



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Kohärenz der Zeilensegmentierung in Untertiteln: Eine empirische Studie zur
Verständlichkeit und Rezeption

verfasst von | submitted by

Andrea Tuswald BA BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 070 331 345

Studienrichtung lt. Studienblatt | Degree pro-
gramme as it appears on the student record
sheet:

Masterstudium Translation Deutsch Französisch

Betreut von | Supervisor:

Univ.-Prof. Dr. Hanna Risku

Inhalt

1 Einleitung	3
2 Die audiovisuelle Übersetzung.....	5
2.1 Geschichte der audiovisuellen Übersetzung	7
2.2 Produktions- und Rahmenbedingungen der audiovisuellen Übersetzung.....	11
2.3 Audiovisuelle Übersetzungsformen	13
3 Theoretische Grundlagen der Untertitelrezeption.....	24
3.1 Multimodale Verarbeitung und semiotische Kohäsion.....	24
3.2 Kognitive Belastung bei der Untertitelrezeption.....	25
3.3 Kohärenz, Kohäsion und Verständlichkeit.....	26
3.4 Erstsprache und Untertitelrezeption	27
4 Studien zur Zeilensegmentierung.....	28
4.1 Die kognitive Effizienz der Untertitelverarbeitung.....	29
4.2 Auswirkungen der Gliederung von Texteinheiten auf die Untertitelung.....	31
4.3 Eine Eye-Tracking-Studie über die Präferenzen der Zuseher:innen.....	33
4.4 Der Einfluss der Textsegmentierung auf das Lesen von Untertiteln	35
4.5 Synthese	36
5 Empirische Untersuchung	38
5.1 Forschungsdesign.....	39
5.2 Forschungsmethode.....	39
5.3 Untersuchungsmaterial.....	40
5.4 Versuchsgruppe.....	41
5.5 Datenbereinigung	41
5.6 Auswertungsverfahren	42
6 Ergebnisse.....	42
6.1 Hauptanalyse	43
6.2 Robustheitsprüfung der Ergebnisse.....	44
6.3 Explorative Ergebnisse nach Erstsprache	45
7 Diskussion	47
8 Fazit.....	49
Literaturverzeichnis.....	51
Abbildungs- und Tabellenverzeichnis.....	54
Anhang.....	55

1 Einleitung

Im digitalen Zeitalter sind audiovisuelle Produktionen fester Bestandteil unseres Medienkonsums und werden über sprachliche und kulturelle Grenzen hinweg genutzt. Dadurch steigt auch der Bedarf an Übersetzungen. Untertitel ermöglichen den Zugang zu fremdsprachigen Inhalten und haben insbesondere mit der Verbreitung von Streaming-Plattformen an Bedeutung gewonnen. Im Unterschied zur Synchronisation bleibt hier das Originalaudio erhalten, während die verbalen Informationen parallel schriftlich vermittelt werden. Es handelt sich somit um eine Form der audiovisuellen Übersetzung, bei der die sprachliche, zeitliche und visuelle Ebene miteinander in Einklang gebracht werden müssen. Eine zentrale Herausforderung liegt dabei in der Strukturierung der Untertitel, vor allem in der Segmentierung des Textes über mehrere Zeilen.

Die Forschung im Bereich der Untertitelung konzentrierte sich zunächst auf deren Produktion und Inhalt, doch mit der Zeit interessierten sich Forscher:innen auch immer mehr für die Rezeption durch das Publikum. Auch die vorliegende Arbeit beschäftigt sich mit der Wahrnehmung von Untertiteln, konkret mit Bezug auf Zeilenumbrüche innerhalb von Untertiteln. Üblicherweise bestehen Untertitel aus maximal zwei Zeilen und müssen am Zeilenende getrennt werden. Gängige Untertitelrichtlinien gehen davon aus, dass eine syntaktisch kohärente Segmentierung, bei der syntaktische und semantische Einheiten nicht auf zwei Zeilen verteilt werden, den Lesevorgang und das Verständnis erleichtert. Im Gegensatz dazu wird angenommen, dass sich eine inkohärente Segmentierung störend auf den Lesefluss auswirken und zu einer erschwerten Verarbeitung des Textes führen kann, was das Seherlebnis beeinträchtigen könnte.

Dieses Thema war bereits Gegenstand mehrerer psycholinguistischer Studien, die zwar Tendenzen, jedoch keine einheitlichen Ergebnisse erkennen lassen. Zwar existieren empirische Studien zu den Effekten von Untertitelsegmentierung in englisch- und romanischsprachigen Kontexten (vgl. z. B. Szarkowska et al. 2018; Perego et al. 2010), doch standen experimentelle Untersuchungen, die sich explizit mit der Rezeption und dem Vergleich von syntaktisch kohärent und inkohärent segmentierten deutschsprachigen Untertiteln beschäftigen, bisher nicht im Mittelpunkt der Forschung.

Ein weiterer bislang wenig erforschter Aspekt betrifft mögliche Unterschiede zwischen Erstsprachler:innen und Nicht-Erstsprachler:innen bei der Rezeption von Untertiteln. Die vorliegende Arbeit berücksichtigt diesen Faktor explorativ, um erste Hinweise auf mögliche gruppenspezifische Effekte zu erhalten. Daraus ergibt sich die zentrale Fragestellung dieser Arbeit: Inwiefern beeinflusst die Segmentierung deutschsprachiger

Untertitel die Wahrnehmung, Verständlichkeit und den Lesefluss bei Erstsprachler:innen und Nicht-Erstsprachler:innen? Zur Beantwortung dieser Fragestellung kommt ein standardisierter Fragebogen zum Einsatz, mit dem die subjektiven Einschätzungen der Teilnehmenden zur Rezeption der Untertitel erfasst werden. Die gewonnenen Ergebnisse werden anschließend im Kontext bisheriger empirischer Forschung diskutiert.

Die Arbeit gliedert sich wie folgt: Zunächst werden in Kapitel 2 ein Überblick über die audiovisuelle Übersetzung gegeben, die Untertitelung darin eingebettet und deren Normen und Richtlinien erläutert, einschließlich der Zeilensegmentierung. Kapitel 3 behandelt theoretische Grundlagen der Rezeption und Verarbeitung von Untertiteln. Darauf aufbauend wird in Kapitel 4 auf die bisherige Forschung zur Segmentierung von Untertiteln eingegangen. Kapitel 5 beschreibt die empirische Untersuchung und beinhaltet das Forschungsdesign, das Untersuchungsmaterial, die Erhebungsmethode und die Stichprobe. In Kapitel 6 werden die Ergebnisse der Befragung dargestellt und ausgewertet, bevor sie in Kapitel 7 interpretiert und im Hinblick auf ihre Relevanz für die Untertitelpraxis eingeordnet werden.

2 Die audiovisuelle Übersetzung

Nachdem für diese Form der Übersetzung und ihre Nuancen im Lauf der Zeit zahlreiche terminologische Begriffe geprägt wurden, hat sich inzwischen der sprachübergreifend verwendete Ausdruck *audiovisuelle Übersetzung* als Standard durchgesetzt. Diese Bezeichnung schließt auch die semiotische Dimension des zu übersetzenden Produkts mit ein und dient als Überbegriff für die verschiedenen im audiovisuellen Bereich angewandten Übersetzungsverfahren (vgl. Díaz-Cintas & Remael 2021: 6). Im Deutschen ist auch der Begriff *Medienübersetzung* (Jüngst 2020: 1) in Gebrauch, wenn nicht direkt die einzelnen Verfahren genannt werden (vgl. Jüngst 2020: 1).

In ihrer heutigen Auffassung umfasst die audiovisuelle Übersetzung jene Praktiken, Prozesse und Produkte, die mit der Übertragung multimodaler und multimedialer Inhalte zwischen Sprachen und/oder Kulturen in Zusammenhang stehen (vgl. Pérez-González 2020: 30). Eine reine Orientierung am Text reicht nicht aus, da alle Bedeutungsebenen berücksichtigt werden müssen. Dieses Spannungsfeld zeigt die Komplexität audiovisueller Übersetzung und wurde von Titford (1982) als *constrained translation* bezeichnet, „in which the text is only one of the components of the message or when it constitutes only an intermediate stage for a speech read aloud or dramatized“ (Mayoral et al. 1988: 356). Dieses Konzept wurde von Mayoral et al. (1988: 356) aufgegriffen, die erklärten, dass die Tätigkeit von Übersetzer:innen durch Bilder, Musik und gesprochene Sprache komplizierter und eingeschränkt wird. Auch Díaz-Cintas verwendet 1998 für die Untertitelung noch den Begriff *traducción subordinada* (vgl. Díaz Cintas 1998), weist jedoch in einer späteren Arbeit darauf hin, dass die Bezeichnungen *subordinate* und *constrained translation* bald aufgrund ihrer einschränkenden Konnotationen kritisiert wurden. Ebenso wurde *film translation* problematisch, weil sich das Forschungsgebiet auf andere Programme und Medien ausdehnte (vgl. Díaz-Cintas & Remael 2021: 6).

Darüber hinaus erfüllt die audiovisuelle Übersetzung eine wichtige inklusive Funktion. Sie ermöglicht auch Menschen mit Sinnesbeeinträchtigungen Zugang zu audiovisuellem Material und umfasst Gebärdensprachdolmetschen, Untertitelungen für taube und hörgeschädigte Menschen sowie Audiodeskriptionen für blinde und sehbehinderte Personen (vgl. Díaz-Cintas & Baños 2017: 313) und auch die immer wichtigere „Leichte Sprache“ (Jüngst 2020: 2).

Innerhalb der Übersetzungswissenschaft gilt die audiovisuelle Übersetzung als vergleichsweise junges Forschungsfeld, das sich inzwischen als eigenständige Disziplin

etabliert hat. Wie Díaz-Cintas & Baños (2017: 321) feststellen, fokussierten sich frühe Studien auf ihre Besonderheiten und Eigenständigkeit, vor allem auf die Analyse der Funktion der Sprache, die Herausforderungen beim Sprachtransfer und die Übersetzungsstrategien zu deren Bewältigung. Die experimentelle Forschung auf der Grundlage empirischer Untersuchungen ist jedoch eine relativ neue Entwicklung in der audiovisuellen Übersetzung. Insgesamt hat die Zahl experimenteller Studien in dem Bereich stark zugenommen, und gerade auch in der Untertitelung werden immer spezifischere Fragestellungen betrachtet, die sich mit anderen Disziplinen überschneiden (vgl. Díaz-Cintas & Baños 2017: 321). Einer der behandelten Aspekte ist die Wahrnehmung von Untertiteln, die auch für die vorliegende Arbeit eine zentrale Rolle spielt.

Anders als bei anderen Formen der Übersetzung wird das Original bei audiovisuellen Inhalten adaptiert und meist ergänzt. Je nach Übersetzungsverfahren bleiben unterschiedliche Teile des ursprünglichen Materials erhalten (vgl. Jüngst 2020: 1). Die Tatsache, dass die Originalversion weiterhin für das Zielpublikum sichtbar und in Abhängigkeit von der Übersetzungsform auch hörbar ist, schränkt den Spielraum der audiovisuellen Übersetzung allerdings ein. Filmische Dialoge sind stets in eine konkrete situative und visuelle Umgebung eingebettet, die für die Rezipient:innen des Ausgangstextes dieselbe ist wie für jene des Zieltextes. Dadurch ergeben sich spezifische Anforderungen an die Übersetzung, und die alleinige Berücksichtigung der linguistischen Dimension reicht nicht aus (vgl. Díaz-Cintas 2009: 9).

Die gesellschaftliche Bedeutung der audiovisuellen Übersetzung hat sich zunehmend erweitert – bezüglich der Zahl der Nutzer:innen, der Art der Verwendung von audiovisuellen Produkten sowie der Programme, die übersetzt werden. (vgl. Díaz-Cintas & Remael 2021: 3) Ursprünglich war die audiovisuelle Übersetzung ein kommerzielles Verfahren, um Spielfilmen eine größere Reichweite zu verschaffen. Heute ist die Situation eine andere, denn die Bandbreite der audiovisuellen Genres, die übersetzt werden, ist praktisch unbegrenzt – sei es für kommerzielle, didaktische oder Unterhaltungszwecke. Dieser Wandel zeigt sich besonders deutlich im digitalen Medienumfeld. Auch Zeitungen haben durch die digitale Revolution eine wesentliche Neuausrichtung erfahren: Sie haben sich eine Online-Präsenz aufgebaut und stellen dort Videos bereit, die in der Regel anhand von Untertiteln übersetzt werden, sofern eine Sprachübertragung erforderlich ist (vgl. Díaz-Cintas & Remael 2021: 3).

2.1 Geschichte der audiovisuellen Übersetzung

In diesem Kapitel soll ein Überblick über die historische Entwicklung der audiovisuellen Übersetzung gegeben werden und die wichtigsten Etappen ihrer wissenschaftlichen und technologischen Etablierung nachzeichnen. Forschung und Praxis haben sich von den frühen Anfängen bis zur Digitalisierung gegenseitig beeinflusst. Diese Veränderungen bilden den Hintergrund für das heutige Verständnis der Untertitelung sowie ihrer formalen und funktionalen Eigenschaften.

2.1.1 Anfänge und Etablierung

Obwohl die audiovisuelle Übersetzung bereits seit den Anfängen des Kinos existiert, erlebte die Forschung nach einem holprigen Start in den späten 1950er-Jahren erst Ende des 20. Jahrhunderts einen beachtenswerten Aufschwung (vgl. Díaz-Cintas 2009: 1). Als Grund für das geringe wissenschaftliche Interesse sah man die strengen formalen Parameter der Untertiteltextualität (vgl. Díaz-Cintas & Remael 2021: 4).

Im Jahr 1957 wurde erstmals ein Werk über die Untertitelung veröffentlicht: *Le sous-titrage de films – sa technique, son esthétique* von Simon Laks. Doch in den 1960er- und 1970er-Jahren herrschte in diesem Bereich Stillstand (vgl. Díaz-Cintas 2009: 2). In den 1980er-Jahren erschienen erste Arbeiten, die technische und kognitive Herausforderungen der Untertitelung systematisch diskutierten und zentrale theoretische Konzepte einführten (siehe z. B. Marleau 1982; Titford 1982).

Die erste *Conference on Dubbing and Subtitling*, die 1987 in Stockholm stattfand, löste ein bis dahin beispielloses Interesse an der audiovisuellen Übersetzung aus. In dieser Phase folgten grundlegende Handbücher und theoretische Arbeiten, die zur Institutionalisierung des Forschungsfelds beitrugen. Ein zentrales Werk dieser Periode ist Ivarssons *Subtitling for the Media: A Handbook of an Art* (1992; siehe exemplarisch auch Delabastita 1989).

Das Goldene Zeitalter der audiovisuellen Übersetzung waren schließlich die 1990er-Jahre, in denen systematischer und mit Fokus auf die translatorische Perspektive geforscht wurde. In dieser Zeit erschienen sowohl Sammelbände (z. B. Gambier 1996; Gambier & Gottlieb 2001) als auch grundlegende Arbeiten zur Untertitelung von Henrik Gottlieb, darunter *Subtitling: Diagonal Translation* (1994) und *Subtitles, Translation & Idioms* (1997). Die audiovisuelle Übersetzung konnte sich als Forschungsfeld etablieren; sie wurde zum Gegenstand zahlreicher wissenschaftlicher Beiträge und fand Eingang in die Lehrpläne von

Universitäten (vgl. Díaz-Cintas 2009: 3). Seit der Jahrhundertwende wurde zudem die Untersuchung und Funktion der audiovisuellen Übersetzung in der Unterstützung von hör- oder sehbeeinträchtigten Personen stark ausgeweitet, was zum Teil auf die technologische Weiterentwicklung, aber auch auf gesetzliche Bestimmungen zurückzuführen ist (vgl. Pérez-González 2020: 30).

2.1.2 Entwicklung zum eigenständigen Forschungsfeld

Zu den Anfängen der Filmindustrie erschienen Untertitel erstmals in Stummfilmen als Zwischentitel (vgl. Georgakopoulou 2012: 79). Um einen Film im Ausland zu zeigen, genügte es, die Zwischentitel zu entfernen und mit Texten in der Zielsprache zu ersetzen (vgl. Pérez-González 2020: 30). US-amerikanische Filme dominierten in dieser Zeit den europäischen Markt und begünstigten so die Verbreitung übersetzter Filmfassungen. Mit dem Aufkommen des Tons gegen Ende der 1920er-Jahre entstand beim Publikum eine Nachfrage nach Filmen in der jeweiligen Erstsprache, die von der amerikanischen Industrie nicht so einfach bedient werden konnte, da die Synchronisation erst ab Mitte der 1930er-Jahre angewandt wurde (vgl. Novell-Smith 1998: 3).

In der Ära des Tonfilms in den 1930er-Jahren nahm die audiovisuelle Übersetzung erstmals wirklich Gestalt an (vgl. Georgakopoulou 2012: 79). In den 1940er-Jahren wurde sie schließlich zu einem Nebenprodukt der Filmproduktion, da es möglich wurde, Bild und Ton zu trennen, und es sich somit finanziell rentierte, Filme für ein ausländisches Publikum zu übersetzen. Daraus gingen Untertitelung und Synchronisation als die bis heute wichtigsten Methoden der audiovisuellen Sprachübertragung hervor (vgl. Georgakopoulou 2019: 516). Im Laufe der Zeit etablierten sich je nach Land unterschiedliche Präferenzen für eine der beiden Formen (vgl. Georgakopoulou 2012: 79).

Die 1970er-Jahre markierten einen wichtigen Meilenstein der audiovisuellen Übersetzung, als die ersten Sendungen mit offenen Untertiteln im amerikanischen Fernsehen ausgestrahlt wurden. Im Vereinigten Königreich wurde zudem der Teletext erfunden, der später von der Europäischen Rundfunkunion als internationaler Standard übernommen wurde. Live-Untertitelung wurde ab den 1980er-Jahren technisch möglich und entwickelte sich hauptsächlich aus Gründen der Barrierefreiheit zu einem Standard (vgl. Georgakopoulou 2019: 516f.).

Die Unterhaltungsindustrie veränderte sich durch neue Technologien in den 1980er-Jahren. Bereits in den späten 1970er-Jahren trug die Einführung der VHS zur schnellen Verbreitung von Heimvideos in der ganzen Welt bei, und die Zunahme von Kabel- und

Satellitenfernsehen resultierte schließlich in einer größeren Programmauswahl und damit auch in einem höheren Bedarf an audiovisuellen Übersetzungen. Zeitgleich revolutionierte der Einzug des Desktop-Computers den Übersetzungsprozess, gerade im Bereich der audiovisuellen Übersetzung. Dies ermöglichte eine weitgehende Eigenständigkeit im Untertitelungsprozess und trug zur Professionalisierung des Berufsbilds bei (vgl. Georgakopoulou 2019: 517).

Fortschritte in der Untertitelungssoftware vereinfachten den Produktionsprozess erheblich, und mit dem DVD-Boom zu Beginn des 21. Jahrhunderts stieg der Bedarf an mehrsprachigen Untertiteln stark an (vgl. Georgakopoulou 2019: 517; 2012: 81). Gleichzeitig prägten staatliche Regulierungen und protektionistische Maßnahmen über Jahrzehnte die Verbreitung audiovisueller Übersetzungsmethoden, bevor sich diese Strukturen mit der fortschreitenden Digitalisierung grundlegend wandelten (vgl. Pérez-González 2020: 31).

2.1.3 Digitalisierung und aktuelle Bedeutung

Der technologische Fortschritt hatte stets große Auswirkungen darauf, wie mit der Übersetzung und Verbreitung audiovisueller Produktionen umgegangen wird. Die Ablösung der analogen durch die digitale Technologie um die Jahrtausendwende erwies sich dabei jedoch als eine der folgenreichsten Umbrüche und richtungsweisend für die weitere Ausrichtung (vgl. Díaz-Cintas & Remael 2021: 2). Diese technologische Umstellung spiegelte sich insbesondere in der Entstehung spezialisierter Untertitelungssoftware wider.

Die ersten Programme für professionelle Untertitler:innen kamen in der zweiten Hälfte der 1970er-Jahre auf den Markt und entwickelten sich nach und nach zu den heute verfügbaren Versionen weiter (vgl. Díaz-Cintas & Remael 2021: 48). Mit der Digitalisierung wurden zunehmend cloudbasierte Systeme und Translation-Management-Plattformen eingesetzt, die kollaboratives Arbeiten, Automatisierung und effizientere Projektabläufe ermöglichten (vgl. Georgakopoulou 2019: 519ff.). Moderne Untertitel-Editoren übernehmen zahlreiche technische Arbeitsschritte wie automatisches Timing, Szenenerkennung und Qualitätskontrollen; auch die Textsegmentierung kann softwaregestützt erfolgen, erfordert aber weiterhin die aktive Mitwirkung der Übersetzer:innen (vgl. Díaz-Cintas & Remael 2021: 11, 243).

Das Angebot an Inhalten wurde immer größer und die Untertitelung auch in den traditionellen Synchronisationsländern populärer. Die Branche sah sich mit einer wachsenden

Nachfrage nach mehrsprachigen Untertiteln bei gleichzeitig verkürzten Produktionsfristen konfrontiert. Es entstand ein neuer Arbeitsablauf für die Produktion von Untertiteln – nämlich die Verwendung von Vorlagen in der Ausgangssprache, die in alle für ein Filmprojekt erforderlichen Zielsprachen übersetzt werden – und schon bald bildeten sich internationale Untertitelungsnormen heraus (vgl. Georgakopoulou 2012: 80f.).

Eine weitere Transformation in Richtung einer globalisierten Lokalisierungsbranche erfuhr die Übersetzungsindustrie durch das Breitbandinternet. Mit steigendem Auftragsvolumen begann der Trend zum Outsourcing (vgl. Georgakopoulou 2019: 519). Infolgedessen erhöhte sich der Druck zur Senkung der Produktionskosten (vgl. Georgakopoulou 2012: 88). Gleichzeitig wandelte sich der Medienkonsum mit dem Aufkommen von DVDs und online verfügbaren Filmen und Serien (vgl. Díaz-Cintas 2009: 3).

Die unmittelbare Verfügbarkeit und die globale Reichweite von audiovisuellen Inhalten haben die Homogenität des nationalen Publikums und die interne Kohärenz der Märkte aufgeweicht, was wiederum dazu geführt hat, dass die Dominanz spezifischer Übersetzungsmodalitäten innerhalb einzelner Länder zu erodieren begann (vgl. Pérez-González 2019: 2). Zugleich verschwimmen zunehmend die Grenzen zwischen Produzent:innen und Rezipient:innen audiovisueller Inhalte.

Mit den digitalen Technologien veränderte sich auch die Art und Weise, wie audiovisuelle Produkte konsumiert werden: Während sie zunächst im öffentlichen Kinoraum angesehen wurden und sich später in den privaten Fernsehbereich verlagerten, erfolgt die Nutzung heute vermehrt individuell über mobile Geräte. Diese Erwartungen nach orts- und zeitunabhängigem Zugang werden von Video-on-Demand-Diensten erfüllt (vgl. Díaz-Cintas & Baños 2017: 315). Damit verbunden sind auch die rasche Verfügbarkeit von Untertiteln sowie die Möglichkeit, diese benutzer:innendefiniert anzupassen.

Georgakopoulou (2019: 516) wies bereits darauf hin, dass sich die audiovisuelle Übersetzung immer stärker in Richtung Cloud-Technologien, Crowdsourcing und datengetriebener Verfahren bewegen werde. Da sie als Nebenprodukt der Neuerungen in Film-, Video- und Rundfunktechnologie entstanden ist, wird sie sich auch in Zukunft parallel dazu entwickeln.

2.2 Produktions- und Rahmenbedingungen der audiovisuellen Übersetzung

Dieses Kapitel erläutert die zentralen Produktions- und Rahmenbedingungen der audiovisuellen Übersetzung, die für die Entstehung und Gestaltung von Untertiteln eine wesentliche Rolle spielen. Neben marktbedingten Faktoren und der Wahl der Übersetzungsmethode sind insbesondere organisatorische und professionelle Rahmenbedingungen relevant. Sie bilden den Kontext, innerhalb dessen Entscheidungen über Form und Struktur von Untertiteln getroffen werden, und sind daher für das Verständnis der praktischen Untertitelungsarbeit von Bedeutung.

2.2.1 Markt und Methoden

Die Medienpolitik beeinflusst maßgeblich die Auswahl der Produktionen, die einem bestimmten Publikum zugänglich gemacht werden. Da auf internationaler Ebene englischsprachige Filme und Serien dominieren, werden diese von Entscheidungsträger:innen aus kleineren Sprachgemeinschaften bevorzugt, weil es ein Risiko darstellt, Inhalte zu erwerben und zu übersetzen, mit deren sprachlichem und kulturellem Hintergrund die Zielgruppe weniger vertraut ist (vgl. Gottlieb 2018: 325).

Ein weiterer Aspekt ist die Frage nach der Übersetzungsmethode, wobei grundsätzlich Synchronisation und Untertitelung miteinander konkurrieren. Die Synchronisation gilt als manipulationsanfälliger und weniger anspruchsvoll für die Rezipient:innen (vgl. Gottlieb 2018: 327). In den meisten Ländern ist entweder die Synchronisation oder die Untertitelung vorherrschend, wobei in Ländern mit einer kleinen Sprachgemeinschaft vorwiegend auf die Untertitelung zurückgegriffen wird (vgl. Jüngst 2020: 23), während größere westeuropäische Sprachgemeinschaften synchronisierte Produktionen bevorzugen (vgl. Gottlieb 2018: 327). Doch auch im deutschen Sprachraum werden untertitelte Inhalte nicht mehr als Ausnahme wahrgenommen. Dies lässt sich unter anderem mit der veränderten Mediennutzung erklären, vor allem mit dem abnehmenden Stellenwert des linearen Fernsehens für jüngeres Publikum sowie mit der breiten Verfügbarkeit von Originalfassungen mit Untertiteln auf Streaming-Plattformen und in Mediatheken (vgl. Jüngst 2020: 23).

2.2.2 Professionelle Praxis

In der heutigen Praxis werden Untertitelprojekte häufig von großen Lokalisierungsanbieter:innen koordiniert, wodurch Entscheidungsprozesse weitgehend

zentralisiert sind. Gleichzeitig schränkt die zunehmende Standardisierung der Untertitelungspraxis durch die Dominanz internationaler Unternehmen den individuellen Gestaltungsspielraum der Untertitler:innen ein, etwa durch die Arbeit mit Vorlagen (vgl. Díaz-Cintas & Remael 2021: 53). Die Untertitelübersetzung ist dabei durch enge technische Vorgaben und häufig sehr kurze Produktionsfristen gekennzeichnet. In einer Studie von Szarkowska (vgl. 2016: 18) werden Zeitdruck und fehlende Kontrollen als Faktoren genannt, die sich potenziell negativ auf die Qualität auswirken.

In den letzten Jahren war die Untertitelungspraxis durch den Einsatz von maschineller Übersetzung und neuronalen Netzwerken Veränderungen unterworfen. Während Untertitel früher überwiegend manuell erstellt wurden, funktioniert die Verarbeitung von Sprache heute in Echtzeit (vgl. Pötzsch 2026). Dafür kommen neue Technologien wie etwa Spracherkennungssysteme (Speech-to-Text) zum Einsatz, die Untertitel und Transkripte automatisch generieren (vgl. ORF 2026) und auf deren Basis die Künstliche Intelligenz (KI) einen Rohentwurf erstellt. Im Zuge dieser Entwicklungen entstand das Berufsbild der Post-Editor:innen, die einen solchen Entwurf anschließend überarbeiten, verfeinern und den kulturellen Gegebenheiten des Zielpublikums anpassen (vgl. Pötzsch 2026). Übersetzer:innen müssen dafür keine Untertitelungssoftware mehr erwerben oder vor Ort anwesend sein, weil sich die Arbeit mittlerweile in die Cloud verlagert hat. Im Unterschied zum analogen Zeitalter werden heute oft alle Arbeitsschritte von einer einzigen Person ausgeführt (vgl. Jüngst 2020: 41, 43). Es ist davon auszugehen, dass die veränderten Rahmenbedingungen indirekt auch Entscheidungen bei der Strukturierung und Segmentierung von Untertiteln beeinflussen.

Auch in der Synchronbranche sorgt die KI für Umbrüche. Der Streamingdienstleister Amazon Prime Video hatte im März 2025 angekündigt, Inhalte mittels KI zu lokalisieren, um sie für weitere Märkte zugänglich zu machen. Ein Großteil dieser KI-Synchronfassungen wurde Ende November 2025 wieder von der Plattform entfernt, nachdem zahlreiche Zuseher:innen die Intonation, den emotionalen Ausdruck und die Betonung kritisierten. Ähnlich erging es im Februar 2025 bereits der polnischen Serie *Murderess* auf Magenta TV (vgl. Szymanski 2026). Netflix führte in diesen Zusammenhang kürzlich eine neue Klausel in seine Verträge ein, die besagt, dass die Stimmen der Synchronsprecher:innen für das Training von KI-Programmen verwendet werden dürfen. Diejenigen, die die Zustimmung verweigerten, bekommen kaum noch Aufträge. Der Verband deutscher Sprecher und Sprecherinnen fürchtet dadurch eine Gefährdung der Existenzgrundlage der Synchronsprecher:innen im deutschsprachigen Raum. 2024 kam mit dem chinesischen Spielfilm „Black Dog – Weggefährten“ bereits der erste durch KI auf Deutsch synchronisierte

Film in die Kinos (vgl. Berg 2026). Es bleibt abzuwarten, ob synthetische Stimmen in Zukunft tatsächlich ein Niveau erreichen, auf dem sie nicht mehr von menschlichen Stimmen zu unterscheiden sind und in der Folge auch vom Publikum akzeptiert werden.

2.3 Audiovisuelle Übersetzungsformen

In diesem Kapitel werden zunächst die spezifischen Charakteristika und Herausforderungen von multimedialen bzw. audiovisuellen Texten vorgestellt. Danach wird auf die verschiedenen Übersetzungsformen eingegangen, um die Untertitelung in diesem Kontext verorten zu können und ihre Besonderheiten im Vergleich zu anderen Methoden zu verdeutlichen. Audiovisuelle Übersetzung umfasst mehrere Methoden, deren technische Umsetzung, Rezeptionsbedingungen und angewandte Übersetzungsstrategien variieren.

Audiovisuelle Texte gehören zu den sogenannten multimedialen bzw. multimodalen Texten, bei denen sprachliche Elemente mit visuellen und auditiven Zeichensystemen miteinander interagieren (vgl. Reiss 1981: 125; Snell-Hornby 2006: 84). Díaz-Cintas (2017: 313) unterscheidet auditiv-verbale, auditiv-nonverbale, visuell-verbale und visuell-nonverbale Modi und beschreibt deren Zusammenspiel als eine der zentralen Anforderungen audiovisueller Übersetzung, da ein neues sinnvolles multisemiotisches Ganzes entstehen muss. Untertitel sind dabei in dieses Gefüge eingebunden und müssen mit den übrigen Bedeutungsebenen des Films in Einklang stehen.

Übersetzer:innen sehen sich mit der Herausforderung konfrontiert, dass die verschiedenen semiotischen Codes zusammenwirken, während in der Regel nur der sprachliche Code direkt bearbeitet werden kann (vgl. Díaz-Cintas 2017: 315). Eine reine Orientierung am Text reicht daher nicht aus, da alle Bedeutungsebenen berücksichtigt werden müssen. Dieses Spannungsfeld zeigt die Komplexität audiovisueller Übersetzung und wird häufig als *constrained translation* (vgl. Titford 1982) beschrieben, da die Übersetzung durch technische und semiotische Faktoren eingeschränkt ist.

Aufgrund der Interaktion zahlreicher kommunikativer Elemente ergeben sich Einschränkungen für die Arbeit der Übersetzer:innen, die je nach Genre und Methode variieren – bei der Synchronisation und Untertitelung sind sie oft technischer Natur, weil die Übersetzung mit dem Originaltext abgestimmt werden muss (vgl. Díaz-Cintas 2017: 315). Während bei der Synchronisation zielsprachliche Äußerungen mit den Mundbewegungen der

Schauspieler:innen synchronisiert werden müssen, wird bei der Untertitelung der Text zeitgleich mit dem gesprochenen Originaldialog eingeblendet.

Grundsätzlich lassen sich zwei Herangehensweisen der audiovisuellen Übersetzung unterscheiden: Entweder wird der mündliche Dialog durch eine neue Tonspur in der Zielsprache ersetzt oder ergänzt, oder er wird in schriftlicher Form wiedergegeben, die zeitgleich mit dem ursprünglichen Dialog am Bildschirm erscheint.

2.3.1 Synchronisation, Voiceover und Audiodeskription

Die Synchronisation ersetzt die Dialogspur einer audiovisuellen Produktion vollständig durch eine neu aufgenommene Tonspur in der Zielsprache. Es soll der Eindruck erzeugt werden, die dargestellten Figuren sprächen dieselbe Sprache wie die Zuseher:innen. Dafür müssen Lippensynchronität, Isochronie und kinetische Synchronität gewährleistet sein. Die Methode ist vor allem in großen Sprachräumen wie den deutschsprachigen Ländern, Frankreich, Italien oder Spanien üblich (vgl. Díaz-Cintas & Remael 2021: 8). Bei der Untertitelung wird im Gegensatz dazu der Originalton beibehalten und die Übersetzung in schriftlicher Form eingeblendet. In Österreich sind sowohl internationale Filmproduktionen im Kino als auch TV-Serien standardmäßig auf Deutsch synchronisiert, während Originalversionen (mit oder ohne Untertitel) nur in wenigen Kinos zu sehen sind.

Beim Voiceover wird ein übersetzter Text auf einer zusätzlichen Tonspur über den Originaldialog gelegt, der zunächst für ein bis zwei Sekunden hörbar bleibt und anschließend so leise weiterläuft, dass er zwar noch wahrgenommen, aber nicht verstanden wird. Die Übersetzung ist nicht lippensynchron und endet meist kurz vor dem Originalton (vgl. Jüngst 2020: 138f.). Häufig wird Voiceover für sachliche Inhalte wie Dokumentationen und Interviews eingesetzt und ist insbesondere in einigen osteuropäischen Ländern im Fernsehen verbreitet (vgl. Díaz-Cintas & Remael 2021: 8). Mit der rasanten Zunahme von Medien und online verfügbaren Inhalten hat sich sein Anwendungsbereich mittlerweile unter anderem auch auf Spiel-, Animations- und Unternehmensfilme ausgedehnt, ebenso wie auf Werbung und Videospiele (vgl. Audiobird Support 2023). Wie die Untertitelung orientiert sich das Voiceover an der Dauer des Originaldialogs, wird jedoch mündlich statt schriftlich umgesetzt und greift dadurch stärker in die Tonspur ein.

Die Audiodeskription erweitert audiovisuelle Inhalte um eine zusätzliche Tonspur, in der visuelle Informationen sprachlich beschrieben werden, um sehbehinderten Rezipient:innen das Verständnis der Handlung zu ermöglichen. Die Kommentare umfassen

etwa Gestik, Mimik, das Aussehen von Charakteren und Schauplätzen. Bei Filmen und im Fernsehen wird Audiodeskription normalerweise aufgezeichnet und dem Originalton hinzugefügt, während sie bei Veranstaltungen live vorgetragen wird (vgl. Díaz-Cintas & Remael 2021: 8f.). In Österreich wird diese Übersetzungsform zum Beispiel beim ORF genutzt, der im Jahr 2025 8,8 Prozent seines gesamten Programms mit Audiodeskription angeboten hat und dessen Aktionsplan 2025–2028 eine Steigerung auf mehr als 4.200 Sendestunden bis 2028 vorsieht (vgl. ORF 2026). Anders als die Untertitelung überträgt die Audiodeskription nicht allein sprachliche Inhalte und dient vor allem der Barrierefreiheit.

Neben Synchronisation, Voiceover und Audiodeskription existieren weitere Formen der audiovisuellen Übersetzung – wie etwa die Filmdolmetschung – auf die im Rahmen dieser Arbeit aber nicht näher eingegangen wird.

2.3.2 Untertitelung

Bei der Untertitelung wird ein schriftlicher Text, der den Originaldialog wiedergibt, für gewöhnlich am unteren Bildrand eingeblendet. Auch andere sprachliche Elemente wie Briefe und Schilder auf der visuellen Ebene oder Lieder und Hintergrundstimmen auf der akustischen Ebene werden einbezogen. Die Textsegmente sind in der Regel gekürzte Übersetzungen oder Transkriptionen der verbalen Inhalte (vgl. Díaz-Cintas 2009: 9). Zu den grundlegenden Arbeitsschritten der Untertitelung zählen Spotting, Splitting und die Textbearbeitung (vgl. Jüngst 2020: 41). Die Komplexität der dafür erforderlichen Fertigkeiten wird von Gottlieb

(2002) bildhaft zum Ausdruck gebracht:

... das musikalische Ohr eines Dolmetschers, die nüchterne Urteilsfähigkeit eines Nachrichtenredakteurs und den Sinn eines Designers für Ästhetik ... Um die Untertitel auch synchron mit dem Leinwandgeschehen zu präsentieren, braucht er überdies die sichere Hand eines Chirurgen und das Zeitgefühl eines Schlagzeugers.

Untertitelte Inhalte bestehen aus drei Komponenten: dem gesprochenen Text, den Bildern und den Untertiteln. Bei der Übersetzung ist das Zusammenspiel dieser Ebenen ebenso zu berücksichtigen wie die begrenzte Lesegeschwindigkeit der Zuseher:innen. Untertitel müssen synchron zu Bild und Dialog erscheinen, semantisch adäquat sein und ausreichend lange eingeblendet werden (vgl. Díaz-Cintas 2009: 9). Die Rezeption untertitelter Filme erfordert eine erhöhte Aufmerksamkeit, weshalb mögliche zusätzliche Verständnishürden vermieden werden sollten.

Die Rezipient:innen von untertitelten Programmen werden von Gottlieb (2018: 335) als „secondary audiences“ bezeichnet, also als hörende Zuseher:innen im Ausland oder

hörbeeinträchtigtes Publikum im Produktionsland. Zugleich stellt er fest, dass vor allem jüngere Zielgruppen untertitelten Inhalten zunehmend offen gegenüberstehen (vgl. Gottlieb 2018: 335).

Da der Originaldialog hörbar bleibt, ist die Übersetzung für die Rezipient:innen unmittelbar überprüfbar. Untertitel werden daher mitunter besonders kritisch hinsichtlich Lesbarkeit, Dauer sowie grammatikalischer und orthographischer Korrektheit bewertet (vgl. Gottlieb 2018: 335). Karamitroglou (1998) weist in diesem Zusammenhang darauf hin, dass leicht erkennbare Elemente wie Eigennamen möglichst beibehalten werden sollten, um Irritationen zu vermeiden.

Mit den technischen Entwicklungen hat sich der Einsatzbereich der Untertitelung erweitert, ebenso wie die Kriterien, nach denen sie klassifiziert wird. Die Einteilung erfolgt meist anhand sprachlicher und technischer Merkmale.

2.3.2.1 Arten der Untertitelung

Aus sprachlicher Perspektive wird zwischen intralingualer und interlingualer Untertitelung unterschieden. Intralinguale Untertitel verschriftlichen gesprochene Inhalte innerhalb derselben Sprache und richten sich primär an taube und hörgeschädigte Personen. Dabei werden nicht nur Dialoge, sondern auch paralinguistische und akustische Informationen wie Intonation, Ironie oder Geräusche in Textform wiedergegeben, die für das Verständnis der Botschaft relevant sind (vgl. Díaz-Cintas & Remael 2021: 12). Auch in diesem Bereich plant der ORF weitere Maßnahmen. 2025 wurden beinahe 70 Prozent seiner Sendungen in untertitelter Version angeboten und bis 2028 soll die Quote auf 82,6 Prozent gesteigert werden (vgl. ORF 2026).

Darüber hinaus finden intralinguale Untertitel auch beim Lehren und Lernen von Fremdsprachen sowie bei schwer verständlichen Akzenten Verwendung (vgl. Díaz-Cintas & Remael 2021: 14, 17). Interlinguale Untertitel zielen hingegen darauf ab, den ursprünglichen Dialog sowie alle anderen verbalen Informationen in eine andere Sprache zu übertragen. Dabei kann es sich um visuell dargestellte Inhalte wie beispielsweise Briefe, Einblendungen und Textnachrichten handeln, aber auch um auditiv übermittelte Elemente wie Lieder (vgl. Díaz-Cintas & Remael 2021: 12). Normalerweise sind sie primär für hörende Rezipient:innen gedacht, werden aber auch von hörgeschädigten Personen genutzt. Viele Regeln überschneiden sich mit jenen der intralingualen Untertitelung (vgl. Jüngst 2020: 37).

Technisch betrachtet lassen sich Untertitel in offene (harte) und geschlossene (weiche) Formen einteilen. Sie sind inhaltlich identisch, jedoch auf unterschiedliche Weise in das Bildsignal eingebettet. Offene Untertitel sind dauerhaft in das Bild integriert und können von den Zuseher:innen nicht deaktiviert werden. Geschlossene Untertitel hingegen werden als separate Spur eingeblendet und lassen sich bei Bedarf ein- oder ausschalten, wodurch auch mehrere Sprachversionen angeboten werden können (vgl. Díaz-Cintas & Remael 2021: 26). Während im analogen Zeitalter interlinguale Untertitel im Kino und Fernsehen meist offen und intralinguale Untertitel geschlossen übertragen wurden, haben sich mit der Digitalisierung und insbesondere durch Streaming- und Video-on-Demand-Dienste geschlossene Untertitel als Standard etabliert (vgl. Díaz-Cintas & Remael 2021: 27). Die beiden Arten unterscheiden sich somit nur in der Distributionsform, nicht aber in ihrer sprachlichen Gestaltung.

Bezüglich des Produktionsmodus gibt es Live- und Semi-live-Untertitelung. Live-Untertitel werden in Echtzeit erstellt, etwa bei Sportsendungen, Nachrichtensendungen oder parlamentarischen Debatten. Bei der Semi-live-Untertitelung von Nachrichten-Updates, Theaterstücken oder Opernproduktionen wird kurz vor der Ausstrahlung ein Skript bereitgestellt. Beide Formen können sowohl intra- als auch interlingual erfolgen (vgl. Díaz-Cintas & Remael 2021: 10, 22).

Für die vorliegende Arbeit sind ausschließlich interlinguale, vorproduzierte Untertitel von Relevanz.

2.3.2.2 Gestaltungsparameter

Untertitel unterliegen bestimmten Gestaltungsprinzipien, die sich aus den technischen Rahmenbedingungen audiovisueller Medien sowie aus den Rezeptionsbedingungen des Publikums ergeben. Ziel ist es, eine hohe Verständlichkeit bei möglichst geringer visueller Beeinträchtigung zu gewährleisten. Karamitroglou (1998) formuliert dies als Anspruch, „to provide maximum appreciation and comprehension of the target film as a whole by maximising the legibility and readability of the inserted subtitled text“.

Die Gestaltung von Untertiteln lässt sich dabei in formale und linguistische Parameter unterteilen. Während formale Parameter die visuelle und zeitliche Gestaltung betreffen, beziehen sich linguistische Parameter auf die sprachliche Strukturierung und Verdichtung der Inhalte. Beide Dimensionen greifen ineinander und haben große Bedeutung für die konkrete Ausführung von Untertiteln.

2.3.2.2.1 Formale Parameter

Zu den räumlichen Parametern zählen insbesondere die maximale Zeilen- und Zeichenanzahl sowie die Positionierung des Textes im Bild. Interlinguale Untertitel sind in der Regel auf zwei Zeilen beschränkt und zentriert am unteren Bildschirmrand positioniert, wo sie am wenigsten störend für das Bildgeschehen sind (vgl. Díaz-Cintas & Remael 2021: 93f.; Karamitroglou 1998).

Auch typografische Aspekte wie Schriftart, -größe und -farbe beeinflussen die Leserlichkeit. In den meisten westlichen Sprachen sind neutrale, wenig auffällige Schriftarten ohne Serifen wie Arial, Verdana oder Helvetica üblich. Häufig empfohlen wird für das lateinische Alphabet Arial in 30 Punkt und die Verwendung von weißer Schrift (vgl. Díaz-Cintas & Remael 2021: 94).

In zeitlicher Hinsicht liegt das Hauptaugenmerk auf der Synchronität der Untertitel mit dem Dialog und dem Bild (vgl. Díaz-Cintas & Baños 2017: 317). Präzises Timing ist für die semiotische Kohäsion der übersetzten Inhalte wesentlich und trägt dazu bei, Sprecher:innenwechsel nachvollziehbar zu machen (vgl. Díaz-Cintas & Remael 2021: 103f.). Bezüglich der Einblenddauer wird angenommen, dass ein Untertitel aus zwei Zeilen mit jeweils etwa 37 Zeichen bequem in sechs Sekunden gelesen werden kann. Als Richtwert gelten 160 Wörter pro Minute oder 15 Zeichen pro Sekunde; ein Untertitel sollte jedoch mindestens rund eine Sekunde sichtbar bleiben, um wahrgenommen werden zu können (vgl. Díaz-Cintas & Baños 2017: 317). Zudem wird empfohlen, Untertitel nicht über einen Szenenwechsel hinaus fortzuführen, da dies zu erneuter Rezeption führen kann (vgl. Díaz-Cintas & Remael 2021: 114).

Die genannten räumlichen und zeitlichen Einschränkungen bilden den formalen Rahmen, innerhalb dessen Entscheidungen zur sprachlichen Verdichtung und Segmentierung getroffen werden.

2.3.2.2.2 Linguistische Parameter

Neben den räumlichen und zeitlichen Rahmenbedingungen sind Untertitel maßgeblich durch linguistische Parameter bestimmt. Eine klare und syntaktisch einfache Formulierung soll dabei die Verarbeitung erleichtern und die Rezeption unterstützen.

Zentral für die Qualität von Untertiteln sind Kohärenz und Kohäsion. Kohärenz trägt dazu bei, dass die Botschaft ankommt, während Kohäsion jene sprachlichen Mittel umfasst, mit denen Kohärenz gefördert werden kann. In der Untertitelung kann Kohärenz durch

Strategien der Verdichtung und Kürzung beeinträchtigt werden, ebenso durch Unterbrechungen, die etwa durch die Aufteilung eines Satzes auf mehrere Untertitel entstehen. Inkohärenzen können beispielsweise durch unklare Referenzen, sprunghafte Übergänge oder unvollständige Satzstrukturen bedingt sein. Werden Informationen ausgelassen, ist daher darauf zu achten, dass die Logik des Ausgangstextes in der Zielversion erhalten bleibt (vgl. Díaz-Cintas & Remael 2021: 168).

Die sprachliche Komprimierung stellt dabei ein wesentliches Charakteristikum der Untertitelung dar. Aufgrund der beschränkten Anzahl an verfügbaren Zeichen kann gesprochener Dialog für gewöhnlich nicht wortgetreu wiedergegeben werden. Dies ist auch nicht erforderlich, da Untertitel stets im Zusammenspiel mit Bild und Ton rezipiert werden. Gekürzt wird entweder durch das Entfernen weniger relevanter Elemente oder durch die möglichst knappe Formulierung zentraler Inhalte (vgl. Díaz-Cintas & Remael 2021: 147). Dies kann beispielsweise Füllwörter, kumulative Adjektive und Adverbien oder Antwortpartikel betreffen (vgl. Karamitroglou 1998).

Darüber hinaus sehen sich Untertitler:innen mit übersetzerischen Herausforderungen wie hybriden Sprachformen oder Mehrsprachigkeit konfrontiert. Dazu kommt der Umgang mit Problemen der sprachlichen Übertragung, die auch bei anderen Formen der Übersetzung auftreten: idiomatische Wendungen, kulturelle Referenzen oder Tabuwörter (vgl. Díaz-Cintas & Remael 2021: 146).

Diese linguistischen Entscheidungen stehen in engem Zusammenhang mit der Frage, wie sprachliche Einheiten innerhalb der gegebenen formalen Begrenzungen strukturiert und segmentiert werden. Die Zeilensegmentierung bildet somit eine Schnittstelle zwischen formalen und linguistischen Parametern und wird im folgenden Abschnitt gesondert behandelt.

2.3.2.2.3 Zeilensegmentierung

Unter Segmentierung wird das Aufteilen eines Satzes in mehrere Teile (Segmente) verstanden, um dessen Lesbarkeit und Verständlichkeit zu fördern. Sie gilt als ein wichtiger Faktor für das rasche Erfassen von Untertiteln, da diese nur kurz eingeblendet werden und isoliert zwischen vorangehenden und nachfolgenden Einheiten erscheinen. Díaz-Cintas und Remael (2021: 170) sehen es daher als eine der goldenen Regeln des Berufs von Übersetzer:innen, Untertitel semantisch und syntaktisch in sich geschlossen zu strukturieren.

Bereits Karamitroglou (1998) plädierte in seinem *Proposed Set of Subtitling Standards* dafür, längere Untertitel auf zwei Zeilen aufzuteilen, da zweizeilige Einheiten vom Auge als

kompakter Textblock wahrgenommen und den Lesevorgang beschleunigen würden. Auch der *Code of Good Subtitling Practice* empfiehlt, dass Untertitel möglichst syntaktisch in sich abgeschlossen sein sollen: „Ideally, each subtitle should be syntactically self-contained“, sowie in Sinneinheiten und grammatikalische Einheiten gegliedert werden: „Subtitle text must be distributed from line to line and page to page in sense blocks and/or grammatical units“ (Ivarsson & Carroll 1998: 1). Als weitere Anforderungen werden Kohärenz und grammatikalische Abgeschlossenheit genannt (vgl. Ivarsson & Carroll 1998: 1). Gemeint sind damit semantische oder syntaktische Wortgruppen, die jeweils innerhalb derselben Zeile stehen sollten.

In der einschlägigen Literatur wird grundsätzlich angenommen, dass eine regelbasierte Segmentierung die Rezeption erleichtert. Idealerweise bildet jeder Untertitel einen vollständigen Satz mit klarer Struktur und ohne unbeabsichtigte Mehrdeutigkeiten. Ist dies nicht möglich, sind Spotting und Segmentierung von entscheidender Bedeutung. Zieht sich ein Satz über mehrere Untertitel hinweg, sollte jeder einzelne in sich geschlossen sein und zugleich signalisieren, dass die Aussage im nächsten Untertitel fortgeführt wird (vgl. Díaz-Cintas & Remael 2021: 170).

Die Segmentierung erfolgt auf zwei Ebenen: Einerseits kann ein Satz auf die zwei verfügbaren Zeilen innerhalb eines Untertitels aufgeteilt werden, andererseits kann sich ein langer Satz über mehrere aufeinanderfolgende Untertitel erstrecken. In beiden Fällen gelten grundsätzlich dieselben Segmentierungsregeln; bei fortlaufenden Sätzen ist allerdings zu berücksichtigen, dass die Gedächtnisspanne der Rezipient:innen begrenzt ist. Komplexe syntaktische Strukturen sollten daher nach Möglichkeit in kürzere Einheiten gegliedert werden (vgl. Díaz-Cintas & Remael 2021: 171).

Hinsichtlich der konkreten Verteilung der Zeichen auf zwei Zeilen finden sich in der Literatur verschiedene Ansätze. Während in älteren Werken betont wird, dass die Augen eher an das Lesen von Text in rechteckigen als in dreieckigen Blöcken gewöhnt sind (vgl. Karamitroglou 1998), wird in der aktuellen Forschung überwiegend die Ansicht vertreten, dass syntaktische und grammatikalische Aspekte Vorrang vor rein visuellen Erwägungen haben sollten (vgl. Díaz-Cintas & Remael 2021: 171). Wenn die Zeilen nicht dieselbe Länge haben, empfehlen Ivarsson & Carroll (1998: 2), vorzugsweise die erste Zeile kürzer zu halten, um das Bild weniger zu überlagern. Empirische Befragungen professioneller Untertitler:innen zeigen, dass zusammengehörige semantische Einheiten bei der Zeilenaufteilung als wichtiger erachtet werden als formale Symmetrie, wenngleich die meisten Untertitler:innen eine kürzere erste Zeile bevorzugten (vgl. Szarkowska 2016: 15).

Eine sorgfältige Segmentierung gilt in der Praxis als Qualitätsmerkmal professioneller Untertitelung, während bei ungünstigen Zeilenumbrüchen häufig mangelnde Kompetenz vermutet wird (vgl. Szarkowska 2016: 22). Segmentierungsregeln werden oft in unternehmensspezifischen Styleguides spezifiziert, jedoch existierten eine Vielzahl an Stilen und unterschiedliche Auffassungen über deren Effizienz (vgl. Szarkowska 2016: 16). Zudem können Kostendruck und eventuell erforderliche Textkürzungen dazu führen, dass Regeln nicht konsequent umgesetzt werden (vgl. Szarkowska et al. 2018: 2).

Die bisherigen empirischen Studien liefern kein eindeutiges Bild hinsichtlich der Frage, ob syntaktisch ausgerichtete Segmentierung die Rezeption verbessert. Trotz fehlender eindeutiger Evidenz wird in den maßgeblichen Untertitelungsrichtlinien weiterhin eine kohärente und semantisch begründete Zeilensegmentierung empfohlen (vgl. Karamitroglou 1998; Carroll & Ivarsson 1998; Díaz-Cintas & Remael 2021: 169–175).

2.3.2.2.3.1 Segmentierung innerhalb von Untertiteln

Die Zeilentrennung innerhalb eines einzelnen Untertitels betrifft die Verteilung des Textes auf maximal zwei Zeilen und orientiert sich idealerweise an syntaktischen und semantischen Einheiten. Karamitroglou (1998) argumentiert, dass bei unvermeidbarer Zeilentrennung die Segmentierung möglichst mit dem „highest syntactic node“ zusammenfallen sollte. Je höher dieser Knoten in der syntaktischen Struktur angesiedelt ist, desto größer ist die gebündelte semantische Informationseinheit, die vor einer Verarbeitungspause abgeschlossen wird. Da jede Segmentierung das Gehirn zu einer kurzen Unterbrechung der sprachlichen Verarbeitung zwingt, sollte diese Pause an einem Punkt erfolgen, an dem bereits eine inhaltlich möglichst vollständige Einheit vermittelt wurde.

Auf dieser Grundlage formulieren Díaz-Cintas & Remael (vgl. 2021: 173) konkrete Richtlinien für die Zeilentrennung.¹ Wörter sollten nicht durch Bindestrich getrennt werden. Bei Untertiteln aus mehreren Sätzen soll – sofern möglich – ein Satz pro Zeile erscheinen. Besteht ein Satz aus zwei Nebensätzen und überschreitet die zulässige Zeilenlänge, soll jede Zeile jeweils einen Nebensatz enthalten. Es ist nicht notwendig, die maximale Zeichenzahl der ersten Zeile auszunutzen, bevor zur zweiten übergegangen wird.

Da die Trennung semantischer Einheiten die sprachliche Verarbeitung verlangsamen kann, sollten zusammengehörige Elemente in derselben Zeile verbleiben. Dazu zählen etwa Adjektive und die dazugehörigen Substantive, Adverbien und Verben, Artikel und

¹Die Richtlinien wurden primär für die englische Sprache formuliert. Aufgrund vergleichbarer syntaktischer Grundstrukturen lassen sie sich jedoch weitgehend auf das Deutsche übertragen.

Substantive, Vor- und Nachnamen sowie das Wort „zu“ und der zugehörige Infinitiv. Auch zusammengesetzte Verben, Hilfsverben, Reflexivpronomen sowie Negationen sollten nicht vom Verb getrennt werden, auf das sie sich beziehen. Ebenso ist es zu vermeiden, ein Verb von seinem direkten oder indirekten Objekt zu trennen (vgl. Díaz-Cintas & Remael 2021: 173). Im Deutschen kann allerdings die durch die Verbklammer bedingte Distanz zwischen finitem und infinitem Verb die konsequente Umsetzung dieser Regel erschweren.

Erstreckt sich ein einfacher Satz über zwei Zeilen, wird empfohlen, das Subjekt in die obere und das Prädikat in die untere Zeile zu setzen.² Reaktionen auf Fragen oder Aussagen sollten vorzugsweise in der zweiten Zeile desselben Untertitels erscheinen, sofern dadurch keine inhaltlichen Vorgriffe entstehen oder ein Szenenwechsel dies verhindert (vgl. Díaz-Cintas & Remael 2021: 173).

2.3.2.2.3.2 Segmentierung über mehrere Untertitel

Nicht immer kann ein vollständiger Satz innerhalb eines einzelnen Untertitels realisiert werden – beispielsweise, wenn eine Äußerung die maximale Zeichenzahl überschreitet, ein Szenenwechsel berücksichtigt werden muss, die Informationsdichte zu hoch ist oder sich die Satzstruktur nicht in unabhängige syntaktisch-semantische Einheiten aufteilen lässt. In solchen Fällen erstreckt sich ein Satz über zwei oder mehrere aufeinanderfolgende Untertitel (vgl. Díaz-Cintas & Remael 2021: 174). Dies gilt insbesondere für das Deutsche, dessen häufig komplexe Satzstrukturen zusätzliche Segmentierungsentscheidungen erforderlich machen können.

Auch bei der Segmentierung über mehrere Untertitel hinweg sollten die zuvor dargestellten Prinzipien der syntaktischen und semantischen Geschlossenheit gewahrt bleiben. Jede Einheit sollte in sich verständlich sein und das Satzende – soweit möglich – im folgenden Untertitel errahnen lassen. Ein Satz sollte sich dennoch nicht über zu viele Untertitel erstrecken, da dies das Verständnis erschweren kann (vgl. Díaz-Cintas & Remael 2021: 175).

Der Arbeitsvorgang, längere Sätze auf mehrere Untertitel aufzuteilen, wird als *Splitting* bezeichnet. Traditionell wird eine Fortsetzung durch einen Gedankenstrich oder durch Auslassungspunkte markiert. Welche Variante verwendet wird, hängt in der Regel von den Vorgaben der Auftraggeber:innen ab. Der nachfolgende Untertitel beginnt üblicherweise mit einem Kleinbuchstaben; bei Verwendung von Auslassungspunkten können diese erneut vorangestellt werden. In der Praxis zeichnet sich jedoch eine Tendenz ab, auf explizite

²Aufgrund der Verbzweitstellung im Deutschen kann die syntaktisch optimale Zeilentrennung von englischsprachigen Empfehlungen abweichen.

Fortsetzungszeichen zu verzichten. Es wird vermutet, dass diese Entwicklung auf das Vorbild der Live-Untertitelung zurückgeht (vgl. Jüngst 2020: 46).

3 Theoretische Grundlagen der Untertitelrezeption

Untertitel sind keine isolierten Texte, sondern in ein audiovisuelles Gesamtgefüge eingebettet, das gesprochene Sprache, Bilder und schriftlichen Text umfasst. Zuseher:innen müssen daher mehrere Informationskanäle parallel verarbeiten. Folglich hängt die Rezeption und Verständlichkeit untertitelter Inhalte nicht ausschließlich von der Qualität der Untertitel ab, sondern auch von kognitiven Verarbeitungsprozessen und formalen Gestaltungsmerkmalen. Um den Einfluss der Zeilensegmentierung auf die Verständlichkeit angemessen einordnen zu können, werden in diesem Kapitel ausgewählte Theorien zur Rezeption und Verarbeitung multimodaler Informationen vorgestellt. Es handelt sich dabei um kognitionswissenschaftliche und translationswissenschaftliche Ansätze, die erklären, wie Untertitel wahrgenommen, verarbeitet und in den Gesamtkontext integriert werden.

3.1 Multimodale Verarbeitung und semiotische Kohäsion

Für das Verständnis der Rezeption von Untertiteln ist die multimodale Beschaffenheit audiovisueller Texte zentral, da diese durch die Interaktion mehrerer semiotischer Modi Bedeutung erzeugen (vgl. Díaz-Cintas & Remael 2021: 64f.). Diese multimodale Struktur bildet im audiovisuellen Kontext eine wichtige Grundlage für Überlegungen zur Synchronität und Lesbarkeit sowie die Abstimmung von Untertiteln mit visuellen Bildelementen (vgl. Díaz-Cintas & Remael 2021: 4, 9, 119).

Modi sind für Kress (2010: 79) sozial geformte und kulturell bedingte semiotische Ressourcen zur Bedeutungskonstruktion. Er führt Bild, Schrift, Layout, Musik, Gestik, Sprache, bewegte Bilder, Soundtracks und 3D-Objekte als Beispiele für Modi an, die in der Darstellung und Kommunikation verwendet werden. In diesem Verständnis ist Sprache nicht mehr das dominante Zeichensystem, sondern einer von mehreren spezifischen Modi der Bedeutungskonstruktion.

Filme sind komplexe multisemiotische Texte, in denen verschiedene Modi zusammenwirken, um für die Zuseher:innen ein kohärentes Ganzes zu schaffen. Bilder und audiovisuelle Inhalte sind jedoch nicht kulturunabhängig zu verstehen. Díaz-Cintas & Remael (vgl. 2021: 64f.) betonen, dass Bilder nicht zwangsläufig universell verständlich, sondern oft kulturell konnotiert sind. Die Notwendigkeit audiovisueller Übersetzung ist daher nicht allein auf das Aufkommen des Tonfilms zurückzuführen.

Eine zentrale Herausforderung audiovisueller Übersetzung besteht darin, bei der Übertragung verbalen Materials alle filmischen Zeichensysteme zu berücksichtigen, um ein

neues multisemiotisches Ganzes zu schaffen. Díaz-Cintas und Remael (2021: 65) unterscheiden im filmischen Text vier grundlegende Modi:

- auditiv-verbal (Filmdialoge, Voiceover, Erzählung, Liedtexte),
- auditiv-nonverbal (Musik und Soundeffekte, z. B. Hintergrundgeräusche),
- visuell-verbal (Text auf dem Bildschirm, Abspann, Briefe, Anzeigen, Zeitungsüberschriften),
- und visuell-nonverbal (Schnitt, Fotografie, Mise-en-scène, Gesten).

Bei fremdsprachlichen Produktionen wird der auditiv-verbale Modus für die Rezipient:innen meist unzugänglich und gefährdet damit das multimodale Zusammenspiel der Zeichensysteme, durch das Sinn konstruiert wird. Untertitelung wird eingesetzt, um diese Lücke zu schließen (vgl. Díaz-Cintas & Remael 2021: 65). Diese Verschiebung des auditiv-verbalen in einen visuell-verbalen Modus verändert nicht nur die sprachliche Verständlichkeit, sondern wirkt sich auch auf die Wechselwirkung der Modalitäten des Films aus und beeinflusst damit die Bedingungen seiner Rezeption.

3.2 Kognitive Belastung bei der Untertitelrezeption

Im Unterschied zu nicht untertitelten Produktionen stellt die Rezeption untertitelter Medien eine zusätzliche kognitive Anforderung dar: Neben der auditiven und visuellen Wahrnehmung müssen Zuseher:innen auch schriftsprachliche Informationen verarbeiten. Das Lesen der Untertitel kommt somit als weitere kognitive Aufgabe hinzu.

Theoretisch lässt sich diese Fragestellung im Rahmen der Kognitiven Theorie des multimedialen Lernens verorten, die auf der Cognitive-Load-Theorie basiert, einer umfassenderen Theorie der menschlichen Kognition (vgl. Mayer & Moreno 2012: 133). Ein zentraler Grundsatz beider Ansätze ist, dass das menschliche Arbeitsgedächtnis nur über eine begrenzte Verarbeitungskapazität verfügt. Dabei wird zwischen mehreren Formen kognitiver Belastung unterschieden: intrinsische Belastung (bedingt durch die Komplexität des Materials), extrinsische Belastung (verursacht durch die Gestaltung der Darbietung bzw. durch lernirrelevante Faktoren) und germane Belastung (kognitive Verarbeitung, die dem Aufbau mentaler Strukturen dient). Problematisch wird die Rezeption dann, wenn diese Belastungsformen gemeinsam die verfügbare Kapazität übersteigen (vgl. Mayer & Moreno 2012: 133f.). Das Ziel für die Gestaltung multimedialer Inhalte ist daher die Bewältigung von intrinsischer, die Reduzierung von extrinsischer und die Förderung von germaner Belastung

(vgl. Mayer 2021: 29). Daraus ergibt sich für die Untertitelung vor allem eine Relevanz der extrinsischen Belastung, da sie durch die Gestaltung der Untertitel beeinflusst werden kann.

Ergänzend verweist die Kognitive Theorie des multimedialen Lernens darauf, dass das kognitive System über zwei teilweise getrennte Kanäle verfügt: einen visuell-bildlichen und einen auditiv-verbale Kanal. Diese verfügen jeweils nur über begrenzte Kapazitäten auf und können nur eine bestimmte Menge an Informationen gleichzeitig verarbeiten (vgl. Mayer 2021: 34, 36). Obwohl Informationen primär über einen spezifischen Kanal aufgenommen werden, besteht bei ausreichenden kognitiven Ressourcen die Fähigkeit, diese mental in den jeweils anderen Kanal zu überführen. So kann beispielsweise visuell dargebotener Text von erfahrenen Leser:innen innerlich in Lautsprache umgewandelt und somit über den auditiven Kanal weiterverarbeitet werden (vgl. Mayer 2021: 36). Vor diesem Hintergrund ist bei der Verarbeitung von Untertiteln zu berücksichtigen, dass visuell präsentierter Text nicht automatisch dem auditiven Kanal zugeordnet wird, sondern zunächst zusätzlich die visuelle Verarbeitung beansprucht.

Multimodale Verarbeitung führt jedoch nicht zwangsläufig zu einer kognitiven Überlastung. Vielmehr zeigen Forschungsergebnisse, dass Untertitel – sofern sie optimal gestaltet sind – das Verständnis sogar unterstützen können. So fanden Kruger, Hefer und Matthew (2013) heraus, dass Untertitel in einem experimentellen Setting die wahrgenommene kognitive Belastung reduzieren können, da sie als weitere Verständnishilfe dienen. Auch eine Studie von Kruger et al. (2016) spricht dafür, dass Untertitel die Aufmerksamkeit lenken und das Eintauchen in das Geschehen fördern können.

Aus diesen Überlegungen lässt sich ableiten, dass auch die Segmentierung von Untertiteln zur Reduzierung von extrinsischer kognitiver Belastung beitragen kann.

3.3 Kohärenz, Kohäsion und Verständlichkeit

Kohärenz gilt als zentrale Eigenschaft gut strukturierter Texte und trägt dazu bei, dass eine beabsichtigte Botschaft erfolgreich vermittelt wird. Kohäsion hingegen bezeichnet die sprachlichen Mittel, mit denen diese Kohärenz hergestellt wird (vgl. Díaz-Cintas & Remael 2021: 168). Während Kohärenz auf der Ebene der inhaltlichen Zusammenhänge angesiedelt ist, betrifft Kohäsion die formale Verknüpfung sprachlicher Elemente im Text.

Nach Beaugrande und Dressler (1981: 5) bezeichnet Kohärenz „die Funktionen, durch die die Komponenten der Textwelt, d. h. die Konstellation von Konzepten (Begriffen) und Relationen (Beziehungen), welche dem Oberflächentext zugrunde liegen, füreinander

gegenseitig zugänglich und relevant sind“. In der Untertitelung können jedoch spezifische Produktionsbedingungen wie Verdichtung und Reduktion sowie Einschränkungen durch Layoutvorgaben oder die Aufteilung syntaktischer Einheiten auf mehrere Untertitel zu Kohärenzproblemen führen. Selbst wenn visuelle und auditive Informationen zum Gesamtverständnis beitragen, ist es daher erforderlich, innerhalb der Untertitel selbst eine möglichst hohe Kohäsion zu gewährleisten, um potenziellen Kohärenzverlusten vorzubeugen. Mangelnde Kohärenz kann sich etwa in unklaren Referenzen, abrupten Übergängen, schlecht strukturierten Sätzen oder unlogischen Segmentierungen manifestieren. Solche Probleme entstehen häufig durch Auslassungen, die mit Umformulierungen einhergehen. Werden Informationen reduziert, muss dennoch sichergestellt werden, dass die Logik des Ausgangstextes innerhalb eines Untertitels sowie über mehrere Untertitel hinweg erhalten bleibt (vgl. Díaz-Cintas & Remael 2021: 169).

Kohäsion wiederum „betrifft die Art, wie die Komponenten des Oberflächentextes, d. h. die Worte, wie wir sie tatsächlich hören oder sehen, miteinander verbunden sind“ (Beaugrande & Dressler 1981: 3f.). Sie beruht auf grammatischen Abhängigkeiten, da die Oberflächenkomponenten durch grammatische Formen und Konventionen strukturell miteinander verknüpft sind. Diese Oberflächenstrukturen sind in vielen Sprachen nur begrenzt veränderbar, ohne Irritation hervorzurufen (vgl. Beaugrande & Dressler 1981: 3f.). Für die Untertitelung ist darüber hinaus die intersemiotische Kohäsion bedeutsam: Sie beschreibt die Weise, in der die sprachlichen Elemente der Untertitel mit der Tonspur und den visuellen Informationen verknüpft sind, sodass sich ein kohärentes sprachlich-visuelles Gesamtgefüge ergibt (vgl. Díaz-Cintas & Remael 2021: 168).

Die Analyse intersemiotischer Kohärenz und Kohäsion ist daher grundlegend für das Verständnis multimodaler Texte. Untertitler:innen sollten nicht ausschließlich die verbale Ebene berücksichtigen, sondern auch stets die Wechselwirkung zwischen sprachlichen und nichtsprachlichen Zeichen im Blick haben (vgl. Díaz-Cintas & Remael 2021: 70).

Zusammenfassend zeigt sich, dass die Verständlichkeit von Untertiteln nicht allein von der sprachlichen Formulierung abhängt. Stattdessen ergibt sie sich aus dem Zusammenspiel von Kohärenz, Kohäsion und der Einbettung in das audiovisuelle Geschehen. Die Herausforderung besteht darin, trotz der formalen und zeitlichen Beschränkungen sprachliche Einheiten zu schaffen, die sowohl innerhalb eines einzelnen Untertitels als auch über mehrere Untertitel hinweg nachvollziehbar bleiben.

Im Hinblick auf die spezifische Fragestellung dieser Arbeit – inwiefern die Qualität der Zeilensegmentierung das Verständnis von Untertiteln beeinflusst – ist festzuhalten, dass bislang nur eine begrenzte Zahl empirischer Untersuchungen dazu vorliegt. Es ist daher noch nicht abschließend geklärt, in welchem Ausmaß die in der Literatur empfohlenen Segmentierungsregeln, insbesondere die Beibehaltung syntaktischer und semantischer Einheiten, tatsächlich eine Rolle für die Verständlichkeit spielen. Vor diesem Hintergrund erscheint eine differenzierte Betrachtung der Segmentierungspraxis angebracht.

3.4 Erstsprache und Untertitelrezeption

In der Forschung wird häufig zwischen Muttersprachler:innen und Nicht-Muttersprachler:innen unterschieden. Der Begriff *Muttersprache* wird jedoch teilweise kritisch betrachtet, da er von eindeutigen Zugehörigkeiten ausgeht. Busch (2021: 49) weist darauf hin, dass Konzepte der Mehrsprachigkeit häufig stillschweigend das Vorhandensein einer klar definierten Muttersprache voraussetzen, obwohl individuelle sprachliche Biografien oft komplexer sind.

In der vorliegenden Arbeit wird daher der Begriff *Erstsprache* verwendet. Als Erstsprachler:innen werden Personen bezeichnet, die Deutsch als Erstsprache erworben haben; Nicht-Erstsprachler:innen verfügen demgegenüber über eine andere Erstsprache als Deutsch.

Für die Rezeption von Untertiteln ist der sprachliche Hintergrund insofern von Bedeutung, als Menschen, die die Sprache der Untertitel als Erstsprache beherrschen, sprachliche Informationen in der Regel schneller und automatisierter verarbeiten können. Personen, die die Sprache erst später erlernt haben, benötigen dafür unter Umständen mehr kognitive Ressourcen. Es ist daher denkbar, dass Unterschiede in der Untertitelgestaltung von beiden Gruppen unterschiedlich wahrgenommen werden.

Die vorliegende Untersuchung erhebt den sprachlichen Hintergrund der Teilnehmenden mittels Fragebogen und unterscheidet explorativ zwischen Personen mit und ohne Deutsch als Erstsprache.

4 Studien zur Zeilensegmentierung

Nach der theoretischen Einordnung der Rezeptions- und Verarbeitungsprozesse widmet sich dieses Kapitel empirischen Studien, die sich explizit mit der Zeilensegmentierung in Untertiteln befassen. Ziel ist es, zentrale Forschungsansätze, methodische Zugänge und Ergebnisse zu beschreiben. In einer abschließenden Synthese sollen die Studien verglichen sowie deren Relevanz für die vorliegende Arbeit eingeordnet werden.

4.1 Die kognitive Effizienz der Untertitelverarbeitung

In einer experimentellen Untersuchung testeten Perego et al. (vgl. 2010: 243) die Hypothese, dass die Verarbeitung von untertitelten Filmen kognitiv effektiv sei. Die Überlegung der Studienleiter:innen war, dass die Kombination von Bild und Untertiteln möglicherweise nicht so komplex sei, weil das Lesen und die visuelle Verarbeitung an sich relativ mühelose und teilweise automatisch ablaufende Prozesse sind. Bilder könnten zudem das Verstehen von Untertiteln erleichtern, weil sie einen entsprechenden Kontext bieten (vgl. Perego et al. 2010: 247). Dieser Ansatz ist für die vorliegende Arbeit insofern interessant, als er andeutet, dass Untertitel nicht zwangsläufig eine zusätzliche kognitive Belastung darstellen. Wenn die Belastung bei der Verarbeitung von Untertiteln eher niedrig ist, könnte daraus folgen, dass auch Unterschiede in der Segmentierung kein großes Gewicht für die Wahrnehmung haben.

Perego et al. (2010: 243) untersuchten außerdem, ob sich eine syntaktisch inkohärente Segmentierung von zweizeiligen Untertiteln störend auf die Informationsverarbeitung und die Erkennungsleistung auswirkt. Eine ähnliche Fragestellung liegt auch der vorliegenden Arbeit zugrunde, in der der Einfluss kohärenter im Vergleich zu inkohärenter Segmentierung auf die Rezeption untersucht wird. Das Ziel von Perego et al. (vgl. Perego et al. 2010: 249) war der Vergleich zweier gegensätzlicher Annahmen: Einerseits könnten inkohärent segmentierte Textzeilen den natürlichen Lesefluss und in der Folge die Konzentration hemmen, wodurch das Verstehen erschwert werden kann. Andererseits konnte davon ausgegangen werden, dass eine suboptimale Zeilensegmentierung den Lesefluss nicht wesentlich stört, weil eine gut erlernte und teilweise automatisierte Fähigkeit wie das Lesen durch geringfügig erhöhte Anforderungen nicht ernsthaft beeinträchtigt werden sollte (vgl. Perego et al. 2010: 249). In Anlehnung an frühere Untersuchungen wurde erwartet, aufgrund des Lesevorgangs mehr Fixationen (d. h. die Verweilzeiten des Blickes auf bestimmten Bildbereichen) auf die Untertitel, aber längere Fixationen auf die Bilder sowie Aufmerksamkeitsverschiebungen von Text zu Bild und umgekehrt zu beobachten (vgl. Perego et al. 2010: 251).

Die methodische Herangehensweise beruhte neben den Daten zu Blickbewegungen auf Worterkennung und dem visuellen Erfassen von Filmszenen. Zusätzlich wurden Fragebögen eingesetzt, um die Aufmerksamkeit für die Untertitel sowie das inhaltliche Verständnis des Films zu überprüfen (vgl. Perego et al. 2010: 253f.). An der Studie nahmen 41 italienischsprachige Personen teil, die über keine Kenntnisse in der Originalsprache des Filmausschnitts verfügten und nach dem Zufallsprinzip in zwei Gruppen eingeteilt wurden, denen jeweils entweder kohärent oder inkohärent segmentierte Untertitel präsentiert wurden. Allerdings wurden die Augenbewegungen aus Zeitgründen nur bei 16 (zufällig ausgewählten) Teilnehmenden aufgezeichnet. Dies wurde mit dem Hinweis kommentiert, dass die Stichprobe bereits größer sei als in den meisten vorangegangenen Studien zur Verarbeitung von Untertiteln, die eine Analyse der Blickbewegungen mit einbezogen (vgl. Perego et al. 2010: 251).

Es wurde entschieden, dass qualitativ hochwertige Untertitel solche sind, bei denen Nominalkonstruktionen (Nomen + Adjektiv, Nomen + Präpositionalphrase, Adjektiv + Nomen, Bestimmungswort + Nomen) nicht getrennt werden, sich also niemals über zwei Zeilen erstrecken. Die inkohärent segmentierten Untertitel wurden so manipuliert, dass diese Blöcke systematisch getrennt wurden und beispielsweise die erste Zeile eines Untertitels mit einem Adjektiv endete, während das zugehörige Nomen an den Anfang der zweiten Zeile gesetzt wurde. Insgesamt enthielt der präsentierte Filmausschnitt 28 Untertitel (vgl. Perego et al. 2010: 252). Auch in der vorliegenden Arbeit orientiert sich die Definition qualitativ hochwertiger Untertitel an diesen Kriterien.

Aus den Fragebögen ging hervor, dass sich die Teilnehmenden beim Verstehen des Films auf die Untertitel verließen, von keinen größeren Schwierigkeiten beim Lesen berichteten und die Handlung gut nachvollziehen konnten. Bezüglich der Erkennung von Wörtern und Filmszenen wurde kein signifikanter Unterschied zwischen den beiden Gruppen festgestellt. Zugleich ergab sich in einem Teil des Films eine signifikant positive Korrelation zwischen Wort- und Szenenerkennung. Die Verarbeitung von Untertiteln ging also nicht zulasten der Erfassung von Szenen – was bestätigte, dass das Ansehen eines untertitelten Films kognitiv effektiv ist. Beim Eye-Tracking waren die Augen der Teilnehmenden zu 67 Prozent der Fixationszeit auf den untertitelten Bereich gerichtet und in der verbleibenden Zeit auf den oberen Bildschirmbereich. Zwar wurden mehr Fixationen auf die Untertitel beobachtet, jedoch waren jene auf das Bildmaterial länger. Kürzere Fixationen auf die Untertitel verteilten sich über den gesamten schriftlichen Text, während sich längere Fixationen auf das Bild tendenziell auf hervorstechende Details fokussierten. Hinsichtlich der

Anzahl der Fixationen fanden sich keine nennenswerten Unterschiede zwischen kohärent und inkohärent segmentierten Untertiteln, nur die mittlere Fixationsdauer war bei inkohärenter Segmentierung geringfügig länger (vgl. Perego et al. 2010: 256–260). Daher wurde vermutet, dass die Qualität der Zeilensegmentierung keinen signifikanten Einfluss hatte und die Teilnehmenden kohärent und weniger kohärent segmentierte Untertitel auf dieselbe Weise und mit denselben Ergebnissen verarbeiten würden. Die Studienleiter:innen schlossen daraus, dass psycholinguistische Bedenken bezüglich der Segmentierung von Untertiteln wahrscheinlich überbewertet seien (vgl. Perego et al. 2010: 263).

Dieses Ergebnis widerspricht somit im Allgemeinen den Leitlinien zur guten Untertitelungspraxis. Einschränkungen in der Interpretation ergeben sich möglicherweise dadurch, dass nur Zeilenumbrüche innerhalb einzelner Untertitel und nicht über mehrere hinweg untersucht wurden, die Zahl der gezeigten Untertitel relativ niedrig war und sich die Manipulation auf Nominalphrasen begrenzte (vgl. Perego et al. 2010: 252). Auch wenn Perego et al. (vgl. 2010: 263) dieser Abweichung der Fixationsdauer nur wenig Bedeutung beimessen, kann sie schließlich doch als Hinweis dafür gesehen werden, dass die kognitive Belastung bei den inkohärent segmentierten Untertiteln höher ist. Die Erkenntnisse dieser Studie liefern Hinweise darauf, inwieweit eine inkohärente Zeilensegmentierung die Verarbeitung von Untertiteln beeinträchtigen kann, und sind damit auch für die vorliegende Arbeit von Interesse. Ein Einfluss der Bildinformation dürfte dabei hingegen gering sein, da im gezeigten Video überwiegend die Vortragende zu sehen ist.

4.2 Auswirkungen der Gliederung von Texteinheiten auf die Untertitelung

Rajendran et al. (2013) untersuchten in einem Eye-Tracking-Experiment, wie sich die Segmentierung von Untertiteln auf die Blickbewegungen auswirkt. Dabei analysierten sie die Effektivität von *Text Chunking* – „the grouping of a block of text into coherent segments“ (Rajendran et al. 2013: 6) – im Rahmen einer Simulation von Untertiteln beim *Respeaking*. Ähnlich wie die Zeilensegmentierung basiert *Text Chunking* auf der Gliederung von Textblöcken in verarbeitbare Einheiten, weshalb auch auf diese Studie näher eingegangen wird.

Als Untersuchungsmaterial dienten Videos mit vier verschiedenen Untertitelungsstilen, die hinsichtlich Lesegeschwindigkeit und Leseverständnis bewertet wurden. Die Untertitel erschienen entweder ohne Segmentierung, Wort für Wort, Phrase für Phrase oder Satz für Satz (vgl. Rajendran et al. 2013: 14). Es nahmen 28 Studierende und

Absolvent:innen an der Studie teil, wobei die Daten von 24 Personen auch tatsächlich ausgewertet werden konnten. Alle Teilnehmenden sahen alle vier Videos und sollten nach jedem Clip ihre Meinung zu den Untertiteln in einem Fragebogen kundtun. Abschließend wurden sie gebeten, die vier Untertitelungsstile in einem Präferenzfragebogen zu bewerten (vgl. Rajendran et al. 2013: 12f.).

In Bezug auf das Verständnis ließ sich kein relevanter Effekt nachweisen, die Eye-Tracking-Daten ergaben jedoch signifikante Abweichungen darin, wie oft die Teilnehmenden zwischen Szene und Untertitel wechselten. Diese Blickwechsel geschahen bei den nach Phrasen oder Sätzen gegliederten Untertiteln im Durchschnitt weniger oft als bei den Wort-für-Wort erscheinenden Untertiteln, bei denen solche Wechsel tendenziell während des Wartens auf das nachfolgende Wort auftraten. Besonders im Vergleich zu den satzweise eingeblendeten Untertiteln wurde dies als weniger günstig bewertet, denn bei der Satz-für-Satz-Segmentierung mussten die Zuseher:innen nicht auf das Erscheinen von Wörtern warten und genossen ein gleichmäßigeres und natürlicheres Seherlebnis (vgl. Rajendran et al. 2013: 14f.). Dieses Ergebnis könnte bedeuten, dass sich Effekte der Segmentierung nicht zwingend im Verständnis niederschlagen, sondern auch in der Art der Verarbeitung sichtbar werden können.

Die nach Phrasen segmentierten Untertitel erforderten die geringste Betrachtungszeit. Die gegenüber den Wort-für-Wort-Untertiteln kürzere Fixationsdauer lässt darauf schließen, dass die kognitive Informationsverarbeitung rascher und einfacher abläuft. Die Daten weisen darauf hin, dass Untertitel in Phrasen die Rezeption am wenigsten beeinträchtigen und sich nicht nachteilig auf das Verständnis auswirken. Im Gegensatz dazu schienen die Wort-für-Wort eingeblendeten Untertitel vergleichsweise das schlechteste Seherlebnis zu bieten (vgl. Rajendran et al. 2013: 16).

Rajendran et al. (2013) sehen damit frühere Erkenntnisse zur Wort-für-Wort-Untertitelung bestätigt, erkennen aber einen Widerspruch zu Studien wie jener von Perego et al. (2010), bei der kaum Einflüsse der Segmentierung festgestellt wurden. Bei Rajendran et al. (2013) liegt der Fokus wiederum stärker auf der Form, wie die Untertitel am Bildschirm präsentiert werden (blockweise, in Sätzen, Phrasen oder Wörtern), und nicht in erster Linie auf der Zeilentrennung, obwohl die Form diese natürlich bedingt. Bei Perego et al. (2010) wurde hingegen nicht die Form, sondern die Wortebene manipuliert (zusammenstehende versus getrennte Nominalkonstruktionen). Ein Vergleich der beiden Studien ist auch deshalb schwierig, weil ihre Untersuchungsbedingungen deutlich voneinander abweichen: Die Untersuchung bei Perego et al. (2010) bestand aus relativ wenigen Teilnehmenden mittels

Eye-Tracking (16), denen zudem nur 28 Untertitel vorgelegt wurden; bei Rajendran et al. (2013) waren dafür die Videos sehr kurz und der Ton war stummgeschaltet.

4.3 Eine Eye-Tracking-Studie über die Präferenzen der Zuseher:innen

Szarkowska & Gerber-Morón (2018) beschäftigten sich mit der Wahrnehmung der Segmentierung von Untertiteln und gingen dabei erstmals auch konkret auf die Präferenzen der Teilnehmenden ein. Ziel der Studie war es herauszufinden, ob es diesbezüglich Unterschiede gibt, die sich auf die Erfahrung der Teilnehmenden mit Untertiteln, auf deren Hörvermögen oder auf die Art der sprachlichen Einheiten zurückführen lassen. In einem zweiten Schritt wurden die Blickbewegungen der Teilnehmenden analysiert, um zu untersuchen, ob inkohärent segmentierte Untertitel mehr Lesezeit erfordern – insbesondere bei Menschen, die mit Untertiteln wenig vertraut waren, und bei tauben Personen. Abschließend wurden die Ansichten der Teilnehmenden auch noch in kurzen semi-strukturierten Interviews erfragt (vgl. Szarkowska & Gerber-Morón 2018: 3).

An der Studie nahmen hörende, schwerhörige und gehörlose Personen aus dem Vereinigten Königreich sowie zusätzlich hörende Teilnehmende aus Polen und Spanien teil. Ihnen wurden 30 Paare von Screenshots präsentiert, wobei jeweils ein Screenshot nach bestimmten Untertitelungsregeln segmentiert war, während im zweiten Screenshot sprachliche Einheiten auf beide Zeilen aufgeteilt wurden. Die Teilnehmenden sollten die ihrer Meinung nach besseren Untertitel auswählen und auch begründen, ob sie ihre Entscheidung aufgrund von sprachlichen Erwägungen oder basierend auf der Form der Untertitel getroffen hatten (vgl. Szarkowska & Gerber-Morón 2018: 3f.).

Als Ergebnis zeigte sich, dass alle hörenden Teilnehmenden die nach syntaktischen Regeln erfolgten Zeilenumbrüche bevorzugten. Am auffälligsten war diese Präferenz bei der Wahl zwischen Untertiteln, bei denen sich ein Satz über zwei Zeilen erstreckte, und solchen, bei denen ein Satz am Zeilenende abgeschlossen war. Die meisten Teilnehmenden zeigten eine Vorliebe für Untertitel, die nach semantischen und syntaktischen Strukturen statt nach formalen Kriterien segmentiert waren. Das Eye-Tracking ergab, dass bei den kohärent segmentierten Untertiteln die Verweildauer bei Nominalphrasen am längsten war, während bei jenen mit inkohärenter Segmentierung die Verbalphrasen die längste Verweildauer aufwiesen. Insgesamt schienen Nominalphrasen mit kohärenten Zeilenumbrüchen schwieriger zu verarbeiten zu sein, und Verbalphrasen in der umgekehrten Bedingung. Bei den spanischen Teilnehmenden, die – wie oben erwähnt – aus einem Synchronisationsland kommen, war generell eine längere durchschnittliche Fixationsdauer als bei den anderen Personen zu

beobachten (vgl. Szarkowska & Gerber-Morón 2018: 12). Diese Befunde sprechen für eine Präferenz kohärent segmentierter Untertitel und stützen damit die Hypothese der vorliegenden Arbeit, in deren Rahmen untersucht wird, ob solche Unterschiede auch in der wahrgenommenen Rezeption erkennbar sind.

Die meisten schwerhörigen und gehörlosen Teilnehmenden favorisierten ebenso bei fast allen sprachlichen Einheiten die syntaktisch segmentierten Untertitel – auch hier vor allem jene, in denen die Sätze nach einer Zeile abgeschlossen waren. Die hörenden und schwerhörigen Personen gaben an, Untertitel auf der Grundlage von semantischen und syntaktischen Phrasen auszuwählen, während für gehörlose Teilnehmende die Form eine größere Rolle spielte (vgl. Szarkowska & Gerber-Morón 2018: 16).

Auch in den Interviews sprach sich mehr als die Hälfte aller Teilnehmenden für die Zeilenumbrüche aus, die syntaktischen und semantischen Regeln folgten. Diejenigen, die umgekehrt Untertitel mit inkohärenter Segmentierung vorzogen, gaben an, dass ihnen diese beim Lesen – insbesondere bei manchen sprachlichen Einheiten – einen Eindruck der Kontinuität vermittelte. Viele Teilnehmende merkten an, dass bei der Segmentierung Syntax und Form in einem ausgewogenen Verhältnis stehen sollten; Untertitel sollten entsprechend den gewohnten Gedankengängen gegliedert werden, damit sie so schnell wie möglich gelesen werden können. Andere erwähnten, dass die Segmentierung für langsame Leser:innen ein wichtiger Aspekt sein könnte. Eine der hörbeeinträchtigten Teilnehmenden war der Meinung, dass Zeilenumbrüche ihren Wert hätten, aber meistens weniger wichtig würden, wenn man schnell lese (vgl. Szarkowska & Gerber-Morón 2018: 16f.).

Im Vergleich zur Studie von Perego et al. (2010) fällt auf, dass in der damaligen Studie keine Unterschiede zwischen syntaktisch kohärent und inkohärent segmentierten Nominalphrasen gefunden wurden, was die Anzahl der Fixierungen oder die Verweildauer anbelangt. Die Fixationsdauer war bei den ungünstigen Zeilenumbrüchen sogar ein wenig länger (vgl. Szarkowska & Gerber-Morón 2018: 17). Somit stehen die Ergebnisse von Perego et al. (2010) nicht nur im Gegensatz zu jenen von Szarkowska & Gerber-Morón (2018), sondern auch zu den Erkenntnissen von Rajendran et al. (2013). Die uneinheitlichen Befunde verdeutlichen, dass der Effekt der Segmentierung stark von den jeweiligen Untersuchungsbedingungen abhängen kann und bilden den Ausgangspunkt der vorliegenden Untersuchung.

4.4 Der Einfluss der Textsegmentierung auf das Lesen von Untertiteln

Szarkowska et al. (2018) befassten sich mit den Auswirkungen der Textsegmentierung auf die Verarbeitung von Untertiteln bei verschiedenen Zusehergruppen. Die Rahmenbedingungen und die Zusammensetzung der Teilnehmenden dieser Studie sind jenen von Szarkowska & Gerber-Morón (2018) ähnlich. Der wesentliche Unterschied besteht darin, dass als Untersuchungsmaterial keine statischen Screenshots, sondern Videos verwendet wurden. In einem der beiden Videos waren die Untertitel – wie auch in der vorliegenden Arbeit – nach syntaktischen Regeln segmentiert, im zweiten nicht. Gemessen wurde die kognitive Belastung der Teilnehmenden (Wahrnehmung von Schwierigkeit, Anstrengung und Frustration auf einer Skala von 1–7), das Verständnis sowie eine Reihe von Eye-Tracking-Variablen (absolute und proportionale Lesezeit der Untertitel, Anzahl der Fixationen). Um beurteilen zu können, welche Rolle dem Hörvermögen und der Erfahrung mit untertitelten Inhalten dabei zukommt, wurden auch hier Menschen mit verschiedenen Erstsprachen (Experiment 1) sowie taube und hörgeschädigte Personen (Experiment 2) einbezogen (vgl. Szarkowska et al. 2018: 4f.).

Die kognitive Belastung wurde bei inkohärenten Zeilenumbrüchen von allen Teilnehmenden in Experiment 1 als höher angegeben, doch nur bei den spanischen Erstsprachler:innen war der Haupteffekt der Segmentierung statistisch signifikant. Generell waren alle drei Aspekte der kognitiven Belastung bei den nicht segmentierten Untertiteln durchwegs stärker ausgeprägt, unabhängig vom sprachlichen Hintergrund. Bezüglich des Verständnisses war kein Haupteffekt der Segmentierung zu beobachten, allerdings schnitt die polnische Gruppe signifikant besser ab als die spanischen Testpersonen (vgl. Szarkowska et al. 2018: 9). Diese Ergebnisse lassen vermuten, dass sich Unterschiede in der Segmentierung zwar auf die wahrgenommene kognitive Belastung auswirken können, aber nicht zwangsläufig auf das Verständnis.

Beim Eye-Tracking wurde festgestellt, dass der Blick der Teilnehmenden während des Videos mit inkohärenter Segmentierung häufiger zu den Untertiteln zurückkehrte (Revisits). Es ergab sich kein Effekt der Segmentierung auf andere Eye-Tracking-Variablen, jedoch ein Haupteffekt der Sprache auf die Lesezeit, die Fixationsdauer und -anzahl. Spanische Erstsprachler:innen hatten ihre Augen im Vergleich zur englischen und polnischen Gruppe signifikant länger auf den Bereich der Untertitel gerichtet. Insgesamt deuten die Ergebnisse darauf hin, dass die Verarbeitung der Untertitel für polnische Teilnehmende am wenigsten anstrengend und für spanische am anstrengendsten war (vgl. Szarkowska et al. 2018: 10).

Im zweiten Experiment (mit hörenden, schwerhörigen und gehörlosen Menschen) bedingten die inkohärenten Zeilenumbrüche ebenfalls bei allen Gruppen eine höhere kognitive Belastung, doch hörgeschädigte Teilnehmende schätzten ihre Frustrationswerte niedriger ein als hörende Personen. Sie wiesen allgemein die geringste kognitive Belastung sowie die höchsten Werte beim Verständnis auf. Entgegen den Vorannahmen zeigte sich ähnlich wie im ersten Experiment kein Haupteffekt der Segmentierung auf das Verständnis, aber auch auf keine der Eye-Tracking-Variablen. Gehörlose Teilnehmende hatten bei den kohärent segmentierten Untertiteln allerdings mehr Revisits als im Video mit inkohärenter Zeilensegmentierung (vgl. Szarkowska et al. 2018: 10ff.).

Im Zuge der Interviews erklärten mehrere spanische, polnische und gehörlose Teilnehmende, dass zusammenstehende Bedeutungseinheiten zur Lesbarkeit von Untertiteln beitragen, während getrennte semantische Einheiten falsche Erwartungen wecken und somit mehr Aufwand bei der Verarbeitung erfordern würden. Die Teilnehmenden passten ihre Lesestrategien an verschiedene Arten von Zeilenumbrüchen an. Einige gaben auch an, keine Unterschiede in der Segmentierung bemerkt zu haben, weil sie nie darauf geachtet hätten (vgl. Szarkowska et al. 2018: 12).

Für zukünftige Forschungen regen Szarkowska et al. (vgl. 2018: 13) an, die Auswirkungen von inkohärenten Zeilenumbrüchen nicht nur innerhalb eines einzelnen, sondern auch über zwei oder mehr aufeinanderfolgende Untertitel hinweg zu untersuchen – dabei könnte sich ein noch stärkerer Effekt auf die kognitive Belastung ergeben. Abschließend wird festgehalten, dass Untertitel mit inkohärenter Segmentierung trotz der höheren kognitiven Belastung unerwarteterweise nicht zu einem schlechteren Verständnis führten. Die Teilnehmenden erzielten unabhängig von der Art der Segmentierung ähnliche Verständniswerte. Auch bei Perego et al. (2010) schlugen sich Untertitel mit inkohärent segmentierten Substantivphrasen nicht negativ auf das Verständnis der Teilnehmenden nieder. Die Ergebnisse sprechen dafür, dass die Segmentierung nicht zwingend das Verständnis beeinträchtigt, sondern sich eher auf anderen Ebenen der Rezeption zeigt. In der vorliegenden Arbeit soll untersucht werden, ob sich vergleichbare Muster unter alltagsnahen Bedingungen in der subjektiven Wahrnehmung nachweisen lassen.

4.5 Synthese

Die vier vorgestellten Studien liefern insgesamt ein gemischtes Bild hinsichtlich der Effekte der Segmentierung auf die Rezeption von Untertiteln. Während Perego et al. (2010)

herausfanden, dass syntaktisch inkohärente Zeilenumbrüche keinen wesentlichen Einfluss auf die Informationsverarbeitung oder das Verständnis haben, weisen die anderen Studien auf bestimmte Effekte hin, die jedoch nicht einheitlich interpretiert werden können. So gelangten Rajendran et al. (2013) zu der Erkenntnis, dass die Segmentierung die Blickbewegungen und damit die Art der Verarbeitung verändert, ohne dass sie eine merkliche Auswirkung auf das Verständnis hat. Zu ähnlichen Ergebnissen gelangten auch Szarkowska et al. (2018), wo inkohärent segmentierte Untertitel mit einer erhöhten kognitiven Belastung in Verbindung stehen, allerdings kein entsprechender Effekt auf das inhaltliche Verständnis nachgewiesen werden konnte. In Ergänzung dazu stellen sie fest, dass kohärent segmentierte Untertitel von den Teilnehmenden bevorzugt werden und sich Abweichungen auch in den Blickbewegungen widerspiegeln. Die Unterschiede zwischen allen vier Studien betreffen somit weniger das Verständnis als vielmehr die Verarbeitung sowie die subjektive Beurteilung von Untertiteln. Diese Befunde lassen sich mit den in Kapitel 3 dargestellten Annahmen in Verbindung bringen, denen zufolge Effekte nicht zwingend im Verständnis sichtbar werden müssen, sondern auch in der Informationsverarbeitung und der Aufmerksamkeit auftreten können.

Methodisch weisen die Studien mehrere Gemeinsamkeiten auf, da sie überwiegend Eye-Tracking-Daten mit Fragebögen kombinieren, um sowohl objektive als auch subjektive Aspekte der Rezeption zu erfassen. Das Untersuchungsdesign variiert hingegen deutlich, etwa bezüglich des verwendeten Materials (statische Screenshots versus Videos), der Art der Präsentation der Untertitel und den spezifischen Manipulationsmethoden (z. B. Zeilenumbrüche im engeren Sinn im Vergleich zu Einheiten wie Wörtern, Phrasen oder Sätzen). Diese Unterschiede erschweren die direkte Vergleichbarkeit der Ergebnisse und erklären möglicherweise auch, warum einige Ergebnisse voneinander abweichen. Auch hier lässt sich ein Zusammenhang mit den in Kapitel 3 diskutierten Einflussfaktoren auf die Untertitelrezeption herstellen, im Speziellen mit der Rolle von Präsentationsform, Lesegeschwindigkeit und kognitiver Belastung.

Im Hinblick auf die Übertragbarkeit auf den Kontext der vorliegenden Arbeit erscheinen insbesondere die Studien von Szarkowska & Gerber-Morón (2018) und Szarkowska et al. (2018) relevant, da sie sich explizit mit Zeilenumbrüchen und deren Wahrnehmung in videobasierten Forschungsdesigns befassen. Die Studien von Perego et al. (2010) und Rajendran et al. (2013) sind dagegen nur bedingt übertragbar, da sie entweder auf einer vergleichsweise kleinen Stichprobe beruhen oder unter stark kontrollierten bzw. unnatürlichen Bedingungen durchgeführt wurden. Abgesehen davon bezieht sich keine der

Studien auf deutschsprachige Untertitel bzw. Rezipient:innen, was die Vergleichbarkeit zusätzlich einschränkt.

Aus der Gegenüberstellung der Studien lässt sich somit eine Forschungslücke ableiten. Zwar deuten die Ergebnisse der bisher durchgeführten Studien darauf hin, dass Segmentierung sich auf die kognitive Belastung, die Blickbewegungen sowie die Präferenzen der Teilnehmenden auswirken kann, jedoch bleibt unklar, inwieweit diese Effekte unter realitätsnahen Bedingungen und in konkreten Anwendungskontexten auftreten. Bislang scheinen nur wenige Untersuchungen vorzuliegen, die längere, zusammenhängende Videoinhalte sowie alltagsnahe Rezeptionssituationen berücksichtigen und dabei auch die subjektive Wahrnehmung stärker einbeziehen. Hier setzt die vorliegende Arbeit an und untersucht den Einfluss kohärenter und inkohärenter Segmentierung aus der Sicht eines deutschsprachigen Publikums.

5 Empirische Untersuchung

Im Folgenden steht die empirische Untersuchung im Fokus, in der analysiert wird, ob und wie sich die Wahrnehmung und das Verstehen von Untertiteln in Abhängigkeit von der Segmentierung ihrer Inhalte verändert. Zunächst werden das zugrunde liegende Forschungsdesign, die Methode und das Untersuchungsmaterial beschrieben; danach die Zusammensetzung der Stichprobe und der Ablauf der Erhebung dargestellt. Abschließend erfolgt die Auswertung der erhobenen Daten.

5.1 Forschungsdesign

Die vorliegende Untersuchung ist als quantitative empirische Studie konzipiert, die zwei experimentelle Untertitelbedingungen (kohärent versus inkohärent segmentiert) vergleicht. Damit wird der Einfluss der Segmentierung auf die Wahrnehmung, die Verständlichkeit und den Lesefluss von Untertiteln untersucht. Das Forschungsdesign ist als experimentelle Studie ohne Messwiederholungen (*Between-Subjects-Design*) angelegt (vgl. Döring 2023: 211), da jede teilnehmende Person ausschließlich einer der beiden Experimentalbedingungen zugeordnet wurde. Eine wiederholte Darbietung wäre mit dem Verlust an Aufmerksamkeit einhergegangen, weil sich die Bedingungen – abgesehen von der Segmentierung – nicht unterscheiden. Die Zuteilung zu diesen Bedingungen erfolgte nicht randomisiert, sondern über eine zeitversetzte Veröffentlichung der beiden Varianten des Fragebogens auf Umfrageplattformen.

Die Einschätzungen der Teilnehmenden in den beiden Untertitelbedingungen wurden systematisch miteinander verglichen. Unterschiede in der wahrgenommenen Verständlichkeit, im Lesefluss sowie in der allgemeinen Rezeption des Videos wurden anhand der Antworten im standardisierten Fragebogen erfasst und zwischen den Gruppen analysiert. Ziel der Untersuchung ist es, zu bestimmen, ob und in welchem Ausmaß eine kohärente Segmentierung zu einer positiveren Wahrnehmung und besseren Verständlichkeit beiträgt.

5.2 Forschungsmethode

Als Erhebungsmethode kam ein standardisierter Onlinefragebogen zum Einsatz, der überwiegend aus geschlossenen Fragen (5-stufige Ratingskalen sowie Ja-/Nein-/Teilweise-Antwortformaten) bestand und durch ergänzende offene Fragen erweitert wurde (vgl. Döring 2023: 399, 403). Die standardisierte Form der Befragung ermöglicht es, beide

Untertitelbedingungen direkt zu vergleichen und mehr Teilnehmende zu erreichen, die die Umfrage orts- und zeitunabhängig ausfüllen können.

Nach den Fragen zu den demografischen Angaben wurden die Teilnehmenden gebeten, sich das Video anzusehen und im Anschluss Fragen zum Verständnis mit Fokus auf die Untertitel zu beantworten. Am Ende bestand die Möglichkeit, in offenen Textfeldern Anmerkungen und etwaige Schwierigkeiten niederzuschreiben, wodurch es möglich war, zusätzliche qualitative Hinweise dazu erhalten, wo Probleme oder Unklarheiten auftraten. Die Bearbeitungszeit lag insgesamt bei etwa 30 Minuten.

Zur Kontrolle der Datengüte (vgl. Döring 2023: 403) wurde eine inhaltliche Frage eingefügt, die auf ein Detail im letzten Drittel des Videos Bezug nimmt. Dabei führte eine falsche Antwort nicht automatisch zum Ausschluss, sondern diente als Indikator für das allgemeine Textverständnis. Insbesondere bei Personen mit einer anderen Erstsprache als Deutsch wurde berücksichtigt, dass die falsche Antwort auf sprachliche Verständnisschwierigkeiten zurückzuführen sein könnte. Bei auffälligen Antwortmustern (durchgehend identische Werte auf der Ratingskala) wurde der Datensatz allerdings ausgeschlossen.

5.3 Untersuchungsmaterial

Als Untersuchungsmaterial diente ein 22-minütiges Video der Non-Profit-Medienorganisation TED (*Technology, Entertainment, Design*) in französischer Sprache, das ursprünglich mit französischen Untertiteln auf YouTube veröffentlicht wurde. Es handelt sich dabei um einen Vortrag mit zusammenhängendem Diskurs, in dem Untertitel für das Textverständnis zentral sind und Effekte der Zeilensegmentierung besonders deutlich hervortreten können. Die Länge des Films erlaubt eine Rezeption über einen längeren Zeitraum, wodurch einzelne Verständnisschwankungen leichter ausgeglichen werden.

Im Vortrag berichtet die Organisatorin der alle vier Jahre stattfindenden Segelregatta *Vendée Globe* von dem Mut der Segler:innen und den Herausforderungen, die sie zu bestehen hatten. Es wurde bewusst ein neutrales Thema mit einem angenehmen und emotionalen Erzählstil gewählt, um eine breite Zielgruppe anzusprechen und deren Aufmerksamkeit möglichst bis zum Ende des Videos zu halten.

Freiwillige Untertitler:innen können bei TED den plattformeigenen Untertitel-Editor nutzen. Ausgehend von den bereitgestellten französischen Untertiteln wurden zwei deutschsprachige Untertitelversionen erstellt (siehe Anhang), die sich nur in der

Segmentierung unterscheiden. In der ersten Version wurden die Untertitel kohärent – d. h. unter Berücksichtigung syntaktischer und semantischer Einheiten – segmentiert. In der zweiten Version erfolgte eine inkohärente Segmentierung, bei der auch innerhalb zusammengehöriger Wort- und Satzgruppen Zeilenumbrüche gesetzt wurden. Inhalt, Einblenddauer und Gestaltung der Untertitel sind in beiden Versionen identisch. Die Untertitel wurden über YouTube Studio in das Video integriert und am unteren Bildrand eingeblendet. Im Fragebogen war immer nur eine der beiden Versionen des Videos verlinkt, um nachvollziehen zu können, wie viele Personen jeweils teilgenommen haben. Die Teilnehmenden wurden gebeten, die Studie am PC/Laptop auszufüllen, um einerseits eine gute Lesbarkeit und einheitliche Darstellung zu gewährleisten, und andererseits auszuschließen, dass die Untertitel nicht durch automatische Übersetzung zusätzlich auf Englisch aufscheinen.

5.4 Versuchsgruppe

Nach der Datenbereinigung konnten insgesamt 608 Datensätze für die Auswertung verwendet werden. Davon sahen 306 Teilnehmende die Version mit inkohärent segmentierten Untertiteln und 302 jene mit kohärent segmentierten Untertiteln. 557 Personen gaben Deutsch als Erstsprache an; die Gruppe der Nicht-Erstsprachler:innen war mit 51 Personen erwartungsgemäß deutlich kleiner. Analysen nach Erstsprache wurden daher explorativ durchgeführt.

Rekrutiert wurden die Teilnehmenden über Umfrageplattformen (hauptsächlich SurveyCircle sowie in geringerem Umfang SurveySwap und PollPool), auf denen die Befragung für alle registrierten Nutzer:innen sichtbar war. Eine finanzielle Vergütung wurde nicht angeboten, da die Umfrageteilnahmen auf Gegenseitigkeit beruht. Die meisten Teilnehmenden (492) waren zwischen 21 und 30 Jahre alt. Zwei Drittel (402) waren Frauen, ein Drittel (201) Männer, und fünf Personen identifizierten sich mit keiner dieser beiden Kategorien.

5.5 Datenbereinigung

Zur Sicherstellung einer hohen Datenqualität wurden vor der Auswertung einige Kriterien definiert (Kontrollfrage, Antwortdifferenzierung, inhaltliche Konsistenz). Uneingeschränkt in die Analyse einbezogen wurden unauffällige und valide Datensätze, in denen die Kontrollfrage richtig beantwortet worden war, differenzierte Werte auf den Ratingskalen

vorlagen (keine Gleichförmigkeit) und inhaltliche Konsistenz gegeben war. Ebenso wurden Datensätze behalten, die leichte Auffälligkeiten aufwiesen (z. B. einzelne extreme Werte oder geringe Differenzierung), jedoch insgesamt inhaltlich nachvollziehbar waren. Datensätze mit falsch beantworteter Kontrollfrage wurden in die Hauptanalyse aufgenommen, sofern dennoch ein differenziertes und konsistentes Antwortmuster erkennbar war. In einigen Fällen ließen die Antworten auf die optionalen offenen Fragen auf eine gewissenhafte Bearbeitung des Fragebogens schließen. Nicht berücksichtigt wurden Datensätze, in denen die Kontrollfrage falsch beantwortet worden war und zusätzlich stark gleichförmige Antworten (*Straightlining*) und/oder inhaltliche Widersprüche festgestellt wurden.

5.6 Auswertungsverfahren

Auf Grundlage der bereinigten Daten erfolgten zunächst deskriptivstatistische Auswertungen (vgl. Döring 2023: 601–603). In der Hauptanalyse wurden mit der Gesamtstichprobe beide Untertitelversionen miteinander verglichen. Zur Überprüfung der Robustheit der Ergebnisse wurden die Analysen zusätzlich nach Ausschluss aller Datensätze mit falsch beantworteter Kontrollfrage wiederholt. Separate Subgruppenanalysen für nicht-erstsprachliche Teilnehmende konnten aufgrund der ungleichen Gruppengrößen nur explorativ durchgeführt werden, um Hinweise auf mögliche Unterschiede in Abhängigkeit von der Sprachkompetenz zu gewinnen.

Die Daten wurden in Microsoft Excel tabellarisch erfasst und ausgewertet; dabei kamen unter anderem Filterfunktionen und Pivot-Tabellen zum Einsatz. Die Antworten auf die Fragen mit Ratingskalen wurden mit den Werten 1–5 codiert und anhand von Mittelwerten und Standardabweichungen deskriptiv analysiert. Für die kategorialen Fragen wurden Häufigkeitsverteilungen berechnet. Zur Überprüfung von Abweichungen zwischen den beiden Versionen wurden t-Tests für unabhängige Stichproben (vgl. Döring 2023: 693f.) bzw. bei kategorialen Fragen ein Chi-Quadrat-Test angewendet (Signifikanzniveau $p < .05$) (vgl. Döring 2023: 935, 943).

Die Ergebnisse der beiden Untertitelbedingungen (kohärent versus inkohärent segmentiert) wurden deskriptiv miteinander verglichen, um Unterschiede in der Verständlichkeit, im Lesefluss und in der Wahrnehmung der Segmentierung zu ermitteln. Offene Antworten wurden ergänzend herangezogen, um Auffälligkeiten oder technische Probleme zu identifizieren, flossen aber nicht systematisch in die quantitative Auswertung ein.

6 Ergebnisse

Im Folgenden werden die Ergebnisse der Online-Befragung deskriptiv dargestellt. Der Fokus liegt auf dem Vergleich der beiden Untertitelbedingungen in Bezug auf das wahrgenommene Verständnis des Videos, die Verständlichkeit der Untertitel sowie den Lesefluss.

6.1 Hauptanalyse

Im Rahmen der Hauptanalyse wurden die beiden Untertitelversionen (A mit inkohärent segmentierten und B mit kohärent segmentierten Untertiteln) hinsichtlich mehrerer Dimensionen miteinander verglichen. Tabelle 1 zeigt die Mittelwerte, Standardabweichungen und Ergebnisse der t-Tests für unabhängige Stichproben.

Tabelle 1: Gruppenvergleiche zwischen den Untertitelversionen (Hauptanalyse). M = Mittelwert, SD = Standardabweichung. Für die skalierten Variablen wurden t-Tests für unabhängige Stichproben durchgeführt; für die kategoriale Variable „schwierige Textstellen“ wurde ein Chi-Quadrat-Test berechnet.

Variable	Version A M (SD)	Version B M (SD)	p
Allgemeines Verständnis des Videos	3,58 (1,18)	3,67 (1,16)	0,23
Verständlichkeit der Untertitel	3,95 (1,17)	4,00 (1,10)	0,57
Passung von Bild, Ton und Text	3,95 (1,03)	3,95 (1,02)	0,94
Lesefluss der Untertitel	3,69 (1,15)	3,77 (1,04)	0,32
Häufigkeit des Zurückspulens	3,69 (1,26)	3,71 (1,26)	0,88
Wahrnehmung unnatürlicher Segmentierung	3,58 (1,14)	3,60 (1,12)	0,82
Schwierige Textstellen	-	-	0,71

Abbildung 1 auf der folgenden Seite veranschaulicht die Mittelwerte der beiden Untertitelversionen im Vergleich über alle Dimensionen.

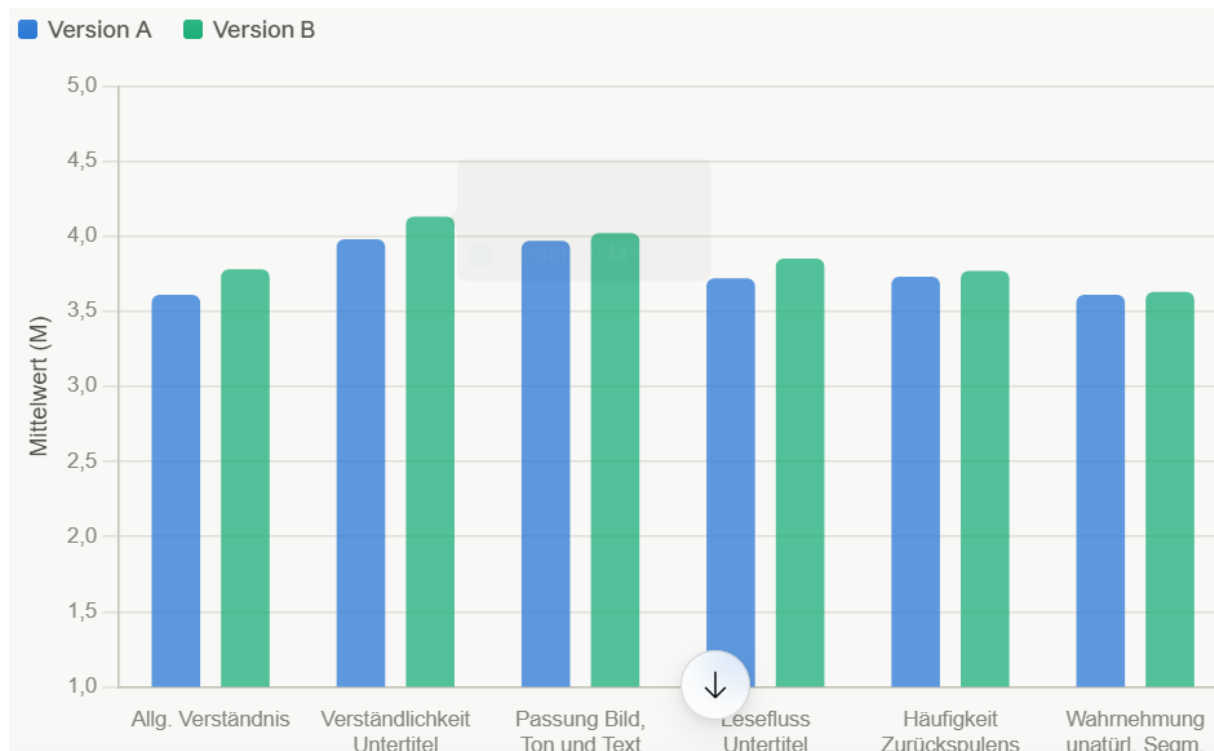


Abbildung 1: Mittelwerte der beiden Untertitelversionen in den untersuchten Dimensionen (Hauptanalyse).

Für keine der untersuchten Variablen ergaben sich signifikante Unterschiede zwischen den Versionen A und B (alle $p > 0,05$), auch wenn die Werte bei Version B in fast allen Dimensionen geringfügig höher sind. Dies betrifft sowohl die skalierten Variablen als auch die kategoriale Variable „schwierige Textstellen“, die mit dem Chi-Quadrat-Test analysiert wurde. Insgesamt lassen sich somit keine Hinweise darauf finden, dass sich die beiden Untertitelversionen hinsichtlich der untersuchten Dimensionen systematisch unterscheiden.

In den offenen Antworten wurden immer wieder Aspekte wie Einblenddauer, Lesefluss, Satzlänge und inhaltliche Komplexität als potenziell erschwerend genannt.

6.2 Robustheitsprüfung der Ergebnisse

Mit der Robustheitsprüfung sollte überprüft werden, ob sich die Ergebnisse der Hauptanalyse durch den Ausschluss potenziell unaufmerksamer Teilnehmender verändern.

Tabelle 2: Gruppenvergleiche zwischen den Untertitelversionen nach Ausschluss der Datensätze mit falsch beantworteter Kontrollfrage (Robustheitsprüfung).

Variable	Version A M (SD)	Version B M (SD)	p
Allgemeines Verständnis des Videos	3,61 (1,18)	3,78 (1,15)	0,09
Verständlichkeit der Untertitel	3,98 (1,18)	4,13 (1,03)	0,11
Passung von Bild, Ton und Text	3,97 (1,03)	4,02 (1,02)	0,59
Lesefluss der Untertitel	3,72 (1,13)	3,85 (1,04)	0,16
Häufigkeit des Zurückspulens	3,73 (1,26)	3,77 (1,25)	0,72
Wahrnehmung unnatürlicher Segmentierung	3,61 (1,14)	3,63 (1,13)	0,79
Schwierige Textstellen	-	-	0,38

Abbildung 2 zeigt die Mittelwerte der beiden Untertitelversionen nach Ausschluss der Datensätze mit falsch beantworteter Kontrollfrage.

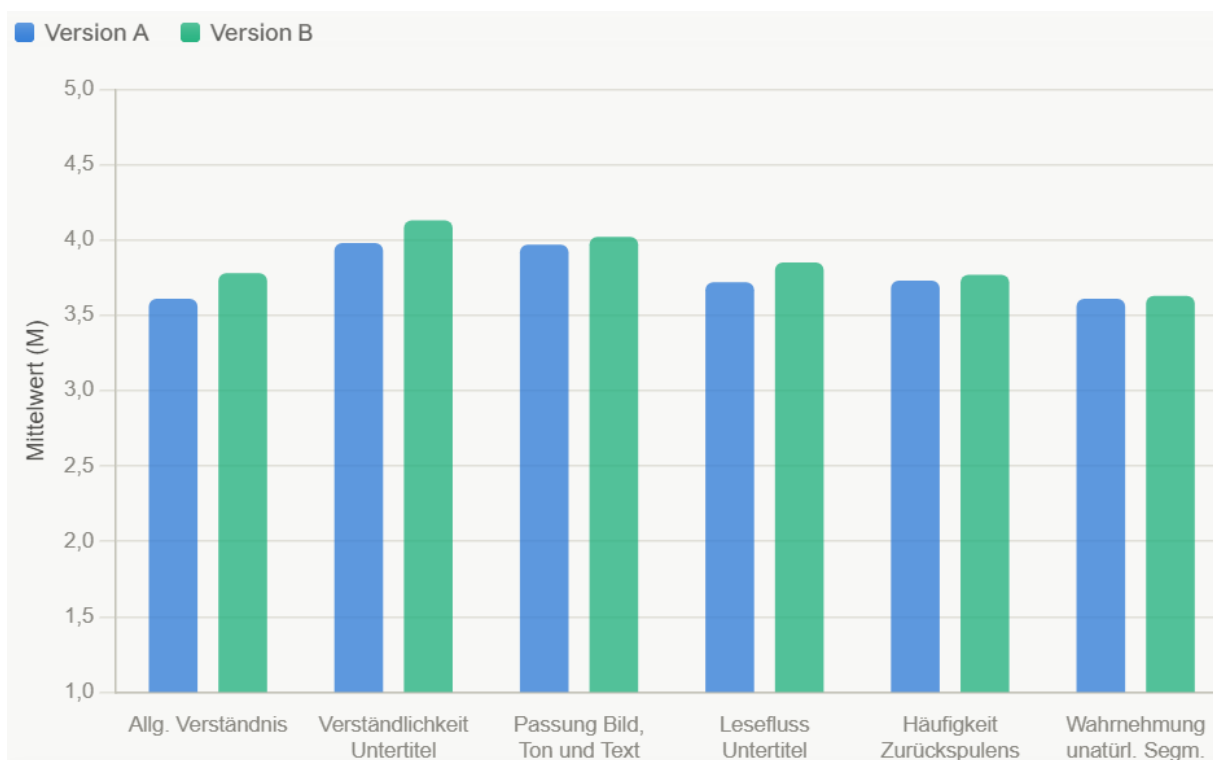


Abbildung 2: Mittelwerte der beiden Untertitelversionen in den untersuchten Dimensionen nach Ausschluss der Datensätze mit falsch beantworteter Kontrollfrage (Robustheitsprüfung).

Deskriptiv zeigen sich auch in der Robustheitsprüfung durchgehend leicht höhere Werte für Version B, wobei die Differenzen gegenüber der Hauptanalyse etwas stärker ausfielen.

Die inferenzstatistischen Analysen bestätigten dieses Bild. Für keine der skalierten Variablen können signifikante Effekte zwischen den beiden Untertitelversionen festgestellt werden (alle $p > 0,05$). Die kategoriale Variable „schwierige Textstellen“ wurde erneut mittels Chi-Quadrat-Test berechnet, der ebenfalls keine signifikante Abweichung in der Verteilung der Antworten zwischen den beiden Untertitelversionen ergab (χ^2 -Test, $p = 0,38$).

Die Ergebnisse der Robustheitsprüfung entsprechen weitgehend jenen der Hauptanalyse und deuten somit darauf hin, dass deren Befunde stabil sind und nicht maßgeblich durch unaufmerksame Teilnehmende beeinflusst werden.

6.3 Explorative Ergebnisse nach Erstsprache

Um eventuelle Hinweise zu Unterschieden zwischen den Untertitelversionen in Abhängigkeit von der Erstsprache zu erhalten, wurde eine Subgruppenanalyse mit ausschließlich nicht-erstsprachlichen Teilnehmenden durchgeführt.

Tabelle 3: Gruppenvergleiche zwischen den Untertitelversionen (Subgruppenanalyse)

Variable	Version A M (SD)	Version B M (SD)	p
Allgemeines Verständnis des Videos	2,93 (1,31)	3,62 (1,16)	0,06
Verständlichkeit der Untertitel	3,43 (1,28)	4,67 (1,19)	0,10
Passung von Bild, Ton und Text	3,77 (1,04)	3,86 (0,79)	0,74
Lesefluss der Untertitel	3,30 (1,32)	3,76 (1,04)	0,19
Häufigkeit des Zurückspulens	3,13 (1,25)	3,57 (1,03)	0,19
Wahrnehmung unnatürlicher Segmentierung	3,40 (1,33)	3,71 (1,06)	0,37
Schwierige Textstellen	-	-	0,005

In Abbildung 3 sind die Mittelwerte der beiden Untertitelversionen für Personen ohne Deutsch als Erstsprache dargestellt.

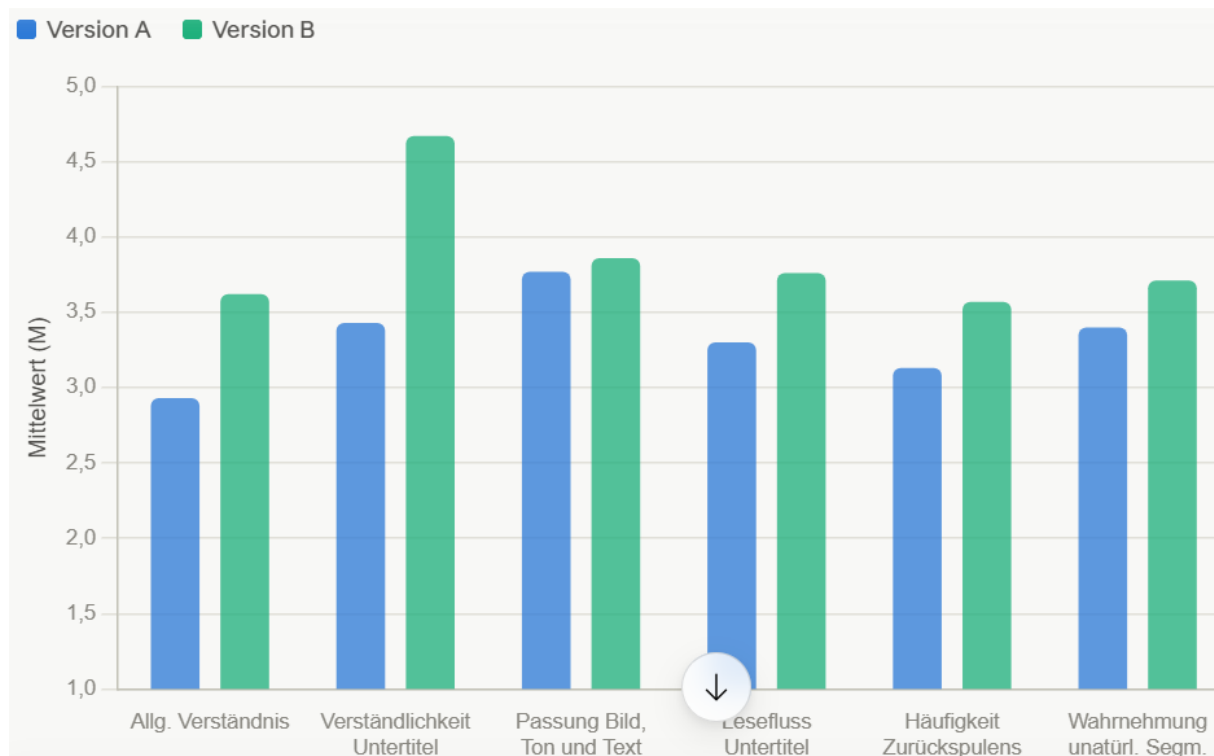


Abbildung 3: Mittelwerte der beiden Untertitelversionen in den untersuchten Dimensionen nach Erstsprache (explorative Subgruppenanalyse)

Die Mittelwerte der untersuchten Variablen weichen in den beiden Untertitelversionen in mehreren Dimensionen voneinander ab. Für Version B wurden insgesamt tendenziell höhere Werte erzielt als für Version A, vor allem im Hinblick auf das allgemeine Verständnis des Videos sowie die Verständlichkeit und den Lesefluss der Untertitel. Geringer fielen die Effekte für die Passung von Bild, Ton und Text sowie für die Wahrnehmung unnatürlicher Segmentierung aus.

Die inferenzstatistischen Analysen ergaben für die skalierten Variablen keine signifikanten Abweichungen zwischen den beiden Versionen (alle $p > 0,05$). Sie liefern daher lediglich begrenzte Hinweise auf Unterschiede zwischen den Untertitelversionen in Abhängigkeit von der Erstsprache. Nur für die Variable „schwierige Textstellen“ zeigte sich in der Subgruppenanalyse ein signifikanter Unterschied in der Verteilung der Antworten zwischen den Versionen A und B (χ^2 -Test, $p = 0,005$). Dies weist auf Unterschiede in der wahrgenommenen Schwierigkeit zwischen den beiden Untertitelversionen hin.

7 Diskussion

Obwohl deskriptiv in beiden Analysen leicht höhere Werte für die kohärent segmentierte Version nachgewiesen wurden, konnten diese Effekte statistisch nicht abgesichert werden. Die Robustheitsprüfung stützte diese Befunde, was für deren Robustheit spricht und darauf schließen lässt, dass die Ergebnisse nicht durch einzelne unsichere Datensätze verzerrt wurden. Daraus folgt, dass die der vorliegenden Arbeit zugrunde liegende Hypothese nicht bestätigt werden konnte. Dieses Ergebnis steht im Einklang mit den Befunden von Perego et al. (vgl. 2010, Kap. 4.1) sowie Szarkowska et al. (vgl. 2018, Kap. 4.4), die ebenfalls keinen bzw. nur einen sehr geringen Einfluss der Zeilensegmentierung auf das Verständnis nachweisen konnten. Gleichzeitig verdeutlichen Rajendran et al. (vgl. 2013, Kap. 4.2), dass sich Effekte eher in der Art der Verarbeitung als im Verständnis selbst niederschlagen.

In der explorativen Subgruppenanalyse fällt auf, dass die kohärent segmentierte Version in mehreren Dimensionen besser abgeschnitten hat. Dies entspricht der ursprünglichen Annahme, dass kohärent segmentierte Untertitel insbesondere für nicht-erstsprachliche Rezipient:innen von Vorteil sein müssten, ist allerdings ebenfalls nicht statistisch abgesichert. Ähnliche Tendenzen berichten auch Szarkowska et al. (vgl. 2018, Kap. 4.4), die insbesondere bei nicht muttersprachlichen Rezipient:innen Unterschiede in der wahrgenommenen kognitiven Belastung feststellen konnten, während sich Effekte auf das Verständnis nur eingeschränkt zeigten. Für die in der vorliegenden Studie beobachteten Unterschiede lässt sich aufgrund der vergleichsweise kleinen Stichprobe keine eindeutige Erklärung ableiten. Die Ergebnisse sind daher mit Vorsicht zu interpretieren und nur als Indiz für mögliche Unterschiede zu betrachten, die in zukünftigen Studien gezielter untersucht werden könnten.

Neben den quantitativen Ergebnissen lassen sich in den Daten auch einzelne Antwortmuster identifizieren, die auf weitere Aspekte der Untertitelrezeption hinweisen. So deuten die deskriptiven Muster in den Daten darauf hin, dass eine hohe wahrgenommene Qualität der Untertitel nicht zwingend mit einem besseren Gesamtverständnis einhergeht. Bei beiden Untertitelversionen konnten einige Teilnehmende dem Inhalt zwar schlecht folgen, bewerteten die Untertitel aber dennoch als gut. Dieses Muster stimmt mit den Ergebnissen von Perego et al. (vgl. 2010, Kap. 4.1) sowie Szarkowska et al. (vgl. 2018, Kap. 4.4) überein, die ebenfalls zeigen, dass Unterschiede in der kognitiven Belastung nicht zwingend mit Veränderungen im Verständnis einhergehen.

Ergänzend dazu sind in den offenen Antworten ebenfalls mehrere sich wiederholende Muster zu erkennen. Häufig wurde die Einblenddauer als zu kurz empfunden, während die Länge und Komplexität einzelner Untertiteleinheiten vor allem bei Sätzen mit hoher Informationsdichte als erschwerend betrachtet wurden. Auch Szarkowska & Gerber-Morón (vgl. 2018, Kap. 4.3) beschreiben vergleichbare subjektive Bewertungen, wobei syntaktisch kohärent segmentierte Untertitel von den Teilnehmenden als lesefreundlicher wahrgenommen wurden. Zahlreiche Anmerkungen in der vorliegenden Studie beziehen sich außerdem auf den Lesefluss und die Segmentierung, die etwa als „abgehackt“ oder „unnatürlich getrennt“ beschrieben wurde. Die Antworten offenbaren jedoch auch, dass Verständnisschwierigkeiten vielfach nicht allein auf die Segmentierung zurückgeführt wurden, sondern auch auf Aspekte wie Fachbegriffe, Informationsdichte oder fehlendes Vorwissen. Diese Beobachtungen legen den Schluss nahe, dass die Verarbeitung audiovisueller Inhalte von einem Zusammenspiel mehrerer Faktoren abhängt und Effekte der Segmentierung nicht isoliert auftreten.

Außerdem finden sich in einzelnen Antworten Inkonsistenzen zwischen der Bewertung des Leseflusses, der Segmentierung und des Gesamtverständnisses. Einige Teilnehmende berichten von gut verständlichen Untertiteln bei gleichzeitig erschwertem Gesamtverständnis oder umgekehrt. Dies weist darauf hin, dass potenzielle Effekte der Segmentierung nicht immer bewusst wahrgenommen oder eindeutig bestimmten Aspekten zugeordnet werden. Gute Segmentierung erleichtert zwar den Lesefluss, garantiert aber nicht automatisch ein hohes inhaltliches Verständnis. Vergleichbare Inkonsistenzen zwischen subjektiver Bewertung und tatsächlichem Verständnis stellen auch Szarkowska et al. (vgl. 2018, Kap. 4.4) fest, wo Unterschiede in der wahrgenommenen Belastung nicht mit entsprechenden Unterschieden im Verständnis korrelierten.

Die Ergebnisse lassen vermuten, dass Segmentierung eine Rolle für die Untertitelrezeption spielt, wobei ihre Bedeutung in Wechselwirkung mit weiteren Faktoren zu betrachten ist. Diese Einschätzung entspricht auch der Synthese der bisherigen Forschung (Kap. 4.5), wonach die Effekte der Segmentierung stark von zusätzlichen Einflussfaktoren wie Lesegeschwindigkeit und Informationsdichte abhängen. Für die Untertitelpraxis bedeutet dies, dass neben einer möglichst kohärenten Segmentierung auch Aspekte wie Lesegeschwindigkeit, Informationsdichte und sprachliche Komplexität berücksichtigt werden sollten.

Zusammenfassend bleibt festzuhalten, dass sich in der Gesamtstichprobe keine signifikanten Unterschiede zwischen kohärent und inkohärent segmentierten Untertiteln

zeigten, während sich für nicht-erstsprachliche Rezipient:innen differenziertere Effekte andeuten könnten.

8 Fazit

Ziel dieser Arbeit war es, die Wirkung von kohärenter und inkohärenter Segmentierung von Untertiteln auf die Rezeption audiovisueller Inhalte zu vergleichen. Dafür wurden zwei Untertitelversionen hinsichtlich mehrerer Dimensionen des Verständnisses und der Verarbeitung untersucht.

Die Befunde der quantitativen Analyse ergaben in der Gesamtstichprobe keine statistisch signifikanten Unterschiede zwischen den beiden Versionen. Dies deutet darauf hin, dass die Segmentierung der Untertitel unter den gegebenen Bedingungen keinen nachweisbaren Effekt auf die Rezeption hatte. Dieses Ergebnis steht in Übereinstimmung mit früheren Studien (vgl. Perego et al. 2010, Kap. 4.1; Szarkowska et al. 2018, Kap. 4.4), die ebenfalls keinen bzw. nur einen sehr eingeschränkten Einfluss der Segmentierung auf das inhaltliche Verständnis feststellen konnten. Gleichzeitig lagen in den deskriptiven Analysen konsistent leicht höhere Werte für die kohärent segmentierte Version vor, die jedoch inferenzstatistisch nicht abgesichert werden konnten.

Die Robustheitsprüfung bestätigte diese Befunde und unterstreicht damit die Robustheit der Ergebnisse. Die explorative Subgruppenanalyse für nicht-erstsprachliche Teilnehmende lieferte hingegen Hinweise auf differenziertere Effekte, die für mögliche Vorteile kohärenter Segmentierung sprechen könnten. Auch Szarkowska et al. (vgl. 2018, Kap. 4.4) finden heraus, dass insbesondere nicht-erstsprachliche Rezipient:innen sensibler auf Unterschiede in der Untertitelsegmentierung reagieren können, wenngleich sich diese Effekte nicht einheitlich im Verständnis niederschlagen. Aufgrund der begrenzten Stichprobe sind die Ergebnisse der vorliegenden Studie allerdings vorsichtig zu interpretieren.

Darüber hinaus legen sowohl die quantitativen als auch die qualitativen Antworten nahe, dass die Segmentierung nicht den einzigen Einflussfaktor auf die Rezeption darstellt. Vielmehr scheint die Verarbeitung audiovisueller Inhalte von einem Zusammenspiel mehrerer individueller und textbezogener Faktoren abhängig zu sein. Diese Erkenntnis deckt sich auch mit der in Kapitel 4.5 dargestellten Forschungssynthese, wonach Effekte der Segmentierung in Wechselwirkung mit weiteren Faktoren wie Lesestrategien und kognitiver Belastung zu betrachten sind.

Insgesamt kann die vorliegende Arbeit dazu beitragen, die Rolle der Untertitelsegmentierung differenzierter zu betrachten. Zwar ließ sich kein eindeutiger Haupteffekt belegen, dennoch zeigen die Ergebnisse, dass kontextuelle und individuelle Faktoren eine wichtige Rolle spielen und in zukünftigen Studien stärker berücksichtigt

werden könnten. Weitere Forschung könnte sich außerdem vermehrt auf nicht-erstsprachliche Rezipient:innen konzentrieren und entsprechende Untersuchungen mit größeren Stichproben durchführen.

Für die Translationswissenschaft bedeuten die Ergebnisse, dass etablierte Normen der Untertitelung – insbesondere die Bevorzugung syntaktischer Kohärenz bei der Segmentierung – nicht in jedem Fall einen messbaren Einfluss auf die Rezeption haben. Es scheint daher notwendig, solche Normen stärker empirisch zu überprüfen und im Kontext multimodaler Rezeptionsprozesse zu betrachten. Zugleich zeigt sich, dass audiovisuelle Übersetzung nicht allein auf der Ebene des Textes analysiert werden kann, sondern stets im Zusammenspiel mit visuellen und auditiven Informationen und unter Berücksichtigung der jeweiligen Zielgruppe verstanden werden muss.

Literaturverzeichnis

- Audiobird Support (2023). *Was ist ein Voiceover? Der ultimative Guide für Medienprofis*. Audiobird Europe. <https://audiobird.com/blog/was-ist-ein-voiceover-der-ultimate-guide-fuer-medienprofis> [25.04.2026]
- Beaugrande, Robert-Alain de & Dressler, Wolfgang Ulrich (1981). *Einführung in die Textlinguistik*. Tübingen: Max Niemeyer Verlag.
- Berg, Boris (2026). *Angst um den Job: Synchronsprecher im Streit mit Netflix*. Bayerischer Rundfunk. <https://www.br.de/nachrichten/deutschland-welt/angst-um-den-job-synchronsprecher-im-streit-mit-netflix> [25.04.2026]
- Busch, Brigitta (2021). Mehrsprachigkeit. 3. Auflage. Wien: Facultas.
- Delabastita, Dirk (1989). Translation and the Mass Media: Film and TV Translation as Evidence of Cultural Dynamics. *Babel* 35 (4), 193–218.
- Díaz Cintas, Jorge (1998): La labor subtituladora en tanto que instancia de traducción subordinada. In: Pilar Orero (Hg.). *Actes del III Congrés Internacional sobre Traducció*. Universitat Autònoma de Barcelona, 83–90.
- Díaz-Cintas, Jorge (2009). *New trends in audiovisual translation*. Briston & Tonawanda: Multilingual Matters.
- Díaz-Cintas, Jorge & Baños, Rocio (2017). Language and translation in film. In: Malmkjær, Kirsten (Hg.). *The Routledge handbook of translation studies and linguistics*. London / New York: Routledge, 313–326.
- Díaz-Cintas, Jorge & Remael, Aline (2021). *Subtitling: concepts and practices*. Abingdon / New York: Routledge.
- Döring, Nicola (2023). *Forschungsmethoden und Evaluation*. 6. Auflage. Berlin, Heidelberg: Springer.
- Gambier, Yves (Hg.) (1996). *Les transferts linguistiques dans les médias audiovisuels*. Lille: Presses universitaires du Septentrion.
- Gambier, Yves & Gottlieb, Henrik (2001). *(Multi) Media Translation: Concepts, practices and research*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Georgakopoulou, Panayota (2012). Challenges for the audiovisual industry in the digital age: the ever-changing needs of subtitle production. *The Journal of Specialised Translation* 17, 78–103.
- Georgakopoulou, Panayota (2019). Technologization of audiovisual translation. In: Pérez González, Luis (Hg.). *The Routledge handbook of audiovisual translation*. London: Routledge, 516–539.
- Gottlieb, Henrik (1994). Subtitling: Diagonal Translation. *Perspectives, studies in translatology* 2 (1), 101–121.
- Gottlieb, Henrik (1997): *Subtitles, Translation & Idioms*. Dissertation. Kopenhagen.
- Gottlieb, Henrik (2002). Untertitel. Das Visualisieren filmischen Dialogs. In: Friedrich, Hans-Edwin. *Schrift und Bild im Film*. Bielefeld: Aisthesis-Verlag.
- Gottlieb, Henrik (2018). The politics of audiovisual translation. In: Fernández, Fruela & Evans, Jonathan (Hg.). *The Routledge handbook of translation and politics*. London: Routledge, 323–342.
- Ivarsson, Jan & Carroll, Mary (1998). Code of Good Subtitling Practice. In: Ivarsson, Jan & Carroll, Mary. *Subtitling*. Simrishamn: TransEdit. <https://esist.org/resources/code-of-good-subtitling-practice> [25.04.2026]
- Jüngst, Heike E. (2020). *Audiovisuelles Übersetzen. Ein Lehr- und Arbeitsbuch*. 2. Auflage. Tübingen: Narr Francke Attempto Verlag.
- Karamitroglou, Fotios (1998). A proposed set of subtitling standards in Europe. *Translation Journal* 2 (2). <http://translationjournal.net/journal/04stndrd.htm> [25.04.2026]

- Kruger, Jan-Louis; Hefer, Esté; Matthew, Gordon (2013). Measuring the impact of subtitles on cognitive load: eye tracking and dynamic audiovisual texts. *ETSA '13. Proceedings of the 2013 conference on eye tracking South Africa*. Konferenzschrift. Kapstadt: ACM, 62–66.
- Kruger, Jan-Louis; Soto-Sanfiel, María T.; Doherty, Stephen; Ibrahim, Ronny (2016). Towards a cognitive audiovisual translology. Subtitles and embodied cognition. In: Muñoz Martín, Ricardo (Hg.). *Reembedding Translation Process Research*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 171–187.
- Laks, Simon (1957). *Le sous-titrage de films: sa technique, son esthétique*. Paris: Propriété de l'auteur.
- Marleau, Lucien (1982). Le sous-titres ... un mal nécessaire. *Meta* 27 (3), 271–285.
- Mayer, Richard E. (2021). *Multimedia learning*. 3. Ausgabe. Cambridge: Cambridge University Press.
- Mayer, Richard E. & Moreno, Roxana (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In: Plass, Jan L.; Moreno, Roxana & Brünken, Roland (Hg.). *Cognitive Load Theory*. New York: Cambridge University Press, 131–152.
- Mayoral, Roberto; Kelly, Dorothy; Gallardo, Natividad (1988). Concept of Constrained Translation. Non-Linguistic Perspectives of Translation. *Meta* 33 (3), 356–367.
- Novell-Smith, Geoffrey (1998). *Hollywood and Europe: economics, culture, national identity: 1945–95*. 1. Auflage. London: British Film Institute.
- ORF (2026). *ORF barrierefrei – Aktionsplan 2025-2028*.
<https://der.orf.at/unternehmen/humanitarian/barrierefreiheit/aktionsplan-barrierefreiheit104.html> [25.04.2026]
- Perego, Elisa; del Missio, Fabio; Porta, Marco; Mosconi, Mauro (2010). The cognitive effectiveness of subtitle processing. *Media Psychology* 13 (3), 243–272.
- Pérez-González, Luis (2019). Rewiring the circuitry of audiovisual translation. Introduction. In: Pérez-González, Luis (Hg.). *The Routledge handbook of audiovisual translation*. London / New York: Routledge.
- Pérez-González, Luis (2020). Audiovisual translation. In: Baker, Mona & Saldanha, Gabriela (Hg.). *Routledge encyclopedia of translation studies*. 3. Auflage. London, New York: Routledge, 30–34.
- Pötzsch, Florian (2026). *Streaming ohne Grenzen: Wie KI-gestützte Untertitel die Sprachbarriere bei internationalen Serien aufbrechen*. Auerbach Verlag und Infodienste GmbH. <https://www.digitalfernsehen.de/news/technik/tech/streaming-ohne-grenzen-wie-ki-gestuetzte-untertitel-die-sprachbarriere-bei-internationalen-serien-aufbrechen-1176970> [25.04.2026]
- Rajendran, Dhevi J.; Duchowski, Andrew; Orero, Pilar; Martinez, Juan; Romero-Fresco, Pablo (2013). Effects of text chunking on subtitling: a quantitative and qualitative examination. *Perspectives, studies in translology* 21 (1), 5–21.
- Reiss, Katharina (1981). Textbestimmung und Übersetzungsmethode. Entwurf einer Texttypologie. In: Wilss, Wolfram (Hg.). *Übersetzungswissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft, 76–91.
- Snell-Hornby, Mary (2006). *The turns of translation studies: New paradigms or shifting viewpoints?* Amsterdam/Philadelphia: John Benjamins.
- Szarkowska, Agnieszka (2016). *Report on the results of an online survey on subtitle presentation times and line breaks in interlingual subtitling. Part 1: Subtitlers*. Centre for Translation Studies, University College London. https://avt.ils.uw.edu.pl/wp-content/uploads/sites/170/2016/10/SURE_Report_Survey1.pdf [29.06.2026]

- Szarkowska, Agnieszka & Gerber-Morón, Olivia (2018). Line breaks in subtitling: an eye tracking study on viewer preferences. *Journal of eye movement research* 11 (3), 1–22.
- Szarkowska, Agnieszka; Gerber-Morón, Olivia; Woll, Benice (2018). The impact of text segmentation on subtitle reading. *Journal of eye movement research* 11 (4), 1–18.
- Szymanski, Przemyslaw (2026). *Amazon zieht Teile der KI-Synchronisation zurück*. Golem Media GmbH. <https://www.golem.de/news/prime-video-amazon-zieht-teile-der-ki-synchronisation-zurueck-2512-202893.html> [25.04.2026]
- TEDx Talks (2022). L'immersion dans le Vendée Globe | Laura Le Goff | TEDxLaBaule. [Videoclip 22'02"] <https://www.youtube.com/watch?v=Be01PQRQtOA&t=231s> [26.06.2026]
- Titford, Christopher (1982). Sub-Titling - Constrained Translation. *Lebende Sprachen* 27 (3), 113–116.

Abbildungs- und Tabellenverzeichnis

Tabelle 1.....	42
Tabelle 2.....	44
Tabelle 3.....	45
Abbildung 1.....	43
Abbildung 2.....	44
Abbildung 3.....	46

Anhang

Anhang 1: Untersuchungsmaterial

Das Untersuchungsmaterial basiert auf dem YouTube-Video:

TEDx Talks (2022). L'immersion dans le Vendée Globe | Laura Le Goff | TEDxLaBaule. [Videoclip 22'02"] <https://www.youtube.com/watch?v=Be0IPQRQtOA&t=231s>
[26.06.2026]

Für die vorliegende Studie wurden zwei deutschsprachige Untertitelversionen erstellt, die sich ausschließlich hinsichtlich der syntaktischen Zeilensegmentierung unterscheiden:

- Version A (inkohärente Segmentierung): <https://youtu.be/89Au1S3eVE0>
- Version B (kohärente Segmentierung): <https://youtu.be/b-wuWEpUkQs>

Anhang 2: Fragebogen - Wahrnehmung von Untertiteln

Liebe:r Teilnehmer:in, vielen Dank, dass du dir für diese Umfrage Zeit nimmst. Ich untersuche, wie unterschiedliche Arten der Untertitelung wahrgenommen und verstanden werden. Das Video dazu ist 22 Minuten lang - bitte sieh es dir aufmerksam und aus technischen Gründen am Laptop an (nicht am Handy). Danach folgt ein kurzer Fragebogen. Dieser ist anonym, deine Angaben werden ohne Rückschluss auf deine Identität und ausschließlich zu wissenschaftlichen Zwecken verwendet und nicht an Dritte weitergegeben.

1. Soziodemografische Angaben

- **Muttersprache:**
 - ☐ Deutsch
 - ☐ Andere
- **Wie gut verstehst du gesprochene Sprache auf Deutsch (nur falls Deutsch nicht deine Muttersprache ist)?**
 - ☐ Sehr gut
 - ☐ Gut
 - ☐ Mittel
 - ☐ Weniger gut
- **Alter:**
 - ☐ unter 20
 - ☐ 21–30
 - ☐ 31–40
 - ☐ 41–50
 - ☐ über 50
- **Geschlecht:**
 - ☐ weiblich ☐ männlich ☐ keine Angabe

2. Video

Bitte kehre nach dem Ansehen des Videos zum Fragebogen zurück: [Link]. Sollten keine oder nicht ausschließlich deutsche Untertitel angezeigt werden, deaktiviere bitte die automatische Übersetzung.

3. Fragen

Wie gut konntest du dem Inhalt des Videos insgesamt folgen?

1 = gar nicht gut ... 5 = sehr gut

Wie verständlich waren die Untertitel für dich formuliert?

1 = gar nicht gut ... 5 = sehr gut

Wie gut haben Bild, Ton und Untertitel zusammengepasst?

1 = gar nicht gut ... 5 = sehr gut

Wie angenehm war der Lesefluss der Untertitel für dich?

1 = sehr unangenehm ... 5 = sehr angenehm

Musstest du häufig zurückspringen oder zurückdenken, um den Sinn zu erfassen?

1 = sehr häufig ... 5 = kaum

Hattest du das Gefühl, dass Wörter oder Sätze unnatürlich getrennt waren?

1 = sehr häufig ... 5 = kaum

Waren bestimmte Textstellen schwerer zu verstehen als andere?

☐ Ja

☐ Teilweise

☐ Nein

Wenn ja: Was hat es schwieriger gemacht? (*offenes Textfeld*)

Hast du Anmerkungen zur Darstellung oder zum Lesefluss der Untertitel?

(*offenes Textfeld*)

Abstract

Deutsch: Diese Masterarbeit untersucht die Wirkung unterschiedlicher Untertitelversionen auf die Rezeption von deutschsprachigen audiovisuellen Inhalten. Im Mittelpunkt steht die Frage, wie sich die Art der Zeilensegmentierung in Untertiteln auf das Verständnis, die kognitive Verarbeitung und die subjektive Wahrnehmung von Videoinhalten auswirkt. Dazu werden bestehende Forschungsergebnisse zusammengeführt und durch eine empirische Analyse ergänzt. Im Rahmen dieser Untersuchung wurden 608 Teilnehmenden zwei Versionen desselben französischsprachigen Vortrags mit deutschen Untertiteln präsentiert, die sich ausschließlich in der Segmentierung unterschieden. Die subjektive Rezeption hinsichtlich Verständlichkeit, Lesefluss und allgemeiner Wahrnehmung wurde mittels standardisiertem Online-Fragebogen erfasst. Die deskriptiv-statistischen sowie inferenzstatistischen Auswertungen ergaben keine signifikanten Unterschiede in der Gesamtstichprobe keine signifikanten Unterschiede. Eine explorative Subgruppenanalyse nicht-erstsprachlicher Teilnehmender deutet jedoch auf mögliche Vorteile kohärenter Segmentierung hin, die in künftiger Forschung mit größeren Stichproben gezielter untersucht werden sollten. Die Ergebnisse stehen im Einklang mit früheren Studien und legen nahe, dass die Segmentierung zwar ein Qualitätsmerkmal professioneller Untertitelung darstellt, ihr Einfluss auf die messbare Rezeption jedoch mit weiteren Faktoren wie kognitiver Belastung, Informationsdichte und sprachlichem Hintergrund zusammenhängt.

Englisch: This master's thesis investigates the effects of different subtitle versions on the reception of German-language audiovisual content. It focuses on the question of how line segmentation in subtitles influences comprehension, cognitive processing, and the subjective perception of video content. To this end, existing research findings are synthesized and complemented by an empirical analysis. In the course of this study, 608 participants were presented with two versions of the same French-language lecture with German subtitles, which differed exclusively in their segmentation. Subjective reception in terms of comprehensibility, reading fluency, and overall perception was measured using a standardized online questionnaire. The descriptive and inferential statistical analyses revealed no significant differences in the overall sample. However, an exploratory subgroup analysis of non-native participants suggests potential benefits of more coherent segmentation, which should be investigated further in future research with larger sample sizes. The results are consistent with previous studies and indicate that, while segmentation is a quality feature of professional subtitling, its impact on measurable reception is associated with additional factors such as cognitive load, information density, and linguistic background.