



Jingcao Chen und Jing Li*

Das multimodale Chinesen-Image im deutschen Dokumentarfilm *Die Neue Seidenstraße – Chinas Griff nach Westen*

The Multimodal Chinese-Image in the German Documentary *The New Silk Road – China's Move on the West*

<https://doi.org/10.1515/ifdck-2023-0011>

Zusammenfassung: Deutschland und China gelten als wichtige Knotenpunkte der Neuen Seidenstraße und werden aufgrund ihrer wirtschaftlichen Bedeutung häufig zum Gegenstand der Medienberichterstattung bezüglich der Neuen Seidenstraße. Trotz Anerkennung der Relevanz von der Multimodalität beim Prozess der Sinnbildung bzw. Bedeutungskonstruktion beschäftigen sich die meisten vorangegangenen Untersuchungen zur Imagekonstitution im Zusammenhang mit der neuen Seidenstraße mit schriftlichen Texten (vor allem mit Medienberichten). In dem vorliegenden Beitrag wird der deutsche Dokumentarfilm *Die Neue Seidenstraße – Chinas Griff nach Westen*, der mit verschiedenen Modalitäten arbeitet, als Untersuchungsgegenstand ausgewählt und auf das multimodal konstituierte Chinesen-Image untersucht. Von relevanten Ansätzen der multimodalen Analyse und linguistischer Diskursanalyse ausgehend versucht dieser Beitrag, die gängigen Herangehensweisen im Rahmen der Erhellung von Sprachgebrauchsmustern auf die Bildmodalität zu übertragen und dadurch ein sprache-bild-integratives Untersuchungsmodell zu erarbeiten. Aus der Untersuchung ergeben sich insgesamt 4 durch die Zusammenwirkung von Sprach- und Bildmodalität konstituierte Imageaspekte der Chinesen: ‚greifende Fremde¹ ‚Gewinner bei der Zusammenarbeit‘

1 In diesem Beitrag werden die Imageaspekte von Chinesen durch einfache Anführungszeichen gekennzeichnet.

Article Note: Dieser Beitrag ist im Rahmen des Forschungsprojektes von Social Science Fund of Beijing „Eine korpusbasierte kontrastive Diskursanalyse über Chinas Seidenstraßen-Initiative“ [北京市社会科学基金项目《基于语料库的中德“一带一路”话语对比研究》] (Projektnummer 18YYC015) entstanden.

***Korrespondenzautorin: Assoc. Prof. Dr. Jing Li**, School of German Studies, Beijing Foreign Studies University, Xisanhuan Beilu 2, 100089 Beijing, China. E-Mail: lijing@bfsu.edu.cn

Jingcao Chen, School of German Studies, Beijing Foreign Studies University, Xisanhuan Beilu 2, 100089 Beijing, China. E-Mail: cjingcao@bfsu.edu.cn. <https://orcid.org/0009-0006-1370-608X>

„fleißige Opfer“ und „Leiter bei der Zusammenarbeit“. Diese Arbeit zeigt auch die methodische Relevanz, bei der Imagekonstruktion verschiedene Modalitäten zu berücksichtigen.

Stichwörter: Multimodalität, Dokumentarfilm, Chinesen-Image, Neue Seidenstraße

Abstract: Germany and China are considered important hubs of the New Silk Road and have garnered significant media coverage due to their economic importance in relation to the New Silk Road. Despite acknowledging the relevance of multimodality in the process of meaning-making and image construction, most previous studies on image constitution in the context of the New Silk Road primarily focus on written texts, especially on media reports. In the present article, the German documentary *The New Silk Road – China’s Move on the West*, which employs various modalities, is selected as the subject of investigation for the examination of its multimodally constituted Chinese image. Building on relevant approaches in multimodal analysis and linguistic discourse analysis, this article seeks to apply common methods used to illuminate language usage patterns to the realm of visual modality, thereby developing a language-image-integrated investigative model. From the study, a total of four image aspects of the Chinese, constituted through the interaction of language and visual modality, are identified: „gripping strangers“, „winners in cooperating“, „hard-working victims“ and „leaders in cooperating“. This study also highlights the methodological significance of considering various modalities in the image construction.

Keywords: multimodality, documentary, image of Chinese, the New Silk Road

1 Einleitung

Deutschland hat einen wichtigen geographischen und wirtschaftlichen Status für die Entwicklung der von China im Jahr 2013 eingeführten Seidenstraßen-Initiative. Seitdem wurden diese Initiative sowie der Initiator China vielfach in verschiedenen Medien thematisiert. Die mediale Präsenz einer Nation gilt „als nicht nur wissenschaftlich-theoretisch interessante Fragestellung, sondern auch als politisch höchst relevant“ (Thimm 2017: 29). Dementsprechend haben zahlreiche Wissenschaftler ihr Augenmerk auf Medienberichte in Bezug auf die Neue Seidenstraße gerichtet (vgl. Gu/Zhao 2018; Wang 2019).

Trotz der hohen Anzahl der Untersuchungen über Konstitutionen der Neuen Seidenstraße in den deutschen Medien werden multimodale Texte, die aufgrund

der digitalen Entwicklung von Massenmedien eine immer wichtigere Rolle bei der Wissensvermittlung und Wissenskonstruktion spielen sollten (vgl. Stöckl 2004: 7; Opilowski 2013: 218), kaum berücksichtigt. Ein Dokumentarfilm gehört zu solchen multimodalen Medienformen. Er arbeitet mit verschiedenen Modalitäten und gilt als wichtige Wissensquelle der Bevölkerung. Eine Untersuchung von Dokumentarfilmen könnte zur Bereicherung der medialen Konstruktion über die Neue Seidenstraße in Deutschland beitragen. Vor diesem Hintergrund konzentriert sich die vorliegende Studie auf den 2019 vom ZDF gedrehten zweiteiligen Dokumentarfilm *Die Neue Seidenstraße – Chinas Griff nach Westen*. Untersucht werden vor allem die sprachlich und bildlich realisierten Zuschreibungen zu den aufgenommenen Chinesen in diesem Dokumentarfilm. Im Mittelpunkt stehen die folgenden Fragestellungen:

- 1) Wie lässt sich aus dem zu untersuchenden Dokumentarfilm durch ein sprachwissenschaftliches Verfahren ein integratives Chinesen-Image aufgrund von Sprach- und Bildmodalität ermitteln?
- 2) Was für ein integratives multimodales Chinesen-Image lässt sich aus dem zu untersuchenden Dokumentarfilm herausarbeiten?

Um diese Forschungsfragen zu beantworten, wird auf Basis der theoretischen Prämissen von multimodalen Forschungen sowie der linguistischen Imageanalyse ein multimodales Untersuchungsmodell erarbeitet, das das Chinesen-Image in beiden Modalitäten ermittelt und zu einem integrativen Image verbindet.

2 Theoretische Prämissen

2.1 Schlüsselbegriffe in multimodaler Forschung

Mit der Entwicklung der modernen Kommunikationstechniken steht nicht nur die Sprache im Zentrum der Texte bzw. der Diskurse. Die Texte aus Zusammenschluss mehrerer unterschiedlicher Zeichenmodalitäten wie Sprache, Bild, Musik und Geräusch mit innerer Kohäsion und Kohärenz treten in den Vordergrund der Kommunikation (vgl. Stöckl 2016: 20). Dementsprechend entstehen in den 1980er bis 1990er Jahren die multimodalen Forschungen mit theoretischen Fundamenten des sozial-semiotischen Ansatzes, die von Kress/van Leeuwen und O’Toole aus Hallidays System-Funktionaler Grammatik entwickelt wurden (vgl. O’Halloran 2011: 122).

Trotz der weltweiten Entwicklung werden der Begriff *Modalität* und die Kategorisierung von Modalitäten seit Beginn der multimodalen Analyse nur diffus abgegrenzt. Der vorliegenden Studie wird das Konzept *Zeichenmodalität* von Stöckl

(2016) zugrunde gelegt, das drei Hauptfacetten von verschiedenen Modalität-Definitionen enthält. Eine Zeichenmodalität, kurz Modalität genannt, lässt sich gleichzeitig aus semiotischer (als Zeichen), psychologischer (als Modus bzw. Wahrnehmungskanal) und medialer (als Medium) Perspektive interpretieren (vgl. Stöckl 2016: 6). Die Modalität Sprache gehört beispielsweise zum sprachlichen Zeichensystem, lässt sich durch verschiedene Medien – aus dem Mund gesprochen oder aufs Papier geschrieben – realisieren und kann daraufhin visuell oder auditiv wahrgenommen werden.

Nach Stöckl (2016: 6) gibt es vier zentrale Modalitäten, welche Sprache, Bild, Musik und Geräusch umfassen. Ein multimodaler Text gilt als „Zusammenschluss mehrerer unterschiedlicher Zeichenmodalitäten zu einem kohäsiven und kohärenten Ganzen“ (Stöckl 2016: 20). Im Dokumentarfilm, der den Untersuchungsgegenstand der vorliegenden Arbeit bildet, werden mehrere zentrale Zeichenmodalitäten zusammengesetzt. Die Bedeutung des Dokumentarfilms wird insofern ebenfalls aus Zusammenspiel verschiedener Modalitäten generiert (vgl. Luginbühl 2011: 259), indem sie sich wechselseitig in einer „Reißverschluss-Relation“ (Holly 2016: 394) transkribieren. In der vorliegenden Studie werden die zwei wichtigsten Modalitäten Sprache und Bild untersucht. Dargelegt werden soll, wie diese beiden Modalitäten gemeinsam zur Bedeutungsgenerierung beitragen und sich zusammen zu einem integrativen Image der Chinesen im untersuchten Dokumentarfilm verbinden.

2.2 Sprachgebrauchsmuster und Imageanalyse im Rahmen der linguistischen Diskursanalyse

Die vorliegende Studie versteht sich als eine linguistische Diskursanalyse (LDA), die sich auf verschiedene Modalitäten stützt, und schließt sich unmittelbar der linguistischen Imageanalyse (LIma) an (Vogel 2010a, 2010b).

Die linguistische Diskursanalyse versteht sich als ein Versuch, die semantische Tiefenstruktur des Diskurses hinter der Oberfläche des Zeichengebrauchs zu beschreiben und dadurch die sozialen Strukturen zu entlarven (vgl. Warnke/Spitzmüller 2008: 22). Verschiedene methodische Herangehensweisen der linguistischen Diskursanalyse haben eins gemeinsam, nämlich, das Musterhafte des Zeichengebrauchs durch die Analyse zu gewinnen, denn das Musterhafte kann die gesellschaftliche Kognition widerspiegeln (vgl. Klug 2016: 178). Nach Bubenhofer (2009: 7) ist die Gewinnung von Sprachgebrauchsmustern die Basis der Diskursanalyse. Das kann auf die These des Kontextualismus von Firth, welcher plädiert, dass die Bedeutung eines Wortes in seinem Kontext liegt, zurückgeführt werden (vgl. Firth 1957: 7, zit. n. Wei 2008: 29). Der Kontext bezieht sich auf die mit diesem Wort zusammen vorkommenden Wörter. Solche Wortkombinationen werden in einer Gesellschaft

so häufig wiederkehrend benutzt, dass sie allmählich konventionalisiert werden. Diese konventionalisierten Sprachgebrauchsmuster können „als Indikatoren für mediale Zuschreibungsmuster“ (Vogel 2010b: 92) dienen. Die durch Massenmedien weit verbreiteten und wiederholten Zuschreibungen fließen schließlich in Stereotypen von einer Sozialgruppe über einen Sachverhalt oder eine andere Sozialgruppe ein (vgl. Vogel 2010a: 348). Wechselseitig spiegeln sich die „Schematisierung und Pauschalisierung von stereotypem Zuschreibungswissen“ (Vogel/Haberer 2017: 230) auch in den wiederkehrenden sprachlichen Mustern wider (vgl. Vogel/Haberer 2017: 230).

Aufgrund der Sprachgebrauchsmusteranalyse hat Vogel eine Methode zur linguistischen Imageanalyse entwickelt. Nach ihm ist Public-Image (auch Medienimage) „Teil global-diskursiver Ereignisse, die sich in Form typischer, d. h. wiederkehrender, ko(n)textsensitiver Sprachmuster manifestieren“ (Vogel 2010a: 350). Durch das Sprachgebrauchsmuster wird ein Referenzobjekt mit bestimmten Attributionen und Prädikationen bzw. Eigenschaftszuschreibungen verknüpft (vgl. Vogel 2017: 49). Aus dem Zusammenschluss wiederkehrender Zuschreibungen lassen sich Themenfelder und Imageaspekte hinsichtlich des Referenzobjektes induktiv erschließen (vgl. Li 2017: 84–85).

2.3 Von Sprachgebrauchsmuster zu Bildgebrauchsmuster

Anders als die gängigen Imageanalysen, die sich vor allem mit Sprachgebrauchsmustern beschäftigen, versucht die vorliegende Studie, die Gewinnung vom Musterhaften von Sprache auf Bild auszudehnen.

Nach Klug (2016: 179) kann das „Muster gesellschaftlichen Wissens auch durch Formen anderer Zeichenmodalität als der Sprache realisiert werden, z. B. durch das Bild“, weil die verschiedenen Zeichensysteme von den gleichen Verstehungsmechanismen abhängen (vgl. Große 2011: 33). Die Notwendigkeit und Durchführbarkeit der Untersuchung vom visuellen Musterhaften wurde bereits bestätigt:

“The idea that concept co-occurrence is an important source of information about the relationships between concepts – can be successfully extended from the language domain to non-verbal visual experience.” (Sadeghi et al. 2015: 58)

Die Kookkurrenz von Objekten im Bild wird in dieser Arbeit in Anlehnung an Sprachgebrauchsmuster als *Bildgebrauchsmuster* genannt. Bubenhofer (2008: 408) definiert die Sprachgebrauchsmuster als „rekurrentes Auftreten von textuellen Einheiten in bestimmten Sprachausschnitten“. Die sogenannten *textuellen Einheiten* beziehen sich bei der linguistischen Imageanalyse vor allem auf die Wortkombination bzw. Kollokation von Wörtern im Umfang eines Satzes (vgl. Vogel 2010a:

353; Bubenhofer 2009: 117). Um die Übertragung von Sprachgebrauchsmuster auf Bildgebrauchsmuster zu ermöglichen, müssen daher zuerst operationalisierbare Einheiten, die der sprachlichen Wort- bzw. Satzebene entsprechen, in der Bildmodalität identifiziert werden. Daraufhin sind das Annotieren der bildlichen Elemente und die Berechnung der Frequenzen von den Kookkurrenzen verschiedener Bildelemente in gewissem Maß operierbar.

Große (2011: 55) hat die vorhandenen Gliederungsmöglichkeiten von bildlichen Ebenen zusammengefasst und die strukturellen Gliederungsebenen von dem statischen ikonischen bildlichen System entwickelt, wobei das sprachliche System als Referenzvorlage gilt. Laut ihr stellen sich *Figura* und *ikonisches Bildsystem* als Pendant jeweils zu *Wort* und *Satz* dar (vgl. Große 2011: 54–57). *Figura* stammt aus dem Lateinischen und bedeutet „das in sichtbare Form Gefasste“ (Große 2011: 54), das sich auf Referenz (*wer/was*) und Prädikation (*wie/was tun*) des propositionalen Aktes einer Handlung bzw. einer Bildhandlung bezieht (vgl. Große 2011: 54). Durch die Zusammensetzung von mehreren *Figuraen* entsteht ein Bildsystem bzw. ein ganzes Bild eines Ereignisses (vgl. Zhang 2009: 29). Aufgrund dieser Entsprechung von sprachlichen und bildlichen Gliederungsebenen und in Anlehnung an die Definition von Sprachgebrauchsmuster kann das Bildgebrauchsmuster bei statischen Bildern als *rekurrentes Auftreten von Kombination bzw. Kollokation von Figuraen (bildlichen Elementen) in bestimmten Mengen von Bildern* definiert werden.

Bilder lassen sich wiederum in statische und bewegte einteilen. Der zu untersuchende Dokumentarfilm ist eine Form des bewegten Bildes. Bei einem Film werden Einzelbilder, die *Figurae* wie Personen oder Gegenstände enthalten, zu einer Einstellung verbunden (vgl. Mikos 2015: 82) und als Ganzes von den Zuschauern wahrgenommen. Diese Einzelbilder haben eine relativ stabile Bedeutung und können durch ein Schlüsselbild repräsentiert werden (vgl. Iedema 2004: 188), wenn innerhalb einer Einstellung zwischen den Grundgestaltungen der Einzelbilder (z. B. Einstellungsgröße, Komposition und Standpunkt) keine deutlichen Unterschiede bestehen. Daraufhin lässt sich die Einstellung, die sich aus den Einzelbildern zusammensetzt und durch ein Schlüsselbild repräsentiert werden kann, als grundlegender Rahmen für die Kookkurrenz von *Figuraen* aus dem entsprechenden Schlüsselbild betrachten.²

² Bei der Analyse von Einstellung ist noch ein Sonderfall zu berücksichtigen. Es gibt Einstellungen, in denen mehrere Handlungen kontinuierlich dargestellt werden. Das heißt: Zwischen den Gestaltungen und somit den durch die Gestaltungen ausgedrückten Bedeutungen der Einzelbilder sind deutliche Unterschiede zu erkennen. In solchen Einstellungen (z. B. bei einem long take) gibt es mehrere Schlüsselbilder. Iedema hat festgestellt: „With long shots, you may decide there is more than one salient frame which needs to be included in the analysis“ (Iedema 2004: 188). In diesem Fall gilt die Einstellung nicht als ein Satz, sondern als eine Summe von Sätzen, die miteinander eng

Bei einem Film, der aus Einstellungsfolgen besteht, bezieht sich das Bildgebrauchsmuster – wiederkehrende Kombination einzelner Figurae – nicht nur auf die Kollokation der Figurae in derselben Einstellung, sondern auch auf ihre Kollokation in den angrenzenden Einstellungen, die durch Montage kombiniert werden, denn die Bedeutung einer Einstellung „durch die davor und dahinter montierten Filmtteile komplett verändert“ (Böhringer et al. 2011: 380) werden kann. Die Kontinuität zwischen den Einstellungen, die von den Zuschauern erwartet wird, wird je nach Zweck des Regisseurs/der Regisseurin durch Montage etabliert. Zugleich werden die Figurae in den angrenzenden Einstellungen miteinander verbunden, obwohl diese Verbindung der Wirklichkeit nicht unbedingt entspricht (vgl. Hickethier 2012: 142). Wenn eine bestimmte Figura immer wieder durch Montage mit einer anderen in angrenzenden Einstellungen verknüpft wird, werden die Zuschauer den Eindruck bekommen, dass es zwischen den beiden Figuraen ebenfalls einen bestimmten Zusammenhang gibt. Dementsprechend kann das Bildgebrauchsmuster in einem Film als *rekurrentes Auftreten von Kombination bzw. Kollokation von Figuraen (bildlichen Elementen) in einzelnen sowie angrenzenden Einstellungen* definiert werden.

3 Methodisches Verfahren

Durch die Untersuchung des Musterhaften beim Zeichengebrauch der Sprach- und Bildmodalität können die rekurrenten Zuschreibungen zu einer bestimmten Person oder einem bestimmten Gegenstand in einem multimodalen Text ermittelt werden. Aus denen lässt sich das Image des entsprechenden Referenzobjektes induktiv erschließen. In dem vorliegenden Beitrag werden bei der Untersuchung des Chinesen-Images die sprachliche sowie die bildliche Ebene kombinierend berücksichtigt, damit das abgeleitete Image vollständiger und umfassender wird (vgl. Abb. 1).

Bei der Imageanalyse in Bezug auf Sprachmodalität werden zuerst alle gesprochenen Sätze in dem Dokumentarfilm transkribiert. Danach werden mithilfe der Software AntConc alle Sätze zusammengestellt, die sich auf Chinesen beziehen und das Lexem *Chinesen* oder die Ausdrücke wie *chinesisch* Kunde/Ingenieur/Tourist*³ oder *Kunde/Arbeiter* aus China/Peking* enthalten. Diese Sätze dienen als Ausgangspunkt einer KWIC-Analyse.⁴ Mithilfe der Software MAXQDA

verbunden sind. Die Entsprechung für Satz in solchen Einstellungen bildet die Teileinstellung, die durch ihr eigenes Schlüsselbild zu repräsentieren ist.

3 Das Zeichen * ist ein Stellvertreterzeichen in AntConc. Es bedeutet, dass nach dem angegebenen Wort keine oder mehrere Buchstaben bestehen können.

4 KWIC steht für keyword in context. Eine KWIC-Analyse untersucht die Bedeutung eines für die Untersuchung relevanten Wortes anhand des Kontextes. Die in den theoretischen Prämissen

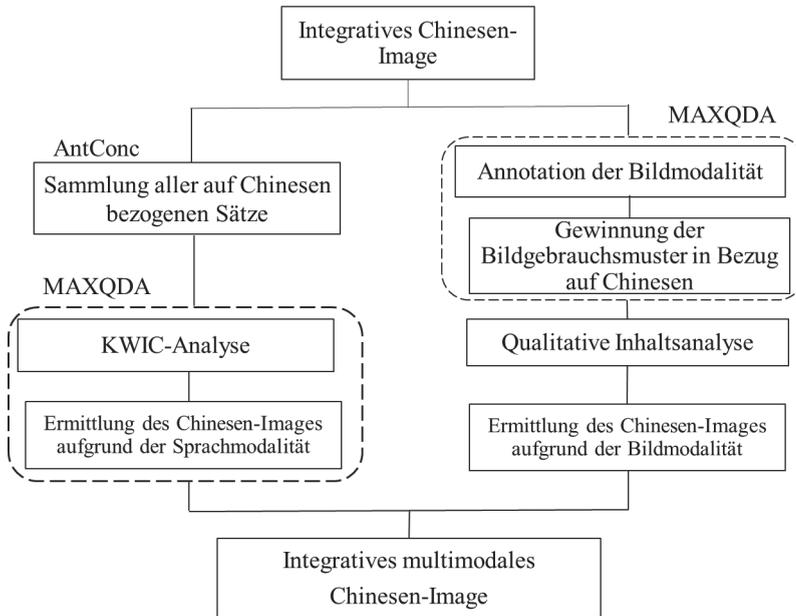


Abb. 1: Forschungsablauf

werden sie unter Berücksichtigung des Kontextes induktiv mit Zuschreibungen zu Chinesen annotiert bzw. klassifiziert, um daraufhin aufschlussreiche Aspekte des Chinesen-Images herauszuarbeiten. Z. B. werden die Aussagen wie „*Die Chinesen haben unsere Stadt vollständig übernommen*“ und „*Diese Arbeiter gehörten zur chinesischen Armee*“ einer Kategorie zugeordnet, in der den Chinesen die Eigenschaft *die Einheimischen verdrängend* zugeschrieben wird. Danach werden sie mit anderen Aussagen, die die Chinesen als *rechtswidrig* und *fragwürdig* für die Einheimischen darstellen, in einer Hyperkategorie zusammengefasst, nämlich das Chinesen-Image von ‚greifenden Fremden‘.

Die Imageanalyse in der Bildmodalität basiert auf einer computergestützten Analyse von Bildgebrauchsmustern. Die wiederkehrenden *Figurae* rund um Chinesen, die sich entweder in derselben oder angrenzenden Einstellung befinden, werden herausgesucht und nach Häufigkeit gewichtet. Die statistische Berechnung setzt voraus, die Einstellungen, die den unmittelbaren Untersuchungsgegenstand Chinesen enthalten, und ihre angrenzenden Einstellungen mithilfe der MAXQDA

genannte Analyse von Musterhaftigkeit gilt für die Analyse der Sprachmodalität des Dokumentarfilms als unzutreffend, denn in dieser Arbeit wird wegen des Zeit- und Arbeitsaufwands nur ein kleiner Korpus mit relativ wenigen sprachlichen Tokens aufgebaut.

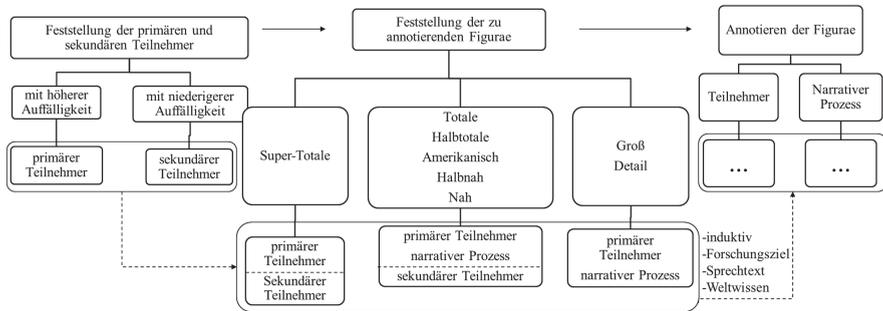


Abb. 2: Annotationsverfahren in Bezug auf die Bildmodalität im Dokumentarfilm

manuell in Bezug auf relevante bildliche Elemente mit zutreffenden Codes⁵ zu annotieren. Zu erwähnen ist, dass es keine geeignete Software gibt, die die Figurae in den Einstellungen automatisch nach Bedarf annotieren kann. Für die manuelle Annotation muss zunächst festgesetzt werden, welche Informationen in den Einstellungen als wichtig gelten, so dass sie annotiert werden sollen. Das heißt: Welche Informationen will der Regisseur/die Regisseurin den Zuschauern zeigen und welche nur als irrelevanter Hintergrund mit aufnehmen. Zur Feststellung und Benennung der wichtigsten, zu annotierenden Figurae wird die von Kress/van Leeuwen entwickelte Visuelle Grammatik als Hilfsmittel zugrunde gelegt, weil sie die Funktionen der Bildgestaltungsmittel systematisch zusammenfasst. Dadurch wird das Annotationsverfahren entwickelt (vgl. Abb. 2).

Die Annotation erfolgt in drei Schritten:

1) Zuerst soll festgestellt werden, welche Teilnehmer (eng. participant)⁶ als primär bzw. sekundär gelten, wobei die gesamte Komposition des Schlüsselbilds einer Einstellung berücksichtigt werden soll. Nach Kress/van Leeuwen (2006: 202) können die Teilnehmer mit dem größeren Platz, im Vordergrund oder mit hoher Schärfe mehr Aufmerksamkeit auf sich ziehen und damit auffälliger und wichtiger als die anderen wirken. Das ähnelt der Akzentuierung beim Sprechen, welche auf bestimmten Silben bzw. Wörtern liegt und den Schwerpunkt eines Satzes bildet. Es wird dadurch zwischen primären und sekundären Teilnehmern differenziert und die akzentuierten Teilnehmer werden als primär angesehen. Die primären Teilneh-

⁵ Der Code bei MAXQDA gilt als ein sprachlicher Zettel, mit dem eine Einstellung annotiert wird. Im folgenden Text werden die Namen der Codes für verschiedene bildliche Elemente in kursiver Schrift gekennzeichnet.

⁶ Mit dem Begriff Teilnehmer wird in der Visuellen Grammatik betont, dass die betroffenen bildlichen Objekte bzw. Elemente an semiotisch bzw. bildlich dargestellten Handlungen teilnehmen (vgl. Kress/van Leeuwen 2006: 47–48). Es bezieht sich auf die Referenz der Figurae (wer/was).

mer werden grundsätzlich annotiert, während die sekundären Teilnehmer je nach Einstellungsgröße annotiert werden oder nicht.

2) Anhand der Einstellungsgröße wird erschlossen, ob die sekundären Teilnehmer sowie die Aktionen⁷ der primären Teilnehmer von den Filmproduzenten betont werden und daher zu annotieren sind. Unter „Einstellungsgröße“ wird „Größe des abgebildeten Menschen im Verhältnis zur Bildgrenze“ (Hickethier 2012: 56) verstanden. Grundsätzlich sind drei Kategorien von Einstellungen zu unterscheiden, nämlich Totale, Halbna- und Großaufnahme (vgl. Hickethier 2012: 59). Detaillierter gesehen ist die Einstellungsgröße wegen ihrer Kontinuität noch in Super-Totale, Totale, Halbtotale, Amerikanisch, Halbna, Na, Groß und Detail zu unterteilen (vgl. Hickethier 2012: 57–59; Mikos 2015: 194–195).

Mit dem Einsatz von einer Super-Totale werden die Zuschauer von dem präsentierten Teilnehmer, der normalerweise auch der primäre Teilnehmer ist, entfernt, sodass sie lediglich einen Überblick über die Umgebung des Teilnehmers bekommen (vgl. Hickethier 2012: 57). In diesem Fall liegt der Schwerpunkt des Bildes nicht auf Aktionen des primären Teilnehmers, sondern auf dem primären Teilnehmer selbst mitsamt seinen Umständen (eng. *circumstances*) bzw. auf die sekundären Teilnehmer. Annotiert werden demzufolge nur die Teilnehmer, aber nicht die konkreten Aktionen.

Bei der Totale bis zur Nahaufnahme steht der präsentierte Teilnehmer viel näher an den Zuschauern, sodass die Zuschauer nicht nur die Teilnehmer selbst, sondern auch die Aktionen des primären Teilnehmers genau ansehen können (vgl. Hickethier 2012: 57–59). Daher sind die primären und sekundären Teilnehmer sowie die Aktion des primären Teilnehmers zu annotieren.

Durch Groß- und Detailaufnahme werden die Teilnehmer so nah aufgenommen, dass Details wie Mimik einer Person oder Einzelteile eines Gegenstandes deutlich dargestellt werden können (vgl. Hickethier 2012: 58–59). Bei solchen Einstellungen ist der Hintergrund (als sekundärer Teilnehmer) oft ganz unscharf und damit unerkennbar. Daher werden nur der primäre Teilnehmer und dessen Aktion annotiert.

3) Im letzten Schritt sollen passende sprachliche Mittel bzw. Codes den zu annotierenden Figuren zugewiesen werden. Sie lassen sich zuerst angesichts ihrer Rolle im Bild nach Kress/van Leeuwen (1996) in *Teilnehmer* und *narrativer Prozess* gruppieren. Dann werden sie – dem Forschungsziel dienend – mithilfe des Sprechtextes sowie des Weltwissens induktiv weiteren Subkategorien zugeordnet. Beispielsweise ist unter der Gruppe *Teilnehmer* zunächst zwischen *Menschen* und *Sonstiges* zu unterscheiden, denn das Forschungsinteresse liegt auf *Chinesen*, die

7 Die Aktionen der Teilnehmer beziehen sich auf die Prädikation der Figuren (wie/was tun).

eine Subkategorie von *Menschen* bilden. Die vielen Objekte in der Gruppe *Sonstiges* werden zuerst mit ihrem Gattungsnamen intuitiv als *Pferde*, *Schaf*, *Autobahn*, und *Eisenbahn* annotiert.⁸ Mit dem Vorantreiben der Annotation lassen sie sich aufgrund ihrer Ähnlichkeiten weiter in Hyperkategorien wie *Tier* und *Straßen* zusammenfügen.

Die annotierten Daten können in MAXQDA automatisch auf die Häufigkeit der Bildelemente überprüft werden, die mit der Figura *Chinesen* kookkurrieren. Erfüllt eine kookkurrierende Figura die Bedingungen der statistischen Auffälligkeit⁹ und des semantischen Interpretationspotentials¹⁰, wird die Kombination von dieser Figura mit *Chinesen* als ein Bildgebrauchsmuster angesehen.

Im Anschluss an die qualitative Inhaltsanalyse nach Mayring (2015) wurden die Bildgebrauchsmuster mit Berücksichtigung des Kontextes induktiv auf entsprechende Zuschreibungen untersucht und auf dieser Basis größere Gruppen erarbeitet. Hierbei kamen zwei grundlegende Techniken der qualitativen Inhaltsanalyse zur Anwendung: Explikation und Zusammenfassung (vgl. Mayring 2020: 497). Im Rahmen der Explikation wurden zutreffende Zuschreibungen durch den Rückgriff auf den filmischen Kontext der jeweiligen Bildgebrauchsmuster bestimmt. Bei dem Bildgebrauchsmuster *Chinesen + Schneeberge* wird z. B. durch den begleitenden Sprechtext „*Es wird eine lange Fahrt durch eine dramatische und gefährliche Bergwelt*“ zu der tiefgehenden Zuschreibung von *Öde und Verlassenheit*, statt zum oberflächlichen Eindruck von *schöner Landschaft entlang der Seidenstraße* gelangt. Bei der Zusammenfassung wurden die bestimmten Zuschreibungen zu größeren, übergeordneten Schlüsselzuschreibungen kombiniert. Beispielsweise wurden Zuschreibungen wie *Öde und Verlassenheit* sowie *Druck und Depression* zu der übergeordneten Kategorie *Schwierige Arbeitsbedingungen* zusammengefügt. Daraus werden Imageaspekte in der Bildmodalität abgeleitet – hierbei der Aspekt von ‚Opfer‘. Diese wurden wiederum mit den Imageaspekten aus der Sprachmodalität zu einem integrativen Ganzen kombiniert.

8 Menschliche kategoriale Wahrnehmung von beobachteten Phänomenen erfolgt in der Regel intuitiv entlang der Kategorie *Gattung (Genus)*. Zum Beispiel wird ein Objekt eher als *Pferd* kategorisiert, weniger als *Tier* oder als *Shetlandpony*. Diese Art der Klassifikation bildet den Ausgangspunkt für das kindliche Erlernen der Taxonomie von Objekten und stellt somit auch den zentralen Knotenpunkt im Hierarchiesystem der Kategorisierung dar (vgl. Lakoff 1987: 32–34).

9 In dieser Arbeit werden die Kombinationen von Figuraen, deren Häufigkeit die Durchschnittshäufigkeit aller Kombinationen übersteigt, als statistisch auffällig angesehen. Diese Kombinationen haben dann das Potenzial, Bildgebrauchsmuster zu sein.

10 Z. B. kommt die Kombination von *Chinesen* und *Pflanzen* zwar oft vor, aber als ein überall zusehendes Objekt treten die Figura von Pflanzen auch häufig mit anderen Figuraen auf, wie ein Funktionswort der Sprache, daher wird diese Kombination nicht als ein Bildgebrauchsmuster weiter analysiert.

4 Forschungsergebnis: Das multimodale Chinesen-Image

Der ausgewählte deutsche Dokumentarfilm ist ein vom ZDF im Jahr 2019 produzierter Dokumentarfilm mit dem Titel *Die Neue Seidenstraße – Chinas Griff nach Westen*. Sie enthält zwei Episoden, die jeweils ca. 45 Minuten dauern. Zwei Reporter besuchten eine Vielzahl von Städten entlang der Neuen Seidenstraße¹¹ und führten Interviews sowohl mit Chinesen als auch mit Einheimischen über ihre Einstellungen zu der Neuen Seidenstraße. Außerdem drücken die beiden Reporter ihre eigenen Meinungen und Kommentare gegenüber Chinesen bzw. China sowie der neuen Seidenstraße aus.

Anhand der genannten methodischen Verfahren werden die beiden Modalitäten Sprache und Bild auf wiederkehrende Zuschreibungen zu Chinesen untersucht. Durch deren Zusammenwirkung ist das multimodale Chinesen-Image zu erschließen.

Der Sprechtext dieses Films umfasst insgesamt 1930 Types und 4336 Tokens. Die Abbildung 3 stellt ein Screenshot des endgültigen Annotationsergebnisses bei der Sprachmodalität mit MAXQDA dar. Aus den Äußerungen in Bezug auf Chinesen lassen sich zuerst verschiedene Zuschreibungen ableiten, welche weiterhin zu vier Imageaspekten zusammengefasst werden, nämlich ‚greifende Fremde‘, ‚Gewinner‘, ‚fleißige Opfer‘ und ‚Leiter‘.

Die Untersuchung der Bildmodalität stützt sich zunächst auf die Interpretation der Bildgebrauchsmuster, die mittels manueller Annotation und statistischer Berechnung, wie im letzten Abschnitt ausgeführt wurde, gewonnen wurden. Nach der Extraktion der Bildgebrauchsmuster lässt sich automatisch eine *Code-landkarte* durch die Anwendung von MAXQDA generieren (vgl. Abb. 4), womit die Beziehungen zwischen *Chinesen* und den mit *Chinesen* häufig innerhalb einer Sekunde¹² kookkurrierenden bildlichen Elementen – in Form von Codes – darge-

¹¹ Seeweg: Shenzhen (China) – Sihanoukville (Kambodscha) – Bagan (Myanmar) – Hambantota (Sri Lanka) – Gwadar (Pakistan) – Sur (Oman) – Duqm (Oman) – Marseille (Frankreich) – Duisburg (Deutschland)

Landweg: Shenzhen (China) – Dunhuang (China) – Kashgar (China) – Islamabad (Pakistan) – Issyk Kul (Kirgisistan) – Samarkand (Usbekistan) – Istanbul (die Türkei) – Belgrad (Serbien) – Budapest (Ungarn) – Duisburg (Deutschland)

¹² Beim Gebrauch der Funktion von „Code-landkarte“ sollte zuerst die Nähe von Codes festgesetzt werden. Unter Berücksichtigung der Zeitdauer vom Kurzzeitgedächtnis, das ohne weitere Stimulation verschwinden wird und normalerweise ca. eine Sekunde dauert (vgl. Liang 2006: 234), können Figuren in Einstellungen, zwischen denen weniger als eine Sekunde besteht, als miteinander verbunden angesehen werden. Dadurch können die bildlichen Kookkurrenzpartner in derselben und angrenzenden Einstellungen identifiziert werden.

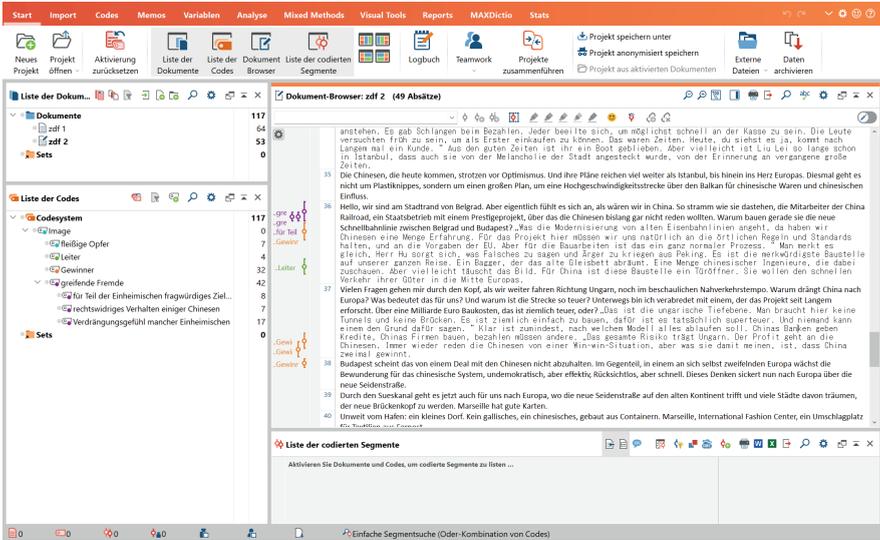


Abb. 3: Screenshot aus MAXQDA (1)

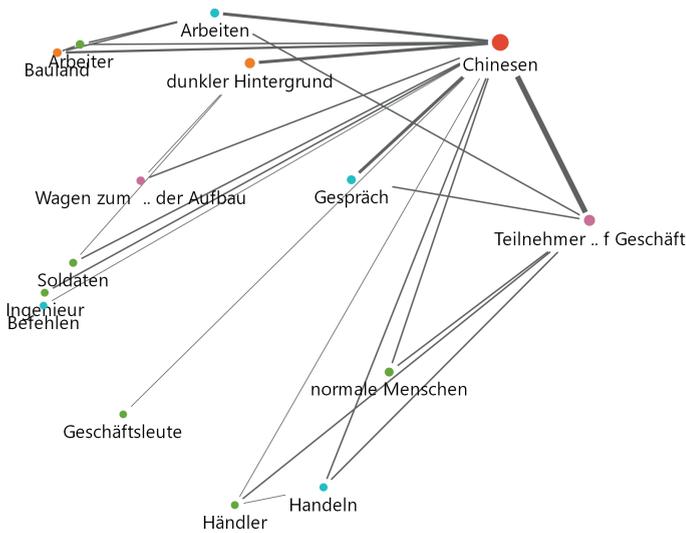


Abb. 4: Screenshot aus MAXQDA (2)

stellt werden. Je häufiger zwei Codes zusammen vorkommen, desto näher stehen sie auf der Karte nebeneinander und desto dicker werden die Verbindungslinien dazwischen. Aus der Codelandkarte ist beispielsweise deutlich zu erkennen, dass die Codes *Teilnehmer in Bezug auf Geschäft, Arbeiten* eng mit *Chinesen* verknüpft sind. Gemeinsam mit Codes wie *Bauland, Handeln* u. ä. akzentuieren sie die Zusammenarbeit der Chinesen mit den Einheimischen, wobei diese Zusammenarbeit vor allem im Bereich des Handels und der Bauarbeit stattfindet. Dadurch werden für die aus der Sprachmodalität ermittelten Imageaspekte wie ‚Gewinner‘ bzw. ‚Leiter‘ authentische Arbeitsszenarien als grundlegender Rahmen spezifiziert. Aus der Analyse der Bildmodalität lassen sich folgende Imageaspekte erschließen: ‚greifende Fremde‘, ‚Zusammenarbeiter‘, ‚Opfer‘ und ‚Leiter‘.

Die Ergebnisse in der Bildmodalität und in der Sprachmodalität überschneiden sich teilweise und ergänzen bzw. spezifizieren sich gegenseitig. Zum Schluss lassen sich aus den beiden Modalitäten vier integrierte Imageaspekte zusammenfassen, die nach ihrer Wichtigkeit bzw. der Häufigkeit wie folgt angeordnet werden können: ‚greifende Fremde‘, ‚Gewinner bei der Zusammenarbeit‘, ‚fleißige Opfer‘ und ‚Leiter bei der Zusammenarbeit‘.

4.1 Chinesen als ‚greifende Fremde‘

Im Mittelpunkt aller Imageaspekte steht zuerst ‚greifende Fremde‘. Am Anfang des Films wird China schon wie folgt beschrieben: „*Stolz und selbstgewiss greift es nach Westen*“ (E1, 01:50)¹³. Die Eigenschaft von *greifend* bezieht sich auf die in dem Dokumentarfilm dargestellte Verbreitung der Chinesen und deren als problematisch dargestelltes Handeln; *Fremde* betont die Distanz zwischen den Chinesen und Einheimischen. Dieser Imageaspekt lässt sich in beiden Modalitäten erkennen, jedoch mit verschiedenen Schwerpunkten.

Mehr als die Hälfte der sprachlichen Beschreibungen über Chinesen in der Filmserie weist darauf hin, dass es zwischen Chinesen und einigen einheimischen Völkern Konflikte gäbe und viele den Chinesen und deren „*Versprechen auf Wohlstand*“ (E2, 43:08) mit der Neuen Seidenstraße eher misstrauen.

Dafür wurden in dieser Filmserie vor allem drei innerlich miteinander verbundene Aspekte in den Vordergrund gerückt: 1) Verdrängungsgefühl mancher

¹³ Im Folgenden werden die konkrete Episode und Zeit aller Ausschnitte aus dem Sprechtext in Klammern angegeben.

Einheimischer; 2) rechtswidriges Verhalten einiger Chinesen und 3) für einen Teil der Einheimischen fragwürdiges Ziel der Initiative.

Mit Wörtern wie „*Armee*“ und „*übernehmen*“ wurden die eigentlich normalen und harmlosen Chinesen direkt oder metaphorisch als Soldaten eines Imperiums porträtiert und die Einheimischen als Opfer, die ihre Heimat verloren haben: „*Die Chinesen haben unsere Stadt vollständig übernommen*“ (E1, 09:29) / „*Diese Arbeiter gehörten zur chinesischen Armee*“ (E1, 39:20). Dieser Aspekt ist auch häufig mit der großen Anzahl der Chinesen im Ausland verknüpft: „*Sie haben keine Lust, nur noch Chinesen um sich rum zu sehen*“ (E1, 09:20) / „*Das ist Mist. Immer und überall nur Chinesen*“ (E2, 39:16). Dadurch wurden von den Filmproduzenten mit Absicht starke Abneigungen kreiert.

Der Meinung der interviewten Einheimischen nach fühlen sie sich nicht nur von der Menschenmenge bedroht, sondern auch vom ungerechten bzw. rechtswidrigen Verhalten mancher Chinesen. Dabei werden oft auch die lokalen Regierungen mitbeschuldigt, die sich am Geld der Chinesen bereichern und sich gar nicht um das Leben der Völker kümmern: „*Für Kambodschaner ist Glücksspiel verboten. Für Chinesen auch, aber nur in China*“ (E1, 12:51) / „*Und manche Chinesen fahren halt betrunken Auto*“ (E1, 12:40). Über den von Chinesen mitgebrachten Wohlstand (vgl. Heng/Po 2017: 2) wird allerdings in dem Film geschwiegen.

Außerdem findet sich eine Reihe von Fragesätzen in dem Voice-Over, die das Ziel der Neuen Seidenstraße in Frage stellen und die Konfrontation zwischen den Einheimischen und den Chinesen sprachlich verstärken: „*Welches Rom wollen die Chinesen erreichen?*“ (E1, 37:48) / „*Schickt Peking heute Arbeiter und morgen Soldaten?*“ (E2, 14:15) / „*Was steckt hinter Chinas großem Plan?*“ (E1, 02:05) Solche unbegründeten Zweifel, die die Feindseligkeit aufflammen lassen, dienen als Manipulationsmittel der Emotionen von Zuschauern, was normalerweise in den heutigen Dokumentarfilmen zu vermeiden ist (vgl. Nie 2016: 316–317).

In der Bildmodalität werden durch verschiedene Bildgebrauchsmuster sowie weitere visuelle Mittel Chinesen als *Fremde* gegenüber den Einheimischen konstituiert. Dabei sind die Kombinationen von *Chinesen + einheimische Völker (Ausländer)* und *Chinesen + Soldaten (Ausländer)* besonders zu berücksichtigen.

Mithilfe der Software MAXQDA ist deutlich zu sehen, dass die Kookkurrenzen von *einheimischen Völkern* mit *Chinesen* meistens in angrenzenden Einstellungen statt in derselben Einstellung stattfinden, d. h. die *Einheimischen* und *Chinesen* werden oft getrennt aufgenommen. Bei den wenigen Bildern, auf denen die *Chinesen* und *Einheimischen* zusammen auftreten, dienen die *Einheimischen* meistens nur als Hintergrund bzw. Umgebung (erste zwei Bilder in Abb. 5) und selten als Kommunikationspartner. Beim Aufnehmen von Gesprächen zwischen *Chinesen* und *Einheimischen* lässt sich eine große räumliche Distanz zwischen den Beteiligten erkennen (drittes Bild in Abb. 5), was die soziale und psychologische Distanz wider-

gespiegelt (vgl. Kress/van Leeuwen 2006: 125).¹⁴ Diese Inszenierungen bewegen die Zuschauer zu den Gedanken, dass es beim Zusammenleben der Chinesen und Einheimischen kaum direkte Kommunikation gibt.



Abb. 5: Beispielbilder für Bildgebrauchsmuster von *Chinese + einheimische Völker (Ausländer)*

Außerdem richtet sich das Bildgebrauchsmuster von *Chinesen + Soldaten (Ausländer)* auch auf instabile Beziehung zwischen *Chinesen* und *Einheimischen* (Abb. 6). Die Existenz der Soldaten spielt auf Konflikt und Misstrauen an und verändert die Bedeutung der Einstellung: Ohne Soldaten werden die chinesischen Ingenieure lediglich als Mitarbeiter wahrgenommen, wohingegen mit Soldaten in der Nähe der Eindruck erweckt wird, dass die *Chinesen* geschützt oder überwacht werden sollen, weil jemand sie angreifen könnte oder weil man ihnen nicht traut. Die Anwesenheit der Soldaten, die als eine Anspielung auf Konflikt und Misstrauen genutzt wird, ist eigentlich auf die Unruhe und die Terroranschläge in Pakistan (vgl. Xie 2016) zurückzuführen.



Abb. 6: Beispielbilder für Bildgebrauchsmuster von *Chinesen + Soldaten (Ausländer)*

¹⁴ Direkte Kommunikationen finden sich fast nur zwischen *chinesischen* und *einheimischen Geschäftsleuten* sowie *Ingenieuren* bei der Zusammenarbeit, wo sie in derselben Einstellung zusammen vorkommen.

4.2 Chinesen als ‚Gewinner bei der Zusammenarbeit‘

‚Gewinner bei der Zusammenarbeit‘ gilt als der zweit wichtigste Imageaspekt. Die Zusammenarbeit zwischen den *Chinesen* und den *Einheimischen* verschiedener Länder entlang der Neuen Seidenstraße findet hauptsächlich im Bereich des Handels und des Infrastrukturaufbaus statt. In der Sprachmodalität werden Chinesen als Handelspartner oder Aufbauer porträtiert und negativ bewertet. In der Bildmodalität wird mit *Chinesen* meisten konkrete Szenarien aus der Zusammenarbeit (Handel, Infrastrukturaufbau) verbunden.

Die Neue Seidenstraße zielt zwar immer darauf ab, durch Kooperation und gegenseitiges Lernen eine Win-Win-Situation zu erreichen (vgl. Li 2016), aber in der Sprachmodalität wird nur einseitig die Ansicht einiger Einheimischer vermittelt, dass die Einheimischen zurzeit noch nicht gewonnen haben, sondern nur die Chinesen: *„Immer wieder reden die Chinesen von einer Win-Win-Situation, aber was sie damit meinen, ist, dass China zweimal gewinnt.“* (E2, 34:27) / *„Der Profit geht an die Chinesen“* (E2, 34:23).

Nur zwei Sätze, bei denen die Chinesen von Einheimischen aus anderen Ländern gelobt werden, werden in dem Dokumentarfilm zitiert. Aber die Hochschätzungen werden gleich durch Voice-Over der Filmproduzenten in Zweifel gezogen. Ein Beispiel davon lautet wie folgt: *„Er schätzt die Partner für Vision und Tatkraft. „Die Chinesen sind sehr schlau und sehr aktiv. Und sie sind schnell. [...] Und dann setzen sie ihre Pläne blitzschnell um. Das beste Beispiel sehen Sie hier:“ Na ja, ehrlich gesagt, wir sehen nichts. Oder fast nichts“* (E2, 19:57). Dies verdeutlicht ebenfalls die selektive Darstellung der Filmproduzenten: Wenn die Interviewten etwas Negatives über Chinesen oder die Neue Seidenstraße äußern, werden ihre ausführlichen Erläuterungen direkt in den Film aufgenommen. Wenn sie aber etwas Positives erzählen, werden ihre hochschätzenden Erläuterungen meistens nicht aufgenommen. Oder es wird – wie das Beispiel hier zeigt – direkt durch einen Kommentar der Filmproduzenten verneint.

In der Bildmodalität wird die Zusammenarbeit zwischen Chinesen und Einheimischen authentisiert. Die am häufigsten mit *Chinesen* zusammenvorkommenden Codes sind verschiedene *Waren*. Außerdem sind z. B. *Baustelle*, *Laden*, *LKW*, *Ingenieure (Ausländer)* und *Geschäftsleute (Ausländer)* Kookkurrenzpartner von *Chinesen* innerhalb einer Sekunde (Abb. 7). Dadurch wird ersichtlich, dass *Chinesen* in der Bildmodalität vor allem mit Arbeitsszenarien (z. B. Handel, Infrastrukturaufbau) zusammen vorkommen.

Anders als die Sprachmodalität bietet die Bildmodalität in diesem Zusammenhang keinen direkten Zugang zur Bewertung von *Chinesen*, vielfach liefert sie anschauliche Fallbeispiele über die enge Zusammenarbeit zwischen *Einheimischen* und *Chinesen*, die durch sprachliche Sprechtexte näher qualifiziert werden. Durch die



Abb. 7: Beispielbilder für Bildgebrauchsmuster über die Zusammenarbeit zwischen Chinesen und Einheimischen

sprachlichen Prädikationen werden die chinesischen ‚Zusammenarbeiter‘ weiterhin als ‚Gewinner bei der Zusammenarbeit‘ spezifiziert, die unfaire Geschäfte mit anderen Ländern durchführen und eine zweifelhafte Zukunft versprechen, was der Tatsache in keiner Weise entspricht (vgl. Zhang/Wu 2022).

4.3 Chinesen als ‚fleißige Opfer‘

Nicht nur die einheimischen Völker entlang der Seidenstraße werden als *Verlierer* bzw. *Opfer* von „Chinas großem Plan“ (E1, 02:05) dargestellt, sondern auch die Chinesen selbst. Dieser Imageaspekt wird vor allem durch folgende Sätze in den Vordergrund gerückt: „Aber auch die Chinesen kommt sie teuer zu stehen.“ (E1, 39:08) / „Ein Friedhof für chinesische Arbeiter, die beim Bau ums Leben kamen.“ (E1, 39:12) / „Drei chinesische Ingenieure sind hier gestorben“ (E2, 13:02).

Auch die Bildmodalität spielt bei dieser Imagekonstruktion ‚Opfer‘ eine große Rolle. Das kommt vor allem durch die Darstellung schlimmer Lebensbedingungen und die Gestaltung einer visuell bedrückenden Atmosphäre zustande. Wiederholt kommt die Kombination von *Chinesen* + *Blehbüchsen* vor. Klein und primitiv sind die aufgenommenen Blehbüchsen, wobei immer mit einem kühlen und dunklen Farbton aufgenommen wird (Abb. 8). Diese Gestaltung vermittelt den Zuschauern ein Gefühl von Druck und Depression (vgl. Jiang 2021: 120).



Abb. 8: Beispielbilder für Bildgebrauchsmuster von *Chinesen* + *Blehbüchsen*

Zudem werden Chinesen oft vor dem Hintergrund von *dunklen Wolken*, *Wüste* und *Schneegebirge* aufgenommen, die den Eindruck von Öde und Verlassenheit ver-



Abb. 9: Beispielbilder für Bildgebrauchsmuster von *Chinesen + dunkler Hintergrund*

mitteln (Abb. 9). Diese depressive Atmosphäre gilt als ein indirektes Mittel zur Konstituierung des Opfer-Images.

Besonders zu erwähnen ist, dass beim Aufnehmen solcher Szenen einige interviewte chinesische Arbeiter ihren Fleiß als eine hochzuschätzende Tugend betont haben. Beispielsweise hat ein Arbeiter stolz behauptet: „*Wir Chinesen sind fleißige Leute*“ (E2, 21:26). Eine interviewte Kauffrau hat auch geäußert: „*Wir arbeiten eben gern und sind stolz darauf*“ (E2, 36:38). Trotz ihrer Hochschätzung wird dieser Fleiß von den Filmproduzenten zu einer Ursache für das Opfersein (um)interpretiert. Das heißt, die Chinesen werden zu Opfern des Plans der Regierung, indem sie fleißig sind oder sein müssen. In folgenden Beispielen zeigt sich die Umleitung der Filmproduzenten von Fleiß und Mitwirkung zu Ausnutzung und Opfersein:

„Morgens erklärt Chefindenieur Wang Hui, wie es weitergehen soll mit der Straße. Seit über zehn Jahren baut er daran. Die Haare sind grau geworden darüber: [...] Und genau das erwartet Peking von Wang Hui, Tag für Tag.“ (E1, 41:10) / *„Opfer muss man bringen, jeder. Wir arbeiten hier ein halbes Jahr. Dann gibt es einen Monat Urlaub.“* (E2, 23:04)

Die Arbeiter sehen sich – trotz schwieriger Arbeitsbedingungen – eigentlich nicht als Opfer und haben beim Interview sogar ihren Stolz hinsichtlich ihrer Mitwirkung beim Aufbau der Seidenstraße ausgedrückt, wie die obigen Beispiele zeigen. Solcher Stolz wird aber durch die Inszenierung der Filmproduzenten ungerecht uminterpretiert, genauso wie der Stolz auf ihre Fleiß-Tugend.

4.4 Chinesen als ‚Leiter bei der Zusammenarbeit‘

Neben ‚Gewinner‘ und ‚Opfer‘ werden Chinesen auch als *Leiter* der Zusammenarbeit dargestellt, die Befehle geben und für die die anderen arbeiten:

„Und wer hat hier das Sagen? Hier auf der Baustelle ich, der Pakistaner. Ich kontrolliere alles. Aber ihr arbeitet doch im Auftrag der Chinesen.‘ Ja klar, der Big Boss ist Chinese.“ (E2, 12:00) / *„Die Chefs kommen aus China [...]“* (E2, 27:45).

Wie bei ‚fleißige Opfer‘ wird der ‚Leiter bei der Zusammenarbeit‘-Image vielfach durch die Bildmodalität untermauert. In den Einstellungen mit dem narrativen Prozess *Befehle* erteilen immer nur die Chinesen Befehle zur Arbeit (Abb. 10 oben), vor allem in den Szenarien auf Baustellen. Es gibt keine einzige Szene, wo einheimische Ingenieure oder Geschäftsleute als Leiter dargestellt werden. Außerdem wird noch eine Szene gefilmt, wo chinesische Ingenieure bei Bauarbeiten zuschauen und selbst nicht daran teilnehmen (Abb. 10 unten), und noch dazu durch den folgenden Sprechertext begleitet werden: *„Es ist die merkwürdigste Baustelle auf unserer ganzen Reise. Ein Bagger, der das alte Gleisbett abräumt. Eine Menge chinesischer Ingenieure, die dabei zuschauen“* (E2, 32:59). Die Ursache solcher Szenen könnte in der Tat aber an der Anwesenheit der Filmproduzenten liegen – diese chinesischen Leiter mussten sie beim Besuch begleiten und die laufende Arbeit vorstellen, daher konnten sie nichts anderes tun, als stehen zu bleiben und aufgenommen zu werden.



Abb. 10: Beispielbilder für den Imageaspekt von ‚Leiter bei der Zusammenarbeit‘

Mit solchen Bildern werden nicht nur die Chinesen als Leiter bei der Zusammenarbeit dargestellt, sondern auch gleichzeitig die einheimischen Arbeiter als niedrige Schicht beim Aufbau der Seidenstraße, die schweigend den Befehlen vonseiten der leitenden Schichten gehorchen und unter schwierigen Bedingungen für Chinesen arbeiten. Dieser visuelle Eindruck wird wiederum in der Sprachmodalität thematisiert bzw. bestärkt: *„Es hat etwas von Strafkolonie, zumindest sieht es so aus. Unter der sengenden Wüstensonne arbeiten sie an Chinas großem Plan, von dem die Männer hier noch nie etwas gehört haben“* (E2, 22:15). Auf diese Art und Weise verstärkt das ‚Leiter bei der Zusammenarbeit‘-Image noch die Konfrontation zwischen *Einheimischen* und *Chinesen* und untermauert damit auch das ‚greifende Fremde‘-Image. Die durch die Neue Seidenstraße neu hervorgebrachten Arbeitsplätze in diesen Ländern (vgl. Rehman et al. 2021: 248) werden im Film nicht erwähnt.

5 Fazit und Ausblick

Der vorliegende Beitrag fokussiert sich auf das multimodal zu ermittelnde Chinesen-Image in dem deutschen Dokumentarfilm *Die Neue Seidenstraße – Chinas Griff nach Westen*. Das gesamte Image ergibt sich aus der Kombination der Sprach- und Bildmodalität, wobei sowohl eine KWIC-Analyse aller auf *Chinesen* bezogenen Sprachbelege im Sprechtext als auch eine Untersuchung vom bildlichen Musterhaften durchgeführt wurde. Die ermittelten vier Imageaspekte können weiterhin in zwei Kategorien zugeordnet werden: gegenüber den Einheimischen werden die Chinesen als ‚Gewinner bei der Zusammenarbeit‘, ‚greifende Fremde‘ und ‚Leiter bei der Zusammenarbeit‘ konstituiert und an sich als ‚fleißige Opfer‘ (vgl. Tab. 2).

Tab. 1: Beziehungen zwischen den konstituierten Imageaspekten

	gegenüber Einheimischen	an sich
Integratives multimodales Chinesen-Image	Gewinner bei der Zusammenarbeit	
	↓ greifende Fremde	fleißige Opfer
	↑ Leiter bei der Zusammenarbeit	

Im Mittelpunkt steht der Imageaspekt ‚greifende Fremde‘. Die Imageaspekte ‚Gewinner bei der Zusammenarbeit‘ und ‚Leiter bei der Zusammenarbeit‘ dienen als zwei wichtige Stützpunkte, warum die Einheimischen und Filmproduzenten Antipathie und Wachsamkeit gegenüber Chinesen zeigen. Beispielsweise folgt der Beschwerde eines Einheimischen „*wir verlieren, sie gewinnen*“ (E1, 21:03), was als ein Beleg für den Imageaspekt ‚Gewinner bei der Zusammenarbeit‘ dient, noch der Kommentar der Filmproduzenten in Form von Voice-Over: „*Der ganze Jademarkt von Mandalay ist in ihrer Hand*“ (E1, 21:11). Dieser Kommentar bezieht sich weiterhin auf den Aspekt ‚greifende Fremde‘. Auch die Aufsicht der chinesischen Bauleiter auf einer Baustelle wird von den Filmproduzenten wie folgt kommentiert: „*Für China ist diese Baustelle ein Türöffner. Sie wollen den schnellen Verkehr ihrer Güter in die Mitte Europas*“ (E2, 33:17), was wiederum mit der Vorstellung von „Chinas Griff“ im Zusammenhang steht. Andererseits werden Chinesen an sich auch als ‚Opfer‘ des großen Bauprojektes konstituiert, wobei die von Chinesen selbst häufig betonte Eigenschaft *Fleiß* ausgenutzt wird.

Aus der empirischen Untersuchung lässt sich ermitteln, dass die deutschen Filmproduzenten versucht haben, durch selektive Berichte und verkehrende Interpretationen die Förderung der Neuen Seidenstraße zu delegitimieren. Beispiels-

weise werden das harmonische Zusammenleben von Chinesen und Einheimischen sowie die mit der Neuen Seidenstraße mit sich gebrachten Profite der Einheimischen den Zuschauern nicht vermittelt (siehe 4.1). Zudem versuchen die Filmproduzenten aufgrund ihrer eigenen Voreingenommenheit lobende Aussagen über die Chinesen umzuinterpretieren oder lehnen sie einfach ab, ohne jede Rücksicht auf die eigentliche Haltung der Interviewten (siehe 4.2 und 4.3). Diese Erkenntnisse tragen auch in gewissem Maße zu den jüngsten Diskussionen über die Objektivität von Dokumentarfilmen bei und bestätigen die Ansicht, dass die Aufzeichnungen in Bezug auf die Sprach- und Bildmodalität von Dokumentarfilmen gleichermaßen von der Perspektive der Produzenten und ihrer Ideologie beeinflusst werden (vgl. Nie 2019: 30).

Durch diese Arbeit wird auch methodisch ersichtlich, dass verschiedene Modalitäten nicht nur als zusätzliche Bestätigung zueinander fungieren, sondern unentbehrliche Bestandteile für die gesamte Imagekonstruktion im multimodalen Text bilden. Selbst bei Überschneidungen beider Modalitäten in manchen Imageaspekten werden auch unterschiedliche Schwerpunkte gesetzt. Holly (2016: 405) hat festgestellt:

„Sprechtexteinheiten nominieren Referenzobjekte und/oder präzisieren Sachverhalte, die man im Bild erkennen kann. Bildelemente authentisieren, dramatisieren und emotionalisieren die sprachlichen Referenzen und Propositionen.“

Während die filmischen Einstellungen authentische Szenarien der Zusammenarbeit zwischen Chinesen und Einheimischen bieten (authentisieren), beurteilt der Sprechtext die Chinesen und beschreibt sie als Gewinner (präzisieren) (siehe 4.2). Der durch die getrennte Aufnahme von Chinesen und Einheimischen herbeigeführte Anschein von fehlender Kommunikation zwischen beiden Seiten verstärkt das Gefühl von Konflikt und Misstrauen (dramatisieren) (siehe 4.1). Zudem zielen viele visuelle Faktoren darauf ab, bei den Zuschauern bestimmte Emotionen zu wecken, beispielsweise wird eine depressive Atmosphäre durch die Auswahl des Farbtons geschaffen (emotionalisieren) (siehe 4.3). Daher ist es dringend notwendig, verschiedene Modalitäten für umfassendere Imagekonstruktionen einzubeziehen. Im vorliegenden Beitrag wird ein exploratives Annotationsverfahren zur Ermittlung von *Bildgebrauchsmuster* entwickelt, das in zukünftigen Forschungen überprüft, modifiziert bzw. vertieft werden sollte.

Literaturverzeichnis

- Böhringer, Joachim/Bühler, Peter/Schlaich, Patrick. 2011. *Kompendium der Mediengestaltung für Digital- und Printmedien*, 5. Auflage. Berlin/Heidelberg: Springer.
- Bubenhof, Noah. 2008. Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse. In: Warnke, Ingo H./Spitzmüller, Jürgen (Hrsg.): *Methoden der Diskurslinguistik. Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*. Berlin: Walter de Gruyter. 407–434.
- Bubenhof, Noah. 2009. *Sprachgebrauchsmuster, Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin/New York: Walter de Gruyter.
- Große, Franziska. 2011. *Bild-Linguistik: Grundbegriffe und Methoden der linguistischen Bildanalyse in Text- und Diskursumgebungen*. Frankfurt am Main: Peter Lang.
- Gu, Jie (顾洁)/Zhao, Chen (赵晨). 2018. Die Untersuchung des chinesischen Images in den Mainstream-Medien nicht-englischsprachiger Länder (非英语国家主流媒体中的中国形象研究). In: *Modern Communication* (《现代传播》) 40(11). 25–31.
- Heng, Kimkong/Po, Sovinda. 2017. Cambodia and China's Belt and Road Initiative: Opportunities, Challenges and Future Directions. In: *UC Occasional Paper Series* 1(2). 1–18.
- Hickethier, Knut. 2012. *Film- und Fernsehanalyse*, 5. Auflage. Stuttgart/Weimar: J. B. Metzler.
- Holly, Werner. 2016. Nachrichtenfilme als multimodale audiovisuelle Texte. In: Klug, Nina-Maria/Stöckl, Hartmut (Hrsg.): *Handbuch Sprache im multimodalen Kontext*. Berlin/Boston: Walter de Gruyter. 392–409.
- Iedema, Rick. 2004. Analysing Film and Television: A Social Semiotic Account of Hospital: An Unhealthy Business. In: Van Leeuwen, Theo/Jewitt, Carey (Eds.): *The Handbook of Visual Analysis*. London: Sage. 183–206.
- Jiang, Yiwen (姜依文). 2021. Fantasie und Realität – Eine kurze Diskussion zur Atmosphärenschaffung im Filmszenendesign (幻想与现实——浅谈电影中场景设计的氛围营造). In: *Home Drama* (《戏剧之家》) 6. 119–121.
- Klug, Nina-Maria. 2016. Multimodale Text- und Diskurssemantik. In: Klug, Nina-Maria/Stöckl, Hartmut (Hrsg.): *Handbuch Sprache im multimodalen Kontext*. Berlin/Boston: Walter de Gruyter. 165–189.
- Kress, Gunther/Van Leeuwen, Theo. 2006. *Reading Images*, 2. Version. London/New York: Routledge.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal About the Mind*. Chicago and London: The University of Chicago Press.
- Li, Baoxu (李保旭). 2016. *Ein Bild zur Veranschaulichung des Rahmenkonzepts der „Belt and Road“* (一图看懂“一带一路”框架思路). <https://www.yidaiyilu.gov.cn/zchj/tjzc/316.htm> (zuletzt aufgerufen am 07.09.2023).
- Li, Jing. 2017. „Reich der Mittel“ – Linguistische Imageanalyse zu Chinas Wirtschaft (2000–2013). In: Vogel, Friedemann/Jia, Wenjian (Hrsg.): *Chinesisch-Deutscher Imagereport. Das Bild Chinas im deutschsprachigen Raum aus kultur-, medien- und sprachwissenschaftlicher Perspektive (2000–2013)*. Berlin/Boston: Walter de Gruyter. 73–101.
- Liang, Ningjian (梁宁建). 2006. *Introduction to Psychology* (《心理学导论》). Shanghai: Shanghai Educational Publishing House (上海教育出版社).
- Luginbühl, Martin. 2011. Vom kommentierten Realfilm zum multimodalen Komplex – Sprache-Bild-Beziehungen in Fernsehnachrichten im diachronen und internationalen Vergleich. In: Diekmannshenke, Hajo/Klemm, Michael/Stöckl, Hartmut (Hrsg.): *Bildlinguistik. Theorie – Methoden – Fallbeispiele*. Berlin: Erich Schmidt. 257–277.
- Mayring, Philipp. 2015. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*, 12. Auflage. Weinheim und Basel: Beltz Verlag.

- Maying, Philipp. 2020. *Qualitative Inhaltsanalyse*. In: Mey, Günter/Mruck, Katja (Hrsg.): *Handbuch Qualitative Forschung in der Psychologie. Band 2: Designs und Verfahren*, 2. Auflage. Wiesbaden: Springer. 495–512.
- Mikos, Lothar. 2015. *Film- und Fernsehanalyse*, 3. Auflage. Konstanz und München: UVK.
- Nie, Xinru (聂欣如), 2016. *Einführung in die Dokumentarfilmkunst* (《纪录片概论》). Shanghai: Fudan University Press (复旦大学出版社).
- Nie, Xinru (聂欣如), 2019. Disputes over the “Objectivity” of Documentaries (纪录片“客观性”之争论). In: *Journal of Shanghai University (Social Sciences Edition)* (《上海大学学报(社会科学版)》) 36(03). 25–34.
- O’Halloran, Kay. 2011. Multimodal Discourse Analysis. In: Hyland, Ken/Paltridge, Brian (Eds.): *Companion to Discourse Analysis*. London/New York: Continuum. 120–138.
- Opiłowski, Roman. 2013. Von der Textlinguistik zur Bildlinguistik. Sprache-Bild-Texte im neuen Forschungsparadigma. In: *Zeitschrift des Verbandes Polnischer Germanisten* 2(3). 217–225.
- Rehman, Zia Ur/Ali, Syed Ahmad/Ahmed, Muhammad/Khattak, Muhammad Arof. 2021. Transition or Change? The Morphosis of One Belt One Road Initiatives in Pakistan: A Study on the Challenges, Prospects and Outcomes of the China-Pakistan Economic Corridor. In: *International Journal of Technological Learning, Innovation and Development* 13(3). 246–282.
- Sadeghi, Zahra/McClelland, James L./Hoffman, Paul. 2015. You Shall Know an Object by the Company it Keeps: An Investigation of Semantic Representations Derived from Object Co-occurrence in Visual Scenes. In: *Neuropsychologia* 76. 52–61.
- Stöckl, Hartmut. 2004. Typographie: Gewand und Körper des Textes – Linguistische Überlegungen zu typographischer Gestaltung. In: *Zeitschrift für Allgemeine Linguistik* 41. 5–48.
- Stöckl, Hartmut. 2016. Multimodalität – Semiotische und textlinguistische Grundlagen. In: Klug, Nina-Maria/Stöckl, Hartmut (Hrsg.): *Handbuch Sprache im multimodalen Kontext*. Berlin/Boston: Walter de Gruyter. 3–35.
- Thimm, Caja. 2017. China im Spiegel der Printmedien – Zwischen Verdammung und Überhöhung? Medieninhalte und Expertenperspektiven zur Berichterstattung in Deutschland. In: Vogel, Friedemann/Jia, Wenjian (Hrsg.): *Chinesisch-Deutscher Imagereport. Das Bild Chinas im deutschsprachigen Raum aus kultur-, medien- und sprachwissenschaftlicher Perspektive (2000–2013)*. Berlin/Boston: Walter de Gruyter. 29–47.
- Vogel, Friedemann. 2010a. Linguistische Imageanalyse (LIma). Grundlegende Überlegungen und exemplifizierende Studie zum öffentlichen Image von Türken und Türkei in deutschsprachigen Medien. In: *Deutsche Sprache (DS). Zeitschrift für Theorie, Praxis, Dokumentation* 4. 345–377.
- Vogel, Friedemann. 2010b. *Ungarn – das Tor zum Westen*. Das Bild Ungarns in deutschsprachigen Medien 1999–2009. Eine linguistische Imageanalyse. In: Masát, András/Tichy, Ellen (Hrsg.): *Jahrbuch der ungarischen Germanistik*. Bonn/Budapest: DAAD. 87–124.
- Vogel, Friedemann. 2017. Linguistische Imageanalyse Chinas. Theoretisch-methodische Grundlagen und exemplarische Analyse. In: Vogel, Friedemann/Jia, Wenjian (Hrsg.): *Chinesisch-Deutscher Imagereport. Das Bild Chinas im deutschsprachigen Raum aus kultur-, medien- und sprachwissenschaftlicher Perspektive (2000–2013)*. Berlin/Boston: Walter de Gruyter. 48–69.
- Vogel, Friedemann/Haberer, Maximilian. 2017. Das China-Image in der deutschsprachigen Werbung. Multimodale Formen und Funktionen eines asiatischen Ethnostereotyps in persuasiven Verwendungskontexten. In: Vogel, Friedemann/Jia, Wenjian (Hrsg.): *Chinesisch-Deutscher Imagereport. Das Bild Chinas im deutschsprachigen Raum aus kultur-, medien- und sprachwissenschaftlicher Perspektive (2000–2013)*. Berlin/Boston: Walter de Gruyter. 229–256.

- Wang, Ziyue (王紫叶), 2019: Das Bild Chinas im Rahmen der „Belt and Road Initiative“ in deutschen Medien (“一带一路” 视域下德媒建构的中国形象). In: *Youth Journalist* (《青年记者》) 14. 30–31.
- Warnke, Ingo H./Spitzmüller, Jürgen. 2008. Methoden und Methodologie der Diskurslinguistik – Grundlagen und Verfahren einer Sprachwissenschaft jenseits textueller Grenzen. In: Warnke, Ingo H./Spitzmüller, Jürgen (Hrsg.): *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*. Berlin: Walter de Gruyter. 3–55.
- Wei, Naixing (卫乃兴). 2008. The Firthian Foundations of Corpus Linguistics (语料库语言学的弗斯学说基础). In: *Journal of Foreign Languages* (《外国语》) 2. 23–32.
- Xie, Guiping (谢贵平). 2016. The Construction of the “China-Pakistan Economic Corridor” and its Cross-border Non-traditional Security Governance (“中巴经济走廊” 建设及其跨境非传统安全治理). In: *Southeast Asian Affairs* (《南洋问题研究》) 3. 23–37.
- Zhang, Delu (张德禄). 2009. On a Synthetic Theoretical Framework for Multimodal Discourse Analysis (多模态话语分析综合理论框架探索). In: *Foreign Languages in China* (《中国外语》) 6(1). 24–30.
- Zhang, Xinyue (张馨月)/Wu, Xinru (吴信如). 2022. Der Armutsbekämpfungseffekt der „Belt and Road Initiative“ für die Länder entlang der Neuen Seidenstraße – Eine Analyse basierend auf dem Mechanismus der „Five Connectivity“ (“一带一路” 倡议对沿线国家的减贫效应——基于 “五通” 作用机制的分析). In: *Shandong Social Sciences* (《山东社会科学》) 11. 169–175.

Angaben zu Personen

Jingcao Chen ist Doktorandin im Bereich interkultureller Kommunikation zwischen Deutschland und China an der Beijing Foreign Studies University (China). Ihre wissenschaftlichen Interessen liegen vor allem in der mediatisierten interkulturellen Kommunikation und multimodalen Diskursanalyse.

Jing Li ist Assoc. Professorin an der Beijing Foreign Studies University (China). Sie hat im Bereich Rechtslinguistik promoviert. Zu ihren Forschungsschwerpunkten gehören Text- und Diskursanalyse, Fremdsprachendidaktik.