

# Text-Bild-Gefüge

## Digital Humanities und der Diskurs der Moderne

**Klemstein, Franziska**

f.klemstein@gmail.com

Bauhaus-Universität Weimar, Germany

Die DH gelten als interdisziplinäres Feld par excellence. Sie bringen geisteswissenschaftliche Disziplinen, wie etwa die Linguistik oder die Geschichtswissenschaft, in enge Verbindung mit Disziplinen wie Computer Science und Information Science. Das daraus entstehende Potential zur übergreifenden Zusammenarbeit zwischen einzelnen geisteswissenschaftlichen Disziplinen wird bereits vielfach genutzt. Dennoch stehen textfokussierte und bildzentrierte Projekte oft unvermittelt nebeneinander. Gleichwohl haben multimodale Untersuchungsmethoden derzeit Konjunktur (Barman et al 2021).

Zumeist findet dort jedoch eine Fokussierung auf sehr spezifische Einzelaspekte statt. So fokussierten Barman et al auf das Training weniger, exemplarisch ausgewählter Klassen, um ihre Netzwerke auf das Erkennen von Todesanzeigen innerhalb eines schweizerisch-luxemburgischen Korpus historischer Zeitungen zu trainieren. Bilder wurden hierbei als auffällige Layout-Bereiche (areas of interests) innerhalb eines strukturierten und sich wiederholenden Gesamtlayouts verstanden, nicht jedoch als eigenständige Bilder innerhalb eines Gesamtgefüges aus Text und Bild.

Unser Forschungsprojekt nimmt dies zum Anlass um Texte und Bilder als gleichwertige Elemente sowohl im Bereich CV als auch im Bereich des NLP zu untersuchen. Innerhalb unseres ersten Projektabschnitts wird jedoch vor allem auf die Layout-Analyse und die Untersuchung des Verhältnisses von Text und Bild zueinander fokussiert. Konkret sollen dabei im ersten Teilprojekt u.a. folgende Fragen beantwortet werden:

- 1.) Wie entwickelt sich in unserem Zeitschriftenkorpus die strukturelle Anordnung von Text und Bild?
- 2.) Welche Rolle spielen dabei Bildunterschriften und wie wird im Text auf die Bilder Bezug genommen?
- 3.) Welche Arten von Bildern und welche Bildinhalte werden in die Texte einbezogen und wann dominieren welche Bildarten in welchen Textgattungen bzw. Untersuchungsdomänen?

Diese hier formulierten Fragen sollen eine systematische Untersuchung größerer Text-Bild-Korpora ermöglichen und damit das Text-Bild-Verhältnis innerhalb der sogenannten Moderne untersuchen. Unsere Hypothese lautet, dass Veröffentlichungen, wie zum Beispiel die Bauhausbücher als Ausdruck und Zeugnis des Diskurses der Moderne verstanden werden, da sie sich in ihrem Text-Bild-Verhältnis in einem Spektrum bewegen, das zwischen einer Verwissenschaftlichung künstlerischer Publikationen bzw. der Kunstpraxis einerseits steht und andererseits experimentelle Layoutversuche offenbart, die die Zweidimensionalität der Seite grundlegend in Frage stellt. Ob diese Hypothese zutrifft und tatsächlich als charakteristisch oder prägend für den Zeitraum der Moderne bezeichnet werden kann, muss durch eine möglichst breite und dennoch tiefgehende Analyse verifiziert werden.

Um diese Hypothese nicht allein in Bezug auf die Bauhausbücher, sondern in einem größeren Kontext mit Blick auf die Veränderungen von Text-Bild-Gefügen innerhalb des Untersuchungszeitraums (1880–1930) untersuchen zu können, umfasst unser

Korpus Publikationen (Zeitschriften, Monographien u.v.m.) aus verschiedenen Themenbereichen. Exemplarisch sind hier zu nennen: die Bauhausbücher, die Zeitschrift „Das Kunstgewerbe“, die „Fliegenden Blätter“, die „Zeitschrift für Psychologie und Physiologie der Sinnesorgane“ oder auch das „Centralblatt der Bauverwaltung“.

Neben der Vorstellung der Untersuchungsdomänen und Textkorpora soll ebenso das methodische Vorgehen erläutert werden. So soll unter anderem der Annotationsprozess zur Erstellung des Datensatzes erläutert werden, auf dessen Grundlage dann zu einem späteren Zeitpunkt mithilfe des PubLayNet-Datensatzes (Zhong et al 2019) ein Mask R-CNN-Transferlernen (He et al 2017) für ein 19-Klassen-Problem (siehe Abb.1) durchgeführt werden soll. Da der Annotationsprozess selbst jedoch bereits zahlreiche nicht-triviale Entscheidungen in sich birgt, soll dieser hier besonders ausführlich beschrieben werden. Hierbei soll einerseits der Entscheidungsprozess für das Computer Vision Annotation Tool (CVAT) (Sekachev et al 2019) und gegen das Aletheia Document Analysis System (Clausner et al 2011) dargelegt werden sowie die Definition der zu annotierenden Klassen, die sich auch im Inter Annotator Agreement widerspiegeln und andererseits erste Analyseergebnisse vorgestellt werden, die auf der Grundlage der ersten Annotationen bereits möglich sind.

Die in CVAT annotierten Digitalisate können in verschiedenen Formaten exportiert und auf diese Weise für verschiedene weitere Anwendungsmöglichkeiten nutzbar gemacht werden. Innerhalb unseres Projektes wird das Format „COCO 1.0“ genutzt, d.h. das nach dem Export ein Ordner mit den Digitalisaten sowie einer json-Datei vorliegen. Diese Datei bildet die Grundlage für erste oberflächliche Analysen, die jedoch bereits erste aufschlussreiche Anhaltspunkte für die Untersuchung verschiedener Textgattungen im Zeitraum von 1880 bis 1930 zulassen. Zu diesen gehören unter anderem folgende Abfragen:

- a) Welche Annotationsklassen sind auf welchen Seiten zu finden?
- b) Welche Seiten haben wie viele Textfelder bzw. Textspalten?
- c) Welche Seiten haben wie viele Abbildungen?
- d) Welche Seiten haben Bildunterschriften?

Ebenso kann das Verhältnis von der spezifischen Klasse zur Gesamtseite oder im Verhältnis zu anderen Klassen ausgewertet werden und vieles mehr.

Durch die Beantwortung dieser Fragen können nicht nur erste Annahmen zur strukturellen Anordnung von Text und Bild systematisch analysiert und hinterfragt werden, sondern zugleich eine Vergleichsebene zwischen verschiedenen Untersuchungsdomänen vorgenommen werden. Im Zentrum steht dabei die Frage, inwiefern sich die von Daston und Galison vorgenommene Ausweitung der Foucaultschen Diskursanalyse auf die (wissenschafts-)historische Untersuchung von Bilddiskursen (Daston/Galison 2007) durch den Gebrauch von verschiedenen digitalen Technologien konkretisieren und damit präzisieren lässt.

Labelclass	Short-Definition
image	is described by 2 categories: a) media type, b) content type
logo	is a graphically designed sign
editorial note	is every information that originates from the editor/publisher
decoration	is a artistic or structuring element
frame	is the border of an advertisement, image, text or page
footer	is a specific area of a page and follows the main text, or body
footnote	is a note or further explanation by the author
page number	is a consecutive numbering
header	is detached from the main text at the top of the text page
dropped capital	is a letter at the beginning of a word that is larger than the rest of the text and is often ornately decorated
heading	is the name for a book or an article within a journal.
subheading	is the name for a section or the sub-head of the heading
author	is the creator or originator of any written work
text	is a thematically and / or functionally oriented, coherent linguistic or linguistic-figurative complex
noise	is an element subsequently added to the published work
column title	is a text usually placed at the top - in the header - of the page or column
advertisement	is a page-region and can include decorations, text, images, logos, frames
caption	is either an explanatory text nearby the image/table eventually including a reference number; a reference number nearby the image/table without any explanatory text; a reference number within the main text or a reference number within the main text, referencing to multiple images
table	is an ordered arrangement of information or data, typically in rows and columns, or possibly in a more complex structure

Abb. 1.

## Bibliographie

**Barman, R. / Ehrmann, M. / Clematide, S. / Oliveira, S. A. / Kaplan, F.:** Combining Visual and Textual Features for Semantic Segmentation of Historical Newspapers, in: Journal of Data Mining & Digital Humanities, DOI: 10.46298/jdmhdh.6107 , arXiv:2002.06144.

**Clausner, C. / Pletschacher, S. / Antonacopoulos, A.:** Alet-heia - An Advanced Document Layout and Text Ground-Tru-thing System for Production Environments, in: International Conference on Document Analysis and Recognition, 2011, pp. 48-52, doi: 10.1109/ICDAR.2011.19.

**Daston, L. / Galison, P.:** Objektivität, Frankfurt/Main 2007.

**He, K. / Gkioxari, G. / Dollár, P. / Girshick, R.:** Mask R-CNN, arXiv:1703.06870.

**Sekachev, B. / Zhavoronkov, A. / Manovich, N.:** Computer Vision Annotation Tool: A Universal Approach to Data Annotation, 2019, <https://software.intel.com/content/www/us/en/develop/articles/computer-vision-annotation-tool-a-universal-approach-to-data-annotation.html> [letzter Zugriff: 13.07.2021].

**Zhong, X. / Tang, J. / Yepes, A. J.:** PubLayNet: largest dataset ever for document layout analysis, arXiv:1908.07836.