



Sprache(n) als Zugang:
Die Anwendung von
Sprache als Mittel der
Zugänglichkeit im digitalen
Zeitalter

MAIJA HIRVONEN/TAMPERE

Zusammenfassung Dieser Aufsatz setzt sich mit der Funktion der Sprache als Mittel der digitalen Zugänglichkeit auseinander. Es wird anhand konkreter Anwendungsbereiche erläutert, wie Sprache als Ersatz von nicht-sprachlichen Informationen fungiert, und darauffolgend veranschaulicht, wie unterschiedliche Sprachen wegen ihrer unterschiedlichen Ausdrucksmöglichkeiten bzw. -konventionen diesen Ersatz gewährleisten. Das sprachliche Beschreiben wird dabei in den Fokus gerückt, weil es einen wesentlichen Zugang zu visuellen und audiovisuellen Inhalten in Bereichen wie ‚deskriptive Metadaten‘ sowie ‚Audio- und Videodeskription‘ leistet. Danach wird durch eine kontrastive Fallanalyse der deutschen, englischen und spanischen Audiodeskription eines Films veranschaulicht, welche Anforderungen die Wahl der Sprache an das Beschreiben von Visualität stellt. Abschließend werden künftige Entwicklungen der Versprachlichung von digitalen Inhalten erörtert und dabei die Perspektiven der mehrsprachigen Versprachlichung ausgelotet.

1. Digitalisierung, Zugänglichkeit und Sprache

Digitalisierung bedeutet eine Umwandlung vieler Lebensbereiche, darunter auch die Kommunikation, von analogen in digital-elektronische Formate. Dabei werden die Möglichkeiten der Menschen, miteinander zu kommunizieren, in dem Sinne neu gestaltet, dass sie es mittels Maschinen bzw. künstlicher Intelligenz tun und dabei eine Vielfalt von Kommunikationsmodalitäten anwenden. Die Digitalisierung bringt einerseits viele Vorteile mit sich: neue Möglichkeiten der Beschaffung und Nutzung von Dienstleistungen, z.B. automatische Translation (Übersetzung und Dolmetschen) in Realzeit. Auch für die Wissenschaft ergeben sich neue Perspektiven: Beispiele dafür sind neue, intelligente Werkzeuge zur Korpuserstellung und -analyse, aber nicht nur Textkorpora im herkömmlichen Sinn sind dabei zu beachten, sondern auch Korpora von audiovisuellen Daten, wie z.B. Film- und Fernseharchive. In diesem Aufsatz werden insbesondere die Möglichkeiten der zuletzt genannten Korpora dargestellt, weil Sprache und Sprachen im Hinblick auf diese eine Schlüsselrolle spielen.

Andererseits treten mit der Digitalisierung auch Barrieren auf, die u.a. mit der Kommunikation zu tun haben und die überwunden werden müssen, um eine gerechte Gesellschaft schaffen zu können. Die Barrierefreiheit bzw. die Zugänglichkeit sind spätestens seit den letzten Jahren in aller Munde. Laut der Richtlinie der Europäischen Union (2016/2102) über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen, die ihrerseits auf der UN-Behindertenrechtskonvention und der UN-Menschenrechtskonvention beruht, müssen öffentliche digitale Dienstleistungen, ob als Mobile-Apps oder als Internetseiten verfügbar, im EU-Raum barrierefrei sein. Das bedeutet, es muss allen – auch jenen mit Sinnes-, physischen oder geistigen Behinderungen – ermöglicht werden, die Dienstleistungen und Informationsquellen durch wahrnehmbare und verständliche Informationen gebrauchen zu können. In den kommenden Jahren wird die Vorschrift der Zugänglichkeit auf den Privatmarkt erweitert, so die neueste EU-Richtlinie (2019/882) über die Barrierefreiheitsanforderungen für Produkte und Dienstleistungen (siehe Maaß/Rink 2019, Kruse 2019, Hirvonen & Kinnunen 2020a). Doch können viele von diesen Barrieren gerade dank der Digitalisierung und der künstlichen Intelligenz überwunden werden, wenn technologische Lösungen die Limitationen natürlicher Fähigkeiten der Menschen überschreiten (siehe Burchardt/Uszkoreit 2018: VIII).

Sprachlichkeit hat eine wichtige, aber keine eindeutige Rolle in der Zugänglichkeit. Sie kann sowohl eine Barriere als auch einen Zugang herstellen. Eine Barriere gestaltet die Sprachlichkeit dann, wenn die Nutzer¹ die Sprache, die verwendete Sprachvarietät bzw. das Register nicht genug beherrschen. Zum Beispiel kann die einheimische Sprache in ausländische Sprachen, die Expertensprache in Laiensprache oder die Standardsprache in Leichte Sprache übersetzt werden, um Informationen für eine breitere Gesellschaft verständlich zu machen. Viele Menschen mit geistigen oder kognitiven Beeinträchtigungen bedürfen Leichter Sprache in alltäglichen Kommunikationssituationen sowie beim Zeitunglesen oder Fernsehen

1 Die männliche Form wird in diesem Aufsatz aus Gründen der besseren Lesbarkeit verwendet.

(Bredel/Maaß 2019). Darüber hinaus wird erwartet, dass die Umsetzung der (oft kompliziert strukturierten) Standardsprache in die (einfacher strukturierte und standardisierte) Leichte Sprache für den Nutzer generell und auch für die künstliche Intelligenz sowie die maschinelle Übersetzung hilfreich sein könnte (siehe Matthiesen 2017). Das bringt neue Möglichkeiten und ein Positivsummenspiel hervor, wenn die Informationen zuerst in Leichter Sprache erstellt werden und danach maschinell übersetzt werden können.

Die Sprache ermöglicht aber vielen Menschen einen Zugang zur Kultur und Kommunikation, aus der sie sonst ausgeschlossen wären. Intermodale Translation bedeutet das Übersetzen und Dolmetschen zwischen Modalitäten, wenn zum Beispiel Rede in Text, Text in Rede oder Sprache in Bilder umgesetzt wird (siehe z.B. Hirvonen 2018: 95–96). Für Menschen mit Sehbeeinträchtigungen kann die Umsetzung von visuellen Informationen in sprachliche Beschreibungen u.a. Film, Fernsehprogramme, Kunst und Theater zugänglich(er) machen.

2. Sprache als Mittel der Zugänglichkeit

Dieser Aufsatz fokussiert auf die Sprache als ein kommunikatives Mittel, das Zugang zu nicht-sprachlichen Informationen und Erlebnissen leistet. In diesem Zusammenhang wird ein besonderes Augenmerk auf die beschreibende Funktion der Sprache gerichtet und ihre Rolle in kommunikativen und technologischen Praktiken diskutiert.

2.2. Sprachliches Beschreiben im Gespräch und in Texten

Neben Narration und Argumentieren gehört die Deskription zu den Kernfunktionen der Sprache (Stutterheim/Kohlmann 2001: 1279). Das Beschreiben ist in vielen menschlichen Kommunikationssituationen elementar, sowohl im Gespräch (z.B. in Wegbeschreibungen) als auch in verschiedenen Textsorten (z.B. in Reiseführern und -geschichten sowie in Drehbüchern). Durch das Beschreiben wird ein Sachverhalt möglichst genau erläutert, um diesen für andere vorstellbar und verständlich zu machen (ebd. 1280, siehe auch Schubert 2009: 5–6). Diese Definition leiht der Sprache

und der Gattung Deskription eine vermittelnde Funktion, die schon längst relevant ist. Ein historischer Überblick (Pujol/Orero 2007) offenbart, dass die Übertragung der Visualität durch Sprache in unterschiedlichen Situationen ausgeübt worden ist; darunter etwa die Ekphrasis, die anhand lyrischer Wortkunst physische und visuelle Grenzen überschritt und Werke der bildenden Kunst mit Worten beschreibt.

Noch heute ist die Deskription in der menschlichen Kommunikation umfassend. In der sozialen Interaktion tritt die Gattung Deskription u.a. in Raumbeschreibungen hervor (z.B. Mondada 2013 zu Beschreibungen in Stadtführungen, siehe auch Mondada 2011: 392 über den Stand der Forschung zu Raumbeschreibungen). Zudem ist das Beschreiben ein wesentlicher Teil von Gesprächen, in denen Patienten oder sonstige Angetroffene über ihre physiologische Lage oder eben über ihre räumliche Lokation berichten (z.B. bei Notanrufen, Mondada 2011). Beschreibungen sind ebenso in vielen Textsorten üblich. Schubert (2009) zeigt anhand eines umfangreichen Korpus englischsprachiger Texte, dass Raumbeschreibungen oder „Blickführung durch Sprache“ sowohl in Reiseführern und -berichten, in historischen und geographischen Texten, in Romanen und Kurzgeschichten als auch in Gedichten, Dramen und Drehbüchern vorkommen (ebd. 469–474).

Beschreibungen sind nicht rein linguistisch zu definieren, sondern sie werden in der multimodalen Kommunikation und Interaktion gestaltet. In literarischen Werken und schriftlichen Texten tragen Bilder und sonstige visuelle Elemente, sowie Typographie und Grafik, zur Beschreibung eines Sachverhaltes bei. In der Interaktion sind Beschreibungen multimodale Konstrukte, die über verschiedene Ressourcen verfügen (z.B. gesprochene und geschriebene Sprache, Gestik, Mimik und Körperhaltung in der Beschreibung von Weg). Außerdem ist zu vermerken, dass die (sprachliche und graphische) Beschreibung von Forschungsdaten zugunsten der Analyse eine bedeutende Rolle in der sozialwissenschaftlichen Forschung spielt (Mair/Kierans 2007), vor allem in der ethnographischen Methodologie (Schwartz 2002). Dabei ist zu berücksichtigen, in welchen Weisen Forschenden dem Forschungsgegenstand durch die Repräsentation von Daten gerecht werden (z.B. die Beschreibung und Transkribierung vom Weinen wird in Hepburn (2004) in den Blick genommen). Dieser Repräsentationscharakter macht aus

der Sprache ein wichtiges Mittel auch in der digitalen Datenverarbeitung, wie als Nächstes gezeigt wird.

2.1. Digitale Anwendungsbereiche des sprachlichen Beschreibens

Deskriptive Metadaten

Das Beschreiben von Daten wird in der Informatik mit der Bezeichnung *descriptive metadata* erörtert. Indem mit *data* auf „ein potenziell informatives Objekt“ hingewiesen wird, sind *metadata* „eine Darlegung und eine Stellungnahme zu den Daten“, d.h. sie sind Beschreibungsansätze über die Daten („*Metadata is a statement about a potentially informative object*“, Pomerantz 2015: 26). Metadaten dienen der Funktion, Informationen durch die Paarung von Elementen mit Werten („*element-value pairs*“) zu finden (ebd. 79), so dass es beispielsweise möglich ist, bestimmte Texte oder Bilder im Internet aufzurufen. Metadaten werden in verschiedenen Typen kategorisiert, u.a. zu beschreibenden, strukturellen und technischen Metadaten (Pomerantz 2015). Unter diesen waren die beschreibenden Metadaten schon immer der Kern der Metadaten, denn sie gewährleisten die effiziente Anwendung von Ressourcen (vor allem von Informationen) in elektronischen Netzwerken. Doch für die breite Akzeptanz ist laut Pomerantz (ebd. 69) eines ausschlaggebend: die Unkompliziertheit bzw. die Durchschaubarkeit der Beschreibung („*simplicity*“). In dem am meisten verbreiteten Standard von deskriptiven Metadaten, im Dublin Core, sollen mit 15 Elementen (Beschreibungstypen) alle Ressourcen beschrieben werden können (ebd. 71–72). Unter diesen 15 Elementen wird ein Element als *Description* bezeichnet, mit dem „*an account of the resource*“ gegeben werden soll (siehe auch den Terminus ‚*content description*‘ in Braun/Starr 2019).

Wie eine Inhaltbeschreibung aussehen soll, lässt Pomerantz (2015) allerdings offen. Es ist aber auf Folgendes zu verweisen: Während es möglich ist, mit wenigen Wörtern bzw. Eigennamen Werte zu Elementen wie *Contributor*, *Creator*, *Publisher* und *Title* anzugeben und mit einem spezifischen Wortschatz *Format*, *Language* und *Subject* zu beschreiben, mag die Darstellung der Ressource eine längere und differenziertere Beschreibung

verlangen. Das wird von Pomerantz aber nicht angesprochen, obgleich die Inhaltsbeschreibung für Medienunternehmen und -archive einen wichtigen Datenträger darstellt (Braun/Starr 2019: 12). Die Inhaltsbeschreibung funktioniert bei ihnen als ein sprachlicher Ersatz für visuellen Inhalt, aus dem textbasierte Suchen und Abrufe durchgeführt werden können (ebd. 18). Einen weiteren Anwendungsbereich bilden die stetig zunehmenden Sammlungen von visuellen und audiovisuellen, digitalen oder digitalisierten Artefakten und Repräsentationen, wie Kunstwerke, Museumsobjekte, Filme, Fernsehprogramme und neuerdings auch Videos. Je mehr Videos und andere audiovisuelle Daten die Menschen herstellen, desto relevanter wird deren inhaltliche Beschreibung, weil sie eine Datenbank für die künftige historische Forschung darstellt (siehe Unesco 2010–2017). In dieser Hinsicht erwähnt auch Pomerantz (2015: 187) die digitalen Sammlungen wie Europeana und The Digital Public Library of America, die historische Bilder sowie Audio- und Videodokumentation beinhalten, als Beispiele für neue und immer relevantere Anwendungsbereiche von Metadaten.

Metadaten sind somit auch eine Art von Zugang: Sie stellen durch die Verbindung von Werten (Informationseigenschaften) mit Elementen (Informationsbezeichnungen) „access points“ (Pomerantz 2015: 79) zur Verfügung. Inhaltsbeschreibungen finden in dieser Hinsicht nicht nur in professionellem Zusammenhang Anwendung, sondern sie sind auch als Textalternative für digitalen Inhalt nützlich, die die Anwendung des Inhalts bei Kommunikationsbeeinträchtigten unterstützen bzw. ermöglichen.

Audio- und Videodeskription

Für das Ersetzen von visuellen und audiovisuellen Inhalten durch die verbale Beschreibung von Bildern und bewegten Bildern (Videos und Filmen) sind sowohl manuelle als auch automatische Methoden entwickelt worden. Die Audiodeskription verweist auf eine gelegentlich manuell (d.h. durch menschliche Arbeit) geleistete Beschreibungstätigkeit und die Video- bzw. Bilddeskription auf automatische Methoden des Beschreibens. Da die Audiodeskription in der sprach- und translationswissenschaftlichen Literatur bereits ausführlich eingeführt ist (siehe dazu aus den deutschsprachigen Quellen u.a. Fix 2005, Benecke 2014), fokussiert dieser Artikel auf die

automatischen Methoden der Videodeskription (siehe die Überblicke von Aafaq u.a. 2019 und Mogadala u.a. 2019).

Über die Audiodeskription lässt sich zusammenfassend sagen, dass sie schon vor der Videobeschreibung, ab den 1970er und 1980er Jahren gezielt als Zugänglichkeitsdienstleistung gebraucht worden ist. Das heutige Modell wurde aus den USA in die westeuropäischen Länder übernommen, und seit den 1990er Jahren wird die Audiodeskription als eine regelmäßige Dienstleistung u.a. in Deutschland angeboten. Heutzutage ist die Audiodeskription neben der Untertitelung für Hörgeschädigte ein Standardangebot bei Filmen und Fernsehprogrammen in vielen Ländern. Die Audiodeskription bedeutet eine hörbare (also gesprochene) sprachliche Beschreibung dessen, was es gerade zu sehen gibt, und ihre Funktion ist es, die multimodale Repräsentation zu ergänzen oder die visuelle Repräsentation zu ersetzen, sodass auch Nicht-Sehende und Sehbeeinträchtigte diese Kommunikationsmodalitäten verstehen können. Die Audiodeskription wird typischerweise für fertige Filme und Fernsehprogramme von einzelnen Filmbeschreibern oder von Beschreiberteams erstellt.

In Anlehnung an Aafaq u.a. (2019) kann die Video- und Bilddeskription grob auf drei Zwecke mit jeweils mehreren Aufgaben aufgeteilt werden:

- 1) das Vermehren barrierefreier Dienstleistungen durch
 - a) die automatische Beschreibung von Videos für Sehgeschädigte,
 - b) die automatische Umsetzung von Gebärdensprache in gesprochene bzw. geschriebene Sprache sowie
 - c) den automatischen Abruf von visuellen Informationen;
- 2) eine verbesserte Human-Computer-Interaktion durch den Einsatz von sprachlichen Beschreibungen, die beispielsweise
 - a) die Handlungen auf Video in vereinfachten Handlungsanweisungen umsetzen oder
 - b) dabei helfen, dass Roboter das menschliche Handeln „verstehen“ können;
- 3) eine Erweiterung der perzeptiven und kognitiven Fähigkeiten der Roboter, die noch mehr in der Lage sein sollen, auf Grundlage der Bilder (d.h. der Bildpunkte/Pixel) Geschichten zu erzählen.

Aafaq u.a. (2019: 2) fassen die automatische Beschreibung von visuellen Inhalten in zwei Phasen zusammen: Zuerst muss die künstliche Intelligenz den Inhalt, der zu beschreiben ist, verstehen, um ihn danach in sprachlich richtigen („grammatically correct“) Sätzen in natürlicher Sprache zu beschreiben. Im Bereich der künstlichen Intelligenz und des maschinellen Lernens wird auf ein menschenähnliches Verstehen gezielt durch das Erlernen von Prozessen und Produkten, die Menschen realisieren bzw. erstellen. So tritt dieses Paradigma beispielsweise in der Entwicklung der maschinellen Übersetzung als „kontextsensitive“ Methodologie des Übersetzens auf (Schmalz 2019). In der Language-Vision-Forschung hat sich in den letzten Jahren die Einbeziehung von extrasprachlichem, vor allem sachlichem und allgemeingültigem Wissen als eine interessante Möglichkeit für die Verbesserung der Maschinenleistung erwiesen (Mogadala u.a. 2019: 71, auch Aafaq u.a. 2019: 23).

Die Fähigkeit zum Verstehen ist eine Kernaktivität sowohl bei Menschen als auch bei Maschinen. Doch sind die Grundlagen unterschiedlich. Bei Menschen erfolgt das Verstehen zwangsläufig individuell, denn zu verstehen ist eine Interaktion zwischen dem Text (oder dem ‚Informationsangebot‘) und dem Leser (oder dem Nutzer des Informationsangebots): Es bedarf einer Integration der stilistisch-formalen Textmerkmale (Wortschatz, Gliederung usw.) mit individuell bedingten Verarbeitungsvoraussetzungen (Leserstil, Vorwissen, Intelligenz und Interesse, darunter auch Emotionen und Motivationen) (Christmann/Groeben 2019: 123–126; zum translationswissenschaftlichen Begriff ‚Informationsangebot‘ siehe Reiss/Vermeer 1984 und eine Diskussion in Hirvonen 2018: 108–110). Bei Maschinen, die kaum zu Individuen zählen, mag das Verstehen ein Textoberflächenphänomen bleiben: es beruht, so Bender und Koller (2020: 5193), auf die sprachliche Form und kann deshalb die Bedeutung lediglich widerspiegeln („they learn some reflection of meaning into the linguistic form“, ebd.). Der Unterschied zwischen ‚Bedeutung‘ und ‚Sinn‘ (*meaning* / *sense*, siehe Risku/Pircher 2008: 157) mag gerade das ausschlaggebende Charakteristikum des menschlichen Verstehens sein, was wiederum wichtige Herausforderungen für die maschinelle Nachahmung stellt: über der enzyklopädisch abzuleitenden Bedeutung hinaus geben Menschen

kontextabhängig und individuell bestimmten Sinn zu Sachen, die sie spüren, hören, sehen, lesen usw. Das Verstehen ist demnach ein komplexes Phänomen, das außer individuellen Prozessen auch interaktiv und kollektiv hervorgeht und das deshalb nicht nur psychologischen und linguistischen aber auch soziologischen Forschungsperspektiven bedarf.

Dennoch hat die Sprache spezifische Eigenschaften, die aus ihr ein angemessenes vermittelndes Material macht. In Anlehnung an Grodal (1997: 22–23) können der Sprache bestimmte semiotische Affordanzen zugeordnet werden, die sie zu einem wesentlichen Gegenstand der menschlichen Intelligenz macht: Die Sprache kann die außersprachliche Welt abstrahieren, sie festigt Wahrnehmungen und verdichtet sie in sprachlichen Konstruktionen und Wörtern, und hält sie zudem mehr oder weniger konstant über Zeit, Raum und Personen (ebd.). Mit Sprache wird ein einmaliges Phänomen (etwa Wahrnehmungen und Vorstellungen einer Person) in ein intersubjektiv und kollektiv verteilbares Verständnis „übersetzt“ (etwa die Kategorisierung, die durch linguistische Zeichen markiert werden). Wie diese Übersetzungen von (visuellen) Wahrnehmungen in konzeptuelle Strukturen in unterschiedlichen Sprachen zustande kommen ist über die generellen Affordanzen der Sprache hinaus eine relevante Frage. In Bezug auf Audiodeskription hat Vandaele (2012) nahegelegt, dass als ein deskriptiver Rahmen der Konzeptualisierung die kognitive Linguistik angewandt werden könnte. Dazu zählen Aspekte wie zum Beispiel die Selektion (welchen Anteil der visuellen Szene die Konzeptualisierung trägt), die Perspektivierung (u.a. die Aufteilung der Szene in Figur(en) und (Hinter)Grund), und die Abstraktion (wie viel Detailinformationen die Konstruktion enthält) (ebd. 96–97). Als Nächstes wird diese Relevanz am Beispiel von der Versprachlichung bzw. die Lexikalisierung von Bewegung in der Audiodeskription eines Filmes in drei Sprachen untersucht.

3. Unterschiedliche Sprachen als Mittel am Beispiel von deutscher, englischer und spanischer Versprachlichung von Bewegung

Bewegung (hier im Sinne von *translocation*, also das Übergehen von einer Lokation in eine andere, zu begreifen) ist in vielerlei Hinsicht ein relevanter Forschungsgegenstand. Erstens ist es für die sprachwissenschaftliche

Forschung von Interesse, wie sich die Sprachen voneinander unterscheiden oder wie sie sich ähneln in Bezug auf die Enkodierung von Bewegung (statt die zahlreichen kontrastiven Studien zu Bewegungsverben aufzulisten, erwähne ich an dieser Stelle die berühmte Theorie von Talmy (2000) und die nachfolgende Entwicklung der Theorie z.B. in Slobin (2004) und Beavers u.a. (2010), sowie die Entwicklungen im Forschungsbereich ‚Raumlinguistik‘, u.a. Berthele (2006) und Doval/Lübke (2014)). Zweitens ist die Erkennung bzw. die Beschreibung von Bewegung in der Vision-Language-Forschung ein echtes Desiderat (Kale/Patil 2016). Drittens zeigt die Robotik ein ständiges Interesse an der menschlichen Bewegung, da sie eines der grundlegenden Prinzipien für die robotische Modellierung ist (Siciliano/Khatib 2016: 4). Zum Beispiel in der Menschen-Roboter-Interaktion stellen Bewegungsverben eine wichtige Ressource für die Aufgabe dar, dass die Roboter nach Weisungen der Menschen verfahren können (Kollar u.a. 2014).

Bei der Versprachlichung von Bewegung müssen folgende Fragen betrachtet werden: Zuerst die Frage danach, welche Eigenschaften des Bewegens die analysierten Sprachen bzw. die sprachlichen Beschreibungen jeweils lexikalisieren oder konstruieren. Durch das ‚Profilieren‘ (siehe Schubert 2009: 108) heben sprachliche Konstruktionen notwendigerweise bestimmte Eigenschaften hervor. Zweitens gibt es die Frage, die mit dem Vorangesagten zu tun hat: Wie lexikalisieren unterschiedliche Sprachen die Bewegung? Es gibt bekannte Unterschiede, die in der kognitiven Typologie von Bewegungsenkodierung beschrieben werden (Stolova 2014: 19–21 nach Talmy 2000). Und drittens stellt sich die offensichtliche Frage, in welcher Beziehung der Ähnlichkeit zum außersprachlichen Ereignis die Versprachlichung steht. Falls sie eine Beschreibung dessen sein sollte, soll die Sprache das Ereignis wiedergeben können. Wie oben bereits erklärt, ist die Sprache als semiotisches System dazu geeignet, Ereignisse nachhaltig zu dokumentieren. Um die Nachhaltigkeit zu gewährleisten, bedarf es freilich einer Abstrahierung, die das Ereignis aus der lokalen Situation mit den konkreten Eigenschaften und Details in eine übergeordnete Kategorie übersetzt, die zwar auf die Merkmale des einzigartigen Ereignisses verweist, die aber auch ohne dessen Wahrnehmung nachvollziehbar ist.

Um die Bedeutung der sprachlichen Unterschiede zu veranschaulichen, habe ich exemplarisch eine linguistisch-kontrastive Analyse der Versprachlichung von Bewegung in drei verschiedenen Audiodeskriptionen des Films *Slumdog Millionaire* (Danny Boyle & Loveleen Tandan, 2008) vorgenommen.² In Bezug auf die Beschreibung ergibt sich demnach die Frage: Wie enkodieren unterschiedliche Sprachen die Bewegung in der Audiodeskription? Untersucht werden die Audiodeskriptionen des Films in deutscher, englischer und spanischer Sprache. Ich habe eine Korpusanalyse durchgeführt, in der die finiten Verben im Verbalkomplex, die das Bewegen von einem Platz zu einem anderen (*translocation*) ausdrücken, in den Audiodeskriptionsskripten erfasst wurden. Metaphorische Bewegungsausdrücke wurden weggelassen. Mit dem Datenverarbeitungsprogramm Atlas.ti habe ich die ermittelten Verben nach der sprachtypologischen Einordnung Talmys (2000) mittels zweier Kategorien kodiert: ‚motion verb (*Path*)‘ und ‚motion verb (*Manner*)‘. Die Kategorien geben an, welcher der zwei Gruppen die analysierten Verben angehören, d.h. ob das genutzte Verb auf den Weg bzw. die Richtung (*Path*) oder auf die Art und Weise (*Manner*, auch als ‚Co-Event‘ bezeichnet) der Bewegung verweist (Talmy 2000: 225). Freilich reicht es nicht, die Enkodierung der Bewegung nur auf der Ebene der Verbwurzel zu berücksichtigen, denn weitere Eigenschaften sind durch den Satzkontext des Verbes interpretierbar – also durch den ganzen Verbalkomplex und die Konstruktion der Bewegungsäußerung mit den eingebetteten Wortelementen (z.B. Adverbien). So kann der Bewegungsausdruck mit mehr Aspekten im sogenannten Satelliten bereichert werden (z.B. im Spanischen *salir corriendo* ‚rennend aus-/weggehen‘). Darüber hinaus wird die Analyse zeigen, dass diese Kategorisierung nicht hinreichend ist und dass es eigentlich vier Kategorien gibt, denn einige Verben verfügen über keine der beiden Aspekte (*Path* oder *Manner*) und manche Verben über beide (siehe auch Beavers/Levin/Tham 2010). Dennoch ist es interessant, die kurzmöglichste Form der Bewegungsbeschreibung besonders im Falle der Audio- und

2 In einigen früheren Studien (siehe Hirvonen 2014) habe ich die Versprachlichung von Raum in filmischer Audiodeskription zwar aus kognitiv-linguistischer Sicht aber eher als ein sprachenübergreifendes Phänomen behandelt.

Videodeskription zu betrachten, weil kurze Versprachlichungen dem Zweck der Sprachökonomie dienen, die für beide Praktiken ausschlaggebend ist (siehe Abschnitt 2).

Insgesamt gibt es in den drei Audiodeskriptionen ein Drittel mehr *Path*-Verben (n=404) als *Manner*-Verben (n=307). Dieses Ergebnis mag vor allem daran liegen, dass die *Path*-Verben die *Manner*-Verben in der spanischen Audiodeskription deutlich überschreiten. Tabelle 1 zeigt das Mengenverhältnis.

Tabelle 1: *Path*- und *Manner*-Verben in der deutschen, englischen und spanischen Audiodeskription von *Slumdog Millionär*

	1 Slumdog_DE_ve... 252	2 Slumdog_EN_v... 192	3 Slumdog_ES_v... 269
◇ motion verb (Manner) 307	135	111	61
◇ motion verb (Path) 404	116	81	207
Totals	251	192	268

Während die Anzahl von *Manner*-Verben in der deutsch- (DE) und englischsprachigen (EN) Audiodeskriptionen dominieren – wenn auch in der deutschsprachigen nur in geringem Maße –, ist die überwiegende Mehrheit der Bewegungsverben in der spanischsprachigen (ES) Audiodeskription mit der Enkodierung von *Path* gekennzeichnet. Das bedeutet, dass die Bewegung von Personen (Filmfiguren) oder Gegenständen in der spanischen Audiodeskription typischerweise mit einem den Bewegungsablauf bezeichnenden Verb beschrieben wird. Die im Korpus am häufigsten verwendeten *Path*-Verben sind in Tabelle 2 aufgeführt.

Tabelle 2: Die meistverwendeten *Path*-Verben im Korpus

Word	Length	Count	%
geht	4	54	13,333
acerca	6	22	5,432
kommt	5	15	3,704
llega	5	11	2,716
marcha	6	9	2,222
gehen	5	8	1,975
dirige	6	8	1,975
approaches	10	8	1,975
va	2	6	1,481
sale	4	6	1,481
llegan	6	6	1,481
entrar	6	6	1,481
arrives	7	6	1,481
steigt	6	5	1,235
springt	7	5	1,235
recorre	7	5	1,235
leaves	6	5	1,235
entran	6	5	1,235
aleja	5	5	1,235
vuelve	6	4	0,988
sigue	5	4	0,988
persiguen	9	4	0,988
folgt	5	4	0,988
entra	5	4	0,988
springen	8	3	0,741
siguen	6	3	0,741
salir	5	3	0,741
salen	5	3	0,741
recorren	8	3	0,741
kommen	6	3	0,741
klettert	8	3	0,741

Die meistverwendeten spanischen *Path*-Verben sind

- a. (se) *acerca* ‚kommt näher‘
- b. *llega* ‚kommt (an)‘
- c. (se) *marcha* ‚geht‘
- d. (se) *dirige* ‚begibt sich‘.

Diese profilieren teilweise unterschiedliche Aspekte: a und d verweisen auf ein Ziel oder eine Richtung der Bewegung (die Annäherung oder die Zuführung), wobei b und c nicht so sehr den konkreten Ablauf offenbaren, sondern eher die Perspektive auf die Bewegung (llega bezeichnet eine Bewegung zur Sprecherin hin und (se) marcha gegenteilig das Entfernen von der Sprecherin). Aus den gesamten *Path*-Verben taucht im Korpus aber das deutsche Verb geht am meisten auf. Das bedeutet, dass die meisten ablaufbezeichnenden Verben den relativ abstrakten Vorgang des Fortbewegens bezeichnen. Auffallend ist dennoch, dass in den beliebtesten *Path*-Verben kaum englischsprachige Lexeme auftauchen, sondern die Kategorie wird von den spanischen und deutschen Verben dominiert. Über die in der Tabelle angezeigten Verben hinaus gibt es im Korpus noch 86 verschiedene *Path*-Verben, die nur 1 oder 2 Mal im Korpus vorkommen und die in der Tabelle nicht aufgeführt sind.

Bei der Datenverarbeitung habe ich das Verb absichtlich in der flektierten Form des finiten Verbes (in diesen Fällen der 3. Person Singular) gelassen, denn sie ist in intermodaler Hinsicht bedeutend: die flektierte Form gibt an, was für eine Einheit (z.B. eine oder mehrere Filmfiguren) sich gerade bewegt. Die Tabelle 3 zeigt dahingegen das Mengenverhältnis der meistverwendeten *Manner*-Verben auf.

Tabelle 3: Die meistverwendeten Manner-Verben im Korpus

Word	Length	Count	%
rennt	5	17	5,466
rennen	6	15	4,823
run	3	13	4,180
laufen	6	13	4,180
corre	5	13	4,180
fährt	5	11	3,537
läuft	5	10	3,215
walks	5	9	2,894
corriendo	9	7	2,251
corren	6	7	2,251
runs	4	6	1,929
drängt	6	6	1,929
drängen	7	6	1,929
correr	6	6	1,929
walk	4	5	1,608
tanzen	6	5	1,608
moves	5	5	1,608
bailando	8	5	1,608
tanzt	5	4	1,286
stürmen	7	4	1,286
steps	5	4	1,286
dashes	6	4	1,286
dance	5	4	1,286
stürzt	6	3	0,965
stürmt	6	3	0,965
sneaks	6	3	0,965
rushes	6	3	0,965
running	7	3	0,965
rast	4	3	0,965
pulls	5	3	0,965
hetzen	6	3	0,965
fahren	6	3	0,965
drives	6	3	0,965
bailar	6	3	0,965

Diese Auflistung macht umgehend eine interessante, die verschiedenen Sprachversionen überschreitende Eigenschaft der Audiodeskription sichtbar. In allen Versionen verweist das meistverwendete Bewegungsverb auf dieselbe Art des Bewegens, nämlich auf das schnelle Fortbewegen mit den Ausdrücken *rennt, rennen, run, ‚rennt‘, laufen, corre, ‚rennt‘*. Die Verben können insgesamt wie folgt in Unterkategorien aufgeteilt werden, die ebenso einen interessanten Schwerpunkt im Hinblick auf die Bewegung im Film aufzeigen:

- a. schnelle bzw. hastige Fortbewegungsarten (*rennt/rennen, run/running, laufen, corre/corren, drängt/drängen, stürmt/stürmen, fährt/fahren, rushes, dashes, rast, hetzen und vielleicht auch tanzen, bailando, tanzt, dance, bailar*)
- b. neutrale bzw. ruhige Fortbewegungsarten (*walks/walk, moves, steps, sneaks, pulls, drives*)

Demnach verweist die überwiegende Mehrheit der meistverwendeten Verben auf eine Schnelligkeit, ja sogar eine Hast der Fortbewegung. Diese Neigung zur Kategorisierung der Bewegung in einer bestimmten Art und Weise erschließt interessante Annahmen im Hinblick auf die Visualität des Films: zum einen die Hypothese, dass die Filmfiguren und auch andere Objekte im Film oft beim Laufen oder Rennen gesehen werden; zweitens die Frage, wie diese Handlung visuell dargestellt wird, um sie als ‚rennen‘ oder ‚stürmen‘ zu kategorisieren und was die ausschlaggebenden Merkmale sind, um diese Unterschiede vorzunehmen. Oder anders gefragt: Worauf in der visuellen Vielfalt wird fokussiert?

Interessant sind auch Verben, die beide oder keine von den zwei Eigenschaften ausdrücken, wie etwa *flüchten / huír* (‚weg davon + hasten‘) und *bewegen / move* (keine besondere Perspektive auf die Bewegungsweise oder -richtung). Über das Satzniveau hinaus müssen auch komplexere Konstruktionen berücksichtigt werden, so etwa Kollokationen der Bewegungsverben sowie ganze Verbkomplexe, um die visuelle Vielfalt zu detaillieren. Zum Beispiel drückt die deutsche Audiodeskription die Bewegungsrichtung mit einer Präposition (*rennt über das Feld/die Straße*), mit einem Adverb (*rennt jmdn. hinterher; rennen davon*) oder mit einem Präfix (*läuft jmdm. nach*) aus zusammen mit dem Vollverb, das die Art und Weise der Bewegung aussagt. In der spanischen Audiodeskription wird dahingegen

die Art und Weise durch ein Adverb oder eine Zusatzkonstruktion zu einem die Bewegungsrichtung beschreibenden Vollverb dargestellt (z.B. *se acerca lentamente* ‚kommt langsam näher‘, *sale corriendo* ‚geht rennend davon‘); jedoch scheint dies nicht oft zu geschehen (siehe auch Ibarretxe-Antuñano (2003), die die Versprachlichung von Bewegung in der spanischen und baskischen Übersetzungen von *Der Herr der Ringe* aus dem Englischen analysiert, und Slobin u.a. (2014) zur Versprachlichung von Bewegungsart im Spanischen und Englischen).

Um Satelliten als Teil der Bewegungsbeschreibung zu berücksichtigen, müssen ganze Sätze und der Kontext analysiert werden. Eine qualitative Sichtweise ist ebenso vonnöten. Was ist der (visuelle) Unterschied zwischen *rennen* und *laufen* oder bei Verben wie *gehen* und *kommen*, die gegenseitige Perspektiven ausdrücken? Das kann qualitativ in einer vergleichenden Betrachtung der Versprachlichung und der entsprechenden visuellen Ausdrucksformen erörtert werden. Um diese Annahmen und Fragen zu beantworten, müssen weitere Studien vorgenommen werden. Ohne Zweifel helfen solche Studien auch dabei, deskriptive Metadaten weiterzuentwickeln, wenn mehr Wissen über die Typologien der Intermodalität gewonnen werden kann. Abschließend betrachten wir noch eine Szene aus dem Film unter qualitativer Perspektive.

Abbildung 1: Bewegungsbeschreibungen einer Szene von *Slumdog Millionär* (die Bilder sind Screenshots aus der Szene 00:05:44–00:06:25 von der deutschsprachigen Synchronversion *Slumdog Millionär*, © Celador Films, Film4)

Legende:

Fettdruck: **Vollverb**

Kursiv: *Satellit*

(...): ausgelassener Teil der Audiodeskription bzw. des Hörfilms



(1)

(1)

Zwei Wachmänner kommen auf Motorrädern.

Two airport security guards arrive on mopeds, shouting: "PRIVATE LAND!"

Dos policías en moto llegan para echarles de la pista de aterrizaje.



(2)

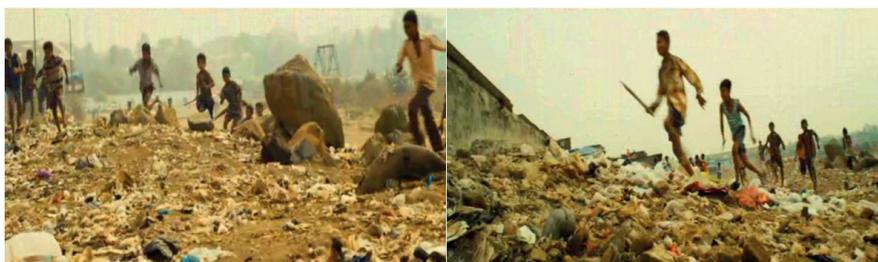
(2)

Die Kinder laufen gehetzt über die schmale Landebahn davon.

The gang of boys suddenly disperse in all directions.

Todos los chavales, alrededor de una veintena, salen corriendo.

(...)



(3)

(3)

Die Kinder *flüchten* über eine große Müllhalde

The boys clamber across a rubbish tip

Los chavales atraviesan corriendo un vertedero. (...)



(4)

(4)
 und rennen in ein riesiges Slumviertel.
 up over corrugated rooftops and
 Todos se internan en una zona chabolista.



(5)

(5)
 Sie hetzen durch die engen verschachtelten Gassen zwischen den Häusern (...)
 down into the narrow alleyways of their slum.
 Saltan sobre los tejados de uralita de las casas.

Im ersten Shot (1) ist zu sehen, wie zwei Filmfiguren auf Motorrädern in Richtung der Kamera fahren. Diese Handlung wird in den drei Audiodeskriptionen sehr ähnlich beschrieben: Alle Sprachversionen deuten mit dem Verb „(an)kommen“ (*kommen, arrive, llegan*) die Bewegungsrichtung und die Kameraperspektive an. Im Hintergrund der Szene sind Kinderjubil, Motorradgebrüll sowie das Schreien der Wachmänner zu hören, und diese Effekte erlauben auditive Rückschlüsse auf die Art und Weise der Bewegung: Es handelt sich um eine hastige Handlung. Die nächste Folge (2) besteht aus mehreren kurzen Shots, die zuerst rennende Kinder in Nahaufnahmen zeigen und dann stufenweise den größeren Umfang durch Großaufnahmen sichtbar machen. Bei den Nahaufnahmen bleibt die Bewegungsrichtung des Rennens zunächst verborgen, und dem wird die deutsche Audiodeskription dadurch gerecht, dass zuerst mit dem Vollverb auf die Art und Weise der Bewegung hingewiesen wird und danach mit dem Satelliten auf die Bewegungsrichtung

(laufen gehetzt über...). Die spanische Audiodeskription verwendet zwar auch eine Satellitenkonstruktion, nimmt die Bewegungsbeschreibung aber umgekehrt vor: zuerst die Richtung und dann die Art und Weise (*salen corriendo* ‚gehen rennend weg‘). Eine umgekehrte Konstruktion ‚*corren saliendo*‘ wäre im Rahmen des korrekten Sprachgebrauchs kaum möglich (eine Google-Suche zu dieser Konstruktion ergibt lediglich 749 Treffer, wohingegen die Konstruktion *salen corriendo* sogar auf 628 000 Treffer ankommt). Die englische Audiodeskription hat sich ebenso für die Versprachlichung der Bewegungsrichtung mit der Bezeichnung *disperse*, ‚auseinandertreiben‘ entschieden.

Die Bilderfolge 3–5 fängt wiederum mit Großaufnahmen an, die das Hinüberrennen der Kinder auf eine Müllhalde aufzeigen. Danach erscheinen im schnellen Tempo kurze Shots, die das Vorgehen der Kinder im Slum aus unterschiedlichen Perspektiven, meistens in Nahaufnahme und mit Kameranachschwenk abbilden. In den Versprachlichungen sind nun bedeutende Unterschiede zu beobachten. Die deutsche Audiodeskription beschreibt diese Handlung weiterhin mit *Manner*-Verben: zuerst als *flüchten*, dann als *rennen* und schließlich noch als *hetzen*. Interessant ist, dass *flüchten* sowohl die Bewegungsrichtung als auch die Art und Weise der Bewegung enkodiert (laut Duden Online (2020) bedeutet es das (plötzliche und sehr eilige) Fliehen). Die spanische Audiodeskription verwendet erneut Verben, die die Bewegungsrichtung andeuten: *atraviesan*, ‚überqueren‘, (*se internan* ‚dringen ein‘, *saltan*, ‚springen‘). Die englischsprachige Audiodeskription weist mit der Verbwahl ebenso auf einen Bewegungsablauf besonderer Art, und zwar wird das Verb *clamber*, ‚mit Mühe klettern‘ verwendet (laut Oxford Dictionary of English (2020): „climb or move in an awkward and laborious way, typically using both hands and feet“).

Bemerkenswert ist zudem, dass die verschiedenen Sprachversionen sich in der syntaktischen Struktur der Beschreibungen sehr stark ähneln. Das Basismuster ist die prototypische Satzkonstruktion: Subjekt (Agent) + Verb (Handlung) + Präpositionalphrase (Lokation). Dies entspricht der typischen Satzstruktur sowohl bei manueller (Audiodeskription) als auch bei automatischer (Videodeskription) Beschreibung.

4. Diskussion

Die vorgestellten Anwendungsbereiche der sprachlichen Beschreibung sowie die Ergebnisse der Fallanalyse legen viele interessante Beobachtungen und Fragestellungen in Bezug auf die Rollen der Sprache(n) als Zugang nahe.

Einer der wichtigsten Punkte betrifft die Frage, was Sprache oder Sprachen als Zugang bzw. Ersatz zum audio-visuellen Inhalt überhaupt vermitteln können. Die Versprachlichung ist stets, wegen ihres abstrahierenden Charakters, zwangsläufig eine Selektion dessen, was Menschen wahrnehmen. Zudem treffen verschiedene Sprachen diese Auswahl teilweise unterschiedlich: Sie differieren in der Wahl der Aspekte ‚Bewegungsrichtung‘ oder ‚Art und Weise‘. Wenn sich diese Selektion in der Lexikalisierung auf dem Wortniveau manifestiert, sodass beispielsweise die Verbsemantik bestimmte Aspekte auslöst, hat dies eine praktische Bedeutung z.B. für die Verwendung der Sprachen. Man könnte sogar behaupten: Je simpler die sprachliche Konstruktion, desto anpassungsfähiger ist sie in Bereichen, wo eine Kürze des Ausdrucks von Vorteil ist (z.B. in Metadaten oder in der Audiodeskription). Insgesamt tendieren die Sprachen, die die Bewegungsrichtung in Satelliten ausdrücken, zu einer höheren lexikalischen Variation von Bewegungsart (Slobin u.a. 2014: 704), was sich wiederum auf die Wahl von sprachlichen Beschreibungen auswirken mag. Das Deutsche soll von den *manner*-Verb-Sprachen wegen seiner deutlichen Prominenz von Ausdrücken von Bewegungsart sogar hervorstechen (De Knop/Gallez 2011).

Weil Sprachen diese vielfältigen Zugänge zum Visuellen bieten, könnten insbesondere die mehrsprachigen Varianten künftig zugunsten einer besseren Versprachlichung eingesetzt werden. Durch die Anwendung mehrerer Sprachen in der Beschreibung desselben Materials können gleichzeitig unterschiedliche Aspekte des Visuellen hervorgehoben werden; die in sich unterschiedlichen Varianten stellen zusammen eine reiche Beschreibung dar und ergänzen sich gegenseitig (siehe auch Aafaq u.a. 2019: 19–20), was bei der Disambiguierung des visuellen Inhalts behilflich sein könnte. Die Berücksichtigung der intersubjektiven Variation unter den Menschen könnte dennoch als Grundlage für menschenähnliches Repräsentieren dienen, was als ein wichtiger Zweck für die künstliche Intelligenz gilt.

Der Ansatz der Mehrsprachigkeit schließt sich der Diskussion über die Machtposition des Englischen als gängige Maschinensprache an. In der Forschung zu automatischen Versprachlichungsmethoden wird die Tatsache, dass Menschen unterschiedliche Sprachen verwenden, noch gering behandelt, obgleich die Forschungsrichtung des maschinellen Paraphrasierens (*machine paraphrasing*) den Zweck hat, einen Sachverhalt in möglichst vielen Weisen maschinell versprachlichen zu können (Chen/Dolan 2011). Das Thema tritt auch im Kontext der multimodalen maschinellen Übersetzung auf, implizit auch in der Diskussion über die Datensätze, die Beschreibungen in unterschiedlichen Sprachen beinhalten können (Mogadala u.a. 2019: 26–29). In der maschinellen Versprachlichung von Bildern und Videos herrscht das Englische vor; andere verwendete Sprachen sind Japanisch, Deutsch, Französisch und Tschechisch (die beiden zuletzt erwähnten wurden aus dem Englischen oder Deutschen übersetzt) (ebd.). Für die Entwicklung der Videobeschreibung steht ein multilingualer Datensatz zur Verfügung, der mit Beschreibungen in vielen Sprachen ausgestattet ist (ebd. 29). Es wird demnach schon eingesehen, dass verschiedene Sprachen benötigt werden, um die Versprachlichung verwendbarer und sinnvoller machen zu können, aber die Mehrsprachigkeit oder die kontrastive Perspektive darauf, inwiefern die Eigenschaften verschiedener Sprachen als Bedeutungsträger und -mittel funktionieren, ist noch kein Gegenstand technischer Forschung. Implizit lässt sich daraus schlussfolgern, dass die Mehrsprachigkeit zu den Herausforderungen des maschinellen Lernens gehört, auch wenn dieser Forschungsbereich sich vorerst mit dem allgemeinen Sprachgebrauch als ein kommunikativer Verstehensprozess hineinfinden muss (siehe Bender/Koller 2020).

Ein Bereich, in dem es bereits kontrastive Untersuchungen zu Bild- oder eher zu Videobeschreibungen gibt, ist die Audiodeskriptionsforschung. Eine der frühen linguistisch-kontrastiven Analysen legten Bourne und Jiménez (2007) vor, in der anhand eines Films lexikalische und syntaktische Eigenschaften sowie Kohärenzmittel der englischen und spanischen Audiodeskription verglichen wurden. Diese Studie hebt unter anderen die Unterschiede in der Versprachlichung von Bewegung und Handlung hervor, die auch für die Fallanalyse in diesem Artikel von Bedeutung sind:

Die spanische Audiodeskription scheint weniger „spezifische“ Verben für die Beschreibung von Handlungen zu verwenden als die englische (va ‚geht‘ – *scampers out* ‚hetzt/trippelt heraus‘), oder das Spanische verbindet das „generelle“ Verb mit einem Adverb, um mehr Spezifität zu gewährleisten (Bourne/Jiménez 2007: 179). Matamala und Rami (2009) haben ihrerseits Unterschiede in der Behandlung von kulturellen Referenzen anhand einer kontrastiven Untersuchung von deutscher und spanischer Audiodeskription belegen können. Dieses Ergebnis führt die Spannweite kontrastiver Analysen vor: Nicht immer liegt das Interesse auf dem Niveau des sprachlichen Systems und dessen Eigenschaften, denn kulturelle Aspekte, die die Sprache widerspiegeln, sind ebenfalls wichtig.

Über die Frage der Darstellungssprache hinaus ist die Frage der *Anwendungssprache* relevant: In welchen Sprachen wollen Nutzer die digitalen Versprachlichungen und Beschreibungen verwenden? In der heutigen Praxis wird mehrsprachig gehandelt: Metadaten und Audiodeskription werden in den jeweiligen Landessprachen erstellt, die maschinelle Videodeskription wird vornehmlich auf Englisch betrieben. Freilich ermöglicht der Dublin Core-Standard, dass verschiedene Sprachen verwendet werden (Pomerantz 2015: 74). Die mehrsprachige Inhaltsbeschreibung desselben Materials wäre demnach durchaus möglich. Ein interessanter Verbindungspunkt zwischen der Erzeugung von neuen digitalen Korpora und den Zugänglichkeitsdienstleistungen kommt aus dem Konzept des ‚Ersatzes‘: In beiden Fällen wird der Ersatz (eine gewisse Translation) nötig, denn sie bedürfen textueller Beschreibungen, um nicht-sprachliche Inhalte zu vermitteln. Es gibt auch eine Wechselwirkung, die künftig durch Forschung erfasst werden muss. Dienstleistungen können neue Daten für Korpora bieten und Korpora ihrerseits können zu Dienstleistungszwecken angewandt werden. Indem das Angebot von Bild- und Videobeschreibungen sich vermehrt, kommen immer mehr als Korpora verwendbare Daten zustande. Oder wenn man bestimmte Datenmengen durch Beschreibungen textuell verstehbar macht, dienen diese möglicherweise als Zugang zum Inhalt für diejenigen, die das Original nicht wahrnehmen können.

Aus der sprachwissenschaftlichen Perspektive wäre nicht nur interessant, sondern auch äußerst wichtig zu untersuchen, auf welchen Modellen der Sprache die maschinelle Versprachlichung beruht und wie die maschinelle Beschreibung mit der menschlichen Beschreibung einhergeht. Bei der menschlichen Beschreibung handelt es sich um eine kommunikative Gattung, die darauf abzielt, ein möglichst genaues Abbild des Sachverhalts beim Hörer zu erzeugen (Stutterheim/Kohlmann 2001: 1279). Hat die maschinelle Beschreibung eine vergleichbare Funktionszuschreibung, die aus der menschlichen Perspektive nötig ist, um kontextsensitive und angemessene Versprachlichungen erzeugen zu können? Das sprachliche Beschreiben dient als Mittel zur Erreichung vieler kommunikativer Ziele und Funktionen. In der Language-Vision-Forschung wird angenommen, dass Beschreibungen in Anweisungen umgesetzt werden können (also die *deskriptive* Funktion in eine *instruktive* umgeformt wird) und dass Roboter dank der beschreibenden Fähigkeit auch eine narrative Fähigkeit besitzen (er könnte auf Grundlage dessen, was er in Bildern sieht, Handlungsabfolgen erkennen und versprachlichen). Die sprachwissenschaftliche Forschung hat jedoch belegt, dass diese Funktionalitäten – oder Grundtypen der Sachverhaltsdarstellung (Stutterheim/Kohlmann 2001: 1279) – unterschiedliche strukturelle, funktionale und kognitive Eigenschaften haben und jeweils eigene Ziele verfolgen. Diese Erkenntnisse sollten in der sprachtechnologischen Forschung zur Versprachlichung berücksichtigt werden.

Ziel dieses Aufsatzes war es, die Funktion der Sprache als Zugang zu digitalen Ressourcen zu diskutieren und für einen interdisziplinären Forschungsbedarf zu plädieren. Die menschliche Versprachlichung einerseits und die maschinelle Versprachlichung andererseits sind schon längst erforscht worden. Nun ist es höchste Zeit, dass sie voneinander lernen und eine notwendige Zusammenarbeit beginnen, um die gebrauchstaugliche Digitalisierung zu fördern, eine humanistische künstliche Intelligenz zu entwickeln und neue Wege für den digitalen Humanismus auszuloten. **N**

XXXX

MAIJA HIRVONEN

TAMPERE

Bibliographie

- AAFAQ, Nayyer u.a. 2019. Video description: A survey of methods, datasets and evaluation metrics (ACM Computing Surveys). Retrieved from: arXiv:1806.00186v1 [cs.CV].
- BEAVERS, John/Levin, Beth/Tham, Shiao Wie 2010. The typology of motion expressions revisited 1. *Journal of Linguistics* 46 (2). 331–377.
- BENDER, Emily M./Koller, Alexander 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. July 5–10, 2020. 5185–5198.
- BENECKE, Bernd 2014. *Audiodeskription als partielle Translation. Modell und Methode*. (mitSprache Translatorische Forschungsbeiträge, Band 4.) Berlin: LIT Verlag.
- BERTHELE, Raphael 2006. *Ort und Weg. Die sprachliche Raumreferenz in Varietäten des Deutschen, Rätoromanischen und Französischen*. Berlin: de Gruyter.
- BOURNE, Julian/Jiménez, Catalina 2007. From the visual to the verbal in two languages: a contrastive analysis of the audio description of *The Hours* in English and Spanish. *Media for All: Subtitling for the deaf, audio description and sign language*, hrsg. von Jorge Díaz-Cintas/Pilar Orero/Aline Remael. Amsterdam: Rodopi. 175–187.
- BRAUN, Sabine/Starr, Kim 2019. Finding the right words: Investigating machine-generated video description quality using a corpus-based approach. *Journal of Audiovisual Translation* 2 (2): 11–35.
- BREDEL, Ursula/Maaß, Christiane 2019. Leichte Sprache. Handbuch für Barrierefreie Kommunikation, hrsg. von Christiane Maaß/Isabel Rink. Berlin: Frank & Timme. 251–272.
- BURCHARDT, Aljoscha/Uszkoreit, Hans 2018. Einleitung. IT für soziale Inklusion. Digitalisierung, künstliche Intelligenz, Zukunft für alle, hrsg. von Aljoscha Burchardt/Hans Uszkoreit. Berlin, Boston: De Gruyter.
- CHEN, David L./Dolan, William B. 2011. Collecting highly parallel data for paraphrase purposes. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon: Association for Computational Linguistics. 190–200.
- CHRISTMANN, Ursula/Groeben, Norbert 2019. Verständlichkeit: die psychologische Perspektive. *Handbuch für Barrierefreie Kommunikation*, hrsg. von Christiane Maaß/Isabel Rink. Berlin: Frank & Timme. 123–146.
- DE KNOP, Sabine/Gallez, Françoise 2011. Manner of motion: A privileged dimension of German expressions. *International Journal of Cognitive Linguistics* (2) 1. 25–40. Duden Online 2020. *Flüchten*. <https://www.duden.de/node/139640/revision/139676> (09.06.2020)
- DOVAL, Irene/Lübke, Barbara (Hrsg.) 2014. *Raumlinguistik und Sprachkontrast. Neue Beiträge zu spatialen Relationen im Deutschen, Englischen und Spanischen*. München:

- Iudicium Verlag. EU-Richtlinie (2016/2102). „Amtsblatt der Europäischen Union 327/1. Richtlinie (EU) 2016/2102 des Europäischen Parlaments und des Rates vom 26. Oktober 2016 über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen“. EU-Richtlinie (2019/882). „Amtsblatt der Europäischen Union 151/70. Richtlinie (EU) 2019/882 des Europäischen Parlaments und des Rates vom 17. April 2019 über die Barrierefreiheitsanforderungen für Produkte und Dienstleistungen“.
- FIX, Ulla (Hrsg.) 2005. *Hörfilm: Bildkompensation durch Sprache. Linguistisch-filmisch-semiotische Untersuchungen zur Leistung der Audiodeskription in Hörfilmen am Beispiel des Films Laura mein Engel aus der Tatort-Reihe*. Berlin: Erich Schmidt.
- GRODAL, Torben 1997. *Moving Pictures. A New Theory of Film Genres, Feelings, and Cognition*. Oxford: Oxford University Press.
- HEPBURN, Alexa 2004. *Crying: notes on description, transcription, and interaction. Research on Language and Social Interaction*. (37) 3. 251–290.
- HIRVONEN, Maija 2014. *Multimodal Representation and Intermodal Similarity – Cues of Space in the Audio Description of Film*. Dissertation. Helsinki: University of Helsinki.
- HIRVONEN, Maija 2018. *Audiodeskription und Sichtdolmetschen: Translation über Sinnesgrenzen. Spielräume der Translation: Sprach- und translationswissenschaftliche Zugänge*, hrsg. von Sigmund Kvam/ Anastasia Parianou/ Jürgen Schopp/Kåre Solfjeld. Münster: Waxmann. 100–124.
- HIRVONEN, Maija/Kinnunen, Tuija (Hrsg.) 2020a. *Saavutettava viestintä: yhdenvertaista yhteiskuntaa edistämässä*. Helsinki: Gaudeamus.
- IBARRETXE-ANTUÑANO, Iraide 2003. What translation tells us about motion: A contrastive study of typologically different languages. *International Journal of English Studies* (3) 2. 153–178.
- KALE, Geetanjali Vinayak/Patil, Varsha Hemant 2016. A study of vision-based human motion recognition and analysis. *International Journal of Ambient Computing and Intelligence* (7) 2. 75–92.
- KOLLAR, Thomas u.a. 2014. Grounding verbs of motion in natural language commands to robots. *EXPERIMENTAL ROBOTICS*, hrsg. von Oussama Khatib/ Vijay Kumar/Gaurav Sukhatme. (Springer Tracts in Advanced Robotics 79). Berlin, Heidelberg: Springer. 31–47.
- KRUSE, Klemens 2019. Einführung: European Accessibility Act. Vortrag in der 6. Sitzung des Expertenkreises, 24.09.2019. <https://www.bundesfachstelle-barrierefreiheit.de/SharedDocs/Downloads/DE/Veroeffentlichungen/vortrag-eea.html?jsessionid=2AACD4D36439D08B895174D3D4C1E852?nn=1108170> (09.06.2020)
- MAASS, Christiane/Rink, Isabel (Hg.) 2019. *Handbuch für Barrierefreie Kommunikation*. Berlin: Frank & Timme.
- MAIR, Michael/Kierans, Ciara 2007. Descriptions as data: developing techniques to elicit descriptive materials in social research. *Visual Studies* (22) 2. 120–136.
- MATAMALA, Anna/Rami, Naila 2009. Análisis comparativo de la audiodescripción española y alemana

- de “Good-bye Lenin”. *Hermenéus, Revista de Traducción e Interpretación* 11. 1–13.
- MATTHIESEN, Aaron J. 2017. *Maschinelle Übersetzung im Wandel. Die Auswirkungen von künstlicher Intelligenz auf maschinelle Übersetzungssysteme mit einer vergleichenden Untersuchung von Google Translate und Microsoft Translator*. Berlin: epubli.
- MOGADALA, Aditya/Kalimuthu, Marimuthu/Klakow, Dietrich 2019. Trends in integration of vision and language research: A survey of tasks, datasets, and methods. Retrieved from: arXiv:1907.09358 [cs.CV].
- MONDADA, Lorenza 2011. Géographies mobiles et divergentes: l'établissement interactionnel des lieux d'intervention dans des appels au secours, *Revue d'anthropologie des connaissances* (5) 2. 390-436.
- MONDADA, Lorenza 2013. Displaying, contesting and negotiating epistemic authority in social interaction: Descriptions and questions in guided visits. *Discourse Studies* (15) 5. 597–626. Oxford Dictionary of English (2020). Clamber. Abgerufen von Kielikone Ltd.
- POMERANTZ, Jeffrey 2015. *Metadata*. (The MIT Press Essential Knowledge Series). Cambridge, MA: The MIT Press.
- PUJOL, Joaquim/Orero, Pilar 2007. Audio description precursors: Ekphrasis, film narrators and radio journalists. *Translation Watch Quarterly* 3 (2). 49–60.
- REISS, Katharina/Vermeer, Hans J. 1984. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Max Niemeyer.
- RISKU, Hanna/Pircher, Richard 2008. Visual aspects of intercultural technical communication: A cognitive scientific and semiotic point of view. *Meta: journal des traducteurs / Meta: Translators' Journal* (53) 1. 154–166.
- SCHMALZ, Antonia 2019. *Maschinelle Übersetzung. Künstliche Intelligenz: Technologie, Anwendung, Gesellschaft*, hrsg. von Volker Wittpahl. (iit-Themenband). Berlin, Heidelberg: Springer Vieweg. 194–211.
- SCHUBERT, Christoph 2009. *Raumkonstitution durch Sprache. Blickführung, Bildschemata und Kohäsion in Deskriptionssequenzen englischer Texte*. Tübingen: Max Niemeyer.
- SCHWARTZ, Howard 2002. Data: who needs it? Describing normal environments. Examples and Methods. *Ethnographic Studies* 7. 7–32.
- SICILIANO Bruno/Khatib Oussama 2016. *Robotics and the Handbook. Springer Handbook of Robotics. 2nd Edition*, hrsg. von Bruno Siciliano/Oussama Khatib. Cham: Springer International Publishing. 1–6.
- SLOBIN, Dan 2004. The many ways to search for a frog. Linguistic typology and the expression of motion events. *Relating events in narrative, volume 2: Typological and contextual perspectives*, hrsg. von Sven Strömquist/Ludo Verhoeven. New York/London: Psychology Press. 219–257.
- SLOBIN, Dan, Ibarretxe-Atuñano Iraide, Kopecka Anetta & Majid Asifa (2014) *Manners of human gait: A cross-linguistic event-naming study. Cognitive linguistics*. 25 (4), 701–741.
- STOLOVA, Natalya I. 2014. *Cognitive Linguistics and Lexical Change: Motion Verbs from Latin to Romance*. Amsterdam: John Benjamins.

- STUTTERHEIM, Christiane von/Kohlmann, Ute 2001. Beschreiben im Gespräch. *Text- und Gesprächslinguistik / Linguistics of Text and Conversation (Halbbd. 2 / Vol. 1)*, hrsg. von Klaus Brinker u.a. Berlin: de Gruyter. 1279–1292.
- TALMY, Leonard 2000. *Toward a cognitive semantics (vol. 1): Concept structuring systems*. Massachusetts, London: MIT Press.
- UNESCO 2010–2017. World Day for Audiovisual Heritage. Archivierte Internetseite. <https://wayback.archive-it.org/10611/20170509024345/http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/archives/world-day-for-audiovisual-heritage/world-day-for-audiovisual-heritage-2010/> (09.06.2020)
- VANDAELE, Jeroen 2012. What meets the eye. Cognitive narratology for audio description. *Perspectives: Studies in Translatology* 20 (1): 87–102.