

Zweitveröffentlichung



Reitbrecht, Sandra

Schreibdidaktische Modellvideos multimodal inventarisieren und transkribieren

Datum der Zweitveröffentlichung: 17.07.2026

Akzeptiertes Manuskript (Postprint), Beitrag in Sammelwerk

Persistenter Identifikator: urn:nbn:de:bvb:473-irb-116221x

Erstveröffentlichung

Reitbrecht, Sandra (2022): Schreibdidaktische Modellvideos multimodal inventarisieren und transkribieren, in: Cordula Schwarze und Sven Grawunder (Hrsg.), Transkription und Annotation gesprochener Sprache und multimodaler Interaktion : Konzepte, Probleme, Lösungen, Tübingen: Narr Francke Attempto Verlag, S. 163–185, doi: 10.24053/9783823394693.

Rechtehinweis

Dieses Werk ist durch das Urheberrecht und/oder die Angabe einer Lizenz geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die für Sie geltende Gesetzgebung zum Urheberrecht und/oder durch die Lizenz erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

Für dieses Dokument gilt eine Creative-Commons-Lizenz.



Die Lizenzinformationen sind online verfügbar:

<https://creativecommons.org/licenses/by/4.0/legalcode>

Schreibdidaktische Modellvideos multimodal inventarisieren und transkribieren

Sandra Reitbrecht

Abstract: Erklärvideos als Bildungsmedien erleben derzeit einen Höhenflug an Relevanz, was auch in der (didaktischen) Forschung entsprechendes Interesse geweckt hat. So gilt es, angemessene forschungsmethodische Zugänge zu Erklärvideos zu finden und zu erproben. Am Beispiel screen-cast-basierter schreibdidaktischer Modellvideos zeigt der Beitrag, wie bestehende Verfahren zur Inventarisierung und Transkription multimodaler Daten auf den „neuen“ Untersuchungsgegenstand angewandt werden können und welchen Mehrwert hierfür gesprächsanalytische Zugänge bereithalten, wenn es um das präzise Aufzeigen räumlich-zeitlicher Koordination beim Erklären bzw. Modellieren geht.

Keywords: Erklärvideos, Inventarisierung, Transkription, Multimodalität, Interaktionalität, digital mediatisierte Handlungen

1 Einleitung

Erklärvideos stehen als Bildungsmedien derzeit hoch im Kurs: Eine Studie des Rats für Kulturelle Bildung (2019: 28) zeigt, dass 47 Prozent der befragten Schüler*innen YouTube für Themen, die in der Schule behandelt werden, als wichtig oder sehr wichtig erachten. Neben dieser individuellen Nutzung von YouTube-Videos für Lernzwecke integrieren auch Lehrwerksverlage vermehrt Videos in ihre Angebote für formelle Lernsituationen. Apps und Softwareprogramme ermöglichen zudem auch Lehrenden ein weitgehend unkompliziertes Erstellen von Erklärvideos (vgl. Dorgerloh & Wolf 2020).

Die didaktische Forschung reagiert auf diese Entwicklung mit Forschungsvorhaben, in denen neben primär beschreibend-analytischen Zugängen (z. B. zur Systematisierung unterschiedlicher Erklärvideoformate und multimodaler Erklärpraktiken) oftmals auch eine wertende Perspektive auf die Erklärqualität

und Wirksamkeit der Videos eingenommen wird (vgl. Findeisen et al. 2019). Erklärvideos als multimodale Forschungsobjekte stoßen in diesem Zusammenhang demnach auch die Diskussion um angemessene forschungsmethodische Herangehensweisen an den Untersuchungsgegenstand an. Erste Annäherungen und Vorschläge liegen bereits vor. Auffällig ist, dass sie sich hinsichtlich der Umsetzung eines Transkriptionsschrittes vor der Datenanalyse bzw. einer Darstellung ihres Vorgehens in Publikationen unterschiedlich verhalten.

Der vorliegende Beitrag knüpft an diesen Ist-Stand an und präsentiert am Genre screencast-basierter schreibdidaktischer Modellvideos ein gesprächsanalytisches Vorgehen für die multimodale Inventarisierung und Transkription der Daten (Abschnitt 4). Dafür werden zunächst die hier im Zentrum stehenden schreibdidaktischen Modellvideos in ihrer Machart und ihren typischen Gestaltungsmitteln als Untersuchungsgegenstand näher beschrieben (Abschnitt 2). Abschnitt 3 widmet sich der Diskussion der hier vorgeschlagenen Verfahren für Inventarisierung und Transkription, und zwar aus zwei unterschiedlichen Blickwinkeln: Zum einen möchte ich mit Blick auf die oben angesprochenen rezenten methodischen Zugriffe auf Erklärvideos die forschungsmethodische Relevanz von Inventarisierung und Transkription der Videodaten herausarbeiten. Zum anderen beleuchtet der Abschnitt angesichts des Rückgriffs auf Verfahren der multimodalen Gesprächs- bzw. Interaktionsanalyse näher, wie sich der hier fokussierte Transkriptionsgegenstand der schreibdidaktischen Modellvideos hinsichtlich der Merkmale Multimodalität und Interaktionalität verhält bzw. näher bestimmen lässt. Ein Fazit (Abschnitt 5) schließt den Beitrag ab.

2 Screencast-basierte schreibdidaktische Modellvideos als Untersuchungsgegenstand

Bei den im Fokus stehenden schreibdidaktischen Modellvideos handelt es sich um ein spezifisches Genre von Erklärvideos: Lehrpersonen modellieren, d. h. sie zeigen vor und kommentieren, wie sie eine (Schreib-)Aufgabe lösen, und legen dabei im Verbalkommentar auch mentale Prozesse (z. B. handlungsleitende Entscheidungen und Begründungen oder metakognitive Regulationsmechanismen) offen, die aus der sichtbaren Handlung nicht oder nur ansatzweise erschlossen werden können. (Kognitives) Modellieren definiert sich damit als eine spezifische, vom eigenen Handeln ausgehende Form multimodalen Erklärens (Klein 2009: 29–30) und ist zugleich als geplante, didaktisch motivierte Lehrhandlung zu verstehen. Das heißt, dass man beim Modellieren zwar vom durch Lautes Denken begleiteten Handeln ausgeht, zugleich aber eine kommunikativ orientierte Erklärhaltung einnimmt, die sich an einer Zielgruppe und

an konkreten Lehrzielen ausrichtet und sowohl das Laute Denken als auch die visuelle Darstellung didaktisch verdichtet (vgl. Philipp 2015: 131, Sturm & Weder 2016: 84–86). Modellvideos¹ ähneln damit *Video-Tutorials* (Wolf 2015: 123) und grenzen sich zugleich von sogenannten *Performanzvideos* ab, „in denen eine beobachtbare Fertigkeit im Sinne einer Dokumentation oder einer Selbstdarstellung ohne weitere didaktische Aufarbeitung gezeigt wird“ (Wolf 2015: 123) (vgl. zu einer differenzierten genretheoretischen Bestimmung von Modellvideos: Reitbrecht 2021: 317–319).

Als genretypische Gestaltungselemente der didaktisch intendierten und verdichteten schreibdidaktischen Modellvideos² konnten bisher folgende Elemente ermittelt werden (vgl. auch Reitbrecht 2021: 311–317), die auf visueller und/oder verbaler Ebene (im Beitrag durchgängig verstanden als gesprochen-verbale Ebene) realisiert werden:

- Strukturierende Elemente: Diese liegen außerhalb des modellierten (Textproduktions-)Prozesses und sind Produkt der didaktisch motivierten Erklärhaltung. Es handelt sich dabei z. B. um Begrüßungspraktiken, Lehrzielformulierungen (► eContent_RE_a; zugleich Bsp. 2 in Abschnitt 4.2.2) oder aber auch Passagen, die Lehr-Lerninhalte, z. B. Merkmale einer Textsorte oder ein Strategienbündel (► eContent_RE_b und ► eContent_RE_c), erklären bzw. darstellen, die für den modellierten Prozess von Relevanz sind.
- (Geraffte) verbale Modellierungen (anhand von Produkten): Dabei handelt es sich um Modellierungen mit zusammenfassendem Charakter. Entweder ausschließlich verbal ohne spezifische Bezugnahme auf visueller Ebene oder anhand eines Zwischenproduktes (z. B. anhand eines erstellten Schreib-

1 Eingesetzt werden Modellvideos im Rahmen von Modelllernverfahren, bei denen die zentrale Phase der Modellbeobachtung durch weitere vorbereitende und nachbereitende didaktische Schritte gerahmt ist (vgl. u. a. Self-Regulated Strategy Development: Graham & Harris 2005; reflexives Modelllernen: Reitbrecht et al. 2019). Eine individuell-informelle Verwendung der Videos zu Lernzwecken ist aber ebenso denkbar.

2 Die Datenbeispiele in diesem Beitrag stammen aus dem AaMoL-Modellvideo 1, erstellt im Rahmen des Forschungsprojektes „Am Modell lernen, als Modell lernen“ (im Zeitraum 04/2018 bis 11/2019 gefördert vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung im Rahmen der Förderschiene Sparkling Science) und in seiner vollständigen Fassung auf der Projekthomepage der Universität Wien abrufbar: www.univie.ac.at/aamol. Das Modellvideo wurde von Brigitte Schatzl entwickelt. Ihr gilt an dieser Stelle mein besonderer Dank. Ebenso danke ich sehr herzlich Nina Hansmann, Esther Schmutte und Lisa Wagner, die im Rahmen eines Seminars im Masterstudium DaF/DaZ an der Universität Wien Modellvideos entwickelt haben und der Veröffentlichung von Screenshots aus ihren Videoprojekten zur weiteren Veranschaulichung der Gestaltungselemente zugestimmt haben (► eContent_RE_b und ► eContent_RE_c).

plans oder eines fertig verfassten Absatzes) werden die zurückliegenden Prozesse gerafft modelliert. Die Prozesshaftigkeit des (kognitiven) Modellierens kann dabei stark reduziert werden. Raffungen dieser Art sind aber angesichts der Dauer von Textproduktionsprozessen sowie der Fokussierung auf spezifische Lehrziele notwendig (► eContent_RE_d).

- Modellierte Prozesse: In diesen Phasen der Modellvideos werden Handlungen tatsächlich sicht-/erfahrbar und im Sinne des (kognitiven) Modellierens verbal kommentiert (► eContent_RE_e; zugleich Bsp. 3 in Abschnitt 4.2.3).

Darüber hinaus ist das hier thematisierte Korpus von schreibdidaktischen Modellvideos durch weitere gestalterische Merkmale gekennzeichnet, die für transkriptionsbezogene Entscheidungen handlungsleitend sind: In ihrer Machart zeichnen sich die Modellvideos durch einen technisch einfachen Produktionsprozess aus. Es handelt sich in den meisten Fällen um Bildschirmaufzeichnungen mit Audiospur für den verbalen Kommentar, aber ohne Webkamera-Einspielung. Man sieht die sprechende Person zumeist also nicht im Video. (Schreib-)Handlungen wie in den oben genannten modellierten Prozessen werden ebenso zumeist ausschließlich in ihrer digital mediatisierten Form beobachtbar. Hinsichtlich des multimodalen Charakters der Videos ist des Weiteren festzuhalten, dass aufgrund des schreibdidaktischen Inhalts zumeist Präsentations-, Textverarbeitungs- oder Readerprogramme zum Einsatz kommen. Aufwändige Animationen sowie Musik-/Geräuscheinspielungen finden nur spärlich Verwendung. Dadurch ist die visuelle Ebene im Vergleich zu anderen Videoformaten oftmals über längere Passagen hinweg statisch. Geräusche und Musik als weitere Modalitäten sind bis auf hörbare Tipp-/Klickgeräusche kaum von Relevanz.

Ziel der Auseinandersetzung mit der soeben skizzierten Datensammlung ist es, beschreibend-analytisch (sowie auch vergleichend-bewertend³ anhand von zu Beginn sowie am Ende einer Lehrveranstaltung erstellten Modellvideos) strukturelle Aspekte von Modellvideos sowie Praktiken des Modellierens in ihrem multimodalen Charakter und ihrer Koordination zu erschließen. Deshalb muss das Transkriptionsverfahren grundsätzlich dem multimodalen Charakter der Daten unter Berücksichtigung paraverbaler Merkmale gerecht werden, zudem „die simultane und sequenzielle Koordination verschiedener multimodaler Ressourcen“ (Deppermann 2018: 57) präzise repräsentieren.

3 Vgl. zu Bewertungen sprachlichen Handelns, die über das rekonstruktive Analyseverständnis der Gesprächsforschung hinausgehen, die aktuelle Diskussion bei Brünner & Pick (2020).

3 Transkriptions- und analysemethodische Zugänge zu Erklärvideos

Betrachtet man vor diesem Hintergrund aktuelle forschungsmethodische Zugänge zu Erklärvideos, so unterscheiden sich diese durchaus in ihrem Umgang mit Transkription sowie hinsichtlich ihrer Passung mit den soeben genannten Anforderungen. Zeigen möchte ich diese Unterschiedlichkeit an drei Verfahren, die als Bezugspunkte – gemeinsam mit dem hier präsentierten gesprächsanalytischen Zugang – einen Raum forschungsmethodischer Handlungsmöglichkeiten für die Auseinandersetzung mit Erklärvideos aufspannen.

Wolf & Kratzer (2015) arbeiten beispielsweise in einem mediendidaktischen Kontext mit einem Analyseraster, das direkt in der (wiederholten) Videobeobachtung angewendet wird. Ein Transkriptionsschritt ist dabei nicht vorgesehen. Neben Kontextinformationen (z. B. Sprache, Dauer des Videos) sowie Filmgestaltungsmerkmalen (z. B. Kameraperspektive) umfasst der Analysebogen als dritten Teil ein differenziertes Kategoriensystem zur Erfassung der didaktischen Gestaltung von Erklärvideos. Damit verkörpert er ein standardisiertes, empirisch erprobtes und vergleichsweise zeitökonomisches Instrument zur Beurteilung des Vorhandenseins sowie der qualitativen Ausprägung struktureller Elemente und bestimmter Merkmale von Erklärvideos (Wolf & Kratzer 2015: 34–36). Berücksichtigung finden dabei unter anderem die Modalitäten der Videos, z. B. in den genannten Erklärmitteln (Bilder, Sprache, Videos, anwesende Person, Schrift) oder in weiteren Kategorien wie Soundeffekten oder flüssiger Sprache, zudem ihre strukturellen Elemente (Intro, Begrüßung, Zielformulierung am Anfang der Erklärung, u. a.). Ebenso wird die Kommunikation mit dem Publikum durch eine Analyse anhand mehrerer Kategorien (direkte Ansprache einer Person, direkte Ansprache einer Gruppe, Interaktion) im Analyseinstrument berücksichtigt (Wolf & Kratzer 2015: 36). Zugleich werden aber auch Grenzen des Zugangs erkennbar, da in der Analyse durch die Kategorienorientierung im Raster das Erklärvideo in seinen Merkmalen und Modalitäten vor allem analytisch dekomponiert wird, darüber hinaus aber kaum Aussagen über das räumlich-zeitliche Zusammenwirken der einzelnen Elemente und Merkmale erschließbar werden (z. B. im Sinne einer modalitätsübergreifenden Kohärenz; vgl. Lobin 2013: 68–71; siehe vergleichbar das oben angeführte Zitat von Deppermann 2018: 57). Damit bleibt auch die Sequenzialität der einzelnen multimodalen Praktiken weitestgehend unerschlossen. Ansatzweise erkennbar wird die Dimension der Zeit nur in wenigen Analysekategorien des Rasters, so z. B. in der oben genannten Kategorie „Zielformulierung *am Anfang* der Erklärung“ oder auch bei „[r]egelmäßige Zusammenfassungen“ und „Erklärung

induktiv, *vom Beispiel zur allgemeinen Regel*“ (Wolf & Kratzer 2015: 36; Kursivierungen S.R.).

Ebel (2021) wählt aus einer sprechwissenschaftlich-phonetischen Perspektive auf die Verständlichkeit von Erklärvideos ein Verfahren, das eine Transkription der verbalen Ebene nach GAT 2 (Selting et al. 2009) sowie Segmentierungen in PRAAT zur Aufbereitung der Daten für messphonetische Zugriffe (z. B. zur Ermittlung der Artikulationsrate) umfasst. Damit wird ein viel präziserer bzw. feinkörnigerer Blick auf verbale Erklärpraktiken in ihrer Sequenzialität und ihrer paraverbalen Qualität und Gestaltung möglich. Die visuelle Ebene wird in die Darstellung der Analyseergebnisse ebenfalls durch Screenshots und Beschreibungen mehrfach einbezogen, genaueren Einblick in die Berücksichtigung der visuellen Ebene bei der Datenaufbereitung und -transkription gibt der Beitrag aber nicht.⁴

Schöne & Wedler (2021) wählen einen textanalytischen Zugang und verstehen ein Erklärvideo als multimodalen Text, der neben weiteren Modalitäten „mindestens ein sprachliches Zeichen enthält (welches meistens durch gesprochene Sprache realisiert wird)“ (Schöne & Wedler 2021: 280). In ihrer beispielhaften multimodalen Textanalyse eines Erklärvideos ziehen sie die von Stöckl (2016: 22–25) vorgeschlagenen Kategorien heran (Gliederung/Abgrenzung, Handlungsstruktur, Themenstruktur, multimodale Verknüpfung und Intertextualität) und erweitern diese um die Kategorie des situativen Kontexts (Schöne & Wedler 2021: 281–287). Damit berücksichtigen die Kategorien des textanalytischen Zugriffs stärker als das oben gezeigte Analyseinstrument von Wolf & Kratzer (2015) die Dimension der Zeit bzw. die zeitliche Strukturierung und Sequenzialität bezogen auf Themen und Handlungen. Die Kategorie der multimodalen Verknüpfung greift zudem das Zusammenwirken der einzelnen Modalitäten im Erklärvideo auf. Hinsichtlich der Datenaufbereitung bzw. eines etwaigen Transkriptionsschrittes geben Schöne & Wedler (2021) in ihrem Beitrag allerdings keine Auskunft. Relevanz und Funktion von Transkription werden im Kontext der multimodalen Textanalyse aber sehr wohl diskutiert (so z. B. auch bei Schöne 2018). Für Schneider & Stöckl (2011: v. a. 28–33) beispielsweise stellt ein Vorgehen ohne Transkription in der textanalytischen Auseinandersetzung mit multimodalen Texten wie Werbespots keine forschungsmethodische Handlungsoption dar: „Will man multimodale Texte – wie in

4 Vgl. diesbezüglich auch Stukenbrock (2009: 148): „Die Integration von aussagekräftigen Standbildern der analysierten Videoausschnitte in den Transkript-Text, die eine holistische Repräsentation des kommunikativen Ereignisses leistet, wird in verschiedenen Publikationen erfolgreich praktiziert, allerdings ohne dass die Kriterien der Standbildgewinnung formuliert würden, so dass der Naivitätsverdacht hinsichtlich des repräsentationalen Status von Standbildern nahe liegt.“

unserem Fall einen TV-Werbespot – unter verschiedenen Gesichtspunkten untersuchen, so *muss* man sie zunächst transkribieren.“ (Schneider & Stöckl 2011: 28; Kursivierung S.R.). Die Umsetzung der Transkription erfolgt in tabellarischer Form, wobei Spalten für die Kategorien Zeit (kumuliert in Sekunden), Kamera (Beschreibung der Kameraperspektive sowie auch zentraler Handlungen im Video), Einstellung (Abbildung eines Screenshots), Sprachtext (Transkription der verbalen Ebene unter Berücksichtigung von Überlappungen) und Ton (Beschreibung von Geräuschen und Musik) vorgesehen sind (vgl. Schneider & Stöckl 2011: 39–44).⁵ Stertkamp & Schüler (2014) wenden eine vergleichbare tabellarische Darstellung auch auf screencast-basierte Daten aus zwei unterschiedlichen Forschungsprojekten an. In einem der beiden Projekte geht es um die multimodale Interaktion (Videochat, Interagieren in einem Videospiel mittels Avataren und Videotelefonie) von Personen in einem Online-Rollenspiel, die füreinander nicht (oder nur in Form ihrer Avatare) sichtbar sind (vgl. Machart der Modellvideos zumeist ohne Webcam-Video in meiner Datensammlung). Die verbale Interaktion wird dabei mit GAT 2 transkribiert, zudem werden unterschiedliche, teils auch nicht-konventionalisierte Verfahren (z. B. Rahmungen) für das Darstellen von Überlappungen bzw. Simultanität (auch zwischen Videochatnachrichten und Interaktionen via Videotelefonie) in der Tabelle angewandt (vgl. Stertkamp & Schüler 2014: 341–352), womit das Verfahren in Summe den am Ende von Abschnitt 2 genannten Anforderungen in vielen Punkten am nächsten kommt.

Der im Folgenden gezeigte gesprächsanalytische Zugang für die multimodale Aufbereitung und Transkription von Modellvideos knüpft hier an und stellt einen weiteren methodischen Zugriff auf Erklärvideos dar, erfüllt dabei die in Abschnitt 2 genannten Anforderungen und zieht gezielt standardisierte und erprobte Verfahren heran: Es sind dies das *Gesprächsinventar* (Deppermann 2008: 32–35), das *Gesprächsanalytische Transkriptionssystem 2* (GAT 2; Selting et al. 2009) sowie die *Conventions for multimodal transcription* (Mondada 2019). Sie erlauben den mit den Konventionen vertrauten Leser*innen einen Nachvollzug der Daten ohne großen Aufwand. Zudem kann durch Verweise auf die entsprechenden Verfahren auch in Publikationen geringeren Umfangs Transparenz gewährleistet werden, da nur mehr etwaige Erweiterungen oder Spezifizierungen der Konventionen ergänzend dargestellt werden müssen. Die folgenden Ausführungen verdeutlichen, dass diese Erweiterungen und Spezi-

5 Vgl. in diesem Zusammenhang auch das Verfahren der Feldpartitur zur Filmtranskription in der Qualitativen Sozialforschung (Moritz 2011), das allerdings für die Aufbereitung und Transkription von Videodaten vordefinierte und damit immer gleich lange Zeiteinheiten (ab einer Dauer von 0,4 Sekunden) vorsieht.

fizierungen in Summe gering ausfallen (vgl. auch die Anwendung von GAT 2 durch multimodale Erweiterungen auf einen Werbespot bei Meer & Pick 2019: 90–100). GAT 2 berücksichtigt zudem die paraverbalen Charakteristika des Sprechens für die hier verfolgten Erkenntnisinteressen angemessen und bietet dahingehend mit Minimal-, Basis- und Feintranskript auch unterschiedliche Ausbaustufen an. Mit dem vorgeschalteten Erstellen von Gesprächsinventaren zu den Videos können bereits wichtige analytische Zugriffe auf makrostrukturelle Organisation und die multimodale Qualität der Datensätze erfolgen. Auch erlaubt dieser Zwischenschritt im Sinne der Zeitökonomie eine Vorsortierung und Auswahl von Videopassagen, die je nach Forschungsfrage/-interesse tatsächlich einer Transkription zugeführt werden. Mondadas Konventionen für die multimodale Transkription (2019) ermöglichen des Weiteren zur Verlaufsdarstellung von Aktivitäten einen Präzisions- und Granularitätsgrad in der Repräsentation multimodaler Koordinationen im Transkript (vgl. Deppermann 2018: 57), der für die Darstellung der kommentierten mediatisierten Handlungen in den Modellvideos eine Bereicherung darstellt (vgl. v. a. das Beispiel Cursorbewegungen in Abschnitt 4.2.2) und dabei den oben vorgestellten forschungsmethodischen Zugriffen überlegen ist.

Damit sind die Gründe für die Wahl der genannten Transkriptionsverfahren offengelegt, sie können mit Blick auf den Untersuchungsgegenstand der Modellvideos als gegenstandsangemessen qualifiziert werden. Nicht geklärt ist allerdings die Frage, wie sich der Untersuchungsgegenstand hinsichtlich der in der multimodalen Gesprächs- bzw. Interaktionsanalyse relevanten Merkmale Multimodalität und Interaktionalität verhält, ob also aus Perspektive der gewählten Verfahren grundsätzlich ebenfalls eine Passung erkennbar ist, worauf ich an dieser Stelle abschließend eingehen möchte.

Hinsichtlich des Multimodalitätsbegriffs ist in der multimodalen Gesprächs- bzw. Interaktionsanalyse oftmals eine Engführung auf nonverbales Verhalten (als Ergänzung zum verbalen) erkennbar. Dies scheint auch durch die zentralen Forschungsinteressen an Interaktionen bei Ko-Präsenz der beteiligten Personen sowie aus einer historischen Perspektive durch das erst nach und nach vermehrte Arbeiten mit Videographie statt ausschließlich Audiographie zur Datengewinnung bedingt zu sein. Grundsätzlich nämlich finden sich in der einschlägigen Fachliteratur durchaus auch Ausführungen, die auf offene Begriffskonzepte hindeuten, wenn beispielsweise laut Birkner, Auer, Bauer & Kotthoff aus konversationsanalytischer Perspektive in multimodalen Transkripten neben dem Gesprochenen auch „andere körperliche bzw. visuelle Ausdrucksressourcen notiert werden“ (2020: 24). Visuelle Ausdrucksressourcen können dabei meines Erachtens auch auf die Gestaltungsmöglichkeiten bei der Videoproduktion sowie auf die digital mediatisierten Handlungen, bei denen die

Personen in ihrem Handeln für das Gegenüber nur mehr eingeschränkt und mittelbar durch das Medium des Modellvideos erfahrbar werden, bezogen werden. Rückendeckung erhält ein offenes Verständnis von Multimodalität in gewisser Weise auch von Imo & Lanwer (2019: 246–248). Sie unterscheiden Multimodalität im engeren Sinne von Multimodalität im weiteren Sinne: Die Interaktionale Linguistik folge dabei zwar einem engen Begriffsverständnis, in dem Multimodalität „das Zusammenspiel von gesprochener Sprache und visuell-körperlichen Ausdrucksmitteln wie Gestik, Mimik usw.“ meint (Imo & Lanwer 2019: 248). Angesichts immer präsenter werdender medial-vermittelter Interaktionen scheint aber ergänzend ein erweiterter Multimodalitätsbegriff zielführend, der z. B. auch das Zusammenwirken von gesprochenen und geschriebenen Sprachnachrichten sowie übermittelten Bildern in der WhatsApp-Kommunikation einschließt (Imo & Lanwer 2019: 248) und damit auch für die Modellvideos Gültigkeit hat. Deppermann (2018: 58) geht aus einem anderen Blickwinkel an die Begriffsfrage heran und erachtet den Begriff „Multimodalität“ an sich als unzureichend spezifisch, wenn es konkret um die „[m]ultimodale Interaktion unter Anwesenden“ geht. Diese sei vorrangig durch „Leiblichkeit“ gekennzeichnet. „„Multimodalität“ hebt dagegen hervor, dass leibliches Handeln Ressourcen verschiedener Modalität benutzt“ (Deppermann 2018: 58), womit im Umkehrschluss der Multimodalitätsbegriff ebenfalls offen für andere als ausschließlich leibliche (oder in ihrer Leiblichkeit unmittelbar wahrnehmbare) multimodale Ressourcen ist.

Die Frage, ob in den Modellvideos darüber hinaus auch das Merkmal der Interaktionalität gegeben ist, hat vermutlich ebenso Potenzial für eine wissenschaftliche Kontroverse. Auch hierzu finden sich aber Positionen in der Fachliteratur, die für ein offenes, erweitertes Begriffsverständnis plädieren und damit auch bei der hier gezeigten Kommunikation via Modellvideos von Interaktionalität sprechen. Konkret tut dies z. B. Schmidt (2018: 29), indem er Interaktion auch dann als gegeben erachtet, wenn wie beispielsweise in Fernsehsendungen „eine Antwort (in welcher Form auch immer) ‚gesucht‘ wird, was sich wiederum im Produkt einschreibt“, und „ein dreizügiger Austausch (Angebot/Annahme/Weitersenden) in formaler und vermittelter Form zustande kommt“ (Schmidt 2018: 29). Für die Modellvideos, die sich durch einen hohen Grad didaktischer Intentionalität auszeichnen (vgl. Abschnitt 2), werden diese Kriterien mit Blick auf ihren Einsatz in Lehr-Lern-Situationen erfüllt: So können z. B. Schüler*innen das Video zum Lernen nutzen und Lehrpersonen in weiterer Folge zu diesem Lernprozess Feedback geben. Ins Produkt eingeschrieben wird dies, wenn z. B. die Lehrperson den Schüler*innen im Video abschließend alles Gute beim Lösen einer vergleichbaren Schreibaufgabe wünscht. Ebenso können Lernende nach der Rezeption des Videos im Unterrichtsgeschehen Fragen an die

Lehrperson richten, welche diese wiederum beantwortet. Darüber hinaus können die Videos aber auch Spuren explizit antizipierter Dreizügigkeit beinhalten, wenn die Lehrperson die Beantwortung potenzieller Schüler*innenfragen im Sinne kompetenten Lehrendenhandelns bereits im Video einlöst (fiktives Beispiel: „Ihr fragt euch nun vielleicht, warum ich mich für diese Vorgehensweise entschieden habe. Nun, das lässt sich ganz einfach beantworten: Ich kann damit beim Schreiben viel schneller vorgehen, muss nicht noch einmal alle Informationen in den Quellen recherchieren.“).

4 Gesprächsanalytisches Vorgehen für Dateninventarisierung und -transkription⁶

4.1 Schreibdidaktische Modellvideos inventarisieren

Gesprächsinventare dienen dem schnellen Überblick sowie einer ersten Gliederung von Gesprächen und stellen eine Grundlage für weitere forschungsmethodische Entscheidungen (z. B. die Auswahl zu transkribierender Passagen) dar (vgl. Deppermann 2008: 32). Sie umfassen ein Deckblatt (► eContent_RE_f, S. 1), welches im Fall der schreibdidaktischen Modellvideos auch den im Video produzierten Text beinhaltet, sowie ein tabellarisch gestaltetes Formblatt. Abb. 1 zeigt in diesem Zusammenhang die von Deppermann (2008: 34) vorgeschlagenen Kategorien der Tabelle, Abb. 2 die adaptierte Form für die hier fokussierte Datensammlung screencast-basierter schreibdidaktischer Modellvideos. Ergänzend kann das in den Kernkategorien vollständige Gesprächsinventar (► eContent_RE_f) digital abgerufen werden.

Zeit	Transk	Sprecher*in	Inhalt-Handlung	Memo	Forschungsfrage
Beginnzeit des Abschnitts laut Aufnahme	Anm., ob Transkript existiert; Angabe v. Seite bzw. Zeilen	Siglen der am Gespräch teilnehmenden Personen	wesentliche Gesprächsthemen und -phasen in einfacher, stichwortartiger Beschreibung	Auffälligkeiten verschiedener Art (z. B. non-verbale Phänomene; Kommen u. Gehen von Personen)	Überlegungen zu Forschungsfragen: zumeist erst bei der Materialbearbeitung befüllt

Tab. 1: Kategorien eines Gesprächsinventars nach Deppermann (2008: 32–35); Erläuterungen auf Basis der Quelle von der Autorin ergänzt

6 Sowohl für die Segmentierung und Annotationen im Gesprächsinventar, die Erstellung der Screenshots als auch die Transkription wurde die Software ELAN verwendet. Die Daten wurden nach dem Export entsprechend weiterverarbeitet.

Tr	Zeit	Inhalt-Handlung	Zeit	Visuelle Ebene	Aufbereitung Bild	Merkmale	Forschungsfrage
	00:00	Begrüßung und Kontextualisierung	00:00	Begrüßungsfolie		stat, klic	
	00:26	Ankündigung der Lehr-Lernziele	00:34	Folie zu Lehr-Lernzielen	v2a (00:34) v2b (00:41) v2c (00:42) v2d (00:56) v2e (01:03)	takt, klic	
	00:43	Erläuterung des 1. Ziels: Auswahl der Informationen aus den Materialien					

Tab. 2: Kategorien des adaptierten Gesprächsinventars für Modellvideos

Die Kategorien „Transkript“ (Transk bzw. Tr) und „Forschungsfrage“ erfüllen auch im adaptierten Inventar die bei Deppermann (2008) genannten Funktionen. Ebenso gilt dies für die Spalte „Inhalt-Handlung“ und die links davon angeführte Zeitspalte. Die Kategorie Sprecher*in entfällt aufgrund des monologischen Charakters der Modellvideos, die entsprechende Information kann dem Deckblatt entnommen werden.

Erweiterungen bzw. Spezifikationen erfolgen mit Blick auf den spezifischen multimodalen Charakter der Modellvideos wie folgt: Neben einer Zeitspalte, die der (verbalen) Handlungslogik des Modellvideos folgt, wird eine zweite Zeitspalte ergänzt, die sich an der Abfolge der visuellen Elemente orientiert. In der Spalte rechts davon werden die einzelnen visuellen Elemente knapp beschreibend genannt. Eine neue Zeile wird immer dann gesetzt, wenn im Video ein Dokumentenwechsel bzw. Schnitt erfolgt. Versetzte Zeilen zeigen Abweichungen in der Taktung der verbalen und visuellen Gestaltungsebene an.

Die Spalte „Aufbereitung Bild“ erlaubt vergleichbar der Spalte „Transkript“ Verweise auf Screenshots zur Aufbereitung der visuellen Gestaltungsebene der Videos. Mittels Hyperlinks können die Dateien ebenso wie Transkripte direkt im Gesprächsinventar hinterlegt werden. Auch ein Einfügen der Screenshots in die Datei als Bilder ist denkbar (vgl. dazu z. B. das Vorgehen bei Schneider & Stöckl 2011: 39–44). Im Gesprächsinventar zum AaMol-Modellvideo 1 (► eContent_RE_f) ist diese Spalte beispielhaft für jene drei Videoabschnitte ausgefüllt, die auch im Abschnitt 4.2 hinsichtlich ihrer Transkription näher vorgestellt werden. Das Basisprinzip ist dabei ein Minimum von einem Screenshot unmittelbar nach Beginn des Abschnitts. Weitere Screenshots werden je nach Gestaltung der visuellen Ebene (z. B. getaktete Präsentation, Cursorbewegungen, Schreibprozesse) ergänzt. Die Zeitmarken dieser weiteren Bildaufnahmen

stimmen mit den Zeitpunkten entsprechender Ereignisse/Veränderungen überein und sind bereits auf die Datentranskription abgestimmt (s. Abschnitt 4.2).

Die Spalte „Memo“ im Gesprächsinventar nach Deppermann (2008) wird in eine Spalte „Merkmale“ umbenannt. Damit wird in Abgrenzung zum Memo-Begriff explizit gemacht, dass in dieser Spalte bereits Ergebnisse aus der analytischen Auseinandersetzung mit dem Videomaterial festgehalten werden, und zwar solche, die sich auf die in den einzelnen Abschnitten erkennbaren (genrespezifischen) Gestaltungselemente beziehen. Dafür empfiehlt sich die Entwicklung eines Codesystems für gängige multimodale Gestaltungsformen der screencast-basierten Modellvideos. Zur Veranschaulichung eines solchen zum aktuellen Zeitpunkt noch unvollständigen Codesystems wurde die entsprechende Spalte ebenfalls für die drei in Abschnitt 4.2 ausgewählten Auszüge befüllt.

4.2 Schreibdidaktische Modellvideos transkribieren

Die Transkripte zu den Modellvideos umfassen bis zu vier Zeilen.⁷ Neben der verbalen Ebene (transkribiert als Basistranskript nach GAT 2; Selting et al. 2009) berücksichtigen die weiteren Zeilen die visuelle Gestaltung des Videos (orientiert an Mondada 2019). Die v-Zeile steht dabei für den „Videohintergrund“. Die beiden weiteren Zeilen bilden auf dem Bildschirm sichtbare Bewegungen (z.B. Cursorbewegungen) und Prozesse (z.B. Schreibprozesse) ab. Klick-/Tip-paktivitäten, die durch den Mediatisierungsprozess ausschließlich zu einem Wechsel des Videohintergrunds (z.B. Folienwechsel bei getakteter Präsentation wie in Bsp. 1 unten) führen, können durch die v-Zeile erfasst werden und werden daher im Sinne der Ökonomie kein zweites Mal abgebildet. Ebenso gibt die v-Zeile das Einsetzen eines neuen Videohintergrunds bereits im präzisen zeitlichen Abgleich mit der verbalen Ebene an, sodass auch von einer eigenen Zeile für Screenshots (vgl. Mondada 2019) abgesehen wird. Die folgenden Beispiele⁸ verdeutlichen das Transkriptionsverfahren anhand von drei häufig vorkommenden multimodalen Praktiken in den Modellvideos.

4.2.1 Getaktete Präsentation

Beispiel 1 (► eContent_RE_a) zeigt die oben bereits erwähnte getaktete Präsentation, bei der auf einer Präsentationsfolie weitere Elemente ergänzt werden.

7 Weitere Zeilen für eingespielte Geräusche/Musik sind ebenfalls möglich, werden hier aufgrund ihres geringen Vorkommens im Korpus nicht näher erläutert. Verweise auf hörbare Tipp-/Klickgeräusche erfolgen in der Merkmalspalte des Gesprächsinventars.

8 Die Beispiele verfolgen das Ziel, das Transkriptionsverfahren zu veranschaulichen. Dass die Auszüge in ihrem Anfang und Ende dabei nur teilweise der Strukturierung im Gesprächsinventar folgen, ist diesem Punkt geschuldet.

Das Transkript umfasst zur Darstellung dieses Prinzips zwei Zeilen: neben der Zeile für das Verbaltranskript auch die v-Zeile.

Bsp. 1: getaktete Präsentation, aus: AaMoL-Modellvideo 1, BS, 00:34-00:55

01 BS *(--) und ZWA:R (---) äm:: beziehen w_äm;;
v *v2a--->



v2a (00:34)

02 BS und zwar berücksichtigen WIR besonders
DR*EI aspekte, (-)*°hh (-)
v ->*v2b----->*v2c--->



v2b (00:41)



v2c (00:42)

03 BS der erste WÄre die auswahl der passenden
informatIOnen aus den SEHR umfangreichen
materIAlIen, °hh

04 BS das HEISST-

05 BS die erGEBnisse °hh äm: der studien und
die meTHOdik der studien;

v ----->>

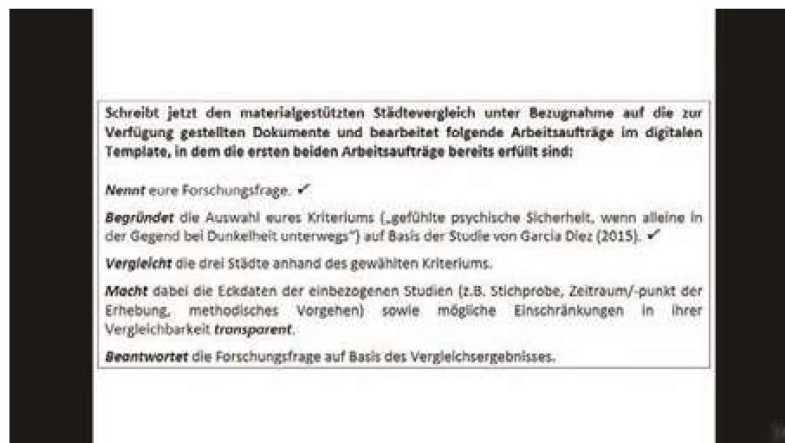
Die ab 00:34 transkribierte Sequenz setzt mit einem Wechsel auf visueller Ebene ein (vgl. Gesprächsinventar: Visuelle Ebene). Zwei Erweiterungen erfolgen durch die getaktete Präsentation auf der gezeigten Folie, die durch die Screenshots v2b und v2c veranschaulicht werden. Die mit * markierten Stellen zeigen den konkreten Moment der Veränderung an. In den Zeilen 03 und 04 wird die v-Zeile aufgrund des kontinuierlichen Fortlaufens von v2c nicht gesondert angezeigt, erst wieder am Ende des Transkriptauszugs (Zeile 05), wobei der doppelte Pfeil am Ende der v-Zeile signalisiert, dass das Bildelement v2c auch noch über den Transkriptauszug hinaus die visuelle Ebene des Modellvideos ausmacht (vgl. Mondada 2019). Bei Präsentationen mit animiertem Wechsel (z. B. Einfliegen neuer Elemente) könnte für diese jeweils kurze animierte Phase die bei Mondada (2019) vorgesehene Konvention für die vorbereitende Phase von Handlungen herangezogen werden (z. B. *v2a--->*.....*v2b--->).

Zeile 02 verdeutlicht dabei das Potenzial des gewählten Verfahrens für den präzisen Nachvollzug der „simultane[n] und sequenzielle[n] Koordination verschiedener multimodaler Ressourcen“ (Deppermann 2018: 57). Das Transkript lässt dank der berücksichtigten paraverbalen Parameter in der Verbaltranskriptzeile sowie der ebenfalls abgebildeten v-Zeile schnell erkennen, dass die Wechsel des Bildhintergrunds als mediatisierte Aktivitäten intrapersonal koordiniert mit einer Akzentuierung bzw. in einer Pause erfolgen. Im weiteren Abgleich mit den abgebildeten Screenshots werden weitere Aspekte des multimodalen Zusammenwirkens erkennbar: So folgt die Lehrperson in Zeile 02 dem Wortlaut der Präsentationsfolie. In Zeile 03 bis 05 zeigen sich hingegen auch Abweichungen. Im gesprochenen Kommentar wird das Adverb SEHR ergänzt und dabei auch betont realisiert. Die Ausführungen in Zeile 04 und 05 erläutern bzw. konkretisieren den Text auf der Folie.

4.2.2 *Cursorbewegungen*

Bespiel 2 widmet sich Cursorbewegungen (mit deiktischer Funktion) auf einem statischen, unveränderten Hintergrund (► eContent_RE_g). Um diese beiden Qualitäten, also den statischen Hintergrund zum einen und die Cursorbewegung zum anderen, zu berücksichtigen, umfasst das Transkript bis zu drei Zeilen. Es setzt einige Sekunden nach dem Beginn der Erläuterungen zur Schreibaufgabe ein.

Bsp. 2: Cursorbewegungen, aus: AaMoL-Modellvideo 1, BS, 01:36-02:02



v3a (01:20)

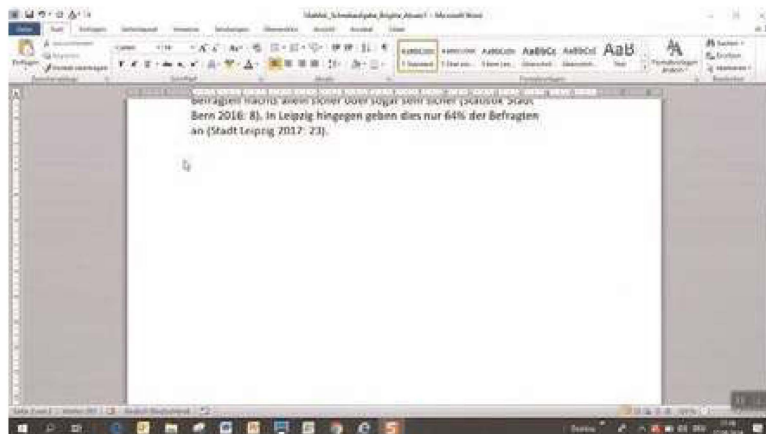
- 01 BS durch die +HÄKchen,+ (1.3)
bs +Cursor geht über rechtes zum
linken Häkchen+Cursor zeichnet beide
Häkchen nach--->
v >>v3a----->
- 02 BS äm:: °hh+ stellt man LEICHT fest;
bs ----->+Cursor kreist über
Arbeitsauftrag 1 und 2--->
- 03 BS was bereits DA ist-+
bs ----->+
- 04 BS +un:d welche +DREI:- (--)
bs +.....+Cursor unterstreicht
Vergleicht, Macht und Beantwortet--->
- 05 BS äm a+rbeitsaufträge noch zu erF+ÜLlen
sind, (1.5)+°hh
bs --->+.....+Cursor
geht nach rechts oben, dann nach rechts
unten in den schwarzen
Rand+,,,,,--->
- 06 BS diese darstellung gibt auch schon
eine- °h+
bs --->,,,,,+
- 07 BS äm: ein sehr gutes beispiel AB- °h
- 08 BS wie g:anz generell SCHREIB (.) planung
aussehen (-) KÖNNte; (--)
- 09 BS oder KANN. °hhh (1.5)
v ----->>

Die in Zeile 01 und 09 ausgewiesene v-Zeile zeigt an, dass v3a als Hintergrund statisch bleibt, bereits vor 01:36 (>>; Mondada 2019) beginnt und erst nach 02:02 (--->>; Mondada 2019) endet. Die Cursorbewegungen werden in der bs-Zeile beschrieben. Die Transkriptionszeichen für vorbereitende Schritte einer Handlung (...; Mondada 2019) werden dabei für die Passagen verwendet, in denen keine klare Gerichtetheit bzw. explizite deiktische Funktion der Bewegungen erkennbar ist. Die Zeichen für nachbereitende Schritte (,,; Mondada 2019) werden für jene Phase ab Zeile 05 genutzt, in welcher der Cursor noch im rechten schwarzen Streifen in Form eines Pfeils sichtbar ist, aber nicht mehr aktiv bewegt/gelenkt wird. Der im Gesprächsinventar vermerkte zweite Screenshot (► eContent_RE_h) ist einem Zeitpunkt entnommen, in dem der Cursor sichtbar ist. Er ist für das Nachvollziehen der Cursorbewegung in zeitlicher Abstimmung mit dem Verbaltranskript nicht weiter relevant, erlaubt aber ein Verständnis der formalen Gestaltung des Cursors (Pfeil), was für einzelne folgende Analyseschritte ebenfalls von Relevanz sein kann.

4.2.3 Produktionsprozess: Auszug eines modellierten Schreibprozesses

Beispiel 3 zeigt einen modellierten Schreibprozess (► eContent_RE_e). Das Transkript umfasst bis zu vier Zeilen (siehe Zeile 06).

Bsp. 3: Produktionsprozess, aus: AaMoL-Modellvideo 1, BS, 09:06-10:07



v12a (09:06)

```

01 BS      *(0.6)+(1.6)+und NU:N:--+ (--)  

bs          +Cursor klickt  

           in Zeile+geht rechts in grauen  

           Rand+,,,,,,,,--->  

v      *v12a----->  

02 BS      äm: (--) schreib ich also den zweiten  

          ABSatz, (1.8)  

03 BS      in dem es um die ECKdaten der d °hh äh  

          der DREI studien GEHT,  

04 BS      diese eckdaten werden nun verGLICHen-  

05 BS      und einander gegenübergestellt.  

06 BS      (1.1)+*% (1.9)      % (1.3) %      (0.6) %  

bs      ---->+  

bs          %BEtrachtet% (1.3) %°[E]{e}man%  

v      ----->*  

07 BS      %beTRACH%Tet man nu:n- (--)  

bs      %      (-)%°nun°die°drei---->  

08 BS      die%drei:: (0.4)%{1.4} STUdien; (---)  

bs      -->%      (---)      %°STudien°[T]{t}  

          hinsichtlich->  

09 BS      hinsich%tlich% (--) ihrer ECK (--) %  

          DATen- (0.5)%{0.7}  

bs      ---->% (--) %°ihrer° Eck (-) daten%  

          (1.1)      %,°so°ergeben-->  

10 BS      so er%GEben (0.1)%{0.6} sich auch DAbei;  

          (0.6)%{0.9}% (2.4)      %{0.6}%  

bs      ---->%      (--)      % °sich°auch°dba[ab]abei  

          %{0.9}%°Untersc (-) hiede%{0.6}%  

11 BS      %UN%terschiede in den d%rei:%{0.7}%{0.6}  

          %STÄ%kten; (0.6)%{0.8}%*(0.3) °h  

bs      %(-)%      °in°den°drei      % (-)%°Städ%{0.6}  

          %ten% (1.2)      % °° %  

v          *v12b---->

```



v12b (09:52)

- 12 BS ich habe mich für diesen EINstieg inem
absatz (-)entschieden==
- 13 BS =oder mir diesen EINstieg in den absatz
überLE:GT, °hh
- 14 BS um (-) die leserin den leser auf den
absatz EINzustimmen;
- 15 BS und auch: (-) etwas auf den INhalt
VORzubereiten. (1.6) *
- v ----->*

Mehrere Verfahren sind bereits aus Beispiel 2 bekannt, so die Darstellung der Cursorbewegungen und die Umsetzung der v-Zeile. Unterbrochen wird die v-Notation hier allerdings für den Zeitraum des Schreibprozesses (Zeile 06 bis 11), da das Hintergrunddokument in dieser Phase einer permanenten Veränderung unterworfen ist. Erst in Zeile 11, als der Prozess mit dem Satzende stoppt und die Schreiberin in weiterer Folge die Funktion des Satzes erläutert, bleibt der Hintergrund v12b bis Zeile 15 unverändert, bevor die Produktion eines weiteren Satzes beginnt (im Ausschnitt nicht mehr ersichtlich). Ein vergleichbares Unterbrechen der Notation auf der v-Zeile bietet sich beispielsweise auch bei Scrollbewegungen an.

Für die Umsetzung des Schreibprozesses müssen einzelne Sonderzeichen und Konventionen ergänzt werden:⁹ Im gezeigten Transkript erfolgt der Abgleich zwischen Schreibprozess und verbaler Kommentierung nach Pausen und weitgehend flüssig-zusammenhängenden Tippphasen. Die Pausennotation entspricht den Prinzipien für die Zeile des Verbaltranskripts. Im Schreibprozess gesetzte Leerräume werden mit ° dargestellt. Löschungen werden in eckige Klammern gesetzt und strikt in der zeitlichen Reihenfolge ihres Auftretens

9 Vgl. zu einem – wenn auch stark am Endprodukt orientierten – Nachvollzug von Textentstehungsprozessen inkl. Textrevisionen die in der Schreibprozessforschung etablierte S-Notation (Severinson Eklundh & Kollberg 1996).

angegeben (z.B. [ab] in Zeile 10: a wird vor b gelöscht). Bei Autokorrekturschritten werden diese ebenfalls genau zum Zeitpunkt ihres Auftretens notiert. Dies ist z.B. in Zeile 06 der Fall, wenn durch das Setzen des Leerraums nach der zeitlichen Pause das zunächst großgeschriebene E automatisch durch ein kleingeschriebenes e ersetzt wird. Die Löschung steht wie manuell vorgenommene Tilgungen in eckigen Klammern, also [E], der eingefügte Kleinbuchstabe in geschwungenen Klammern, also {e}. Geschwungene Klammern können ebenso für manuell gesteuerte Einfügungen verwendet werden.

5 Fazit

Das Gesprächsinventar sowie die drei Transkriptbeispiele verdeutlichen, dass bereits bestehende Konventionen und Verfahren zur gesprächsanalytischen Aufbereitung und Transkription multimodaler Daten (Deppermann 2008, Mondada 2019, Selting et al. 2019) für den Untersuchungsgegenstand der screencast-basierten schreibdidaktischen Modellvideos zielführend angewandt werden können. Der Transkriptionsaufwand steigt dabei mit dem Vorkommen von dynamischen Elementen/Prozessen, v. a. Schreibprozessen, stark an. Für das AaMoL-Modellvideo 1, welches in seiner Gestaltung als typisch für die hier thematisierte Datensammlung bezeichnet werden kann, fällt dieser Aspekt allerdings hochgerechnet auf eine Transkription des gesamten Videos in überschaubarem Ausmaß ins Gewicht: So erstrecken sich Schreibprozesse nur auf eine Dauer von knapp 2,5 Minuten des insgesamt ca. 19,5 Minuten langen Videos. Für andere Erklärvideoformate (z.B. mit höheren Anteilen an Animationen) muss das hier vorgestellte Vorgehen aber im Sinne einer Kosten-Nutzen-Rechnung am Untersuchungsgegenstand und an den Forschungszielen geprüft werden. In Kapitel 3 wurde diesbezüglich versucht, einen Überblick über ausgewählte aktuelle Forschungsprojekte und ihre forschungsmethodischen Zugänge zu Erklärvideos als Anhaltspunkte für eigene forschungsmethodische Entscheidungen zu geben.

Zu bedenken gilt es hinsichtlich der Wahl des Transkriptionsverfahrens allerdings auch, dass allein ein vergleichsweise aufwändiges Verfahren wie das hier gezeigte der präzisen Repräsentation multimodaler Koordination gerecht werden kann (vgl. dazu auch die Erläuterungen zum ersten Beispiel in Abschnitt 4.2.1). Zugleich gilt es hier aber auch auf Grenzen des Verfahrens hinzuweisen, da beispielweise bei Atemgeräuschen aufgrund unterschiedlicher Hörbarkeit die Granularität des Abgleichs dennoch Einschränkungen unterliegt. Auch gilt es in diesem Zusammenhang im Sinne einer reflektierten Forschungspraxis anzumerken, dass man als Transkribent*in in diesen sehr feinkörnigen Bereichen vermutlich verstärkt zum Darstellen von Simultanität

tendiert, z.B. zwischen dem Abschließen einer Cursorbewegung und dem Einsetzen eines Atemgeräuschs. Auch gilt es, das mögliche Vorhandensein von Medieneffekten (z.B. Zeitverzögerungen zwischen Manipulation via Tastatur und Erscheinen am Bildschirm durch Überlastung des Rechners) als Einflussfaktor zu berücksichtigen (vgl. z.B. die Diskussion zu Medieneffekten bei der Arbeit mit Telekonferenzen bei Bliesener 2014).

Die Datentranskription sollte aber nicht nur einen festen Platz in der Umsetzung von Forschungsprojekten, sondern auch in deren Darstellung in Publikationen haben. Dass dabei nämlich die Datenaufbereitung und Transkription inkl. der Auswahl der Standbilder vermutlich oftmals nach wie vor jene Teilaspekte sind, die angesichts begrenzter Zeichenzahlen ausgeblendet werden, bleibt kritisch zu sehen. Mit dem Rückgriff auf bestehende Konventionen und Verfahren kann in diesem Zusammenhang mit vergleichsweise geringer Zeichenzahl dennoch Transparenz gewährleistet werden.

Bezüglich der Ausführungen zu Multimodalität und Interaktionalität in Abschnitt 3 möchte ich abschließend hervorheben, dass anzunehmen ist, dass Kommunikationssituationen, in denen man das Gegenüber hören sowie durch digital mediatisierte Handlungen erleben, nicht aber in seiner Leiblichkeit unmittelbar wahrnehmen kann, zunehmen (werden). Man denke z.B. an die aktuell (April 2021) sehr häufig stattfindenden Videokonferenzen, in denen der Bildschirm geteilt, aber die Webkameras aller Beteiligten möglicherweise abgeschaltet sind. Es bleibt zu beobachten, wie sie in der Forschung aufgegriffen und zur Diskussion von Begriffskonzepten beitragen werden (vgl. dazu u. a. die aktuellen Forschungsbestrebungen zu multimodaler Interaktion am IDS Mannheim, online; ebenso die Diskussion zu Multimodalität als eigenständiger Disziplin bei Wildfeuer et al. 2020).

Literaturverzeichnis

- Birkner, Karin, Auer, Peter, Bauer, Angelika & Kotthoff, Helga. 2020. *Einführung in die Konversationsanalyse*. Berlin & Boston: De Gruyter.
- Bliesener, Thomas. 2014. Transkription synchroner multimedialer rechnerbasierter Telekonferenzen. In Christine Moritz (Hrsg.), *Transkription von Video- und Filmdaten in der Qualitativen Sozialforschung. Multidisziplinäre Annäherungen an einen komplexen Datentypus*, 283–310. Wiesbaden: Springer VS.
- Brünner, Gisela & Pick, Ina. 2020. Bewertungen sprachlichen Handelns und good practice in der Angewandten Gesprächsforschung. *Zeitschrift für Angewandte Linguistik* 72, 63–98.
- Deppermann, Arnulf. 2018. Sprache in der multimodalen Interaktion. In Arnulf Deppermann & Silke Reineke (Hrsg.), *Sprache im kommunikativen, interaktiven und*

- kulturellen Kontext* (Germanistische Sprachwissenschaft um 2020 3), 51–85. Berlin u. a.: De Gruyter.
- Deppermann, Arnulf. 2008. *Gespräche analysieren. Eine Einführung*. 4. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dorgerloh, Stephan & Wolf, Karsten D. (Hrsg.). 2020. *Lehren und Lernen mit Tutorials und Erklärvideos*. Weinheim & Basel: Beltz.
- Ebel, Alexandra. 2021. Einfluss sprachlicher und sprecherischer Merkmale auf die Verständlichkeit von Lernvideos auf YouTube. In Katharina Staubach (Hrsg.), *Multimodale Kommunikation in den Hypermedien und Deutschunterricht. Theoretische, empirische und unterrichtspraktische Zugänge*, 250–275. Baltmannsweiler: Schneider Verlag Hohengehren.
- ELAN (Version 6.0). 2020. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>.
- Findeisen, Stefanie, Horn, Sebastian & Seifried, Jürgen. 2019. Lernen durch Videos – Empirische Befunde zur Gestaltung von Erklärvideos. *MedienPädagogik*, 16–36.
- Graham, Steve & Harris, Karen. 2005. *Writing better: Effective strategies for teaching students with learning difficulties*. Baltimore: Paul Brookes.
- IDS Mannheim. Projekt „Multimodale Interaktion“. <https://www1.ids-mannheim.de/prag/interaktion/multimodal.html> (zuletzt abgerufen am 23.04.2021).
- Imo, Wolfgang & Lanwer, Jens Philipp. 2019. *Interaktionale Linguistik. Eine Einführung*. Berlin: J. B. Metzler.
- Klein, Josef. 2009. Erklären-Was, Erklären-Wie, Erklären-Warum. Typologie und Komplexität zentraler Akte der Welterschließung. In Rüdiger Vogt (Hrsg.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven*, 25–36. Tübingen: Stauffenburg Verlag.
- Lobin, Henning. 2013. Visualität und Multimodalität in wissenschaftlichen Präsentationen. *Zeitschrift für Germanistische Linguistik* 41(1), 65–80.
- Meer, Dorothea & Pick, Ina. 2019. *Einführung in die Angewandte Linguistik. Gespräche, Texte, Medienformate analysieren*. Berlin: J. B. Metzler.
- Mondada, Lorenza. 2019. *Conventions for multimodal transcription*. https://344cc026-c96f-49aa-b4bc-071b454d3061.filesusr.com/ugd/ba0dbb_986ddd4993a04a57acf20ea06e2b9a34.pdf (zuletzt abgerufen am 10.02.2021).
- Moritz, Christine. 2011. *Die Feldpartitur. Multikodale Transkription von Videodaten in der Qualitativen Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Philipp, Maik. 2015. *Lesestrategien. Bedeutung, Formen und Vermittlung*. Weinheim & Basel: Beltz Juventa.
- Rat für Kulturelle Bildung (Hrsg.). 2019. *Jugend/YouTube/Kulturelle Bildung. Horizont 2019*. Essen: Rat für Kulturelle Bildung e.V. https://www.rat-kulturelle-bildung.de/fileadmin/user_upload/pdf/Studie_YouTube_Webversion_final.pdf (zuletzt abgerufen am 10.02.2021).
- Reitbrecht, Sandra. 2021. Schreibdidaktische Modellvideos: Lehrer*innen schreiben für ihre Schüler*innen. In Katharina Staubach (Hrsg.), *Multimodale Kommunikation in den*

- Hypermedien und Deutschunterricht. Theoretische, empirische und unterrichtspraktische Zugänge*, 303–326. Baltmannsweiler: Schneider Verlag Hohengehren.
- Reitbrecht, Sandra, Dawidowicz, Marta, Flotzinger-Aigner, Silvia & Schatzl, Brigitte. 2019. Schreibkompetenzen in der Fremdsprache Deutsch durch Modelllernen fördern. *German as a foreign language* 2, 85–109. <http://www.gfl-journal.de/2-2019/reitbrecht.pdf> (zuletzt abgerufen am 10.02.2021).
- Schmidt, Axel. 2018. Interaktion und Kommunikation. In Dagmar Hoffmann & Rainer Winter (Hrsg.), *Mediensoziologie. Handbuch für Wissenschaft und Studium*, 15–38. Baden-Baden: Nomos.
- Schneider, Jan Georg & Stöckl, Hartmut. 2011. Medientheorien und Multimodalität: Zur Einführung (inkl. Transkripte). In Jan Georg Schneider & Hartmut Stöckl (Hrsg.), *Medientheorien und Multimodalität. Ein TV-Werbespot – Sieben methodische Beschreibungsansätze*, 10–44. Köln: Herbert von Halem Verlag.
- Schöne, Justine. 2018. Potenziale von Erklärvideos als Bildungsmedium. Eine linguistische Untersuchung. <https://sway.office.com/QXZkttzaK0iZOhr?ref=Link> (zuletzt abgerufen am 22.04.2021).
- Schöne, Justine & Wedler, Katharina. 2021. Erklärvideos in der LehrerInnenausbildung: Textanalytische und produktionsorientierte Zugänge. In Katharina Staubach (Hrsg.), *Multimodale Kommunikation in den Hypermedien und Deutschunterricht. Theoretische, empirische und unterrichtspraktische Zugänge*, 276–302. Baltmannsweiler: Schneider Verlag Hohengehren.
- Selting, Margret et al. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353–402.
- Severinson Eklundh, Kerstin & Kollberg, Py. 1996. Computer tools for tracing the writing process. From keystroke records to S-notation. In Gert Rijlaarsdam, Huub van den Bergh, & Michel Couzijn (Hrsg.), *Current research in writing. Theories, models and methodology*, 526–541. Amsterdam: Amsterdam University Press.
- Stertkamp, Wolf & Schüler, Lisa. 2014. Transkription multimodaler Gefüge: Herausforderungen bei der Untersuchung interaktiver Prozesse am PC. In Christine Moritz (Hrsg.), *Transkription von Video- und Filmdaten in der Qualitativen Sozialforschung. Multidisziplinäre Annäherungen an einen komplexen Datentypus*, 311–358. Wiesbaden: Springer VS.
- Stöckl, Hartmut. 2016. Multimodalität – Semiotische und textlinguistische Grundlagen. In Nina-Maria Klug & Hartmut Stöckl (Hrsg.), *Handbuch Sprache im multimodalen Kontext*, 3–35. Berlin: De Gruyter.
- Stukenbrock, Anja. 2009. Herausforderungen der multimodalen Transkription: Methodische und theoretische Überlegungen aus der wissenschaftlichen Praxis. In Karin Birkner & Anja Stukenbrock (Hrsg.), *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung*, 144–169. Mannheim: Verlag für Gesprächsforschung.
- Sturm, Afra & Weder, Mirjam. 2016. *Schreibkompetenz, Schreibmotivation, Schreibförderung. Grundlagen und Modelle zum Schreiben als soziale Praxis*. Stuttgart: Klett Kallmeyer.

- Wildfeuer, Janina, Pflaeging, Jana, Bateman, John A., Seizon, Ognyan & Tseng, Chiao-I. 2020. Multimodality: Disciplinary Thoughts and the Challenge of Diversity – Introduction. In Janina Wildfeuer, Jana Pflaeging, John Bateman, Ognyan Seizov & Chiao-I Tseng (Hrsg.), *Multimodality. Disciplinary Thoughts and the Challenge of Diversity*, 3–38. Berlin/Boston: De Gruyter.
- Wolf, Karsten D. 2015. Video-Tutorials und Erklärvideos als Gegenstand, Methode und Ziel der Medien- und Filmbildung. In Anja Hartung, Thomas Ballhausen, Christine Trültzsch-Wijnen, Alessandro Barberi & Katharina Kaiser-Müller (Hrsg.), *Filmbildung im Wandel*, 121–131. Wien: new academia press.
- Wolf, Karsten D. & Kratzer, Verena. 2015. Erklärstrukturen in selbsterstellten Erklärvideos von Kindern. In Kai-Uwe Hugger, Angela Tillmann, Stefan Iske, Johannes Fromme, Petra Grell & Theo Hug (Hrsg.), *Jahrbuch Medienpädagogik 12: Kinder und Kindheit in der digitalen Kultur*, 29–44. Wiesbaden: Springer VS.