Einleitung

Möchte man die gegenwärtige Kommunikation in einem Wort fassen, erfolgt das am besten über den Begriff *multimodal*. Dieses aus dem Englischen stammende Wort, das vorrangig mit der Entwicklung der sog. Neuen Medien[1] assoziiert wird, ist in letzter Zeit zu einem gängigen Schlagwort geworden, mit dem die Vielschichtigkeit sowie die steigende Komplexität des menschlichen Lebens charakterisiert werden. Multimodalität und die damit verbundenen multimodalen Texte sind jedoch

> selbstredend keine Erfindung des ausgehenden 20. und beginnenden 21. Jahrhunderts. Sie haben eine lange Tradition, die von mittelalterlichen Dokumenten über barocke Emblematik bis zur später aufkommenden, bebilderten Presse reicht (Raith 2014: 26).

Multimodalität bezieht sich auf alle kommunikativen Handlungen und/oder Situationen, die durch mehrere semiotische, sich wechselseitig kontextualisierende Zeichen geprägt sind und nicht nur auf jene, die über die Neuen Medien generiert werden. Die aktuelle mediale Kommunikation, verstärkt durch die Verbreitung neuester elektronischer Geräte, hat diese multimodale Kommunikation jedoch noch weiter intensiviert:

> Auch wenn wir anspruchsvolle Kombinationen aus unterschiedlichen Ausdrucksformen durchaus bereits seit längerem kennen, scheinen sie sich mithilfe neuerer oder neu

---

1 Wegen der Mehrdeutigkeit sowie Unklarheit des Begriffs „Neue Medien" in der Gegenwart erscheint es legitim, an dieser Stelle seine genaue Spezifizierung vorzunehmen. Nach Arnold/Neuberger (2005) zeichnen sich die Neuen Medien vor allem durch ihre Fähigkeit zur technischen Konvergenz und Digitalisierung aus. Zu den Alten Medien zählen daher Zeitung, Hörfunk, Fernsehen und Telefon, während zu den Neuen Medien gehören Computer, Tablets, Smartphones sowie das Internet (vgl. Arnold/Neuberger 2005: 11–17). In jüngster Zeit ist jedoch zusätzlich, insbesondere aufgrund von Levinson (2009), ein neuer Begriff aufgetreten, nämlich „Neue Neue Medien", mit dem der Autor versuchte, soziale Medien zu bezeichnen. Levinson (2009: 1) verdeutlicht: „And what, exactly, do I mean by ‚new new media'? The current roster is listed in the chapter titles of this book: blogging, YouTube, Wikipedia, Digg, MySpace, Facebook, Twitter, Second Life, and podcasts". In der vorliegenden Arbeit werden unter dem Begriff Neue Medien nach Arnold/Neuberger (2005) moderne technische Geräte sowie das Internet verstanden.

> entstehender Medien sowie neuer Produktions- und Zugangstechnologien noch stärker auszubreiten und in alle Bereiche unserer Kommunikation vorzudringen (Wildfeuer u. a. 2020: 23).

Die ursprünglichste Form der Multimodalität ist das Gespräch zwischen zwei Menschen, da es neben der sprachlichen Modalität, auch andere Ausdrucksformen wie Gestik, Mimik, Tonfall oder Intonation umfasst (vgl. Stöckl 2004b: 9; Maćkiewicz 2017: 33; Simon 2018: 59; Wildfeuer u. a. 2020: 11). Nicht ohne Grund definiert Stöckl (2004a: 9) Multimodalität als „a phenomenon which is as old as representation itself and crucial to an understanding of almost all forms of communication […]". Mit der natürlichen Entwicklung der Menschheit veränderten sich auch die Formen der multimodalen Kommunikate. Im Laufe der Zeit wurde ihnen in den verschiedenen Bereichen des zwischenmenschlichen Lebens zunehmend Aufmerksamkeit geschenkt. Bereits in der Antike standen die im Begriff Multimodalität eingeschlossenen Wechselbeziehungen zwischen verschiedenen semiotischen Zeichen im Fokus der philosophischen Betrachtungen:

> Diese Wechselbeziehungen [zwischen verschiedenen Kunstformen: Literatur, Malerei, Musik – KS] wurden immer erkannt – der Ausspruch des antiken Denkers Simonides von Keos, dass Malerei eine stumme Poesie und Poesie eine stumme Malerei sei, wurde über Jahrhunderte hinweg in Überlegungen über Kunst herangezogen (Poprzęcka 1998: 122; Übersetzung – KS).

Im 19. Jahrhundert waren die multimodalen Wechselwirkungen ein Schlüsselelement des künstlerischen Schaffens. Die Grundlagen der sogenannten Korrespondenz der Künste, die auf die Dichter und Philosophen der deutschen Romantik zurückzuführen sind, beruhten auf folgender Annahme:

> Das Universum der Analogien leiteten sie aus der Idee der allgemeinen Angemessenheit von Geist und Materie ab, aus der metaphysischen Vision des Universums, verbunden durch unsichtbare Kräfte der ‚Sympathie' zwischen dem Allgemeinen und dem Besonderen, zwischen Wort und Bild, Ton und Farbe (Poprzęcka 1998: 123; Übersetzung – KS).

Auch in der Literatur lassen sich Spuren des in dem Begriff Multimodalität eingeschlossenen Zusammenwirkens zwischen Bild und Sprache erkennen. Besonders hervorzuheben ist hierbei die Literaturgattung Ekphrasis, die schon im antiken Griechenland praktiziert wurde. Unter diesem Terminus versteht man kurze sprachliche Texte, die die durch ein bildliches Kunstwerk oder einen anderen visuellen Gegenstand hervorgerufenen Emotionen beschreiben (vgl. Radziewicz 2017: 167).

Was die Fremdsprachendidaktik betrifft, ist Multimodalität dort ebenfalls kein neues Phänomen. Als Pionier und Initiator eines solchen (multimodalen) Ansatzes kann

Johann Amos Comenius (1592–1670) angesehen werden, der im Jahr 1658 das erste bebilderte Fremdsprachenwörterbuch – *Orbis Sensualium Pictus* (siehe Anhang, Anlage 1 und 2) – veröffentlichte. Dieses Werk, übersetzt als *Die sichtbare Welt*, konzentrierte sich auf Themen der Natur und der Umwelt und enthielt neben lateinischen Phrasen und ihren deutschen Übersetzungen auch illustrierte Abbildungen. Das Ziel dieser Abbildungen war es, den Lernenden das Fremdsprachenlernen durch Anschaulichkeit zu erleichtern (vgl. Cieśla 1974: 68; Hecke/Surkamp 2010: 19).

Die Welt unterliegt einem ständigen Wandel, und seit einigen Jahrzehnten hat die rasante Entwicklung der Technologie einen bedeutenden Einfluss darauf. Neue Technologien eröffnen neue Möglichkeiten für die Erzeugung, Bearbeitung und Verbreitung von Kommunikation. Die Allgegenwärtigkeit multimodaler Texte im öffentlichen Diskurs durch die weit verbreiteten Neuen Medien hat dazu geführt, dass man begann, diese auch im Prozess des Fremdsprachenlernens und -lehrens zu berücksichtigen. Ihr enormes didaktisches Potenzial, das aus der Verknüpfung sprachlicher (geschriebene/gesprochene Sprache), bildlicher (Bild, Grafik, Abbildung) sowie auditiver (Ton, Musik) Elemente resultiert, bietet neue Möglichkeiten zur Unterstützung des Fremdsprachenlernens und -lehrens. Besonders deutschsprachige Forscher*innen haben in ihren Arbeiten zunehmend die Themen und Vorteile bestimmter Formen multimodaler Texte untersucht (vgl. Abraham 2009, 2013; Leitzke-Ungerer 2009; Schmäling 2014; Wittstruck 2016, 2018; Simon 2018; Herberichs 2018). Die bisherigen Untersuchungen behandeln jedoch das Phänomen der Multimodalität und seinen Einfluss auf den Fremdsprachenlern- und -lehrprozess nur fragmentarisch. Es mangelt an einer umfassenden Untersuchung dieses Themas in diesem Forschungsbereich. Die vorliegende Arbeit setzt sich zum Ziel, diese Lücke zu füllen, indem sie untersucht:

1. wie solche multimodalen Texte von Studierenden aufgenommen werden,
2. ob und wie sie die Entwicklung ihrer sprachlichen und nichtsprachlichen Kompetenzen[2] beeinflussen, und schließlich
3. ob sie die Motivation der Lernenden zum Fremdsprachenlernen erhöhen können.

---

2 „Kompetenzen" werden in der vorliegenden Arbeit nach Weinert (2002: 27 f.) verstanden als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können". Kompetenzen sind also vielfältig, erlern- und veränderbar, situativ bedingt und umfassen kognitive Fähigkeiten und Fertigkeiten, die zur Problemlösung dienen. Sie lassen sich weiter in drei Hauptgruppen unterteilen, nämlich 1) fachliche Kompetenzen, die Kompetenzen physikalischer, fremdsprachlicher sowie musikalischer Art umfassen; 2) fachübergreifende Kompetenzen, die sich auf das Problemlösen sowie Teamfähigkeit beziehen; 3) Handlungskompetenzen, die „neben kognitiven auch soziale, motivationale, volitionale und oft moralische Kompetenzen enthalten und es erlauben, erworbene Kenntnisse und Fertigkeiten in sehr unterschiedlichen Lebenssituationen erfolgreich, aber auch verantwortlich zu nutzen" (Weinert 2002: 28).

Die Dissertation besteht aus vier Kapiteln. Die drei ersten Kapitel sind den theoretischen Fragen gewidmet. Im Mittelpunkt des ersten Kapitels, das als Ausgangspunkt für die weiterführenden Überlegungen der Arbeit dient, steht das Thema der Multimodalität, ihrer Vielfältigkeit und Omnipräsenz im öffentlichen Diskurs. Das erste Unterkapitel setzt sich zum Ziel, den Terminus Multimodalität selbst zu beleuchten. Im zweiten Unterkapitel wird die Rolle und Bedeutung multimodaler Texte in der heutigen Zeit untersucht. Da diese Texte schon als „ein Normalfall des Kommunizierens" (Stöckl 2011: 47) und/oder „eine schon immer geltende Norm" (Wildfeuer u. a. 2020: 18) bezeichnet wurden, wird an dieser Stelle untersucht, ob die Beherrschung sowie die korrekte Anwendung der sogenannten multimodalen Kompetenz (vgl. Baacke 2007: 98 f., Stöckl 2011: 45, Müller 2012: 30) gegenwärtig bereits eine Notwendigkeit darstellen. Das letzte Unterkapitel präsentiert verschiedene Ansätze zur Betrachtung des Phänomens Multimodalität. Es beschreibt, wie multimodale Texte in diesem Kontext untersucht und analysiert werden können.

Das zweite Kapitel widmet sich dem Konzept der Bildlinguistik, einschließlich der logovisuellen Didaktik, die innerhalb dieser linguistischen Subdisziplin entwickelt wurde. Diese Didaktik reagiert hauptsächlich auf die Notwendigkeit, den kritischen und reflektierten Umgang mit multimodalen Texten im schulischen Kontext zu fördern. Nach der Darlegung der bisherigen theoretischen und methodischen Überlegungen zu dieser Didaktik wird in diesem Kapitel aufgezeigt, wie Text-Bild-Kombinationen die Entwicklung sprachlicher und nichtsprachlicher Kompetenzen im DaF-Unterricht unterstützen können.

Im dritten Kapitel verlagert sich der Schwerpunkt des Interesses auf die kognitiv-affektive Dimension des Lern- und Lehrprozesses im Fremdsprachenunterricht. Zunächst werden die beiden Begriffe Kognition und Emotion definiert, anschließend wird ihre Interaktion im Lern-Lehr-Prozess behandelt. Danach werden ausgewählte Modelle dargestellt, die das Thema des multimodalen Lernens im Fremdsprachenunterricht unter Einbeziehung kognitiver und kognitiv-affektiver Faktoren beleuchten.

Das vierte, empirische Kapitel enthält eine detaillierte Darstellung des Forschungsansatzes, der Forschungsorganisation sowie der Durchführung der eigenen Forschung. Diese gliedert sich in zwei Hauptteile: die Vor- und die Hauptstudie. Zuerst werden die Ergebnisse der Vorstudie (unkontrollierte Beobachtung) präsentiert, welche das Verhalten der Studierenden bei der Arbeit mit multimodalen Texten im Unterricht erfassen. Danach folgen die Ergebnisse der Hauptstudie: Zunächst werden die aus dem Fragebogen erhobenen Daten analysiert, die die Präferenzen der Studierenden im Hinblick auf das Fremdsprachenlernen sowohl im Bildungswesen als auch in der Freizeit sowie deren Meinungen zur Arbeit mit multimodalen Texten im Fremdsprachenunterricht betreffen (erste Phase der Hauptstudie). Anschließend werden in der zweiten Phase die Daten aus Partnerinterviews untersucht, die detaillierte Meinungen zum Einsatz multimodaler Texte im DaF-Unterricht beleuchten (zweite Phase der Hauptstudie). Der empirische Teil schließt mit der Diskussion der Ergebnisse und

einem Ausblick. Am Ende der Dissertation werden Schlussfolgerungen präsentiert. Die Arbeit enthält außerdem ein Abbildungsverzeichnis, ein Diagramm- und Tabellenverzeichnis sowie eine Bibliografie. Zusätzlich wird ein Anhang hinzugefügt.

1 Multimodalität, multimodale Texte und deren Charakteristik

„Multimodality – the late discovery of the obvious"
Stöckl (2004a: 9)

Da in der Literatur mehrere miteinander verwandte, aber trotzdem unterschiedliche Konzeptionen des Begriffs Multimodalität vorzufinden sind, wird in diesem Kapitel ein umfassender Überblick über unterschiedliche Multimodalitätsauffassungen gegeben, um eine gewisse Ordnung in die Terminologie zu bringen und auf diese Weise eine genaue Definition, die in der vorliegenden Arbeit verwendet wird, herauszuarbeiten. Außer der terminologischen Klärung werden an dieser Stelle die Relevanz und die Notwendigkeit, die sog. multimodale Kompetenz zu beherrschen, thematisiert und schließlich wird das Augenmerk auf einige Konzeptionen des multimodalen Textverstehens, die für die Arbeit mit multimodalen Kommunikationsformen unerlässlich sind, gerichtet.

1.1 Zur Begrifflichkeit des Phänomens Multimodalität

Der Begriff Multimodalität leitet sich von dem englischen Wort *multimodality* ab, das in der englischsprachigen Literatur zum ersten Mal an der Wende vom 20. zum 21. Jahrhundert verwendet wurde[3] (vgl. Maćkiewicz 2016: 19). Den Kern bildet das Wort *mode,* das wegen seiner Mehrdeutigkeit und Ungenauigkeit schwer eindeutig in andere Sprachen zu übersetzen ist. Żebrowska (2013), die sich dafür im Deutschen des Äquivalentes Modalität bedient, betont:

---

3 In Anlehnung an Michalewski (2009: 7) weist Maćkiewicz (2016: 19) darauf hin, dass man diesem Wort in Form eines Adjektivs noch früher begegnen konnte, z. B. in Verbindungen wie: der multimodale Text (englischsprachige Literatur) oder am frühesten – schon in den 70er Jahren des 20. Jahrhunderts – in der Verbindung: multimodale Kommunikation (französischsprachige Literatur).

> Wichtig ist dabei auch der Begriff „Modalität", dem mehrere Bedeutungen zugewiesen werden: Sinneskanal, Sinnesmodalität, die Art und Weise der Ausführung der Kommunikation, Zeichentyp, Wahrnehmungsqualität, Verarbeitungsmodus in der Zeichenrezeption (z. B. auditiv, visuell) (Żebrowska 2013: 105).

Daraus wird ersichtlich, dass in dem Wort *mode* mannigfaltige Bedeutungen enthalten sein können. Aus diesem Grunde erscheint es wünschenswert, in der vorliegenden Arbeit zunächst verschiedene Perspektiven bei der Betrachtung dieses Phänomens, das grundsätzlich den Ausgangspunkt für die ganze Konzeption der Multimodalität bildet, zu beleuchten und kurz zu analysieren, um anschließend auf die Frage der Multimodalität näher einzugehen.

Stöckl (2004a, 2004b) unternimmt den Versuch, sich mit der Bedeutung des Wortes *mode* sowohl im Englischen als auch im Deutschen auseinanderzusetzen. Der Betrachtungsweise von Jakobson (1971) folgend nimmt der Autor an, dass:

> ‚Mode' would seem to be an easily definable term. Following Jakobson's idea of ‚code' (Jakobson 1971), it is often glossed as ‚sign system', ‚sign-repertoire' or ‚semiotic' from which communicators can pick their signs to realise their communicative intentions (Stöckl 2004a: 11).

Darüber hinaus ergänzt er:

> Dabei ist zunächst davon auszugehen, dass Modalitäten Kodesysteme sind, die über bestimmte Zeichenrepertoires und Verwendungsregeln verfügen, derer sich Textproduzenten in konkreten Situationen entsprechend ihrer kommunikativen Absichten bedienen und so in kommunikativen Ereignissen sozialen Sinn stiften (Stöckl 2004b: 16).

Wie diesen Worten zu entnehmen ist, vereinfacht Stöckl dieses Problem und sieht in diesen drei Wörtern – *mode/code/sign system* (dt. Modalität/Kodalität/Zeichensystem) – lediglich Äquivalente. Die Sache wird jedoch komplizierter, wenn man Überlegungen dazu anstellt, was genau zu den Zeichensystemen gerechnet werden kann. Um zu präzisieren, welche Elemente in multimodalen Kommunikaten gekoppelt werden können und wie sie zu hierarchisieren sind, spezifiziert Stöckl folgende Gruppen:

1. *sensory channels* (Sinneskanäle), z. B. visuell, auditiv, kinästhetisch, olfaktorisch, gustatorisch,
2. *core modes* (zentrale Modalitäten), z. B. das Bild, die Sprache,
3. *medial variants* (mediale Varianten), z. B. ein statisches/dynamisches Bild,
4. *sub-modes* (Submodalitäten), z. B. die Farbe,
5. *features* (Merkmale), z. B. der Farbton oder die Farbsättigung (Stöckl 2004a: 11–16).

Diese Einteilung erweitert Stöckl (2004b) noch um eine weitere, sechste Gruppe. Dies erfolgt anlässlich der Thematisierung von Typographie[4], die der Autor als einen Untersuchungsgegenstand der Textlinguistik zu etablieren versucht, da sie in den linguistischen Arbeiten nicht genügend Beachtung findet. Die Typographie, also der „Körper des Textes – seine unverzichtbare Lebensgrundlage, der Ort und Stoff seiner Existenz" (Stöckl 2004b: 15) ist seiner Meinung nach immer abhängig von Sprache, bildet aber gleichzeitig ein eigenständiges Ausdruckssystem, das in der „Variabilität der Form von Schriftzeichen-Typen, d. h. Serifen, Strichstärken, Größenverhältnissen, Rundungen, Ornamenten etc." (Stöckl 2004b: 12) in einem Gesamttext zum Ausdruck kommen kann. Die Typographie „[verfügt] über Form und Bedeutung und [entfaltet] sich erst im kontextsensiblen Gebrauch zwischen Textproduzent und -rezipient" (Stöckl 2004b: 15). Ist die Typographie also ein eigenständiges, aber immer mit Sprache in Verbindung stehendes System, müsse sie auf irgendeine Weise von dieser differenziert werden. Dies geschieht unter Trennung der Modalitäten in zentrale und periphere Modalitäten:

> Periphere Modalitäten zeichnen sich dadurch aus, dass sie nur durch die mediale Realisierung einer zentralen Modalität ins Spiel kommen und eng mit dieser Kernmodalität verbunden bleiben. So sind Typographisches, Paraverbales und Nonverbales periphere Zeichenmodalitäten der zentralen Modalität Sprache, die sie materialisieren helfen, begleiten und überformen (Stöckl 2004b: 16).

Stöckl (2004b) betrachtet somit alle Texte, die mindestens aus sprachlichen Elementen bestehen, als multimodal, weil sie sowohl von der zentralen Modalität (Sprache) als auch von der peripheren Modalität (Typographie) gebildet werden[5]. Die Einteilung in zentrale bzw. periphere Modalitäten sagt dabei nichts über die Wichtigkeit aus (vgl. Stöckl 2004b: 16). Um die Korrelationen, die zwischen den zentralen/peripheren Zeichenmodalitäten, ihren Gestaltungsdimensionen (Sub-Modalitäten) und Gestaltungsressourcen (Merkmale) in multimodalen Texten auftreten können, zu veranschaulichen, erstellt Stöckl ein Schema, das auch in der vorliegenden Arbeit als Beilage hinzugefügt wird (siehe Anhang, Anlage 3).

Zusammenfassend ist festzustellen, dass Stöckl (2004a, 2004b) unter dem Begriff *mode* Zeichen jeglicher Art in kommunikativen Situationen versteht. Diese können weiter in sechs Gruppen unterteilt und so klassifiziert werden. Multimodalität wird

---

4 Typographie, also „[die] Form, Auswahl und Verwendung von Schrift im Text sowie [die] gesamt[e] graphisch[e] Gestaltung eines Schriftstückes oder Dokuments" (Stöckl 2004b: 5).

5 Dies ähnelt den Betrachtungsweisen von Kress/van Leeuwen (1998) und Simon (2018), die sagten, dass: „all texts are multimodal" (Kress/van Leeuwen 1998: 186) und: „Alle Kommunikationsformen sind als multimodal zu verstehen, auch das Sprechen aufgrund der begleitenden para- und nonverbalen Faktoren; alle Texte sind multimodal, auch schriftliche Kommunikation, wenn man beispielsweise Layout und typographische Gestaltung berücksichtigt" (Simon 2018: 59).

von ihm einfach als das vernetzte System dieser Elemente angesehen (vgl. Stöckl 2004a: 9).

Eine andere Perspektive auf das Verständnis des Begriffs *mode* präsentiert Holly (2009). Ihm zufolge stellen die beiden zwei Wörtern *mode* und *code* völlig verschiedene Erscheinungen dar. Der Terminus „kodal" bezieht sich nämlich auf die materiale Qualität der Zeichenoberflächen, d. h. auf die Herstellungsweisen von Zeichenarten wie Lautsprache, Schriftsprache oder auch Gebärdensprache. Das Wort „modal" hingegen steht für ein bestimmtes Rezeptionspotenzial von Zeichenarten, das mit den Sinnesorganen Ohr und Auge verknüpft ist. In diesem Sinne wird von den auditiv wahrnehmbaren (Sprachlaute, Töne, Musik) sowie den visuell wahrnehmbaren Zeichenarten (Gestik, Mimik, Bild, Film, Skulptur, Schrift) gesprochen (vgl. Holly 2009: 391).

Auch nach Żebrowska (2013: 100 f.) und Opiłowski (2013: 221 ff.[6]) werden die beiden Termini *mode* und *code* nicht einheitlich verwendet, obwohl sie sehr eng miteinander korrelieren. *Modes* (dt. Modalitäten) begreifen sie als Sinneskanäle, *codes* (dt. Kodalitäten, bzw. Kodes) als Zeichensysteme. Makowska (2014) betont:

> In Sprache-Bild-Texten werden Informationen auf bimodale und selektiv wahrzunehmende Weise präsentiert. Codes, die mit den einzelnen Modalitäten verbunden sind, gelten als Zeichensysteme, die bestimmen, wie Zeichen zusammengesetzt werden müssen, damit sich intendierte Bedeutungen ergeben (Makowska 2014: 229 f.).

Am weitesten fassen den Begriff *mode* jedoch Kress und van Leeuwen (1996, 2001) sowie Kress (2010). Die Autoren, die sich auf Hallidays (1978) Konzeption der Metafunktionen von Sprache (engl. *Systemic Functional Linguistics* – SFL) stützen, verstehen unter dem Terminus *mode* (Abkürzung von *signing modes*) jedes semiotische Zeichensystem, das die folgenden drei Funktionen erfüllt:

1. *ideational metafunction* (dt. repräsentative Funktion) – die auf der Schilderung bzw. Interpretation der Welt beruht:

   > Any semiotic system has to be able to represent, in a referential or pseudo-referential sense, aspects of the experiential world outside its particular system of signs. In other words, it has to be able to represent objects and their relations in a world outside the representational system (Kress/van Leeuwen 1996: 40 f.).

---

6 Opiłowski (2013: 221 ff.) spricht im Kontext der Massenmedien und der komplexeren, multimodalen Texte, die dadurch generiert werden, von Kodes (z. B. Sprache, Bild) und Subkodes (Schriftgestaltung, Farben, Strukturen und entsprechende Sprachregister).

2. *interpersonal metafunction* (dt. interpersonelle Funktion) – die die sozialen Beziehungen zwischen den Gesprächspartnern in einer kommunikativen Situation herstellt:

   > Any semiotic system has to be able to project the relations between the producer of a sign or complex sign, and the receiver/reproducer of that sign. That is, any semiotic system has to be able to project a particular social relation between the producer, the viewer and the object represented (Kress/van Leeuwen 1996: 41).

3. *textual metafunction* (dt. textuelle Funktion) – die mit den strukturellen Aspekten der Wortorganisation in einem Text verbunden ist:

   > Any semiotic system has to have the capacity to form texts, complexes of signs which cohere both internally and with the context in and for which they were produced (Kress/van Leeuwen 1996: 41).

Den genannten Autoren zufolge betreffen diese Funktionen nicht nur die Zeichenmodalität Sprache (wie es anfangs von Halliday 1978 konzipiert wurde), sondern sie können auf weitere semiotische Elemente erweitert werden[7]. Erfüllt ein Element alle drei Funktionen, kann es als *mode* (Zeichensystem) betrachtet werden. Diese Funktionen sind laut Kress (2010) um eine weitere zu ergänzen, und zwar die sog. soziokulturelle Funktion (engl. *social and cultural aspect*). Damit wird zusätzlich der kulturelle Kontext, in dem ein Kommunikat auftritt, berücksichtigt. Kress (2010: 79) äußert sich dazu folgendermaßen:

> Mode is a socially shaped and culturally given semiotic resource for making meaning. Image, writing, layout, music, gesture, speech, moving image, soundtrack, and 3D objects are example of modes used in representation and communication.

Des Weiteren fügt er hinzu: „What a community decides to regard and use as mode is mode" (Kress 2010: 79). Für Kress und van Leeuwen ist also jede Art der semiotischen Ressource, die in irgendeiner Weise zur Herstellung von Bedeutungen verwendet werden kann und von der Gesellschaft als solche betrachtet wird, als *mode* zu bezeichnen. Die Grenzen der *mode*-Anerkennung sind dabei derart fließend, dass sie sogar innerhalb einer Gruppe von Menschen variieren können. Kress u. a. (2000) betonen:

> […] the question of whether X is a mode or not is a question specific to a particular community. As laypersons we may regard visual image to be a mode, while a professional

---

7 Sie verdeutlichen: „In the form in which we gloss them here they [the functions – KS] are not specific to any one semiotic mode: for instance, they are not specific to the linguistic" (Kress/van Leeuwen 1996: 40).

> photographer will say that photography has rules and practices, elements and materiality quite different from that of painting and that the two are distinct modes (Kress u. a. 2000: 43; zitiert in Wildfeuer u. a. 2020: 135).

Wird das Wort *mode* im wissenschaftlichen Diskurs unterschiedlich aufgefasst, scheint auch der Terminus Multimodalität in der Literatur vielfältig und unpräzise verwendet zu werden. Im Folgenden werden mehrere Definitionen dieses Begriffs präsentiert, um die Diversität des Phänomens offenzulegen und schließlich für die vorliegende Arbeit eine möglichst genaue Definition zu geben.

Einer der prominentesten Vertreter der deutschen Medienlinguisten ist Hartmut Stöckl. Seine Untersuchungen widmet er vor allem dem Konzept der Multimodalität, in dem er die Beziehungen zwischen unterschiedlichen Zeichensystemen analysiert. Ihm zufolge „[ist] Multimodalität ein vielleicht hochtrabender und vager Begriff für ein vergleichsweise einfaches und alltägliches Phänomen. Er bezeichnet Texte und kommunikative Handlungen, die mehrere verschiedene Zeichensysteme (Sprache, Bild, Ton) beinhalten" (Stöckl 2011: 45). Dieser Begriff gilt als vage, weil er unterschiedliche Assoziationen auslösen kann. Welche genau dazu gehören, erklärt Stöckl (2011: 47) folgendermaßen:

1. Multimodalität ist die Kopräsenz und wechselseitige Verknüpfung mehrerer Zeichenmodalitäten auf verschiedenen Ebenen (z. B. Semantik, Handlungsfunktion etc.) in einem Gesamttext.
2. Multimodalität ist eine gesamtkulturelle Kompetenz und eine individuelle Intelligenz, die das Erschließen von Sinn durch „Transkribieren", d. h. Kommentieren, Erklären oder Paraphrasieren von Botschaften eines Zeichensystems/Mediums durch ein anderes beinhaltet.
3. Multimodalität ist eine Regeln folgende und Muster bildende semiotisch-kognitive Aktivität in der Produktion und im Verstehen von Texten.

Mit dem Terminus Multimodalität bezeichnet man demzufolge nicht nur die Verknüpfung mehrerer Zeichenmodalitäten in einem Gesamttext. Er wird zugleich mit der Fähigkeit assoziiert, den Sinn solcher Handlungen korrekt zu verstehen bzw. zu erklären und mit ihnen kooperieren zu können.

Dem von Stöckl (2011) vorgeschlagenen Konzept der Multimodalität folgt Siever (2014, 2015). Die Autorin verweist darauf, dass „Kommunikation [...] im 21. Jahrhundert häufig in den digitalen Medien [stattfindet], und Kommunikation in den digitalen Medien wiederum [...] heutzutage multimodaler denn je [ist]" (Siever 2014: 381). Um den Begriff Multimodalität zu definieren, plädiert sie für die Unterscheidung zwischen Sinnesmodalitäten und Kodalitäten:

> Unter Multimodalität wird die Kombination von mehreren Sinnesmodalitäten (visuell, auditiv, olfaktorisch, gustatorisch, taktil) oder Kodalitäten (Sprache, Bild, Ton) verstanden (Siever 2014: 382).

Siever (2014) unterscheidet zwei Arten der Multimodalität: die Multimodalität im engeren Sinne (i. e. S.) und die Multimodalität im weiteren Sinne (i. w. S.). Mit der ersten ist gemeint, dass mehrere Kodalitäten und Sinnesmodalitäten gleichzeitig vorhanden sind, die zweite findet dagegen bei einer Kodalität und mehreren Modalitäten, bzw. einer Modalität und mehreren Kodalitäten statt (vgl. Siever 2014: 382). In der Praxis hieße das für den Fremdsprachenunterricht Folgendes: Über Multimodalität i. e. S. könnte z. B. dann gesprochen werden, wenn der Lehrer bei Einführung neuer Wörter bzw. Redewendungen diese laut ausspricht, an die Tafel schreibt und sie mit einem graphischen Beispiel illustriert. In diesem Falle wird das geschriebene Wort mit dem gesprochenen, und die auditive Sinnesmodalität mit der visuellen verknüpft. Multimodalität i. w. S. liegt dagegen bei der Arbeit mit einem ausschließlich aus verbalen Elementen bestehenden Text vor. Das stille/laute Lesen bzw. Schreiben erfolgt mit Hilfe einer Kodalität (Sprache) und zweier Sinnesmodalitäten (der visuellen und der auditiven) (vgl. Makowska 2018: 219).

Dass im Bereich der Multimodalität immer noch keine terminologische Einheitlichkeit herrscht, zeigt die von Siever (2015) erstellte Tabelle (siehe Tab. 1), in der sie präsentiert, welche Begriffe Werner Holly (WH), Ellen Fricke (EF) und Christina Margrit Siever (CMS) verwenden, wenn sie Fragen der Multimodalität thematisieren.

Der Begriff Multikodalität steht nach Holly (2010: 2203) für eine „verschiedene Zeichenarten enthaltend[e]" Kommunikationsform. Dagegen beschreibt er mit dem Begriff Multimodalität ein Kommunikat, das mehrkanalig ist und verschiedene Sinne anspricht. Für Holly (2010) ist es also wichtig, ob sich ein Kommunikat aus vielen Zeichensystemen zusammensetzt oder auf mehrere Sinneskanäle einwirkt. Wie der Tabelle 1 zu entnehmen ist, fehlt in Hollys Differenzierung ein Terminus für den Fall, dass mehrere Kodalitäten und mehrere Modalitäten auf einmal in einem Kommunikat vorkommen.

Im Fall der mehrkanaligen Kommunikationsformen kann laut Holly (2011: 233) eine noch weitere Differenzierung der Begriffe vollzogen werden. Unter Verwendung von audiovisuellen Formen (gemeint ist hier sowohl die sprachliche als auch die schriftliche Kommunikation) unterscheidet er zwischen der „primären" und der „sekundären" Audiovisualität. Da die Beteiligung von mindestens zwei Sinnen in einer kommunikativen Situation – so Holly (2010) – es rechtfertigt, von einer multimodalen Natur der Dinge zu sprechen, lässt sich die Differenzierung ebenso auf den Begriff der Multimodalität anwenden. In diesem Zusammenhang kann also von einer primären und einer sekundären Multimodalität die Rede sein. Der erste Typ der Multimodalität wird die mikrokommunikativen Face-to-Face-Situationen umfassen (dadurch wird