



# Kompositionalität von Bedeutung

Ein kunsthistorischer Blick, informatisch und  
computerlinguistisch gelenkt

Peter Schmidt

Auszug aus dem Jahresbericht  
„Marsilius-Kolleg 2015/2016“



# Kompositionalität von Bedeutung

## Ein kunsthistorischer Blick, informatisch und computerlinguistisch gelenkt

---

„Bedeutung“ ist ein großes Wort. In den Geisteswissenschaften ohnehin. Seiner gefühlten Schwere steht die Tatsache gegenüber, dass keineswegs selbstverständlich ist, worüber man genau spricht, wenn man es benutzt – konkret: worüber Vertreter verschiedener Fächer und Denktraditionen sprechen, wenn sie mit dem Begriff arbeiten.

Genau das aber hatten wir uns vorgenommen, als wir – Björn Ommer als Informatiker, Stefan Riezler als Computerlinguist und ich als Kunsthistoriker – uns über Kompositionalität von Bedeutung auszutauschen begannen. In den Kulturwissenschaften im Allgemeinen und in der Kunstgeschichte im Speziellen ist „Bedeutung“ ein zentraler Begriff. Sieht doch die Kunstwissenschaft seit ihren Anfängen als akademisch institutionalisierte Disziplin ihre Aufgabe nicht zuletzt im *Deuten*. Das setzt die Annahme voraus, dass sich die Bedeutung historischer Artefakte in zeitlicher und/oder kultureller Distanz keineswegs mehr selbstverständlich (im Wortsinn) erschließt. Sie muss in der Regel rekonstruiert bzw. decodiert werden, so die Übereinkunft.

In der Methodengeschichte der Kunstwissenschaft ist der Begriff „Bedeutung“ verbunden mit dem Werk eines Gelehrten, dessen Verdienste unter anderem darin liegen, seinem Fach seit den 1930er Jahren ein neues Methodenbewusstsein und damit ein geschärftes Profil unter den Geisteswissenschaften vermittelt zu haben. Die letzte Fassung seines mehrfach überarbeiteten Aufsatzes, dessen erste Fassung 1932 unter dem Titel *Zum Problem der Beschreibung und Inhaltsdeutung von Werken der bildenden Kunst* erschienen war, publizierte der von den Nationalsozialisten ins

amerikanische Exil getriebene Erwin Panofsky 1955 in einem Sammelband mit dem bezeichnenden Titel *Meaning in the Visual Arts*. Erstmals wurde damit ein strukturiertes Modell zur Analyse von „Bedeutung“ vorgeschlagen. Es gehört bis heute zum propädeutischen Lehrstoff im universitären Studium der Kunstgeschichte.

Panofsky formulierte ein dreistufiges Schema: Die erste Schicht bei der Beschreibung eines Bildes, die er den „Phänomensinn“ nennt, ist auf der Basis „vitaler Daseins-erfahrung“ zu leisten, also durch das Erkennen von dem Betrachter bekannten Gegenständen und Handlungen. Auf einem spätmittelalterlichen Gemälde könnte man etwa eine Frau mit Krone sowie einem Schwert und einem Rad in den Händen erkennen und beschreiben. Damit jedoch hätte man nicht die Schicht der Bedeutung erfasst, die das Bild einer solchen Frau für den Produzenten und Rezipienten des 15. Jahrhunderts relevant gemacht hatte. Diese zweite Schicht wäre der von Panofsky so genannte „Bedeutungssinn“: Um ihn zu erfassen, bedarf es Informationen, die außerhalb des Bildes selbst liegen – vor allem in textlichen Quellen. Mit Hilfe der Kenntnis von Heiligenlegenden wäre zum Beispiel festzustellen, dass es sich bei der gekrönten Frau um eine heilige Katharina handeln muss, in deren Martyrium nämlich das versuchte Rädern und schließlich die Enthauptung mit einem Schwert die zentrale Rolle spielte. Auf diese Schicht setzt schließlich Panofskys „Dokumentsinn“ oder „Wesenssinn“ als dritte Stufe der Inhaltsdeutung auf. Was die Verehrung dieser Heiligen in diesem konkreten Gemälde in seinem spezifischen historischen Kontext bedeutet hatte, wäre darunter etwa zu fassen.

Was notwendig ist, um den „Bedeutungssinn“ aufzudecken, benennt Panofsky ganz klar: literarisches Wissen. Erst der Text der Heiligenlegende lässt im Rad mehr erkennen als ein defunktionalisiertes Wagenrad, nämlich ein Zeichen ihres Martyriums, und erst das lässt die Frau als eine konkrete Person identifizieren, was wiederum erst den ganzen religiösen Kontext öffnet und die Wirkmacht dieses Bildes für den mittelalterlichen Betrachter erklärt.

Der Begriff „Bedeutung“ ist damit in der Disziplin Kunstgeschichte mit der Vorstellung des Decodierens verbunden. Das Rad verweist eben nicht auf seinen herkömmlichen alltäglichen Kontext, sondern der von einem Text – der Heiligenlegende – gelieferte Schlüssel führt zu der als relevant erachteten Sinnschicht. Das Verständnis (oder Missverständnis) von Panofskys Modell als Decodiermaschine, mit der man als grundsätzlich codiert aufgefassten mittelalterlichen und frühneuzeitlichen Artefakten

ihre „hidden meaning“ entlocken kann, hat seinem Urheber im späten 20. Jahrhundert verstärkt Skepsis eingebracht.

Angesichts dieser Verständnistradition von „Bedeutung“ erwies es sich als produktiv, im Kontakt mit den Kollegen aus dem Bereich des maschinellen Text- und Bildverstehens auf eine basale Ebene der Konstitution von „Bedeutung“ zurückzugehen. Denn schon „Rad“ ist aus der Perspektive der Computer Vision eine Bedeutung, die ein Computer mit Hilfe von Algorithmen und Lernprozessen in einer Ansammlung von Bildpunkten in bestimmten Formen erst einmal erkennen muss. Bedeutung konstituiert sich nach dem von dem Mathematiker und Philosophen Gottlob Frege (1848-1925) formulierten Prinzip grundsätzlich kompositionell: Die Bedeutung eines aus Teilausdrücken zusammengesetzten Ausdrucks (oder aus visuellen Elementen zusammengesetzten Bildes) wird durch die Bedeutungen seiner Einzelteile und die Art ihrer Zusammenfügung bestimmt.

Ausgehend von diesem Prinzip der Kompositionalität von Bedeutung lässt sich versuchsweise eine Anregung aus der Computer Vision und der Computerlinguistik an einem mittelalterlichen Artefakt wie dem als Beispiel genannten Gemälde einer heiligen Katharina erproben: die Verwendung von textuellen und visuellen Phrasen bei der Detektion von Bedeutung. Eine visuelle Phrase kann aus nur zwei visuellen „Wörtern“ bestehen, die in signifikanter Kombination auftauchen. Ein Algorithmus kann bei einer bestimmten Konstellation von Formen auf einem Foto die Bedeutung „Fahrrad“ erkennen, bei einer anderen die Bedeutung „Mensch“. Beide tauchen auffällig häufig zusammen auf – nämlich als Fahrradfahrer –, ihre Bedeutung ist dann kompositionell eine neue und mehr als die Addition ihrer Einzelelemente. Computer Vision macht sich solche visuellen Phrasen zunutze, um Objekterkennung zu verbessern (vgl. Beitrag von Stefan Riezler in diesem Band).

Zum Beispiel der heiligen Katharina zurückgekehrt, kann die Konstitution von Bedeutung von bildlichen Artefakten aus dieser Perspektive anders angegangen werden, als nach dem herkömmlichen Panofskyschen Verfahren: Die visuellen Wörter Frau, Krone, Rad und Schwert bilden eine visuelle Phrase, die in dieser Komposition häufig in einem historischen Bildcorpus zu finden sein wird. Gleichzeitig sind die Wörter dieser visuellen Phrase auch in größer komponierten visuellen wie auch textuellen Phrasen zu detektieren: etwa in narrativen Bildern oder Bilderzyklen zum Martyrium der heiligen Katharina und in Texten wie den erwähnten Heiligenlegenden.

Panofskys Verfahren der Bedeutungsfindung setzt eine nicht näher systematisierbare Kenntnis eines nicht genau eingrenzbaaren Corpus außerbildlicher (Text-)Quellen als Schlüssel voraus – man mag das Bildung des Interpreten nennen. Versuche automatisierter Detektion von ikonographischer Bedeutung können nicht mit humanistischer Bildung arbeiten, wohl aber – so die gegenüber den Computer Sciences ausgesprochene Hoffnung des Kunsthistorikers – mit dem Versuch, Entsprechungen bestimmter visueller Phrasen in großen Bildcorpora zu finden und zu textuellen Phrasen in Textcorpora in Beziehung zu setzen.



Illustrierte Heiligenlegenden haben sich als ein gutes Studienfeld erwiesen, visuelle Phrasen in der Praxis der Bildkomposition zu analysieren. Heiligenlegenden sind stark standardisierte Texte, und sie sind häufig modular aufgebaut: Bestimmte narrative Strukturen wiederholen sich variierend bei verschiedenen Heiligen. Folglich tauchen in den Illustrationen relativ ähnliche visuelle Phrasen in den Bildern verschiedener Heiliger auf. Gleichzeitig bestehen die Bilder in solchen spätmittelalterlichen Handschriften und Drucken oft aus visuellen Mikroerzählungen, auf deren narratologische Analyse das Konzept visueller Phrasen in dem von den Computer Sciences vorgedachten Sinne noch übertragen werden könnte bzw. sollte – eine Perspektive, die die Diskussionen in unserer Gruppe im Marsilius-Kolleg erst aufgezeigt haben.

Die Tatsache, dass ich als Wissenschaftler, der sich ausgerechnet mit mittelalterlicher Kunst befasst, schon vor einigen Jahren mit einem Spezialisten der Computer Vision wie Björn Ommer in Kontakt gekommen bin, mag auf den ersten Blick vielleicht verwundern. Auf den zweiten hat es seine Logik: In der Kunst des Mittelalters gibt es einen vergleichsweise hohen Normierungsgrad und ein ausgeprägtes bildsystematisches Denken, das die Objekte dieser Zeit als Trainingsmaterial für maschinelle Bilderkennung besonders geeignet machen. Weiterer Standardisierungsdruck auf die Bilder dieser Zeit kann medial bedingt sein: Die Durchsetzung des Buchdrucks hatte aus technischen wie auch rezeptionspraktischen Gründen eine Tendenz zu einfachen

und klar strukturierten Illustrationen in Büchern zur Folge. Das hat uns auch dazu geführt, den illustrierten Druck eines Heiligenlebens aus dem Jahr 1488 als Anschauungsmaterial heranzuziehen, um das Operieren mit dem Konzept visueller und textueller Phrasen, das ich von meinen beiden Arbeitsgruppenpartnern gelernt hatte, zu diskutieren.

## Fazit

Die Anwendung automatischen Text- und Bildverstehens spielt unter dem Stichwort *digital humanities* für die Geisteswissenschaften in den letzten Jahren eine immer größere Rolle, um auch in diesem Bereich *big data* zu erschließen oder überhaupt beherrschbar zu machen. Unsere aus Kunstgeschichte, Computer Vision und Computerlinguistik zusammengesetzte Marsilius-Gruppe zielte aber dezidiert nicht unmittelbar auf praktische Anwendungen in diesen Sinne. Kulturwissenschaften und Computerwissenschaften stehen sich im Rahmen der *digital humanities* allzu häufig im Verhältnis bzw. Verständnis von Materialanbietern und Dienstleistern gegenüber. Uns war es dagegen wichtig, in Hinblick auf die Grundlagen des Verstehens unterschiedlicher Denk- und Herangehensweisen unserer Fächer voneinander zu lernen. Konkret anwendungsbezogene Perspektiven in Hinblick auf automatisierte Bedeutungserschließung wurden dabei ganz nebenbei erkennbar. Eine gemeinsam von Björn Ommer und mir betreute Dissertation mag ein Indiz dafür sein. Profitiert habe ich persönlich nicht nur von diesem so intensiven wie freundschaftlichen Austausch innerhalb der Kompositionalitäts-Gruppe, für den das Marsilius-Kolleg den Rahmen geschaffen hat. Die wöchentlichen Treffen aller Marsilius-Fellows haben mich durch die von den beiden Direktoren manchmal schwer zu bändigende Diskussionslust mit Widerspruch und Anregungen aus unerwarteten Richtungen konfrontiert, mit denen ich in keinem anderen Forum konfrontiert worden wäre.