabschnitte in Form von Sätzen sind zu diesem Zeitpunkt nicht das Ziel der Transkription.
- *Worttrennung am Zeilenende*: Sofern ein Trennungszeichen vorliegt (etwa in Form zweier Kommata o. ä.) wird dieses als Gleichheitszeichen ‚=‘ transkribiert. Falls kein Trennungszeichen in der Vorlage erkennbar ist, wird auch keines gesetzt, auch wenn es sich offenkundig um eine Worttrennung handelt.
- *Zeichen, die auf Ergänzungen am Textrand verweisen* (oftmals einem ‚F‘ oder ‚^‘ ähnlich): werden vereinheitlicht als ‚*‘ transkribiert (wie früher im Schulaufsatz) und später in der Endredaktion durch entsprechende Sonderzeichen ersetzt. In Ausnahmefällen, falls etwa auf einer Seite mehrere Ergänzungen mit unterschiedliche Zeichen markiert werden, können weitere Sonderzeichen wie ‚#‘ oder ‚^‘ verwendet werden. Es sollten keine Zeichen verwendet werden, die – sei es auch in anderer Schreibweise bzw. Verwendung – regulär im Text vorkommen (können), wie etwa besagtes ‚F‘.
- *Streichungen (Unterstreichungen, Durchstreichungen)*: Eindeutige Streichungen im Text werden mittels der in der Mittelleiste der Bedienoberfläche zur Verfügung gestellten Textannotation-Buttons ausgewiesen.

Diese Liste ist vorläufig und wird ggf. im Laufe des Projekts erweitert. Sollten Ihnen beim Transkribieren Zeichen, Wendungen oder andere Textphänomene auffallen, deren Transkription unklar ist und die hier nicht behandelt werden, schreiben Sie uns gerne, möglichst unter Verweis auf konkrete Beispiele an: info@fgho.eu

