

ULRICH SCHMITZ

Kann und soll Linguistik angesichts multimodaler Kommunikation eine eigene Disziplin sein?

Resümee

Sprache tritt immer in multimodalen Kontexten auf. Doch durch zunehmende gesellschaftliche Komplexität und dank technischer Erfindungen wird das Zusammenspiel von Sprache und anderen Modi, insbesondere technisch erzeugten stehenden und bewegten Bildern, immer enger und wirksamer. Eine Linguistik, die Sprache, Äußerungen und Texte nicht nur (wie früher meistens) an und für sich selbst, sondern in ihrem multimodalen Gebrauchskontext untersuchen möchte, steht vor erheblichen methodischen Herausforderungen. Angesichts der Fülle und Vielfalt multimodaler Daten sollte Sprachwissenschaft sich auf die Frage konzentrieren, ob und in welcher Weise Sprache eine kommunikative Arbeitsteilung mit anderen Modalitäten eingeht und wie sie dadurch in ihrer Funktion und Form beeinflusst wird. In diesem Aufsatz wird an zwei unterschiedlichen Beispielen (einem Musikvideo und einem wissenschaftlichen Erklärvideo) untersucht, in welcher Weise Sprache bei dynamischen Sehflächen andere Formen annimmt und andere Funktionen erfüllt als sonst üblich. Beide Beispiele sind (wie vermutlich alle dynamischen Sehflächen) mehr oder weniger intensiv ästhetisch komponiert. Dabei zeigen beide Beispiele, dass der Wortlaut sprachlicher Texte (Lexik und Grammatik) in Videos sehr stark auf ihre multimodale Umgebung abgestimmt ist.

Language always occurs in multimodal contexts. However, due to increasing social complexity and thanks to technical inventions, the interplay between language and other modes, especially technically generated still and moving images, is becoming ever closer and more effective. A linguistics that wants to study language, utterances and texts not only (as was mostly the case in the past) in and of themselves, but in their multimodal context of use, faces considerable methodological challenges. In view of the abundance and diversity of multimodal data, linguistics should focus on the question of whether and how language enters into a communicative division of labor with other modalities and how this influences its function and form. This paper uses two different examples (a music video and a science explainer video) to examine the way in which language in dynamic visual surfaces takes on different forms and functions than usual elsewhere. Both examples (like presumably all dynamic visual surfaces) are more or less intensely aesthetically composed.

Both examples show that the wording of linguistic texts (lexis and grammar) in videos is very closely matched to their multimodal environment.

1 Multimodalität – unüberschaubar?

Dieser Jubiläumsband 100 der *Osnabrücker Beiträge zur Sprachtheorie* überfliegt mit einer geistigen Drohne einige weiße Flecken auf der Landkarte der Sprachwissenschaft. Die verschiedenen Beiträge gehen manchem von dem nach, was in der Sprachwissenschaft bisher weder gewusst noch hinreichend gefragt wurde. In diesem Aufsatz wird untersucht, vor welchen Herausforderungen und Gefahren eine Linguistik steht, die Sprache¹, Äußerungen und Texte nicht nur an und für sich selbst, sondern in ihrem multimodalen Kontext untersuchen möchte.

Menschliche Kommunikation geschieht immer multimodal.² Auch Sprache tritt nie monomodal (also als Sprache allein) auf, sondern immer in einem sowohl materiellen als auch multimodalen, situativen, gesellschaftlichen Zusammenhang. „Will man Sprache so beschreiben, wie sie gebraucht wird, lässt sie sich nicht aus ihren multimodalen Kontexten lösen.“ (Klug 2016, 168) In der Sprachwissenschaft hat man aber genau das bis vor Kurzem fast immer gemacht. Und zwar aus guten Gründen. Man kann nicht alles gleichzeitig und mit gleicher Präzision untersuchen. So lässt sich die Geschichte der Sprachwissenschaft verstehen als ein bis heute unentschiedener Kampf zwischen zwei legitimen widersprüchlichen Interessen. Einerseits erwartet man von Wissenschaft methodisch gesicherte und nachprüfbar Resultate. Das gelingt umso besser, je homogener, also je enger umzirkelt der Untersuchungsbereich ist. Andererseits sollen diese Resultate brauchbar, also der kulturellen Wirklichkeit möglichst angemessen sein. Diese ist aber umfassend und unendlich feinkörnig; alles hängt mit allem zusammen.

Den radikalsten Versuch, mit diesem Dilemma fertig zu werden, unternahm Ferdinand de Saussure. In der Lesart seiner Schüler ist die Sprache „ein Ganzes in sich und ein Prinzip der Klassifikation“ (Saussure 1931, 11) und kann folglich unabhängig von anderen Faktoren, insbesondere unabhängig vom „*Leben der Zeichen im Rahmen des sozialen Lebens*“ untersucht werden (ebd., 19). Der authentische Saussure quälte sich zeitlebens damit ab, dass es keinerlei festen „Ausgangspunkt oder Orientierungspunkt in der Sprache [langue] gibt“ (Saussure 2003, 101), und zwar genau weil sprachliche Zeichen „*einzig in einem Prozeß sozialer Zirkulation existieren*“ (Saussure 1997, 209).

Dieser Prozess allerdings ist eben sozial, kulturell, gesellschaftlich, von unüberschaubar vielen und oft unvorhersehbaren Faktoren beeinflusst. Wir können die vielfältigen Bemühungen Saussures, „*dem Spiel der Zeichen*“ (Saussure 2003, 96) auf die Schliche zu kommen, und zwar innerhalb einer eigenständigen Wissenschaft der Sprache, hier

1 „man muß sich von Anfang an auf das Gebiet der Sprache [langue] begeben“ (Saussure 1931, 11).

2 Kress (2010, 1) bezeichnet „‘multimodality’ as the normal state of human communication“.

nicht näher verfolgen.³ Es sei aber darauf hingewiesen, dass Saussures wissenschaftliche Motivation aus einer schonungslosen Abrechnung mit der bisherigen Geschichte der Sprachwissenschaft entspringt, die ihren Gegenstand nicht klar definieren könne (Saussure 1931, 1-7).

Mit Ludwik Fleck könnte man diese Geschichte als eine Abfolge verschiedener Denkkollektive verstehen, die jeweils gemeinsame Anschauungen und Denkstile teilten. „Sie sind überholt worden, nicht weil sie falsch waren, sondern weil sich das Denken entwickelt.“ (Fleck 1980, 85).⁴ Saussure (1931, 1-7) nannte in dieser Reihenfolge Grammatik, Philologie, vergleichende Grammatik mit jeweils lehrreichen Irrtümern und schließlich Junggrammatik. Auf seinen Namen berufen sich dann Strukturalisten; in Abgrenzung dazu beherrscht eine Weile die generative Grammatik zahlreiche linguistische Forschungen; dann folgen (sehr vergrößernd in Erinnerung gerufen) Textlinguistik, zahlreiche interdisziplinäre Bindestrich-Linguistiken (Soziolinguistik, Psycholinguistik ...), mit der so genannten pragmatischen Wende Sprachgebrauchsforschung, in der Folge Gesprochene-Sprache-Forschung, angesichts neuer Techniken Korpuslinguistik, Medienlinguistik, Bildlinguistik⁵ und schließlich aktuell Multimodalitätsforschung. Noch einmal mit Saussure (1931, 8): Der Gesichtspunkt schafft den wissenschaftlichen Gegenstand. Oder mit Fleck (1980, 85): „Die Abhängigkeit der wissenschaftlichen Tatsache vom Denkstil ist evident.“

Das ist aber unbefriedigend. Flecks Denkstiltheorie ist die historisch-soziologische Konsequenz aus Kants Erkenntnistheorie, derzufolge „die Vernunft nur das einseht, was sie selbst nach ihrem Entwurfe hervorbringt“ (Kant 1956, 23 = B XIII). Kant zufolge muss Naturforschung „die Natur nötigen [...], auf ihre Fragen zu antworten“ so wie ein Richter, „der die Zeugen nötigt, auf die Fragen zu antworten, die er ihnen vorlegt“ (ebd.). Für die Sprachwissenschaft dürfte das gleichermaßen gelten. Es kommt also darauf an, gute Fragen zu stellen. Müssen sie vom gerade herrschenden Denkstil abhängen? Kann man sich von solchem „Denkzwang“, wie Fleck (1980, 85) auch sagt, befreien?

Wir versuchen es. Der gegenwärtige Denkstil (oder zumindest einer von mehreren) in der Sprachwissenschaft legt großen Wert auf die materielle Grundlage und multimodale Einbettung von Sprache. Es gibt auch gute Gründe für den aktuellen Hype um Multimodalität.⁶ Erstens wird hier ein Feld betreten, das in der früheren Geschichte der Sprach- und Kommunikationswissenschaft entweder keine oder aber nur eine randständige Rolle

3 Vgl. Saussure 1931, 1997, 2003; Jäger 1975, 2010.

4 Drei Jahrzehnte vorher hatte Max Weber den gleichen Gedanken, sah die Ursache von Änderungen aber weniger im Denken als vielmehr in der Kultur: „Aber irgendwann wechselt die Farbe: die Bedeutung der unreflektiert verwerteten Gesichtspunkte wird unsicher, der Weg verliert sich in die Dämmerung. Das Licht der großen Kulturprobleme ist weiter gezogen. Dann rüstet sich auch die Wissenschaft, ihren Standort und ihren Begriffsapparat zu wechseln und aus der Höhe des Gedankens auf den Strom des Geschehens zu blicken.“ (Weber 1904, 87) Heute wird dieser Strom des Geschehens vorrangig von technischen Entwicklungen angetrieben.

5 Saussure (1931, 5) würde diesen Widerspruch in sich einen „irrigen“ Begriff nennen.

6 An wissenschaftlicher Literatur seien exemplarisch nur genannt: Bateman 2008, Deppermann/Linke (Hrsg., 2010), Forceville/Urios-Aparisi (Hrsg., 2009), Fricke 2012, Jewitt (Hrsg., 2014), Klug/Stöckl

spielte. Zweitens werden (seit Buchdruck, Zeitung und Telefon) immer mehr Modi der Kommunikation technisiert und schließlich digitalisiert, so dass heute ein erheblicher und immer noch wachsender Anteil an Kommunikation über bi- oder multimodale Medien verläuft. Drittens lassen eben diese technischen Mittel neue Aufzeichnungs-, Speicher-, Dokumentations-, Recherche- und Analysewerkzeuge zu, mit denen empirische Daten ganz anders und reichhaltiger zugänglich sind als früher.

Und viertens könnte man sich vorstellen, dass durch nochmalige Erweiterung des Blickwinkels in einer multimodal orientierten Sprachwissenschaft sämtliche bisherigen sprachbezogenen Erkenntnisse in ein neues Licht gerückt werden. Wenn man von der Geschichte der Sprachwissenschaft fortschreitende Erkenntnis erwartet, dann wünscht man sich nicht nur eine größer werdende Ansammlung einzelner Beobachtungen, Deutungen und Theorien, sondern auch eine parallel dazu wachsende Architektur möglichst aufeinander aufbauender und insgesamt immer systematischerer Erkenntnisse. In dieser Sicht müsste sich die Logik der bisherigen Geschichte der Sprachwissenschaft (Geschichte – System – Gebrauch) immer mehr in Richtung Multimodalität bewegen. Denn einerseits führen einige der früher fruchtbaren methodischen Abstraktionen heute in erkenntnisarme Sackgassen. Und andererseits verfügen wir seit einigen Jahrzehnten über digitale Techniken, die viele dieser Abstraktionen überflüssig machen. Um nur ein Beispiel zu nennen: Die bahnbrechende Untersuchung über Sprecherwechsel in Gesprächen (Sacks/Schegloff/Jefferson 1974) konnte sich auf einfache Wort-für-Wort-Transkriptionen mit wenigen Extra-Konventionen (ebd., 731-734) beschränken. Heute gibt es Software für erheblich komplexere Analysen gesprochener Sprache (z. B. EXMARaLDA, Praat) sowie erste Konventionen für multimodale Transkription (Mondada 2014), und man hat die kompletten multimodalen Rohdaten jederzeit reproduzierbar zur Hand.

Allerdings wird Sprache in der aktuellen Multimodalitätsforschung oft (wenn überhaupt) eher nur am Rande berücksichtigt⁷ – und das, obwohl bei multimodalen Kommunikaten Sprache fast immer mit im Spiel ist. Das dürfte auch an methodischen Schwierigkeiten liegen. Denn wer Sprache komplett in ihren stets multimodalen Gebrauchskontexten betrachten will, sieht sich schier unüberschaubaren Herausforderungen gegenüber. Keine Einzelperson beherrscht das gesamte Feld aller beteiligten Disziplinen. Bei interdisziplinärer Forschung in Arbeitsgruppen gibt es unterschiedliche Termini, Methoden und Denkweisen, so dass wechselseitige Missverständnisse naheliegen. Außerdem: Was tun, um nicht in der Fülle der heterogenen Daten zu versinken? Und schließlich: Wie in allen Wissenschaften ist es außerordentlich schwer, wenn nicht unmöglich, multifaktorielles Geschehen in Gänze zu überblicken. Meist dient es der Erkenntnis, von mehreren oder vielen solcher Faktoren zu abstrahieren. Sie werden dann zu für das Untersuchungsziel un-

(Hrsg., 2016), Krause/Schmitz (Hrsg., 2021), Kress 2010, Müller et al. (Hrsg., 2013/2014), Norris (Hrsg., 2016), Norris 2020, Wildfeuer et al. (Hrsg., 2019), Wildfeuer/Bateman/Hiippala 2020.

⁷ Siehe aber Deppermann (2020), Klug/Stöckl (Hrsg., 2016), Stöckl 2019, Wildfeuer/Bateman 2018, Wildfeuer/Bateman/Hiippala 2020.

erheblichen Randbedingungen erklärt.⁸ Die Präzision der so erzielten Erkenntnisse ist aber erkauft mit Scheuklappen gegenüber Zusammenhängen: wissenschaftliche Intelligenz im sorgfältig abgezielten Bereich, Dummheit gegenüber dem großen Ganzen.

Wie kann man aus diesem Dilemma herauskommen? Wir gehen nicht als allgemeine Multimodalitätsforscher, sondern als Linguistinnen und Linguisten an die Sache, grenzen die Frage ein und interessieren uns ausschließlich dafür, in welcher Weise andere Modalitäten Sprache (Wortlaut, Grammatik) in multimodaler Umgebung beeinflussen bzw. in welcher Weise andere Modalitäten mit ihr eine kommunikative Arbeitsteilung eingehen.⁹ In dieser Hinsicht sind Beziehungen zwischen (meist schriftlicher) Sprache und statischem Bild durchaus untersucht worden¹⁰, weniger aber zwischen (meist mündlicher) Sprache und bewegten Bildern¹¹. Deshalb verzichten wir fürs Erste auf Perfektion und versuchen es experimentell, abduktiv und risikobereit. Wir orientieren uns an Lakatos, der meinte, unter manchen Umständen müsse eine allzu strenge Methodologie den Ideen eine Atempause geben:

But some of the greatest scientific research programmes progressed on inconsistent foundations. Indeed in such cases the best scientists' rule is frequently: *'Allez en avant et la foi vous viendra'*. This anti-Popperian methodology secured a breathing space both for the infinitesimal calculus and for naive set theory when they were bedevilled by logical paradoxes. (Lakatos 1971, 113)

Legen wir also einfach los.

2 Sprache bei dynamischen Sehflächen

„Sehflächen sind Flächen, auf denen Texte und Bilder in geplantem Layout gemeinsame Bedeutungseinheiten bilden.“ (Schmitz 2011, 25) Dynamische Sehflächen sind Sehflächen, auf (und akustisch bei) denen sich Zeichen im zeitlichen Ablauf bewegen. Die technische Grundlage dafür sind heute meistens Bildschirme (Fernsehen, Computer, Smartphones); es können aber auch Kino-Leinwände oder Displays an Geräten sein. Für die Linguistik stellt sich hier die Frage, in welcher Weise Sprache in ihre technisierte-multimodale Umgebung eingebettet ist. Etwas dramatischer formuliert: Nimmt Sprache auf dynamischen Sehflächen andere Formen an und erfüllt sie andere Funktionen als sonst üblich?

Diese Frage soll im Folgenden an zwei unterschiedlichen Beispiel-Analysen verfolgt werden. Beide Beispiele, ein Musikvideo und ein wissenschaftliches Erklärvideo, reprä-

⁸ „Die organische Abgeschlossenheit jeder Denkgemeinde geht parallel einer stilgemäßen Beschränkung der zugelassenen Probleme: es müssen immer viele Probleme unbeachtet oder als unwichtig oder sinnlos abgewiesen werden.“ (Fleck 1980, 137)

⁹ Eine solche „Denkstilveränderung – d. h. Veränderung der Bereitschaft für gerichtetes Wahrnehmen – gibt neue Entdeckungsmöglichkeiten und schafft neue Tatsachen.“ (Fleck 1980, 144)

¹⁰ Z. B. Diekmannshenke/Klemm/Stöckl (Hrsg., 2011, Schmitz 2010; 2017, Stöckl 2004; 2015.

¹¹ Z. B. Holly 2004; 2007; 2016 (mit weiterführender Literatur).

sentieren zwei sehr unterschiedliche (von zahllosen) multimodalen Textsorten.¹² Beide teilen aber drei gemeinsame Merkmale. (1) Sie sind dramaturgisch sorgfältig geplant und ästhetisch intensiv durchgearbeitet. (2) Alle Texte wurden schriftsprachlich detailliert vorbereitet und werden akustisch vorgetragen. (3) Die so gesprochenen Texte sind ausschließlich monologisch – im ersten Fall von zwei verschiedenen Handlungsbeteiligten (Sohn und Mutter), im zweiten Fall von einem einzigen handlungsunbeteiligten Sprecher.



Abb. 1: Standbild bei 0:39 Minuten aus „Einfach klein sein“¹³

Als erstes Beispiel dient das offizielle Musikvideo der Gruppe „Deine Freunde“ zu ihrem Kinderlied „Einfach klein sein“ aus dem Jahre 2012. Bis Juni 2022 wurde es über 1,2 Millionen Mal aufgerufen. Es dauert 4:51 Minuten. Es geht um einen etwa neunjährigen Jungen, der von seinen übermäßig leistungs- und erfolgsorientierten Eltern zu höchsten Leistungen auf allen nur erdenklichen Gebieten angetrieben wird, während er doch nur gerne einfach Kind sein möchte.

Im zweiten Video aus dem Jahr 2000 wird ein physikalisches Schlüsselexperiment erklärt, nämlich das Doppelspaltexperiment, das in der Quantenphysik häufig zur Demonstration des Welle-Teilchen-Dualismus dient. Es steht seit 2014 im Internet und wurde bis Juni 2022 über 1,8 Millionen Mal aufgerufen. Das gesamte Video dauert 26:36 Minuten; wir betrachten hier aber nur die ersten 4:51 Minuten.

¹² Zur Verwendung dieses Begriffs vgl. Schmitz 2016. Der hier spezifischere Ausdruck *Videosorten* ließe zu sehr Nichtsprachliches assoziieren.

¹³ Deine Freunde – Einfach klein sein, <www.youtube.com/watch?v=FJDTtGhsoTk>.



Abb. 2: Standbild bei 1:44 Minuten vom Doppelspaltexperiment¹⁴

Eine exakte deskriptive Analyse jedes der beiden Videos würde zu einem Drehbuch führen, nach dessen Anweisung ein mehr oder weniger identisches Video gedreht werden könnte. Das kann aber kein sinnvolles Ziel sein. Auch der wissenschaftliche Geograph will die Erdoberfläche nicht noch einmal erschaffen; und selbst der praktisch orientierte Kartograph verwendet nur ausgewählte Daten für ein abstraktes Diagramm, das bestimmten Zwecken dient. Entsprechend sollten Wissenschaftler, die multimodale Beziehungen untersuchen, eine klare Frage formulieren, mit denen sie wie eine Wünschelrute ihre Objekte zu bestimmten, abstrahierenden Antworten zwingen. Entweder hat man die Frage bereits, bevor man die Untersuchungsobjekte anschaut. Das führte zu einer eher induktiven, eher positivistischen Methode. Hier läuft man Gefahr, dass Frage und Material nicht in einer Weise zusammenpassen, die interessante Ergebnisse zulässt. Oder aber man hat zuerst die Objekte und entwickelt aus deren Betrachtung abduktiv die Frage, ein eher hermeneutischer Weg. Hier läuft man Gefahr, möglicherweise interessante Eigenschaften der Untersuchungsobjekte zu übersehen. Am meisten Erfolg verspricht eine Kombination beider Wege. Dabei wiederum entsteht das Problem eines unendlichen Zirkels: Je genauer ich das Material kenne, je mehr Daten ich daraus gewinne, desto umfangreicher und schließlich endloser wird mein Forschungsprojekt. (Und wenn man darüber nicht süchtig geworden ist, kann man schließlich froh sein, dass die Zeit und/oder die Geldmittel begrenzt sind.)

¹⁴ Quantenmechanik – Doppelspalt, Verschränkung und Nichtlokalität, <www.youtube.com/watch?v=7BV0Fs4eM0I>. Eine Produktion der Gerald Kargl Ges.m.b.H. (Filmproduktion Wien) 2000. Im Auftrag des Ministeriums für Bildung, Wissenschaft und Kultur, Wien und FWU Institut für Film und Bild in Wissenschaft und Unterricht, Grünwald. Buch & Regie: Gerald Kargl.

Wir haben in unserem Fall, der ja nur der Demonstration und Diskussion dient, zunächst die gesprochenen und geschriebenen Texte beider Videos transkribiert. Das ist nicht schwer, weil es sich ausschließlich um offensichtlich vorbereitete und monologisch vorge-tragene Schriftsprache handelt. Deswegen sind die üblichen Transkriptionsregeln aus der Gesprächsanalyse hier nicht relevant. Wichtig ist aber, die zeitliche Abfolge der einzelnen Syntagmen (hier kleinster Sinneinheiten) zu notieren, damit ihr Zusammenspiel mit Bild und Ton erkennbar wird.

Sodann müssen die optischen Anteile in ihrer zeitlichen Abfolge sprachlich so dokumentiert werden, dass man auch ohne Betrachtung des Films wissen kann, worum es im Einzelnen geht. Bei der Segmentierung einzelner Einheiten kann man sich an den (ggf. ebenfalls zu notierenden) verschiedenen Kameraeinstellungen und -schnitten orientieren.

Schließlich gilt es, in gleicher Weise auch Musik, Töne und Geräusche sprachlich und chronologisch zu dokumentieren.¹⁵ Hier sei nur der Beginn der Partituren für beide Beispiele gezeigt:

0:00	blendet langsam auf Doppelbett mit Aufschrift	Manchmal will ich einfach klein sein	Choral im Hintergrund;
0:04	„DEINE FREUNDE / einfach klein sein“	Und ich weiß nicht, wie das geht.	Text wird gesungen
0:07	Kamera fährt nach links über schlafenden Jungen	Manchmal kann der Tag gemein sein,	
0:10		Ohne dass man was erlebt.	
0:12	Schlafzimmertür, daneben großer Bildschirm mit tickender Uhrzeit: „06 59 58“	Manchmal will ich einfach klein sein	
0:15	auf dem Schirm erscheint Wecker (7 Uhr), Mutter joggt ins Schlafzimmer des Jungen	Und ich weiß nicht, wie das geht.	
0:18	liegender Jungenkopf von hinten mit Folie auf dem Hals, auf der ein Startknopf-Icon steht	Manchmal kann der Tag gemein sein...	
0:20	Mutter drückt diesen Startknopf		
0:21	auf der Folie leuchtet ein Akku-Diagramm auf		
0:24	Gesicht des aufwachenden Jungen	Meine Eltern möchten, dass ich mal Karriere mach’.	Choral im Hintergrund endet
0:27	Mutter jogt aus dem Schlafzimmer raus	Deswegen soll ich tausend Sachen machen jeden Tag.	Text jetzt als Sprechgesang mit rhythmischer Musik untermalt
0:30	Junge setzt sich im Bett auf	Ich soll die Dinge lernen, um die es geht im Leben.	
0:33		Deswegen muss ich Schach spielen und Chinesisch reden.	
0:36	Mutter turnt im Vorraum (Diele), Junge tritt ein	Wozu ich Lust hab? Tut mir leid, mir fällt nix ein,	
0:37	Vater tritt telefonierend aus einer anderen Tür		
0:39	und geht in eine dritte; Mutter turnt immer noch	Denn ich muss gerade wieder irgendwo der Beste sein.	
0:43	Vater kommt wieder raus und geht in die 4. Tür	Und wie es aussieht, wird sich das nie wieder ändern.	

15 Je nach Forschungsinteresse kann das unterschiedlich detailliert erfolgen. Natürlich können auch Stimmführung und weitere Parameter notiert werden, um das Zusammenspiel verschiedener Modi, die Verteilung relevanter Informationen darauf und die Folgen für die sprachliche Formulierung zu untersuchen.

0:45		Andere haben Spaß,	
0:47	Junge schaut auf seinen Terminkalender	ich hab 'n Terminkalender.	
0:48	Eltern am Frühstückstisch, Vater telefoniert	Sie scheuchen mich, erst dorthin, dann hierhin.	
0:51	Schwenk weiter am Frühstückstisch zum Jungen, der über aufgetürmten Schulbüchern liest	Und alles angeblich nur, weil ich so talentiert bin.	
0:55	Blick aus dessen Perspektive auf beide Eltern	Und das muss ich euch jetzt jeden Tag beweisen	
0:56	Junge zieht unter einem Schulbuch ein Blatt her- vor	Und aufzählen, wie irgendwelche Nebenflüsse heißen.	
1:00	Junge zieht einen Bleistift hervor	Mein alter Spielplatz, du warst so far- benfroh.	
1:03	Junge zeichnet	Keine Zeit mehr für dich, der Ernst des Lebens wartet schon.	
1:06	Mutter kommt hinzu, schlägt Buch zu, blickt streng	Dabei will ich doch nur 'ne Runde auf dir hocken	
1:08		Und später dann nach Hause kom- men mit schmutzigen Klamotten.	
1:12	Szenenwechsel: Junge sitzt hinten im Auto	Manchmal will ich einfach klein sein,	Text wird gesun- gen (wie oben)
1:16		Und ich weiß nicht, wie das geht.	
1:18	Junge joggt neben joggender Mutter	Manchmal kann der Tag gemein sein,	
1:22	joggender Junge allein im Kopfprofil	Ohne dass man was erlebt.	
1:24	Junge sitzt wieder hinten im Auto	Manchmal will ich einfach klein sein,	
1:28		Und ich weiß nicht, wie das geht.	
1:31		Manchmal kann der Tag gemein sein...	
1:33	das Akku-Diagramm leert sich von grün auf rot		

Abb. 3: Partitur „Einfach klein sein“

0:00	Berglandschaft von weit oben		Vogelgezwitscher
0:05	aus der Ferne rauscht ein Flugzeug her- an		nahender Flugzeug- lärm
0:06	Schnitt, Düsenjäger rast jetzt von links durchs Bild		lauter Flugzeuglärm
0:08		Warum kann ein Flugzeug fliegen?	
0:10	Bergsteiger neben Gipfelkreuz betrach- tet die Landschaft, Düsenjäger (DJ) rast von links vorbei,		(weiter Vogelzwit- schern und Lärm)
0:13	DJ fliegt erneut von links vor gesamter Landschaft	Das Zusammenspiel bestimmter physika- lischer Umstände macht es möglich,	
0:18	DJ in Nahaufnahme wirbelt von unten nach oben	viele Tonnen schweren Stahls leicht wie ei- ne Feder	starkes Flugzeugmo- torgeräusch
0:21	DJ von oben betrachtet wirbelt über Berglandschaft	hoch in die Lüfte zu heben.	
0:26	(weiter so)	Ursache und Wirkung sind im Voraus be- rechnet.	verwandelt sich in Wirbelton
0:29		Denn sie sind ein bestimmendes Prinzip, das unsere Welt beherrscht.	Motorrauschen
0:33	Luftbild der weiten Landschaft mit kreuzendem DJ		
0:34	(weiter so)	Sie sind ein Grundsatz der klassischen Physik, mit dessen	Flugzeug braust her- an

0:38	Bergsteiger neben Gipfelkreuz vor rasendem DJ	Hilfe wir uns die Natur nach logischen Kriterien erklären können.	
0:43	Bild blendet aus, schwarzer Grund	Doch – diese Regel gilt nicht immer.	Flugzeuggeräusche hören auf
0:48	von rechts wandert ein weißer Lichtspalt herein, dann eine flackernde weiße Lichtquelle	Ein simples physikalisches Experiment mit Licht eröffnete uns vor etwa einhundert Jahren,	Projektorgeräusche beginnen
0:54	weiter flackerndes Licht	dass diese Logik nicht überall anzuwenden ist.	
0:57	(weiter so)	In der Welt der kleinsten Dimensionen herrschen Gesetze, die uns seltsam und absurd erscheinen,	
1:03	(weiter so)	und die dennoch zu einer neuen, grundsätzlichen Theorie über die Natur geführt haben:	Projektorgeräusche hören auf
1:05	eine Schrifttafel blendet auf: „Gibt es den Mond wenn niemand hinsieht?“	der Quantentheorie.	
1:10	zusätzlich: „Die geheimnisvolle Welt der Quanten“		
1:11			magische Ein-Ton-Musik beginnt
1:13	Schrift verschwindet, Mondfoto blendet auf		
1:18	Mondfoto wird überblendet durch Lichtquelle		Ein-Ton-Musik endet
1:19	Lichtquelle flackert, Kamera fährt allmählich zurück	Dieses Experiment mit Licht, das alle wesentlichen Elemente der neuen Physik aufzeigt,	Projektorgeräusche setzen ein
1:25	technisches Gerät wird allmählich sichtbar	ist der sogenannte Doppelspaltversuch.	zusätzlich beginnt ein Brummen
1:29	Kamera fährt weiter zurück,	Sein Aufbau ist einfach.	
1:31	so dass die Lichtquelle von der Seite erscheint	Eine Lichtquelle schießt Photonen auf	nur das Brummen bleibt
1:34	schwarzer Detektor mit auffunkelten Lichtpunkten	einen gegenüberliegenden Detektor.	
1:37	ein anderes technisches Gerät erscheint	Dazwischen wird eine Trennwand geschoben.	leise ertönt auch der Projektor
1:40	Photonenquelle und Trennwand werden sichtbar	Diese besitzt zwei Spalte, die unabhängig voneinander geöffnet und geschlossen werden können.	
1:47	Kamera bewegt sich weiter über die Geräte		nur das Brummen bleibt
1:50		Welche Geheimnisse offenbart uns nun dieser simple technische Aufbau?	magische Ein-Ton-Musik beginnt

Abb. 4: Partitur Doppelspaltexperiment

In beiden Fällen sind alle drei Modi (Film, Text, Geräusche) offenkundig klar aufeinander abgestimmt. Beispielsweise erscheint die weckende Mutter in der zwölften Sekunde des Klein-Sein-Videos auf die Zehntelsekunde exakt um 07.00.00 Uhr, während der Liedtext gerade von einer Traumvorstellung („Manchmal will ich einfach klein sein“) auf die realen Hindernisse übergeht („Und ich weiß nicht, wie das geht.“). Und – nur als zweites von zahlreichen Beispielen – im Doppelspalt-Video verstummen erstmals nach 65 Sekunden alle Geräusche, kurz bevor der Sprecher bei 1:10 die Worte „der Quantentheorie“ spricht, während punktgenau gleichzeitig die Schrift „Die geheimnisvolle Welt der Quanten“ aufscheint und unmittelbar danach ein geheimnisvoller länger anhaltender Klang ertönt.

Damit ist die Einleitung beendet und das Thema der nun folgenden Erklärung erstmals genannt.

An dieser Stelle kann nur als These vermutet werden, dass sämtliche dynamischen Sehflächen mehr oder weniger intensiv ästhetisch komponiert wurden. Denn es erfordert viel Bedacht und Können, derartige multimodale Kommunikate zu bauen, und zwar allein schon deshalb, weil (nicht ganz einfache) technische Geräte benutzt werden müssen. Das ist bei mündlichem Sprechen nicht der Fall und bei Schreiben – handschriftlich oder per Smartphone oder Computer – sehr viel weniger. Unsere beiden Beispiele weisen aufgrund ihrer multimodalen Textsorte (Kinderlied und Erklärvideo) und der professionellen Studioproduktion ein besonders starkes Maß an bewusst geplantem ästhetischem Design auf. Aber auch profanere Erzeugnisse etwa von Journalisten (z. B. Nachrichten-Videos) sowie trivialere Beispiele von Laien (z. B. Urlaubsvideos) sind ästhetisch komponiert, und zwar sehr viel aufwendiger als monomodale Texte.¹⁶

Was bedeutet das aber für das Kerngebiet der Sprachwissenschaft? Sind sprachliche Texte in multimodalen Kommunikaten (hier auf dynamischen Sehflächen) anders geformt als außerhalb? Auch hier lautet die Antwort-These: Ja. Denn wenn das Gesamtkommunikat mehr oder weniger bewusst ästhetisch gestaltet wird, muss das notwendigerweise auch Konsequenzen für den Text haben. Im Falle von Liedern war das schon immer so: Bei gesungenen Liedern müssen Silbenzahl und Metrum des Textes zu Melodie und Rhythmus der Musik passen. Oft reimen sich auch die einzelnen Verse; so, wenngleich meist unrein, oft auch in unserem „Klein-Sein“-Beispiel. Der kommunikative Effekt schon des bloß gesungenen Liedes ohne Video wäre erloschen, wenn man den gleichen Sinn in lässig gesprochener Prosa mitteilte. Der hinzukomponierte Film verstärkt seine Wirkung aber ungemein, ohne dass der durchschnittliche Hörer die Kunstfertigkeit dahinter klar bemerkte.

Auch im zweiten Beispiel, dem Erklärvideo, müssen Text und Bild einander angepasst sein. Auch hier wäre der Text für sich allein weitgehend verständlich, aber wenig effektiv. Er wäre sogar kontraproduktiv. Denn es wird zunächst überhaupt nicht klar, was Flugzeuge mit Quantentheorie zu tun haben könnten. An einigen Stellen ist der Text sogar fragwürdig: Nicht „Ursache und Wirkung“ sind „ein bestimmendes Prinzip“ und auch kein „Grundsatz der klassischen Physik“, sondern allenfalls deren hier nicht näher bestimmtes Verhältnis. Auch dass ein Flugzeug „leicht wie eine Feder hoch in die Lüfte“ gehoben werden kann, würde in einem Schulaufsatz als doppelter Ausdrucksfehler gelten. Der zugehörige Film scheint aber genau diesen Wortlaut zu illustrieren. Und die zahlreichen Bewegungen im Film samt interessanter Ton-Untermalung wecken das Interesse des Zuschauers. Kurzum: Hier sind es didaktische Rücksichten, die dazu führen, dass der Text sich dem Bild anpassen muss. Das Bild soll den Text unterstützen.

16 Van Leeuwen (2020, 21) bemerkt: „Like other multimodal researchers [...] I have noted the increasing importance of aesthetics, even in non-artistic texts and other semiotic artefacts“.

In beiden Fällen macht das Bild den Text unterhaltsam. Beim Lied wird durch das Zusammenspiel von Sprache und anderen Modalitäten eine alltäglich familiäre Lebenswelt so eindrucksvoll lebendig inszeniert, dass deren skurril-übertriebenen Momente nicht unglaublich, sondern komisch wirken, woraus kritisches Potential entsteht. Die ästhetische Komposition lässt sogar vergessen, dass es gar nicht der Junge ist, der das Lied parallel zu seinen Handlungen singt, sondern vielmehr eine Männerstimme seine Phantasie überzeugend ausdrückt. Beim Erklärvideo bliebe der Text ohne seine audiovisuelle Einbettung arg langweilig. Ein durchschnittlicher Schüler im Physikunterricht dürfte schnell abschalten.

Beide Texte würden wohl auch ohne ihren multimodalen Kontext (Bild und Ton) weitgehend funktionieren, das heißt: bei entsprechender Motivation verständlich sein – umgekehrt aber nicht (wie man am Computer selbst ausprobieren kann). Das „Klein-Sein“-Video ohne Lied zeigt verschiedene Szenen aus dem Erleben eines Kindes; deren Zusammenhang und Bedeutung bliebe aber unklar, und die vom Gesamtvideo tatsächlich intendierte Bedeutung käme nicht zur Geltung. Das Erklärvideo ohne Text wäre vollends unverständlich.

Dieser Befund illustriert die zentrale Bedeutung von Sprache in vielen multimodalen Kontexten. Je nach multimodaler Textsorte, Thema und ästhetischem Arrangement wird der Anteil von Sprache an der Gesamtbedeutung unterschiedlich groß sein. Man vergleiche etwa schriftliche Theatertexte mit der multimodalen Theateraufführung, PowerPoint-Folien mit dem tatsächlichen Vortrag oder auch textärmere multimodale Textsorten wie Diagramme, Displays an Geräten oder Stummfilme.

Schauen wir nun den Wortlaut der beiden Texte genauer an. Das Lied umfasst 615 Wörter, das Erklärvideo (bis 4:51) 450 gesprochene und 12 geschriebene Wörter. Beim Lied hat man also pro Sekunde durchschnittlich 2,1 Wörter wahrzunehmen, beim Erklärvideo 1,6. Das ist auch sinnvoll, weil das Lied trotz seines ernsten Inhalts mit seinen grotesken Bildern eher spielerisch daherkommt, während der langsamere Sprecher des anderen Textes sich um einen spürbar erklärenden Gestus bemüht.

Das Lied formuliert die Gedanken und Gefühle des Jungen, unterstützt durch die im Bild gezeigte Lebenswelt, durch die er getrieben wird. Der Text ist klar aufgebaut. Im ersten, längsten Teil (bis 3:11) umfassen drei identische Siebenzeiler je vier Vierzeiler, gefolgt von vier identischen Refrainstrophen mit je vier kurzen, apodiktischen Zeilen. Darauf folgt ein textloses Flötenvorspiel des Jungen. In der anschließenden Pause (ab 3:46) hält die ehrgeizige Mutter einen Prosamonolog vor anderen Gästen. Schließlich (ab 4:01) wird der bekannte Siebenzeiler zwei Mal wiederholt und klingt allmählich aus. Alles in allem umfasst es 61 Sätze (mit durchschnittlich je 10 Wörtern). Davon sind 55 (90 %) recht einfach gebaut (nämlich nur Hauptsatz 24, parataktischer Hauptsatz + Hauptsatz 6, hypotaktischer Haupt- und Nebensatz 14, kombiniert Hauptsatz + Hauptsatz + Nebensatz 11). Die restlichen sechs folgen einem etwas komplizierteren Aufbau (z. B. HS + NS + NS). Schnittfolge und Architektur der Bilder sind dem Textverlauf exakt angepasst.

Auch der Text des Erklärvideos folgt einem durchdachten, allerdings wesentlich einfacheren Aufbau. Zuerst wird an klassische Physik erinnert (bis 0:43), um davon die Quantentheorie abzusetzen (bis 1:11). Im Rest werden der Aufbau und ein erstes erstaunliches Ergebnis des Experiments erläutert (hier nur betrachtet bis 4:51). Das gelingt in zusammen 36 Sätzen (mit durchschnittlich je fast 13 Wörtern). Davon sind 30 (83 %) einfach gebaut (nämlich nur Hauptsatz 13 oder hypotaktischer Haupt- und Nebensatz 17), wobei sich diese beiden Konstruktionen häufig unmittelbar abwechseln. Satzlänge und Komplexität nehmen im Laufe des Textes tendenziell zu; die sechs etwas komplizierteren Konstruktionen erscheinen erst in der zweiten Hälfte. Das dürfte didaktisch so gewollt sein, um das Publikum allmählich in seinen Bann zu ziehen und zunehmend lernen zu lassen. Die parallel gezeigten Bilder machen den gesprochenen Text im ersten Teil anschaulich, unterstützt durch eindeutig aus dem Bild erschließbare Geräusche, und werfen im zweiten Teil auch Rätsel auf, hier unterstützt durch seltsame Geräusche.

Wie sieht es mit dem Wortschatz aus? Im Liedtext kommen 38 verschiedene (zusammen 72) Substantive vor. Mit vielen davon lässt sich die hier erlebte Welt schon recht gut charakterisieren (*Cembalo, Chancen, Eltern, Goldkind, Hause, Informationen, Karriere, Kind, Mama, Spiel, Stock, Talentscout, Terminkalender, Ziele*). Unter den 53 verschiedenen (zusammen 139) Verbformen stechen dagegen nur wenige besonders hervor (*aufzählen, beweisen, erlebt, füttern, hocken, kümmern, lernen, scheuchen, trainier, üben*). Auch unter den 15 verschiedenen (zusammen 31) Adjektiven fallen nur wenige auf (*gemein, farbenfroh, goldene, schmutzigen, stocklosen, talentiert*), die aber zusammen mit den genannten Substantiven und Verben eine einigermaßen gute Vorstellung der kleinen Erzählung geben können. Fast alle anderen benutzten 373 Wörter (Adverbien, Präpositionen etc.) sind im Alltag häufig gebrauchte Allerweltswörter.

Im Erklärvideo dagegen kommen 68 verschiedene (zusammen 95) Substantive vor. Erwartungsgemäß sind darunter allerlei Fachwörter (z. B. *Detektor, Doppelspaltversuch, Lichtteilchen, Photonen, Quantentheorie*), aber auch andere im Alltag seltene Wörter (z. B. *Abschlussbedingungen, Apparatur, Impuls, Trennwand, Windrichtung*) sowie Wörter, die man hier kaum erwartet hätte (z. B. *Feder, Flugzeug, Geheimnisse, Lüfte, Mond, Pistolenkugeln, Stahls*). Schon daran lässt sich – wie ja auch an den gewählten Bildern – erkennen, dass es sich um einen didaktisch orientierten Text über ein eher schwieriges physikalisches Thema handelt. Die 58 verschiedenen (zusammen 81) Verbformen dagegen sind – anders als im Lied – kaum prägnant (z. B. *ändern, erfassen, gibt, herrschen, öffnet, treffen, verteilen*). Auch daraus lässt sich schließen, dass es sich wohl kaum um eine Erzählung handelt. Ähnliches gilt für die 40 verschiedenen (zusammen 47) Adjektive (z. B. *beide, einfache, gleichen, kleinsten, schweren, wesentlichen*), wobei einige wenige auf einen Fachtext deuten (z. B. *gleichbleibender, logischen, physikalischen*). Fast alle anderen 239 Wörter sind wie beim Lied häufig gebrauchte Allerweltswörter.

Der Anteil der Substantive im Erklärvideo-Text liegt fast doppelt so hoch wie im Liedtext (20,6 % vs. 11,7 %). Ähnlich verhält es sich bei den Adjektiven (10,2 % vs. 5,0 %). Dafür nutzt das Erklärvideo weniger Verben als das Lied (17,5 % vs. 22,6 %). Auch alle

übrigen Wortklassen (Adverbien, Präpositionen etc.) zusammen kommen im Erklärvideo relativ weniger vor als im Lied (51,7% vs. 60,7%). Auch darin schlägt sich der Unterschied zwischen einem eher fachlichen und einem eher alltäglichen Sprachstil nieder.

Dem entspricht auch die multimodale Aufbereitung. Der von einem ausgebildeten Sprecher ruhig und mit Nachdruck vorgetragene fachdidaktische Text ist eingebettet in anschauliche Bilder und nimmt teils unmittelbaren Bezug darauf. In der Eingangssequenz, die Aufmerksamkeit für das eher spröde Thema wecken soll, zeigen sie, wie ein Betrachter die Bewegungen eines Flugzeugs im hellen und offenen Raum verfolgt; passende natürliche und technische Geräusche untermalen die Szenerie. Danach (ab 0:43) gehen Text und Bild über zu einem physikalischen Experiment, dessen Aufbau in einem völlig abgedunkelten Innenraum und dessen überraschende Resultate im Bild gezeigt und im Text beschrieben werden; die optische und akustische Kulisse suggeriert Geheimnisvolles.

Der gesungene Liedtext hingegen formuliert Gedanken und Gefühle des Jungen, während das Bild entweder ihn in seiner Umgebung oder diese Lebenswelt aus seinem Blickwinkel zeigt. Auch hier nehmen Text und Bild häufig unmittelbaren Bezug aufeinander. So blickt der Junge bei „ich hab ‘n Terminkalender“ (0:47) für eine halbe Sekunde auf seinen vollen Terminkalender; die Dirigierbewegungen des Flötenlehrers passen exakt auf den stark betonten Rhythmus des Textes („richtige Dinge kümmern“, „vernünftige Ziele“ etc., 1:57-2:03); und während „dass ich überall ‘ne goldene Medaille krieg“ gesungen wird, sieht man die dominante Mutter in der Küche vor einem aufgehängten Schild mit der Aufschrift „Herrin der Lage“ einen Goldpokal blitzblank putzen (2:11).

Bei derart ausgefeilter Korrespondenz zwischen Text und Film in beiden Fällen darf man vermuten, dass nicht zuerst ein komplett fertiger Text formuliert und anschließend bebildert wurde. Vielmehr dürfte nach einer ursprünglichen Idee eine Art Drehbuch entstanden sein, so dass ab einer bestimmten Stufe in der Produktion sämtliche Modi (mündliche Sprache, Film, ggf. schriftliche Sprache im Film, Musik, Geräusche) gezielt aufeinander abgestimmt wurden. Da von den Texten grammatisch vollständige Sätze und im Falle des Liedes auch ein bestimmter Rhythmus sowie teilweise unreine Reime erwartet werden, liegt in der letzten Produktionsphase die Hauptlast auf dem korrekten Schnitt der verschiedenen Filmszenen, damit alles gut übereinstimmt.

3 Linguistik und Multimodalitätsforschung

Beide Beispiele zeigen, dass der Wortlaut sprachlicher Texte (Lexik und Grammatik) in Videos sehr stark auf ihre multimodale Umgebung abgestimmt ist. Wir sehen und hören ein Geschehen, das nach einem bestimmten Regieplan zeitlich so abläuft, dass sämtliche Modi synchron zueinander passen. Ähnliches geschieht auch außerhalb von Videos und Filmen im realen Alltag: Auch dort passen in der Regel (bei nichtpathologischer Kommunikation) alle Einzelheiten der Kommunikation zueinander. Doch dort wirken sämtliche Akteurinnen und Akteure gemeinsam am Ablauf der Geschehnisse, so dass die jeweils nächsten Ereignisse nicht unbedingt vorhersehbar sind. In Filmen und Vi-

deos, die ja auf identische Reproduktion hin angelegt sind, ist alles im Vorhinein von den Produzentinnen und Produzenten erdacht und festgelegt, und zwar aufgrund der technischen Bedingungen erheblich strikter als etwa im Theater. Die notwendigerweise vorausgedachte Planung erzwingt ein bewusstes ästhetisches Design, das auch auf das Zusammenspiel von Text, Bild und Ton achtet, auch wenn das bei unterschiedlichen multimodalen Textsorten unterschiedlich streng bzw. locker gehandhabt werden mag.¹⁷

Hier konnte nur an Video-Beispielen knapp vorgeführt werden, wie eine linguistisch orientierte Untersuchung von Sprache in technisierten dynamischen multimodalen Umgebungen angegangen werden kann. Dabei sollte es vor allem darum gehen, ob und in welcher Weise Sprache eine kommunikative Arbeitsteilung mit anderen Modalitäten eingeht und wie sie dadurch in ihrer Funktion und Form beeinflusst wird.

Je weniger Beispiele herangezogen werden, desto stärker hermeneutisch geprägt wird der Zugang sein. Naheliegende Verallgemeinerungen werden umso überprüfbarer ausfallen, je mehr und je unterschiedlichere Exemplare analysiert werden. Dabei dürfte und sollte nach und nach eine immer ausgefeiltere Methodik zur Untersuchung dynamischer Sehflächen entwickelt werden. Der Aufwand ist beträchtlich, sowohl bei der Erstellung sinnvoller Korpora als auch bei der Auswahl und Zuordnung von Metadaten sowie bei der quantitativen und qualitativen Auswertung. Das dürfte mit ein Grund dafür sein, dass die Sprachwissenschaft sich bisher erst vergleichsweise zögerlich mit dem „Wirbel der Zeichen“ (Saussure 1997, 355) eben auch innerhalb technisierter multimodaler Umgebungen, hier speziell innerhalb dynamischer Sehflächen, beschäftigt hat. In der Tat sollte das Verhältnis von Aufwand und Ertrag gut bedacht sein, denn es gibt innerhalb und außerhalb der Linguistik viele andere bedenkenswerte Probleme.

So gesehen ist es letztlich nicht von großer Bedeutung, ob Linguistik angesichts multimodaler Kommunikation eine eigene Disziplin sein kann und soll. Ihr über Jahrhunderte erarbeitetes theoretisches Niveau und ihre methodische und thematische Vielfalt sollten aber nicht in einem modischen Denkstil untergehen, der allzusehr auf viele Spielarten von Multimodalität fixiert ist, ohne die meist zentrale Rolle von Sprache dabei hinreichend zu berücksichtigen.

Literatur

- Bateman, John A. (2008) *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke/New York: Palgrave Macmillan.
- Boeckmann, Barbara/Dorothee Meer et al. (2019) Multimodale Produktbewertungen in Videos von Influencerinnen auf YouTube: Zur parainteraktiven Konstruktion von Warenwelten. In: *Zeitschrift für Angewandte Linguistik* 70, 139-171.

¹⁷ Vgl. z. B. für Influencer-Videos Boeckmann/Meer et al. 2019.

- Deine Freunde – Einfach klein sein, <www.youtube.com/watch?v=FJDTtGhsoTk>, eingesehen am 9.6.2022.
- Deppermann, Arnulf (2020) Sprache in der multimodalen Interaktion. In: Deppermann, Arnulf/Silke Reineke (Hrsg.) *Sprache im kommunikativen, interaktiven und kulturellen Kontext*. Berlin/Boston: De Gruyter, 51-85.
- Deppermann, Arnulf/Angelika Linke (Hrsg., 2010) *Sprache intermedial. Stimme und Schrift, Bild und Ton*. Jahrbuch des Instituts für Deutsche Sprache 2009. Berlin/New York: De Gruyter.
- Diekmannshenke, Hajo/Michael Klemm/Hartmut Stöckl (Hrsg., 2011) *Bildlinguistik. Theorien - Methoden - Fallbeispiele*. Berlin: Erich Schmidt.
- Fleck, Ludwik (1980) *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv* [1935]. Frankfurt/M.: Suhrkamp.
- Forceville, Charles J./Eduardo Urios-Aparisi (Hrsg., 2009) *Multimodal Metaphor*. Berlin/New York: De Gruyter Mouton.
- Fricke, Ellen (2012) *Grammatik multimodal. Wie Wörter und Gesten zusammenwirken*. Berlin/Boston: De Gruyter.
- Holly, Werner (2004) Sprechsprache und bewegte Bilder: Audiovisualität. In: Holly, Werner/Almut Hoppe/Ulrich Schmitz (Hrsg.) *Sprache und Bild I und II*, 122-134 [= Mitteilungen des Deutschen Germanistenverbandes 51.1 und 51.2].
- Holly, Werner (2007) Audiovisuelle Hermeneutik. Am Beispiel des TV-Spots der Kampagne „Du bist Deutschland“. In: Hermanns, Fritz/Werner Holly (Hrsg.) *Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens*. Tübingen: Niemeyer, 387-426.
- Holly, Werner (2016) Nachrichtenfilme als multimodale Texte. In: Klug/Stöckl (Hrsg.), 392-409.
- Jäger, Ludwig (1975) *Zu einer historischen Rekonstruktion der authentischen Sprach-Idee F. de Saussures*. Dissertation Universität Düsseldorf.
- Jäger, Ludwig (2010) *Ferdinand de Saussure zur Einführung*. Hamburg: Junius.
- Jewitt, Carey (Hrsg., 2014) *The Routledge handbook of multimodal analysis* [2011]. 2. Aufl. London etc.: Routledge.
- Kant, Immanuel (1956) *Kritik der reinen Vernunft* [1781]. Werke, Bd. II (Hrsg. Wilhelm Weischedel). Wiesbaden: Insel.

- Klug, Nina-Maria (2016) *Multimodale Text- und Diskurssemantik*. In: Klug/Stöckl (Hrsg.), 165-189.
- Klug, Nina-Maria/Hartmut Stöckl (Hrsg., 2016) *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter.
- Krause, Arne/Ulrich Schmitz (Hrsg., 2021) *Linguistik und Multimodalität / Linguistics and Multimodality*. Osnabrücker Beiträge zur Sprachtheorie (OBST), Bd. 99. Duisburg: Universitätsverlag Rhein-Ruhr.
- Kress, Gunther (2010) *Multimodality. A social semiotic approach to contemporary communication*. Milton Park: Routledge.
- Lakatos, Imre (1971) History of Science and Its Rational Reconstructions. In: Buck, Roger C./Robert S. Cohen (Hrsg.) *PSA: Proceedings of the Biennial meeting of the Philosophy of Science Association 1970. In Memory of Rudolf Carnap*. Dordrecht: D. Reidel, 91-136 [= Boston Studies in the Philosophy of Science, 8].
- Mondada, Lorenza (2014) *Conventions for multimodal transcription*. Basel: Universität Basel, <https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf> (eingesehen am 6.6.2021).
- Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva Ladewig/David McNeill/Sedinha Teßendorf/Jana Bressemer (Hrsg., 2013/2014) *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*. 2 Bde. Berlin/Boston: De Gruyter Mouton.[= HSK 38]
- Norris, Sigrid (Hrsg., 2016) *Multimodality*. London, New York: Routledge.
- Norris, Sigrid (2020) *Multimodal Theory and Methodology. For the Analysis of (inter)action and identity*. New York: Routledge.
- Quantenmechanik – Doppelspalt, Verschränkung und Nichtlokalität, <www.youtube.com/watch?v=7BV0Fs4eM0I>, eingesehen am 09.06.2022.
- Sacks, Harvey/Emanuel A. Schegloff/Gail Jefferson (1974) A Simplest Systematics for the Organization of Turn-Taking for Conversation. In: *Language* 50 (4), 696-735.
- de Saussure, Ferdinand (1931) *Grundfragen der allgemeinen Sprachwissenschaft* [frz. 1916]. Berlin: De Gruyter.
- de Saussure, Ferdinand (1997) *Linguistik und Semiologie. Notizen aus dem Nachlaß. Texte, Briefe und Dokumente*. Frankfurt/M: Suhrkamp.
- de Saussure, Ferdinand (2003) *Wissenschaft der Sprache. Neue Texte aus dem Nachlaß*. Frankfurt/M.: Suhrkamp.

- Schmitz, Ulrich (2010) Schrift an Bild im World Wide Web. Articulirte Pixel und die schweifende Unbestimmtheit des Vorstellens. In: Deppermann, Arnulf/Angelika Linke (Hrsg.), 383-418.
- Schmitz, Ulrich (2011) Sehflächenforschung. Eine Einführung. In: Diekmannshenke, Hajo/Michael Klemm/Hartmut Stöckl (Hrsg.), 23-42.
- Schmitz, Ulrich (2016) Multimodale Texttypologie. In: Klug, Nina-Maria/Hartmut Stöckl (Hrsg.), 327-347.
- Schmitz, Ulrich (2017) Randgrammatik und Design. In: *IDS Sprachreport* 33 (3), 8-17.
- Stöckl, Hartmut (2004) *Die Sprache im Bild – Das Bild in der Sprache. Zur Verknüpfung von Sprache und Bild im massenmedialen Text. Konzepte, Theorien, Analysemethoden*. Berlin/New York: De Gruyter.
- Stöckl, Hartmut (2015) Bewegung auf der Titelseite – Ausdifferenzierung und Hybridisierung durch Sprache-Bild-Texte. In: Hauser, Stefan/Martin Luginbühl (Hrsg.) *Hybridisierung und Ausdifferenzierung. Kontrastive Perspektiven linguistischer Medienanalyse*. Bern etc.: Peter Lang, 235-259.
- Stöckl, Hartmut (2019) *Linguistic Multimodality – Multimodal Linguistics: A State-of-the-Art Sketch*. In: Wildfeuer et al. (Hrsg.), 41-68.
- van Leeuwen, Theo (2021) Multimodality as a multi-disciplinary practice. In: Krause, Arne/Ulrich Schmitz (Hrsg.), 15-31.
- Weber, Max (1904) Die „Objektivität“ sozialwissenschaftlicher und sozialpolitischer Erkenntnis. In: *Archiv für Sozialwissenschaft und Sozialpolitik* 19 (1), 22-87.
- Wildfeuer, Janina/John A. Bateman (2018) Theoretische und methodologische Perspektiven des Multimodalitätskonzepts aus linguistischer Sicht. In: *IMAGE – Zeitschrift für interdisziplinäre Bildwissenschaft* 28, 5-46<www.gib.uni-tuebingen.de/own/journal/pdf/IMAGE%2028a.pdf>.
- Wildfeuer, Janina/John A. Bateman/Tuomo Hiippala (2020) *Multimodalität. Grundlagen, Forschung und Analyse – Eine problemorientierte Einführung* [engl. 2017]. Berlin/Boston: De Gruyter.
- Wildfeuer, Janina/Jana Pflaeging/John A. Bateman/Ognyan Seizovn/Chiao-I Tseng (Hrsg., 2019) *Multimodality – Disciplinary Thoughts and the Challenge of Diversity*. Berlin: de Gruyter.