universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Visuelle Aufmerksamkeit während des Simultandolmetschens mit durch automatische Spracherkennung (ASR) erzeugten visuellen Inputs: Eine experimentelle Studie mit Eyetracking

verfasst von | submitted by

Sara Pasini BA

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

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 070 348 331
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Translation Italienisch Deutsch
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Franz Pöchhacker

Danksagung

An dieser Stelle möchte ich allen, die mich auf dem Weg zur Fertigstellung dieser Masterarbeit unterstützt haben, meinen herzlichen Dank aussprechen.

Ich möchte meinem Betreuer, Univ.-Prof. Mag. Dr. Franz Pöchhacker, meinen aufrichtigen Dank aussprechen. Für das konstruktive Feedback, die Unterstützung und das Vertrauen, das er mir entgegengebracht hat, gebührt ihm mein besonderer Dank.

Bedanken möchte ich mich insbesondere bei den Teilnehmerinnen des Experiments für ihre Bereitschaft, an meiner Forschung teilzunehmen. Danke auch denjenigen, die sich die Zeit genommen haben, meine Arbeit durchzulesen und zu verbessern.

Vorrei dedicare un ringraziamento anche a tutti coloro che mi hanno sostenuta e incoraggiata durante questo percorso. Ai miei genitori Antonella e Luca, mio fratello Andrea e le mie nonne Renata e Renata, per il loro ineguagliabile supporto. Alle mie amiche Caterina, Julia, Rachele, Arjeta, Chiara e Laura, per esserci sempre state. A Saverio, per tutto. Grazie per aver contribuito indirettamente ma significativamente alla stesura di questa tesi.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Einleitung

Die automatische Spracherkennung (auf Englisch: Automatic Speech Recognition, abgekürzt ASR) hat in den letzten Jahrzehnten zunehmend an Bedeutung gewonnen und stellt einen zentralen Bestandteil vieler sprachgesteuerter Softwares und technologischer Geräte dar. Grundlegend dafür waren die immer genaueren Anwendungen, die von den Fortschritten im Bereich Deep Learning und neuronale Netze ermöglicht wurden. Diese haben dazu geführt, dass die automatische Spracherkennung viel Aufmerksamkeit auch in Bereichen bekommen hat, in denen sie vorher undenkbar war, wie zum Beispiel in der Dolmetschwissenschaft. Die große Anzahl der Publikationen, die vor allem in den letzten Jahren über das Thema automatische Spracherkennung und Dolmetschen veröffentlicht wurden (z. B. Cheung & Li 2022; Defrancq & Fantinuoli 2021; Frittella 2022), zeigt, dass ein hohes Forschungsinteresse in diesem Bereich besteht. Die automatische Spracherkennung kann vor allem für die Vorbereitung von Dolmetschaufträgen, für die Terminologieverwaltung und für den Austausch von Informationen zwischen Kolleg*innen verwendet werden (vgl. Fantinuoli 2017: 25).

CAI-Tools (auf Englisch: Computer-Assisted Interpreting Tools) wurden mit dem Ziel entwickelt, unterschiedliche Prozesse vor, während und nach dem Dolmetschauftrag zu beschleunigen bzw. zu automatisieren, jedoch waren die ersten Tools nur rudimentär und erfreuten sich aufgrund einiger Schwachstellen keiner großen Beliebtheit. Die Integration von automatischer Spracherkennung in Dolmetschtechnologien entstand aus dem Bedürfnis, ein bestehendes Problem zu behandeln und zu lösen. Einer der Kritikpunkte der CAI-Tools und im Allgemeinen der technologischen Lösungen, die von Dolmetscher*innen als Hilfestellung während des Dolmetschens vor allem für die Suche nach Termini verwendet werden, war das manuelle Eintippen, um Wörter in den Glossaren während der Dolmetschaufträge zu suchen. Das kann eine besondere Herausforderung darstellen. Einerseits nimmt das Tippen auf der Tastatur eine gewisse Zeit in Anspruch, andererseits muss die Aufmerksamkeit von der Dolmetschleistung auf das Tippen verlagert werden. Da das Dolmetschen viel Konzentration und ein schnelles Verarbeitungsvermögen voraussetzt, musste eine neue Lösung entwickelt werden, welche die Bedürfnisse der Dolmetscher*innen berücksichtigt und die kognitive Last während der Dolmetschung gering hält (vgl. Fantinuoli 2017: 25). Wie Hansen-Schirra (2012: 217) vorschlägt, kann die automatische Spracherkennung verwendet werden, um die Aufmerksamkeits- und Verarbeitungskapazität von

Dolmetscher*innen zu verbessern. Zahlen oder Eigennamen, die das Gedächtnis zusätzlich beanspruchen würden, werden automatisch erkannt und angezeigt. Sie übernimmt also die Rolle eines*r Kabinenpartners*in.

Die automatische Spracherkennung ist ein vergleichsweise junges Forschungsfeld. ASR-gestützte Tools haben sich in der Praxis und in der Aus- und Weiterbildung von Dolmetscher*innen noch nicht stark etabliert (vgl. Rodriguez et al. 2022: 79), und die Forschung über die Wirkung dieser Tools hat sich bis dato fast ausschließlich auf Dolmetschprodukte, wie zum Beispiel die Untersuchung von Dolmetschleistungen oder das Design und die Entwicklung der Tools, fokussiert (vgl. Frittella 2023: 55).

Die vorliegende Arbeit beschäftigt sich mit dem Simultandolmetschen mit ASR-gestützten Anwendungen. Besonderer Wert wird auf die kognitiven Aspekte gelegt, die in einem multimodalen Dolmetschsetting berücksichtigt werden müssen. Die visuellen Inputs der automatischen Spracherkennung müssen gleichzeitig mit den mündlichen Inputs der Ausgangsrede von den Dolmetscher*innen verarbeitet werden. Das könnte einerseits als Hilfestellung für die Dolmetscher*innen dienen, da Inhalte auf zwei Kanälen verfügbar sind. Andererseits können mehr Inputs zu einer erhöhten kognitiven Belastung führen, welche sich negativ auf die Dolmetschleistungen auswirkt. Ausgehend von dieser Feststellung wurde ein Experiment konzipiert, das die Verteilung der visuellen Aufmerksamkeit von Simultandolmetscher*innen mit der Eyetracking-Methode untersucht. Ein Eyetracker wurde verwendet, um das Blickverhalten der Studienteilnehmer*innen während des Dolmetschens auf unterschiedliche Elemente auf einem Bildschirm zu untersuchen und mit den Ergebnissen der Dolmetschleistungen zu vergleichen.

Zu Beginn der Masterarbeit werden die theoretischen Grundlagen zu den kognitiven Aspekten des Simultandolmetschens erläutert. Im Mittelpunkt stehen zwei theoretische Modelle, die in der Simultandolmetschforschung in Bezug auf die kognitiven Leistungen angewendet werden können, um diese zu veranschaulichen. Wichtig ist auch der Aspekt der Mulimodalität und der multimodalen Verarbeitung beim Simultandolmetschen, die auch in den theoretischen Modellen berücksichtigt werden müssen. Im zweiten Kapitel werden unterschiedliche technologische Lösungen für Dolmetscher*innen beschrieben. Dabei wird der Fokus auf CAI-Tools gelegt und wie diese mit ASR integriert wurden. Danach werden die Forschungsmethode des Eyetrackings in der Konferenzdolmetschforschung und die Messungsindikatoren vorgestellt, die in Zusammenhang mit dieser Methode evaluiert werden können. Des Weiteren werden die

Fragestellung und Methode der Arbeit beschrieben. Neun Studienteilnehmerinnen dolmetschten eine Rede vor einem Bildschirm, worauf das Video der Rednerin und die von einer ASR-Software erkannten Zahlen angezeigt wurden. Die Dolmetschleistungen, die Eyetracking-Daten und die Antworten der Teilnehmerinnen in einem Fragebogen wurden ausgewertet und miteinander verglichen, um die Wirkung einer ASR-Software für die Erkennung von Zahlen auf das Blickverhalten und auf die Dolmetschleistungen von Simultandolmetscher*innen zu beschreiben.

1 Kognitive Aspekte in der Simultandolmetschforschung

In der Forschung rund um Kognition wurde das Simultandolmetschen in den letzten Jahrzehnten Schwerpunkt unterschiedlicher Untersuchungen. In den 1960er Jahren befassten sich Psycholog*innen und Psycholinguist*innen damit, wie das Simultandolmetschen funktioniert, während in den 1970ern, mit dem Aufkommen der Kognitionswissenschaft, das Interesse für diese Tätigkeit rasant wuchs (vgl. Ahrens 2017: 445). Da die Dolmetschtätigkeit durch eine große Vielfalt und eine hohe Komplexität gekennzeichnet ist, gibt es eine Reihe von kognitionswissenschaftlichen Forschungsschwerpunkten, die in der Dolmetschwissenschaft und in verwandten Disziplinen herangezogen wurden. In diesem Kapitel werden zunächst die einzelnen Themenbereiche und dann die Methoden vorgestellt, die in der Dolmetschwissenschaft zur Erforschung von kognitiven Aspekten bis dato verwendet wurden. Als nächstes werden zwei Modelle, nämlich das Kapazitätenmodell und das Cognitive Load Model (Modell der kognitiven Belastung), miteinander verglichen und ein kognitives Modell für multimodale Inputs vorgestellt. Abschließend werden Problemauslöser beim Simultandolmetschen vorgestellt, darunter Zahlen, die einer der Forschungsschwerpunkte dieser Masterarbeit sind und eine Schlüsselrolle bei der experimentellen Untersuchung spielen.

1.1 Entwicklung und Methoden der Kognitionsforschung beim Simultandolmetschen

Die ersten Forschungsbemühungen fokussierten sich fast ausschließlich auf den Simultanmodus des Dolmetschens, da die gleichzeitige Verarbeitung von zweisprachigem Input und Output als eine außergewöhnliche Fähigkeit angesehen wurde. Das Interesse lag vor allem auf der Zeitverzögerung zwischen Ausgangs- und Zieltext, die auf Englisch als Ear-Voice-Span (abgekürzt EVS) bekannt ist, sowie auf dem Konzept von Simultanität und auf dem Redetempo der Ausgangsrede. David Gerver war einer der ersten Forscher*innen, die die Wirkung von Veränderungen im Redetempo auf das Verständnis des Ausgangstextes und auf die zielsprachige Produktion erforschten (vgl. Ahrens 2017: 450).

Seitdem steht der Dolmetschprozess im Mittelpunkt unterschiedlicher Untersuchungen. Kognitive und psycholinguistische Aspekte, wie die kognitive Belastung und die Verarbeitungskapazität, wurden Forschungsgegenstand vieler Studien (vgl. Ahrens 2017: 445). Zentral waren die Konzepte des Arbeitsgedächtnisses und der Aufmerksamkeit sowie der Sprachverarbeitung und des Sprachverständnisses. Die Sprachverarbeitung bezieht sich auf Teilprozesse, bei denen die Analyse von phonologischen, lexikalischen, syntaktischen, semantischen und pragmatischen

Merkmalen durchgeführt wird, um Worte zu erkennen und zu analysieren, Informationen zu extrahieren und zu interpretieren, usw. Das Sprachverständnis bezieht sich hingegen auf die semantischen und syntaktischen Zusammenhänge, die zur Erstellung eines mentalen Modells führen (vgl. Seeber 2011: 180f.). Auch Dolmetschstrategien waren Schwerpunkt zahlreicher Forschungsarbeiten und wurden aus unterschiedlichen Perspektiven untersucht: aus der prozess-orientierten Perspektive, um den Umgang mit der kognitiven Belastung zu erforschen, oder aus der produktorientierten Perspektive, um eine effektive Kommunikation zu erzielen (vgl. Ahrens 2017: 447).

Die kognitive Belastung spielt eine wichtige Rolle bei der Forschung rund um die kognitiven Aspekte des Dolmetschens. Das Interesse für die kognitive Belastung beruht einerseits auf dem Versuch, die anspruchsvolle Aufgabe des Dolmetschens zu erfassen, und andererseits auf dem Ziel, herauszufinden, wie Dolmetscher*innen damit umgehen (vgl. Chen 2017: 641). Studien zum Thema kognitive Belastung fokussierten sich auf das Dolmetschen in eine Fremdsprache, auf die Verarbeitung unterschiedlicher Teilaufgaben beim Dolmetschen oder auf Modelle der Einschätzung der kognitiven Belastung (vgl. Chen 2017: 641f.). Chen (2017) schlägt folgende Definition von kognitiver Belastung in Bezug auf das Dolmetschen vor: “That portion of an interpreter’s limited cognitive capacity devoted to performing an interpreting task in a certain environment.” (Chen 2017: 643) Die kognitive Belastung wird also als ein mehrdimensionales Konzept gesehen, das zwei Kategorien von Variablen enthält: aufgaben- bzw. umgebungsbezogene Merkmale (auf Englisch: Task and environmental characteristics) und Merkmale des*der Dolmetschers*in (auf Englisch: Interpreter characteristics) (Abb. 1).

Aufgabenbezogene Merkmale sind der Dolmetschmodus (simultan oder konsekutiv), die Sprachkombination und -richtung, Merkmale der Ausgangsrede (wie z. B. Länge, Redetempo, Thema, usw.), Merkmale des*der Redners*in, die Erwartungen, die verfügbare Zeit für die Aufgabe, die Vorbereitung, das Risiko und das Niveau an Bekanntheit der Aufgabe (vgl. Chen 2017: 643f.). Diese Faktoren können sich auf die kognitive Belastung beim Dolmetschen auswirken.

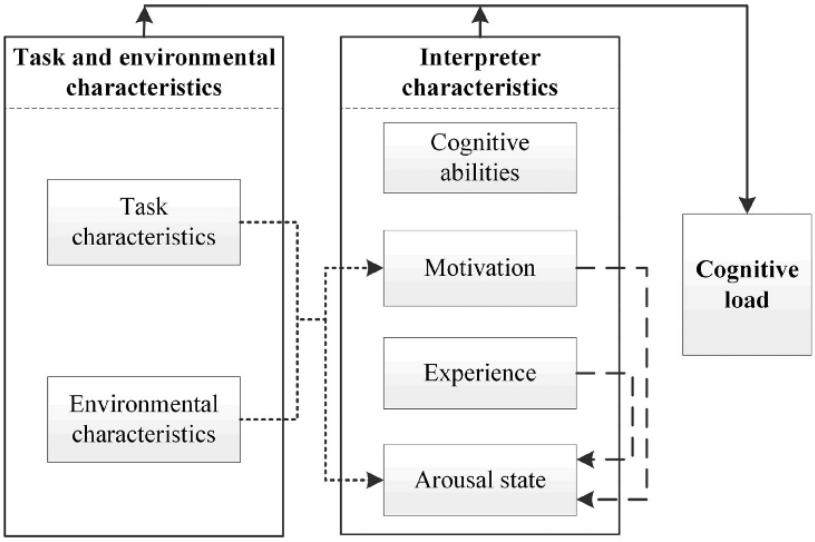

Abb. 1: Graphische Darstellung des Konzepts der kognitiven Belastung nach Chen (2017: 644)

Umgebungsbezogene Merkmale können auch eine wichtige Rolle bei der kognitiven Belastung während der Dolmetschung spielen. Diese sind unter anderem räumliche Bedingungen (Standort, Sitzplätze, Geräusche, Beleuchtung, Temperatur und Belüftung), der visuelle Zugang zu den Redner*innen bzw. zum Publikum, und die technische Ausstattung (vgl. Chen 2017: 645). Zu den Merkmalen des*der Dolmetschers*in zählen kognitive Fertigkeiten, wie Dolmetschfertigkeiten, Erfahrung und Eignung; die Motivation, die sich auf den Fokus und den Grad an Bemühungen von Dolmetscher*innen auswirken kann; die Berufserfahrung bzw. Ausbildung und der physische Erregungsgrad eines*r Dolmetschers*in (vgl. Chen 2017: 646). Die gestrichelten Linien im Schaubild zeigen die Verbindungen zwischen den einzelnen Komponenten, die sich gegenseitig in Bezug auf den kognitiven Bedarf beeinflussen (Abb. 1). Diese graphische Darstellung legt die Grundlagen zur Festlegung der Messungsindikatoren, die bei der Erforschung der kognitiven Belastung verwendet werden können. Unterschiedliche beobachtbare und messbare Indikatoren, die auf die kognitive Belastung hinweisen, können verwendet werden, um dieses theoretische Konstrukt zu erforschen. Das subjektive Gefühl der Belastung kann mit subjektiven Messungsindikatoren erfasst werden; die Dolmetschleistung kann auch verwendet werden, um Verbesserungen bzw. Verschlechterungen in der Wiedergabe zu erkennen, die auf die kognitive Belastung zurückzuführen sind. Die physiologische Erregung des*der Dolmetschers*in kann anhand von physiologischen Indikatoren untersucht werden, während aufgabenbezogene

Merkmale eine vorab vorgenommene Schätzung der kognitiven Belastung auf Basis der Aufgabenkomplexität ermöglichen (vgl. Chen 2017: 467).

Der Zugang zu kognitiven Prozessen war immer eine der größten Herausforderungen in der Wahl der Forschungsmethodik, da mentale Prozesse nicht direkt beobachtbar sind (vgl. Ahrens 2017: 448). In den Anfängen der Dolmetschwissenschaft wurden vor allem Selbstbeobachtungen und Retrospektion verwendet. Selbstbeobachtungen von Dolmetscher*innen und die Analysen ihrer Leistungen sind produktorientierte Methoden, aber es ist kaum möglich, daraus allgemeine Rückschlüsse auf das Verhalten von Dolmetscher*innen zu ziehen. Die Retrospektion ist hingegen eine prozessorientierte Methode und bietet Zugang zu den Handlungen und Gedanken der Dolmetscher*innen, obwohl dies nur eingeschränkt möglich ist. Da sie nur nach der Dolmetschung erfolgen, werden Informationen ausgelassen, und sie können keine unbewussten Prozesse aufdecken (vgl. Ahrens 2017: 448). Um Rückschlüsse auf das kognitive Verhalten zu ziehen und auf die Art und Weise, wie das Gehirn beim Dolmetschen funktioniert, werden neurophysiologische Methoden verwendet. Diese umfassen unter anderem das dichotische Hören und die Elektroenzephalographie (EEG) sowie bildgebende Techniken wie die funktionelle Magnetresonanztomographie (fMRI) und die Positronen-Emissions-Tomographie (PET) (vgl. Ahrens 2017: 448). Auch das Eyetracking und die Pupillometrie werden als diagnostische Messverfahren für die Erforschung von kognitiven Prozessen eingesetzt. Das Eyetracking wird im Labor eingesetzt, um die Verteilung von Aufmerksamkeit auf visuelle Inputs zu bestimmen. Es erfasst den Blickpunkt und die Augenbewegung, während die Pupillometrie die Veränderung der Pupillengröße während kognitiv herausfordernden Aufgaben misst. Da diese Methoden körperliche Reaktionen auf die kognitive Belastung untersuchen, beeinflussen sie die Durchführung von Aufgaben nicht und ermöglichen Rückschlüsse auf kognitive Aspekte der Dolmetschtätigkeit (vgl. Ahrens 2017: 449).

1.2 Zwei theoretische Modelle zum Simultandolmetschforschen

Neben den Konzepten des Arbeitsgedächtnisses, der kognitiven Belastung und der Aufmerksamkeit- bzw. Ressourcenaufteilung haben auch unterschiedliche theoretische Modelle viel Aufmerksamkeit in der Welt der kognitiven Dolmetschforschung erregt. Im folgenden Kapitel werden zwei solche Modelle vorgestellt, nämlich Giles (2008) Kapazitätenmodell für das Simultandolmetschen (Effort Model of Simultaneous Interpreting) und Seebers (2011) Cognitive Load Model. Sie stellen einen Versuch dar, den Dolmetschprozess als eine Multitasking-Tätigkeit

darzustellen, und beide interpretieren Probleme beim Dolmetschen als Schwierigkeit der Aufteilung der begrenzten kognitiven Ressourcen zwischen konkurrierenden Aufgaben.

In Prandi (2022) werden zahlreiche Gründe aufgelistet, warum diese beiden Modelle gegenüber anderen kognitiven, psycholinguistischen und neurolinguistischen-neurophysiologischen Modellen des Simultandolmetschens bevorzugt werden sollten. Sie sind kognitive Modelle aus dem Bereich der kognitiven Dolmetschforschung, in dem Prandi (2022) und die vorliegende Masterarbeit angesiedelt sind. Sie gelten als theoretische Referenzmodelle in der CAI-Forschung und ermöglichen, schriftliche Ressourcen in den Dolmetschprozess miteinzubeziehen. Außerdem eignen sie sich für Laborexperimente und ermöglichen es, das Simultandolmetschen unter relativ natürlichen Bedingungen zu untersuchen (vgl. Prandi 2022: 81). Wichtig ist es, an dieser Stelle anzumerken, dass diese Modelle die Komplexität und Vielschichtigkeit des Dolmetschprozesses belegen. Aus diesem Grund kann geschlussfolgert werden, dass kein Modell sich alleine als Erklärung für das Dolmetschen als Ganzes bewähren kann (vgl. Pöchhacker 2022: 101). In den folgenden Kapiteln werden die zwei Modelle präsentiert und miteinander verglichen, um Gemeinsamkeiten und Unterschiede darzustellen und das Konzept der Multimodalität beim Simultandolmetschen darzulegen.

1.2.1 Giles Kapazitätsmodell

Wie schon erwähnt, war die Gleichzeitigkeit beim Simultandolmetschen einer der ersten Schwerpunkte, die das Interesse verschiedener Forscher*innen wecken. Der Versuch, den Konflikt bzw. die Wechselwirkung von mehreren gleichzeitigen kognitiven Leistungen zu erklären, führte zu einem Modell, das zu einem der bekanntesten Beiträge zur Dolmetschforschung geworden ist: das Kapazitätsmodell (Effort Model). Das Effort Model wird als Erklärungsmodell, didaktisches Modell und konzeptioneller Rahmen für empirische Forschung verwendet (vgl. Gile 2008: 59).

Die Grundlage für die Entwicklung des Modells waren zwei intuitive Gedanken, die auf Beobachtung und Selbstbeobachtung während des Dolmetsches beruhen. Die Dolmetschtätigkeit bedarf einer gewissen geistigen Kraft, die nur begrenzt verfügbar ist; sie verbraucht fast die Gesamtheit dieser Kraft und erfordert manchmal mehr, als zur Verfügung steht. Das führt zu einer Verschlechterung der Dolmetschleistung (vgl. Gile 2009: 159). Erkenntnisse aus der kognitiven Psychologie haben gezeigt, dass alle Handlungen in automatische und nichtautomatische Handlungen unterteilt werden können. Während automatische Handlungen keine Aufmerksamkeit erfordern und schnell geschehen, bedürfen nichtautomatische Handlungen mehr Aufmerksamkeit

und nehmen mehr Zeit in Anspruch. Auf Basis dieser Differenzierung und der Annahme, dass alle Bestandteile einer Dolmetschung einen gewissen Aufwand erfordern, entstand das Effort Model (vgl. Gile 2009: 159f.). Als ‚Efforts' werden drei Teilleistungen bezeichnet, welche das Dolmetschen prägen. Wie der Name impliziert, benötigen sie einen bestimmten Aufwand. Sie erfordern bewusstes Handeln, um Entscheidungen zu treffen, und bestimmte Ressourcen, die währenddessen verbraucht werden (vgl. Gile 2009: 160).

Die erste Aufgabe (Listening and Analysis Effort, abgekürzt L) bezieht sich auf den Prozess des Verstehens und Analysierens des gehörten Textes, um die Bedeutung der Wörter zu erfassen. Sie umfasst ebenfalls die Überprüfung der Plausibilität sowie die Antizipation der folgenden Inhalte. Akustische Signale müssen analysiert und mit den im Langzeitgedächtnis gespeicherten Mustern verglichen werden, um Wörter zu erkennen. Die Wörterkennung erfolgt automatisch. Zu den nichtautomatischen Handlungen zählen die Plausibilitätsprüfung und Antizipationen (vgl. Gile 2009: 161f.).

Die zweite Aufgabe (Production Effort, abgekürzt P) bezieht sich auf die Wiedergabe von Inhalten bei der Dolmetschung. Diese enthält auch die mentale Auslegung und Planung der Dolmetschung sowie Selbstüberwachung bei der Sprachausgabe und eventuell die Korrektur von Fehlern. Diese Kapazität fällt in die Kategorie der nichtautomatischen Tätigkeiten (vgl. Gile 2009: 163).

Die dritte Aufgabe (Memory Effort, abgekürzt M) umfasst die Speicherung von Inhalten im Gedächtnis. Das Kurzzeitgedächtnis wird aufgrund der Verzögerung zwischen dem Moment, in dem der Ausgangstext vorgetragen wird, und dem Moment, in dem die Dolmetschung ausgesprochen wird, beansprucht. Einzelne phonetische Segmente werden gespeichert und analysiert, um Worte zu erkennen. Außerdem müssen Informationen im Gedächtnis gehalten werden, während die Sprachproduktion erfolgt, und zwar während geeignete Wörter und syntaktische Strukturen ausgewählt und formuliert werden. Auch Elemente wie die Informationsdichte, ungewöhnliche sprachliche Strukturen oder der Akzent des*r Sprechers*in können zur Verzögerung der wiedergegebenen Inhalte beitragen. Alle Handlungen, welche die Speicherung von Informationen beinhalten, gehören zur Kategorie der nichtautomatischen Handlungen (vgl. Gile 2009: 165f.). Basierend auf diesen Grundlagen wurde das Kapazitätsmodell für Simultandolmetschen entwickelt, das diese drei Fähigkeiten sowie die koordinierte Anwendung

derselben (Coordination Effort, abgekürzt C) einschließt. Die Aufgaben können also wie folgt abgebildet werden:

$$SI = L + P + M + C$$

Das Simultandolmetschen (SI) besteht aus den oben genannten Aufgaben, die während der Dolmetschung von Sprachsegmenten zur Anwendung kommen (vgl. Gile 2009: 168). Allerdings dient diese Formel nur zur Veranschaulichung des Dolmetschprozesses und enthält einige Vereinfachungen. Die Linearität des Dolmetschprozesses, wobei einzelne Segmente eine nach dem anderen zunächst gehört, dann analysiert, im Gedächtnis gespeichert und abschließend gedolmetscht und ausgesprochen werden, entspricht nicht der Komplexität der Realität. Mehrere Aufgaben können gleichzeitig zum Einsatz kommen und unterschiedliche Sprachsegmente verarbeiten. Die Summierung ist außerdem nicht im streng arithmetischen Sinne zu verstehen. Da alle Efforts gleichzeitig stattfinden und stark miteinander verbunden sind, werden die Anforderungen nicht nur durch individuelle Efforts, sondern auch durch ihre Interaktion bestimmt. Daher kann man annehmen, dass die Gleichzeitigkeit von zwei Efforts insgesamt mit einer höheren Belastung als ein einzelner Effort verbunden ist, und die Gleichzeitigkeit von drei Efforts mit einer höheren Belastung als zwei Efforts verbunden ist (vgl. Gile 2009: 169). Im Effort Model wird vorausgesetzt, dass den Dolmetscher*innen nur eine gewisse Verarbeitungskapazität zur Verfügung steht, die nicht überschritten werden kann. Aus diesem Grund muss diese sorgfältig auf die verschiedenen Efforts verteilt werden, um nicht über die Grenze hinauszugehen. Wenn die Summe der Verarbeitungskapazität von einzelnen Efforts die gesamte Menge an verfügbarer Verarbeitungskapazität überschreitet, dann können Probleme in der Dolmetschleistung entstehen (vgl. Gile 2009: 170).

Neben dem Kapazitätsmodell können auch andere Modelle erwähnt werden, wie zum Beispiel das Kapazitätsmodell des Vom-Blatt-Dolmetschens und des Simultandolmetschens mit Text. Diese Modelle enthalten einen zusätzlichen Aufwand während des Dolmetschens, da gelesen werden muss. Das Lesen stellt somit eine Aufgabe dar. Das Vom-Blatt-Dolmetschen (Englisch: Sight Translation) besteht darin, einen ausgangssprachlichen Text in der Zielsprache vorzulesen. Somit wird die Aufgabe vom Listening and Analysis Effort durch den Reading Effort ersetzt (vgl. Gile 2009: 179):

Sight translation =
Reading Effort + Memory Effort + Speech Production Effort + Coordination

Das Simultandolmetschen mit Text setzt hingegen voraus, dass ein Text von einem*r Redner*in vorgelesen wird. Dolmetscher*innen bekommen Zugang zum Text und dolmetschen ihn, während er vorgelesen wird. Die Vorteile bestehen darin, dass Dolmetscher*innen Zugang zu mündlichen Signalen der Redner*innen bekommen, wie zum Beispiel Intonation und Pausen. Die schriftliche Version der Rede führt zu einer Verringerung des Gedächtnisaufwands und unterstützt die Rezeptionsleistung. Die Nachteile sind die hohe Informationsdichte bei schriftlichen Texten, die erhöhte Gefahr von Interferenzen vom Ausgangstext und die erhöhte kognitive Belastung, die aus der gleichzeitigen Verarbeitung von schriftlichen und mündlichen Informationen entsteht. Es hängen zwei Risiken mit diesem Dolmetschmodus zusammen: der Versuch, alle Inhalte zu dolmetschen, obwohl das Sprechtempo des*r Redners*in zu schnell ist, und der Verlust von Informationen, wenn sie vom schriftlichen Text abweichen, da der Fokus von Dolmetscher*innen auf dem schriftlichen Text liegt (vgl. Gile 2009: 181f.). Das Simultandolmetschen mit Text kann also wie folgt veranschaulicht werden (vgl. Gile 2009: 181):

Simultaneous Interpreting with text =
Reading Effort + Listening Effort + Memory Effort + Production Effort + Coordination Effort

Die Relevanz des Kapazitätsmodells zeigt sich erst in Verbindung mit der sogenannten Tightrope-Hypothese. Diese Hypothese besagt, dass Dolmetscher*innen die meiste Zeit an der Grenze ihrer kognitiven Sättigung arbeiten. Das betrifft entweder die gesamte Verarbeitungskapazität oder die verfügbaren Kapazitäten für einzelne Efforts (vgl. Gile 2009: 182). Diese kognitive Sättigung entsteht aus der Tatsache, dass der kognitive Bedarf höher als die verfügbaren Kapazitäten ist. Aus diesem Grund treten Fehler während der Dolmetschungen aus. Diese sind nicht nur auf Mängel in der Sprachkompetenz oder im Allgemeinwissen zurückzuführen, sondern passieren auch aufgrund mangelnder Verarbeitungskapazität (vgl. Gile 2009: 182).

Wichtig ist auch anzumerken, dass die Kapazitätsmodelle für pädagogische Ziele und nicht für Forschungszwecke entwickelt wurden sind. Sie dienen hauptsächlich dazu, wiederkehrende Schwierigkeiten beim Dolmetschen zu erklären und den Dolmetschstudierenden didaktische Empfehlungen anzubieten. Sie sind nicht als eine Theorie zu sehen, auf deren Grundlagen Vorhersagen über in der Praxis beobachtete Phänomene getroffen werden können, sondern eher als

konzeptioneller Rahmen. Sie stützen sich jedoch auf Erkenntnisse, die in der Kognitionspsychologie und Psycholinguistik geprüft wurden (vgl. Gile 2009: 188).

1.2.2 Seebers Cognitive Load Model

Giles Kapazitätenmodell beruht auf dem Konzept der sogenannten Single-Source-Theorie. Diese besagt, dass Probleme im Dolmetschprozess entstehen, wenn der gesamte Kapazitätsbedarf der vier Kapazitäten die verfügbaren Gesamtkapazitäten übersteigt. Das Multiple-Resource-Modell hingegen zeichnet sich dadurch aus, dass es, um Probleme bzw. Schwierigkeiten beim Simultandolmetschen zu beschreiben, den Schwerpunkt auf diejenigen Komponenten verlagert, die auf dieselben Ressourcen bzw. Strukturen zurückgreifen (vgl. Seeber 2007: 1382). Wickens' (2002) Multiple-Resource-Theorie kann angewendet werden, um komplexe Sprachverarbeitungstätigkeiten zu beschreiben, und wurde daher als Grundlage für das Cognitive Load Model verwendet.

Wickens (2002) beschäftigte sich mit den Konzepten von Ressourcen, Aufmerksamkeit und Belastung, um die Interferenz zwischen unterschiedlichen Leistungen auszuwerten. Das Multiple-Resource-Modell geht davon aus, dass es vier Dimensionen gibt. Die erste Dimension betrifft die Verarbeitungsstufen. Es wird zwischen Ressourcen differenziert, die zur Wahrnehmung und Kognition dienen, und Ressourcen, die für Reaktionsprozesse verwendet werden. Die zweite Dimension betrifft die Verarbeitungscodes, die entweder visuell/akustisch oder räumlich/verbal sind. Die weiteren Dimensionen betreffen die Modalitäten von Inputs und Reaktionen, die akustisch oder visuell sind, und die visuelle Verarbeitung. Diese umfasst das fokale und das umgebende Sehen und ist in die visuelle Modalität eingebettet (vgl. Wickens 2002: 163). Dieses Modell kann verwendet werden, um die Leistung von zwei oder mehr Handlungen vorauszusehen, die gleichzeitig stattfinden, genauer gesagt, inwieweit sie sich gegenseitig beeinflussen bzw. beeinträchtigen. Zwei Leistungen, welche sich auf die gleiche Kategorie in einer Dimension stützen, beeinflussen sich gegenseitig mehr, als zwei Leistungen, die sich auf unterschiedliche Kategorien stützen. (vgl. Wickens 2002: 167).

Seeber (2007: 1383) ergänzte das Multiple-Resource-Modell mit einer Konfliktmatrix. Diese dient der eindeutigen Bestimmung der kognitiven Ressourcen, die bei komplexen kognitiven Aufgaben notwendig sind. In der Konfliktmatrix werden unterschiedliche Werte den Bedarfsvektoren zugewiesen, die für eine bestimmte Handlung zum Einsatz kommen. Den einzelnen Bedarfsvektoren wird der Wert 1 zugewiesen, und ihr Konfliktkoeffizient wird

entsprechend berechnet. Der Basiskonfliktkoeffizient beträgt 0,2 und zeigt generell die Möglichkeit von zwei Aufgaben, gleichzeitig zu geschehen. Der höchste Konfliktkoeffizient ist 1 und stellt die Unmöglichkeit dar, dass zwei Aufgaben gleichzeitig stattfinden, zum Beispiel zwei Aufgaben, die eine verbale Reaktion erfordern (vgl. Seeber 2007: 1383). Von diesem Ausgangspunkt aus wurde das Cognitive Load Model entwickelt. Dieses Modell berücksichtigt die Summe der kognitiven Belastungen, die durch einzelne Handlungen entstehen. Das Simultandolmetschen ist also als eine Kombination von Sprachverständnis und Sprachproduktion zu interpretieren. Die Bedarfsvektoren sind unter anderem: die perzeptive, auditive, verbale Verarbeitung von Input und Output, und die kognitiv-verbale Verarbeitung von Input und Output. Dafür wird dann der Konfliktkoeffizient berechnet, der aus der Wechselwirkung von gleichzeitigen Handlungen entsteht (vgl. Seeber 2011: 187).

Seeber (2011: 191) wendete dieses Modell in der Praxis an und verglich die Unterschiede in der kognitiven Belastung zwischen Dolmetschungen von symmetrischen und asymmetrischen Satzstrukturen. Asymmetrische Satzstrukturen führten zu erhöhten Niveaus von kognitiver Belastung, wie auch in einer späteren Studie anhand von Pupillenmessungen bestätigt wurde (vgl. Seeber & Kerzel 2012). Diese Ergebnisse zeigten, dass es einen sogenannten Spillover-Effekt bei asymmetrischen Satzstrukturen gab, wie Gile (2008) ebenfalls erwähnte. Giles (2008) Tightrope-Hypothese wurde aber im Rahmen dieses Experiments widerlegt. Schwankungen in der kognitiven Belastung waren gering, und die höchsten Niveaus wurden nur für eine geringe Zeit verzeichnet. Das bedeutet, dass Dolmetscher*innen die meiste Zeit unter diesen Niveaus arbeiteten. Außerdem konnte nicht nachgewiesen werden, dass die höchsten Werte die absolute Grenze der kognitiven Belastung darstellen. Dies wiederum stellte einen Gegensatz zur Tightrope-Hypothese dar (vgl. Seeber 2011: 197).

Das Cognitive Load Modell wurde vom Giles (2008) Kapazitätsmodell inspiriert. Die zwei Modelle teilen dasselbe Ziel, kognitive Anforderungen beim Simultandolmetschen darzustellen. Das Kapazitätsmodell stützt sich auf die Single-Source-Theorie, welche einen gemeinsamen kognitiven Bestand für alle Ressourcen voraussetzt. Dabei wird jedoch nicht beachtet, dass Phänomene wie das perfekte Time-Sharing der Handlungen während des Dolmetschens bereits ausführlich belegt sind und dass die Neuaufteilung der Ressourcen zwischen zwei oder innerhalb einer Aufgabe nicht wissenschaftlich bewiesen wurde (vgl. Seeber 2011: 189). Beide Modelle bezeichnen das Simultandolmetschen als eine Kombination von Aufgaben zum Sprachverständnis

und zur Sprachproduktion, und beide legen nahe, dass die kognitive Gesamtbelastung dadurch erhöht wird. Das Cognitive Load Model versucht zusätzlich die kognitive Belastung zu berechnen. Dabei werden sowohl Input als auch Output der Dolmetschung berücksichtigt (vgl. Seeber 2011: 189).

1.3 Multimodalität beim Simultandolmetschen

Die Komplexität der Dolmetschtätigkeit besteht aus der Notwendigkeit, unterschiedliche Informationen aus unterschiedlichen Kanälen zu verarbeiten (vgl. Seeber 2017: 461). Das Konzept, dass das Simultandolmetschen geteilte Aufmerksamkeit voraussetzt, da verschiedene kognitive Aufgaben gleichzeitig ausgeführt werden müssen, ist nicht neu. Schon in den frühen 2000er Jahren wurde die Aufmerksamkeit während der zusätzlichen Aufgabe des Lesens beim Dolmetschen analysiert, um die Wirkung von visuellen Informationen auf die Dolmetschleistung zu untersuchen. Am Beispiel von Lambert (2004) konnten Unterschiede in der Verarbeitung von Informationen anhand von Dolmetschleistungen in den Dolmetschmodi Vom-Blatt-Dolmetschen (Sight Translation), Simultandolmetschen mit Text (Sight Interpreting) und Simultandolmetschen analysiert werden. Die Wechselwirkung von visuellen und akustischen Inputs war Gegenstand unterschiedlicher Studien, jedoch konnte kein Zusammenhang nachgewiesen werden. Das bedeutet aber nicht zwangsläufig, dass der Dolmetschprozess nicht davon betroffen ist (vgl. Seeber 2017: 466). Als noch wichtiger erweist sich heute, den Aspekt der Multimodalität beim Simultandolmetschen zu berücksichtigen. Traditionelle Textformen wie Transkripte der Reden wurden durch Software-Anwendungen mit visuellen Elementen ergänzt, auf die genauso während der Simultandolmetschung zugegriffen werden kann, wie zum Beispiel digitale Glossare oder CAI-Tools.

Das Konzept der multimodalen Verarbeitung beim Simultandolmetschen wird zunächst präsentiert, um ein kognitives Modell für multimodale Inputs zu beschreiben. Dieses neue Modell zeigt den Bedarf an kognitiven Ressourcen beim Simultandolmetschen mit visuellen Inputs und mit Text. Die überarbeitete Konfliktmatrix ermöglicht außerdem eine spezifische Berechnung des gesamten Interferenzwertes für die Suche in digitalen Glossaren und mit ASR während des Simultandolmetschens.

1.3.1 Multimodale Verarbeitung beim Simultandolmetschen

Es ist nicht ungewöhnlich, dass Informationen bei Konferenzen auch über den visuellen Kanal vermittelt werden. Dies kann zum Beispiel in Form von Parasprache, Gestik, Körperhaltung oder Mimik geschehen. Wenn der visuelle Kanal verwendet wird, um Informationen aus dem auditiven Kanal zu wiederholen, können Dolmetscher*innen von der Redundanz der bimodal dargestellten Inhalte profitieren (vgl. Seeber 2012: 343). Ein Beispiel dafür sind Zahlen, die typischerweise bei Konferenzen mit verschiedenen Modalitäten gleichzeitig vermittelt werden: akustisch-verbal (eingebettet in der Ausgangsrede), visuell-räumlich (anhand von Gestik), oder visuell-verbal (als Namen oder Zahlwörter) (vgl. Seeber 2012: 343). Audiovisuelle Informationen, die von Redner*innen vermittelt werden, können also in ergänzende und redundante Informationen unterteilt werden. Als ergänzend werden solche Informationen bezeichnet, die Unzulänglichkeiten ausgleichen und Unklarheiten beseitigen, während Redundanz ein typisches Merkmal von Reden oder Präsentationen ist, in denen Informationen mehrfach wiederholt werden. Während redundante Informationen aus unterschiedlichen Kanälen mit demselben Code zu besseren Ergebnissen in den Leistungen führen, wie im Fall von akustisch-verbalen (schriftlichen) und visuell-verbalen (mündlichen) Inhalten, können unterschiedliche Codes zu einer Verschlechterung der Leistung führen, wie für visuell-verbal und visuell-räumlich vermittelte Informationen in Form von Texten und Grafiken (vgl. Seeber 2017: 463). Bei multimodalen Inhalten spielen auch Synchronität und Abweichungen eine bedeutende Rolle. Synchronität ist grundlegend, um Inhalte aus unterschiedlichen Kanälen zu kombinieren. Kurze Verzögerungen werden jedoch in der Regel als synchron wahrgenommen und erfolgreich miteinander verknüpft (vgl. Seeber 2017: 464).

1.3.2 Ein kognitives Modell für multimodale Inputs

Seebers (2007) Cognitive Load Model stellte die Konfliktmatrix bei unterschiedlichen Teilaufgaben des Simultandolmetschens dar, und zwar, wie diese sich gegenseitig beeinflussen. Dieses Modell setzte implizit Laborbedingungen voraus, wobei visuelle Inputs nur für das Vom-Blatt-Dolmetschen berücksichtigt werden, während beim Simultandolmetschen nur akustische Inputs vorhanden sind. Da Dolmet-scher*innen unterschiedlichen Inputs ausgesetzt sind, entspricht dies aber nicht der Realität der Dolmetschsettings. Diese enthalten auch visuelle Inputs: sowohl Mimik, Gestik und Körpersprache als auch visuelle Hilfsmittel, Präsentationsfolien, usw. (vgl. Seeber 2017: 467). Die visuelle Modalität muss entsprechend auch in theoretischen Modellen

berücksichtigt werden, und zusammen mit der akustischen Modalität in die kognitive Verarbeitungsphase und teilweise in die mündliche Reaktion (die Dolmetschung) eingebettet werden. Das setzt eine komplexere Konfliktmatrix und Ressourcenverwendung voraus (Abb. 2).

In der überarbeiteten Konfliktmatrix wird also ein gesamter Interferenzwert von 11,6 erreicht, der höher als der Interferenzwert von 9 bei der vorherigen Version des Modells ist. Dieses Modell enthält aber nur einen Bezug auf visuell-räumliche Informationen wie zum Beispiel Gesichtsausdruck, Gestik und Körpersprache, während visuell-verbale Informationen wie Texte nicht berücksichtigt werden (vgl. Seeber 2017: 468).

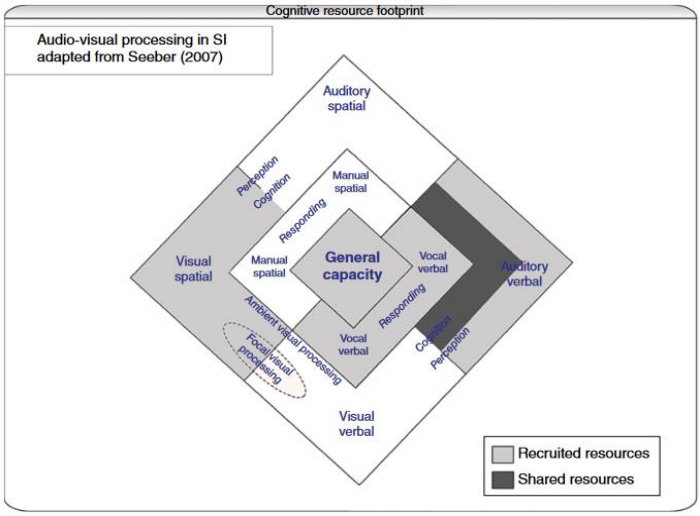

Abb. 2: Bedarf an kognitiven Ressourcen beim Simultandolmetschen mit visuellen Inputs nach Seeber (2017: 468)

Die Überarbeitung des Modells umfasst auch Änderungen der den Koeffizienten zugewiesenen Werte. Sind keine Informationen in einer Modalität vorhanden, wird ein Bedarfsvektor mit dem Wert 0 zugewiesen. Wenn Informationen nur in einer Modalität verfügbar sind, wird ein Bedarfsvektor mit dem Wert 1 festgelegt, während bei Informationen, die in einer Modalität verfügbar sind und eine andere Modalität ergänzen, der Wert 0,5 zugewiesen wird (vgl. Seeber 2017: 469). Das Cognitive Load Model des Simultandolmetschens mit visuellen Inputs setzt voraus, dass visuelle und akustische Informationen einander in der Phase des Sprachverständnissens ergänzen. Beide verbrauchen Ressourcen, aber ihre Komplementarität verringert den Gesamtbedarf, was sich in einem reduzierten Bedarfsvektor von 0,5 widerspiegelt. Dieses Modell stellt also eine umfassendere Beschreibung der Inputs dar, die beim Simultandolmetschen wahrgenommen werden (vgl. Seeber 2017: 469f.).

Visuelle Inputs im Laufe einer Konferenz können aber nicht nur visuell-räumliche Informationen sein, sondern auch Texte. In der traditionellen Form des Simultandolmetschens mit Text bekommen Dolmetscher*innen im Voraus ein Manuskript mit dem vollständigen Text, der von einem*r Redner*in vorgelesen wird. Die Verwendung von Präsentationsfolien mit visuellen Informationen während einer Konferenz ist jedoch auch üblich. Diese Situation weicht von der traditionellen Form vom Simultandolmetschen mit Text ab, da Präsentationsfolien unterschiedliche Arten von Informationen aufweisen, zum Beispiel Bilder, Texte und Zahlen (vgl. Seeber 2017: 470). Bilder werden in der visuell-räumlichen Modalität verarbeitet, während Texte in Form von Schlüsselwörtern, Zitaten oder vollständigen Sätzen in der visuell-verbalen Modalität verarbeitet werden. Zahlen, die in Form von Grafiken und Diagrammen dargestellt werden, werden als eine Mischung von visuell-verbaler und visuell-räumlicher Modalität verarbeitet. Ein weiterer Unterschied zwischen Folien und Manuskripten besteht in der Tatsache, dass Präsentationsfolien in der Regel keine exakte Wiederholung der mündlich gelieferten Informationen sind, sondern diese ergänzen (vgl. Seeber 2017: 470).

Anhand der traditionellen Version des Simultandolmetschens mit Text wurde ein neues Cognitive Load Model für das Simultandolmetschen mit Text entworfen. Die visuell-verbale Modalität wird dabei auch miteinbezogen, denn diese beeinflusst die Phasen der Wahrnehmung und der kognitiven Verarbeitung. Das Leseverstehen wird in die Verstehens- und Produktionsphase eingebracht, da es mit dem Sprachverständnis verglichen werden kann. Das führt dazu, dass in der Phase des Sprachverständnisses visuell-verbale, akustisch-verbale und akustisch-räumliche Informationen wahrgenommen werden. Danach erfolgt die kognitiv-räumliche und kognitiv-verbale Verarbeitung der Informationen (vgl. Seeber 2017: 471f.) (Abb. 3).

Die visuell-räumliche und akustisch-verbale Verarbeitung teilen sich einen Bedarfsvektor, während die visuell-verbale Verarbeitung einen Bedarfsvektor von 1 hat. Produktion und Monitoring ähneln dem Szenario beim Simultandolmetschen ohne Text, und die Phasen der Reaktion, Kognition und Wahrnehmung weisen in der verbalen Modalität jeweils einen Bedarfsvektor auf. Eine ausführliche Konfliktmatrix veranschaulicht die Interferenzen und die Wechselwirkung zwischen einzelnen Aufgaben beim Simultandolmetschen mit Text (Abb. 4).

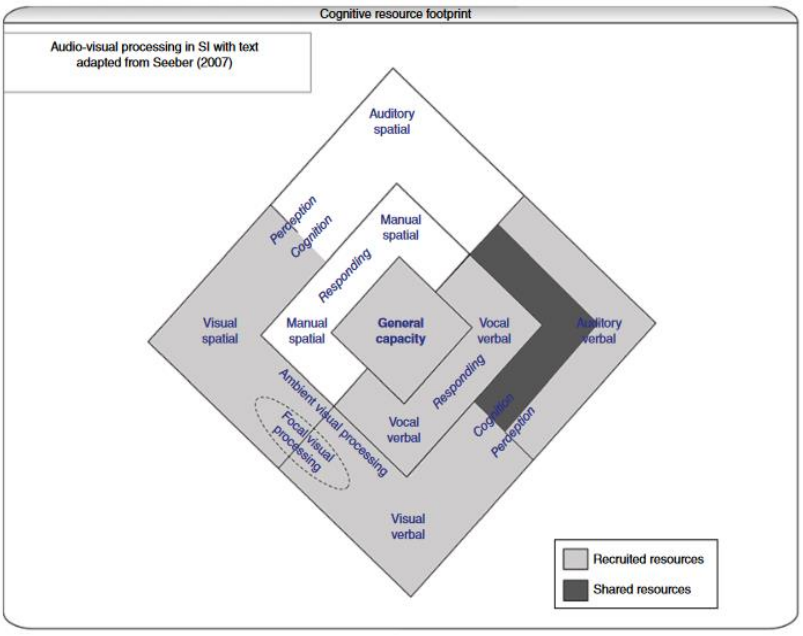

Abb. 3: Bedarf an kognitiven Ressourcen beim Simultandolmetschen mit Text nach Seeber (2017: 471)

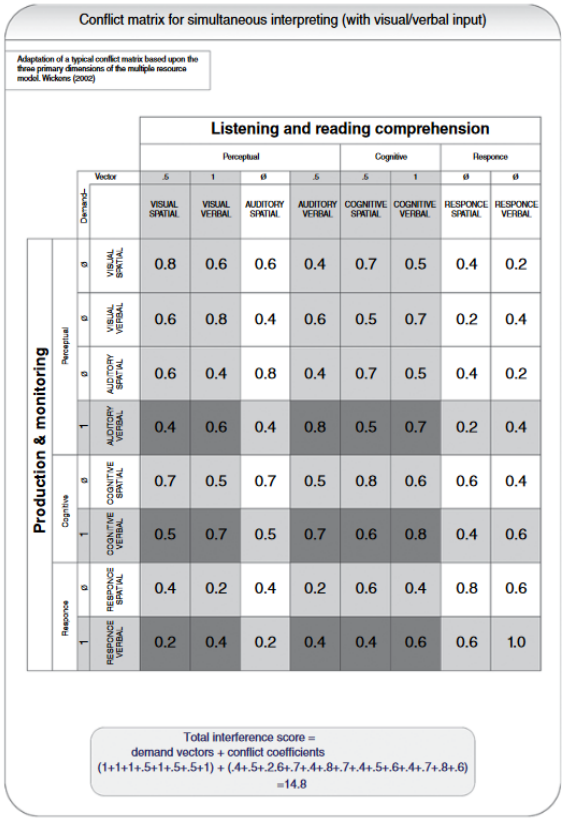

Abb. 4: Konfliktmatrix beim Simultandolmetschen mit Text nach Seeber (2017: 472)

Die Bedarfsvektoren und Konfliktkoeffizienten ermöglichen eine genaue Berechnung des Gesamtinterferenzwerts bei einer Dolmetschsituation mit multimodaler Verarbeitung. Die Summe der Bedarfsvektoren einschließlich des Konfliktkoeffizienten der einzelnen Verarbeitungsphasen und Modalitäten ergibt einen Gesamtinterferenzwert von 14,8. Das liegt deutlich über dem Wert des Simultandolmetschens ohne Text (vgl. Seeber 2017: 473).

Dieses überarbeitete Modell für die multimodale Verarbeitung beim Simultandolmetschen zeigt, dass das Simultandolmetschen mit Text einen höheren Gesamtinterferenzwert im Vergleich zum Simultandolmetschen ohne Text hat. Das Modell lässt also eine Zunahme des Gesamtinterferenzwerts beim Simultandolmetschen mit Text vorhersagen, was einer höheren kognitiven Belastung entspricht (vgl. Seeber 2017: 473). Obwohl der Zugang zu zusätzlichen visuelle Informationen während des Dolmetschens die Genauigkeit der Dolmetschleistung erhöhen könnte, wird dies nicht durch empirische Beweise gestützt. Dieses Ergebnis entspricht dem Eindruck vieler Dolmetscher*innen, dass das Dolmetschen mit Text anspruchsvoller als das Dolmetschen ohne Text ist (vgl. Seeber 2017: 473).

Prandi (2017, 2022) argumentiert, dass das Cognitive Load Model auch im Rahmen der CAI-Forschung eingesetzt werden kann. Der Bedarf an kognitiven Ressourcen beim Simultandolmetschen mit Text nach Seeber (2017) kann ergänzt werden, um die Suche nach Terminologie mit einem CAI-Tool oder mittels elektronischer Glossare einzuschließen. Manuell-räumliche Ressourcen werden in der Reaktionsphase verwendet, um die manuelle Suche an der Tastatur durchzuführen. Visuell-räumliche Ressourcen werden in der Wahrnehmungsphase eingesetzt, um Wörter auf dem Bildschirm zu finden, und in der Kognitionsphase, um Informationen zu verarbeiten. Das kann vor allem während der Suche nach Terminologie zu einer erhöhten kognitiven Belastung führen, die aber nach der Suche, und vor allem im Fall einer erfolgreichen Suche, verringert wird (vgl. Prandi 2017: 5f.). Die automatische Suchfunktion von CAI-Tools führt hingegen dazu, dass weniger manuell-räumliche und visuell-räumliche Aufgaben durchgeführt werden müssen (vgl. Prandi 2017: 9). Ein Vergleich des Bedarfs an kognitiven Ressourcen beim Simultandolmetschen mit manueller Suche (in CAI-Tools oder in digitalen Glossaren) und mit ASR zeigt, dass keine manuell-räumlichen Ressourcen bei der Verwendung einer ASR-Software herangezogen werden müssen. Dennoch werden visuell-räumliche und visuell-verbale Ressourcen verwendet, um die auf dem Bildschirm angezeigten Ausdrücke zu finden und zu verarbeiten (Abb. 5) (vgl. Prandi 2022: 102).

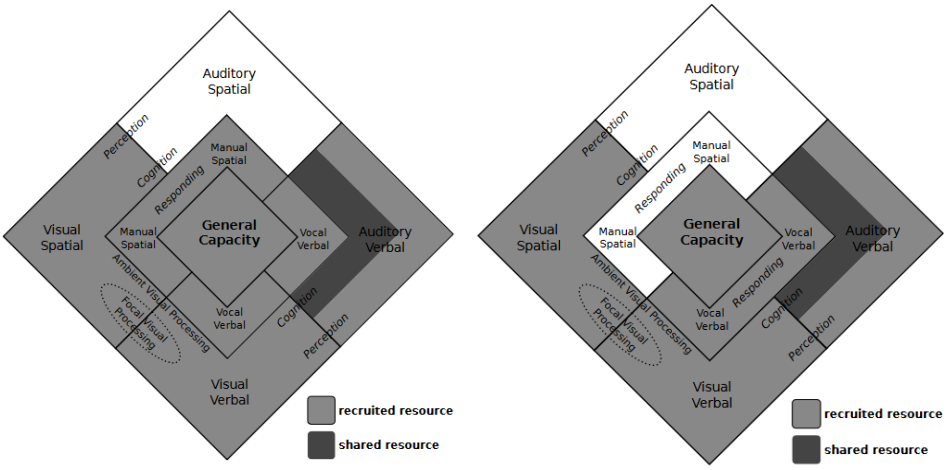

Abb. 5: Bedarf an kognitiven Ressourcen beim Simultandolmetschen mit manueller Suche (links) und mit ASR (rechts) nach Prandi (2022: 102f.)

Die Überarbeitung des Cognitive Load Models kann also verwendet werden, um die Wirkung terminologischer Hilfsmittel auf die kognitive Belastung beim Simultandolmetschen zu erforschen. Ein Beispiel hierfür ist die Forschungsarbeit von Prandi (2022). In dieser Studie wird ein Mixed-Method-Forschungsdesign verwendet, um produkt- und prozessorientierte Forschung im Bereich Computer Assisted Simultaneous Interpreting (abgekürzt: CASI) durchzuführen. Die Ergebnisse zeigen, dass ASR-gestützte CAI-Tools mit der automatischen Suche nach Terminologie zu einer niedrigeren kognitiven Belastung im Vergleich zur manuellen Suche nach Terminologie in Glossaren führen (vgl. Prandi 2022: 236).

1.4 Problemauslöser beim Simultandolmetschen

In der Beschreibung des Kapazitätenmodells und des damit verbundenen Bedarfs an Verarbeitungskapazität werden zudem jene Problemauslöser erwähnt, die mit einem zusätzlichen Verarbeitungsbedarf assoziiert werden können. Diese Problemauslöser sind vielen Dolmetscher*innen bewusst und finden in der Literatur häufig Erwähnung. In der Vergangenheit wurden sie jedoch nicht anhand eines konzeptionellen Rahmens analysiert (vgl. Gile 2009: 171). Gile (2009: 171) erwähnt folgende Merkmale einer Ausgangsrede, die als Problemauslöser gelten: Namen, Zahlen, Aufzählungen, schnelles Sprechtempo, starke Fremd- oder Regionalakzente, geringe Sprachlogik und schlechter Ton.

Die Komplexität von Problemauslösern hat vor allem im letzten Jahrzehnt das Interesse von Akteur*innen in der Dolmetschforschung und -praxis geweckt. Das Aufkommen neuer Spracherkennungssoftwares auf dem Markt hat zu neuen, innovativen Anwendungen geführt,

welche auf den Umgang mit Problemauslösern abzielen (vgl. Fantinuoli 2017: 26). Es ist daher von grundlegender Bedeutung, die Problemauslöser beim Simultandolmetschen genau zu untersuchen, um gezielte Lösungen zur Behebung der damit verbundenen Herausforderungen zu entwickeln.

1.4.1 Klassifizierung der Problemauslöser beim Simultandolmetschen

Kognitive Problemauslöser beim Simultandolmetschen können in unterschiedliche Kategorien unterteilt werden, je nachdem, auf welche Kapazität während des Simultandolmetschens sie sich auswirken. Sie entstehen durch einen Anstieg der erforderlichen Verarbeitungskapazität. Darunter zählen die hohe Informationsdichte, das schnelle Redetempo, unbekannte Namen, syntaktische Unterschiede zwischen Sprachen, die geringe Vorhersehbarkeit des Ausgangstextes und externe Faktoren (vgl. Gile 2009: 192f.).

Die hohe Informationsdichte und das schnelle Redetempo wirken sich sowohl auf die Rezeptionsleistung (Listening and Analysis Effort) als auch auf die Wiedergabeleistung (Production Effort) aus. Eine hohe Informationsdichte kann aber auch bei langsam vorgetragenen Reden verzeichnet werden, vor allem wenn Aufzählungen vorhanden sind (vgl. Gile 2009: 193). Auf die Rezeptionsleistung wirken sich auch externe Faktoren, zum Beispiel mangelnde Tonqualität, Hintergrundgeräusche oder störende Geräusche aus. Zur Erschwerung dieser Leistung zählen auch starke Akzente, grammatikalische und syntaktische Fehler und ein ungewöhnlicher Sprachstil (vgl. Gile 2009: 193).

Auf die Gedächtnisleistung wirken sich vor allem unbekannte Namen und gewisse Sprachkombinationen aus. Kognitive Sättigung findet bei Sprachkombinationen statt, in denen Sprachen syntaktisch stark voneinander abweichen und Informationen länger im Gedächtnis gespeichert werden müssen, bevor sie in den Zieltext eingebettet werden können. Ein Beispiel ist die Sprachkombination Englisch-Deutsch, in der die Satzstrukturen der beiden Sprachen nicht übereinstimmen (vgl. Gile 2009: 192f.).

Schwierigkeiten beim Dolmetschen entstehen aber nicht nur aus erhöhtem Kapazitätsbedarf sondern können auch auf einzelne Sprachsegmente zurückgeführt werden, die eine besondere Anfälligkeit für Probleme aufweisen. Diese Segmente können sehr kurz sein, geringe Redundanz haben oder beim Zuhören schwer zu erkennen sein, wie zum Beispiel Zahlen, kurze Namen oder Akronyme (vgl. Gile 2009: 194). Außerdem sind einzelne Problemauslöser nicht der einzige Grund für erhöhte kognitive Belastung. Die importierte bzw. exportierte kognitive Belastung aus anderen

Sätzen und die Verteilung der Informationen entlang einzelner Sätze können auch der Grund einer erhöhten kognitiven Belastung sein. Diese erklären, warum manche Fehler nicht bei der Wiedergabe von Problemauslösern, sondern in deren Nähe auftreten (vgl. Gile 2008: 62f.).

1.4.2 Wirkung von Problemauslösern auf das Simultandolmetschen

Die verfügbare Kapazität an kognitiven Ressourcen während des Simultandolmetschens ist stark von Problemauslösern beeinflusst. Wenn der kognitive Bedarf die verfügbare Kapazität übersteigt, kann es zu Problemen bei der Aufmerksamkeitsverteilung kommen, die wiederum zur Verschlechterung der Dolmetschqualität führen. Diese Schwierigkeiten beziehen sich entweder auf die Verarbeitung der Ausgangsrede, was zu Fehlern und Auslassungen führt, oder auf die Qualität der Dolmetschleistung, was zu Mängeln in der sprachlichen Wiedergabe, in der Stimmführung oder Intonation führt (vgl. Gile 2009: 171).

Wenn Probleme bzw. Schwierigkeiten entstehen, sollten Dolmetscher*innen trotz der Problemauslöser bzw. der begrenzten Verarbeitungskapazität versuchen, die größtmögliche Menge an Informationen wiederzugeben, die negativen Auswirkungen auf andere Sprachsegmente zu begrenzen und die kommunikative Wirkung der Rede zu gewährleisten (vgl. Gile 2009: 211f.). Es gibt unterschiedliche Strategien, um diese Ziele zu erreichen.

Um Verständnisprobleme zu bewältigen, kann zum Beispiel die Wiedergabe der Dolmetschung einige Sekunden verzögert werden, um Zeit zum Nachdenken zu ermöglichen. Der Kontext der Rede kann auch verwendet werden, um den Sinn zu erschließen. Die Unterstützung von Kabinenpartner*innen oder der Zugriff auf Ressourcen in der Kabine können dabei hilfreich sein, Schwierigkeiten zu beheben (vgl. Gile 2009: 201ff.). Zu den Vorbeugungsstrategien zählen das Notieren von Zahlen oder Namen, die Verlängerung oder Verkürzung der Zeitverzögerung (Timelag), die Segmentierung von Informationen und die Änderung der Reihenfolge bei einer Aufzählung (vgl. Gile 2009: 204f.). Abschließend sind Strategien zur Umformulierung der Inhalte unter anderem die Ersetzung von Ausdrücken durch einen übergeordneten oder allgemeineren Begriff, das Paraphrasieren und die Wiedergabe des in der Ausgangssprache Gehörten (vgl. Gile 2009: 206f.).

1.4.3 Zahlen als Problemauslöser beim Simultandolmetschen

Wie aus dem vorigen Kapitel geschlussfolgert werden kann, stellen Zahlen eine große Herausforderung für Dolmetscher*innen dar. Sie beeinträchtigen die Qualität der Wiedergabe in Bezug auf Vollständigkeit, Genauigkeit, Plausibilität, Flüssigkeit und Wirkung der Dolmetschleistung (vgl. Frittella 2019: 37). Diese Schwierigkeiten sind auf die inhärenten kognitiven Prozesse zurückzuführen, die für die Verarbeitung und Übersetzung von Zahlen aus einer Ausgangssprache in eine Zielsprache notwendig sind, aber auch auf die Funktion von Zahlen in Texten. Aufgrund ihrer kommunikativen Funktion und ihrer semantischen, pragmatischen und kognitiven Struktur heben sich Zahlen als Elemente von besonderer Schwierigkeit in Texten ab (vgl. Frittella 2019: 37).

1.4.3.1 Kognitive Verarbeitung von Zahlen beim Dolmetschen

In Bezug auf numerische Systeme ist es zunächst wichtig, zwischen Zahlen und Zahlwörtern zu differenzieren. Zahlen sind arithmetische Gegenstände, während Zahlwörter die sprachlichen Ausdrücke sind, die diese beschreiben (vgl. Pinochi 2009: 34). Zahlwörter sind eine schriftliche oder mündliche Darstellung von Zahlen und setzen sich aus bestimmten Strukturen zusammen, um numerische Einheiten sprachlich auszudrücken. Sie umfassen syntaktische Elemente, die unterschiedlich kombiniert werden können, um eine unendliche Anzahl von Zahlen in jeder Sprache wiederzugeben (vgl. Braun & Clarici 1996: 85). Die kognitive Verarbeitung von Zahlen besteht aus verschiedenen Phasen (Verständnis und Produktion), Codes (mündlich und arabisch) und Modalitäten des Ausdrucks (schriftlich und mündlich). In der Verständnisphase werden Zahlen als eine Reihenfolge von Ziffern oder Wörtern verarbeitet, um eine mentale Darstellung der Zahl zu schaffen, während in der Produktionsphase der Prozess umgekehrt abläuft, um Zahlen in Sprachform zu äußern (vgl. Braun & Clarici 1966: 85). Die kognitive Verarbeitung von Zahlen kann außerdem im arabischen oder im verbalen Code erfolgen, und zwar mit Ziffern oder Wörtern. Der arabische Code wird nur schriftlich dargestellt, während der verbale Code entweder schriftlich oder phonologisch ausgedrückt werden kann (vgl. Braun & Clarici 1996: 86). Während des Simultandolmetschens erfolgt zusätzlich, zwischen der Verständnis- und Produktionsphase, auch die Übersetzung der Zahlen, die in der mündlichen, phonologischen Modalität wiedergegeben werden.

Um zu verstehen, wie Zahlen in eine Dolmetschung eingebettet werden können, ist es zunächst wichtig, die kognitive Verarbeitung von Zahlen näher zu erläutern. Die syntaktischen und lexikalischen Repräsentationen von Zahlen sind die zwei getrennten Mechanismen, um Zahlen zu verarbeiten. Zunächst findet die syntaktische Repräsentation statt, welche ein allgemeines Schema für die Größenordnung der Zahl aufstellt, wie zum Beispiel Tausend, Hundert, Zehner, usw. Die lexikalische Repräsentation enthält hingegen Informationen über die genauen Mengenangaben. Bei der Dolmetschung von Zahlen müssen also diese zwei Bestandteile berücksichtigt werden, um die richtigen numerischen, phonologischen und orthographischen Merkmale wiederzugeben (vgl. Braun & Clarici 1996: 87).

Die Verarbeitung von Zahlen beim Simultandolmetschen unterscheidet sich grundlegend von der Verarbeitung von Textinhalten, und die Hör- und Verarbeitungsstrategien beim Dolmetschen müssen entsprechend angepasst werden. Während für Textinhalte in der Regel die allgemeine Bedeutung verstanden wird und die nachfolgenden Stellen antizipiert werden können, muss bei Zahlen eine wörtliche Transposition durchgeführt werden, da diese nicht aus dem Kontext abgeleitet werden können. Dies führt zu Unterbrechungen der ausgeübten geistigen Tätigkeit beim Dolmetschen, um die Art und Weise der Verarbeitung von Informationen umzustellen. Zu den Elementen, welche eine ähnliche Verarbeitung erfordern, zählen Namen, Akronyme und Fachbegriffe, die beim Dolmetschen schwer antizipiert werden können und wofür eine wörtliche Übersetzung bzw. Entsprechung verfügbar ist (vgl. Braun & Clarici 1996: 87f.). Die Herausforderungen bei der Wiedergabe von Zahlen beim Dolmetschen führen zu Problemen in folgenden Bereichen: Verständnis (zum Beispiel in der Dekodierung und Transkodierung von Zahlen in der Ausgangs- und Zielsprache), Gedächtnis (zum Beispiel Fehler bei der Speicherung von lexikalischen Elementen), Ausgangssprache (zum Beispiel Fehler in der akustischen Wahrnehmung), Informationsdichte und subjektive Fähigkeiten von Dolmetscher*innen (vgl. Frittella 2019: 81ff.).

1.4.3.2 Studien über die Dolmetschung von Zahlen

Die inhärente Schwierigkeit bei der Dolmetschung von Zahlen hat diese zu einem weit verbreiteten Forschungsschwerpunkt gemacht. Zahlen werden in allen Dolmetschmodi als eine besondere Herausforderung gesehen, was sich in der Genauigkeit ihrer Wiedergabe widerspiegelt, wie mehrere Studien zeigen.

In vielen Forschungsarbeiten wird die von Braun & Clarici (1996) vorgenommene Unterscheidung von Fehlerkategorien für Zahlen verwendet (z. B. Arzik Erzurumlu & Demir 2022; Bereznaya 2022; Canali 2019; Desmet et al. 2018; Fantinuoli & Pisani 2021; Mazza 2001; Penta 2022). Diese Fehlerkategorien beschreiben typische Fehler in der Wiedergabe von Zahlen beim (Simultan)dolmetschen und umfassen Auslassungen, Annäherungen, lexikalische Fehler, syntaktische Fehler, Fehler bei Umstellungen, strukturell fehlerhafte Zahlen, Fehler bei der phonetischen Wahrnehmung, Selbstkorrekturen und sonstige Fehler (vgl. Braun & Clarici 1996: 92). Auf Basis dieser Fehlerkategorien führten Braun & Clarici (1996) eine Untersuchung durch. Die Ergebnisse zeigten, dass die zwölf untersuchten Dolmetschstudierenden eine durchschnittliche Fehlerquote bei Zahlen von 69,49 % bei den Simultandolmetschungen von acht Texten auf Deutsch und Italienisch verzeichneten (vgl. Braun & Clarici 1996: 92). Die Mehrheit der Fehler wurde folgenden Fehlerkategorien zugeordnet: Auslassungen (47,6 %), Annäherungen, lexikalische und syntaktische Fehler.

Auf diese Studie folgte eine Reihe von Veröffentlichungen mit vergleichbaren Ergebnissen. Mazza (2001) stützte sich auf diese Untersuchung, um eine ähnliche Studie durchzuführen. Durchschnittlich war die Fehlerquote der Zahlen bei den zwei gedolmetschten Texten in dieser Studie jeweils 54,9 % und 50,1 % (vgl. Mazza 2001: 96). Pinochi (2009: 45) präsentierte ähnliche Ergebnisse in einer Studie mit der Sprachkombination Deutsch-Italienisch und Englisch-Italienisch, die jeweils durchschnittliche Fehlerquoten von 40,6 % und 41,2 % verzeichneten. Desmet et al. (2018) stellten fest, dass die durchschnittliche Fehlerquote bei der Wiedergabe von Zahlen durch Dolmetschstudierende bei 43,5 % lag. Die häufigsten Fehlerkategorien waren dabei Auslassungen (61 %), lexikalische Fehler, Annäherungen und syntaktische Fehler (vgl. Desmet et al. 2018: 22ff.).

Alle diese Studien befassten sich mit Dolmetschleistungen von Dolmetschstudierenden. Kajzer-Wietrzny et al. (2021) fokussierte sich hingegen auf aufgezeichnete Dolmetschleistungen professioneller Dolmetscher*innen beim Europäischen Parlament. Obwohl die Genauigkeitsquoten höher als bei anderen Studien waren und Dolmetscher*innen im Europäischen Parlament möglicherweise Zugang zu Unterlagen in der Kabine hatten, wurden bis zu 25 % der Zahlen ausgelassen und bis zu 7 % falsch wiedergegeben (vgl. Kajzer-Wietrzny et al. 2021: 482). In einer späteren Studie wurde festgestellt, dass vor allem zwei Phänomene zu Unterbrechungen im

Redefluss führten: das Auftreten von mehr als einer Zahl in einem Satz und Zahlen in Kombination mit Namen (vgl. Kajzer-Wietrzny 2023).

Aufgrund ihrer inhärenten Schwierigkeit und der hohen Fehlerquoten bei der Wiedergabe dieser Problemauslöser wurden Zahlen bald zu einem Schwerpunkt bei der Entwicklung von Ad-hoc-Anwendungen zur Unterstützung von Dolmetscher*innen in der Kabine. Automatische Spracherkennungssoftwares zählen zu diesen Lösungen und wurden als Schwerpunkt dieser Arbeit ausgewählt, um einen Beitrag zur Forschung über das Thema der Wiedergabe der Zahlen zu leisten. Aus diesem Grund werden auf die CAI-Tools, ihre Entwicklung und ihr Potenzial im nächsten Kapitel eingegangen, und ASR-gestützten CAI-Tools ausführlich vorgestellt.

2 Technologische Lösungen für Dolmetscher*innen

Das Simultandolmetschen ist seit seiner Geburtsstunde durch den Einsatz von Technologie gekennzeichnet. Die technische Ausrüstung zur Sprachübertragung ist die Grundlage, um den Konferenzteilnehmer*innen die Dolmetschungen in Echtzeit zu übermitteln, und genau die Einführung von patentierten Systemen war der Ausgangspunkt für die Entstehung des Simultandolmetschens als Dolmetschmodus. Seit dem Aufkommen von Computern und dem World Wide Web wurde es außerdem auch für Dolmetscher*innen unerlässlich, computergestützte Lösungen in ihren Arbeitsalltag einzuführen, um dem Zeitgeist gemäß zu arbeiten und die steigende Geschwindigkeit und Komplexität der Informationsverwaltung zu bewältigen.

Trotz der zunehmenden Verbreitung der Technologie in der modernen Welt hat die Informations- und Kommunikationstechnologie (IKT) die Dolmetschtätigkeit im Vergleich zu anderen Berufen nicht maßgeblich verändert. Das wird gezeigt durch die Tatsache, dass die Dolmetschtätigkeit seit der Einführung vom Simultandolmetschen praktisch unverändert geblieben ist (vgl. Fantinuoli 2018a: 154). Laut Fantinuoli (2018b) können drei technologische Wenden in der Geschichte des Dolmetschens erkannt werden. In den 1920er Jahren fanden die ersten Versuche statt, verkabelte Systeme für die Sprachübertragung zu verwenden. Diese haben zur Entstehung des Simultandolmetschens geführt. Durch die Nürnberger Prozesse (1946-1949) haben diese Systeme an Sichtbarkeit gewonnen und seit damals werden von allen internationalen Organisationen verwendet (vgl. Fantinuoli 2018b: 2). Die zweite technologische Wende fand in den 1990er Jahren statt, als das World Wide Web entstand. Sein Aufkommen hat den Zugang von Dolmetscher*innen zu Informationen erleichtert und einen Einfluss auf die Vorbereitung von Dolmetschaufträgen, auf die Verwaltung großer Mengen an Konferenzunterlagen und auf die terminologische Suche in der Kabine gehabt (vgl. Fantinuoli 2018a: 154). Die aktuelle dritte technologische Wende wirkt sich vor allem auf das Berufsbild und die damit zusammenhängenden sozioökonomischen Aspekte. Diese dritte Wende wurde von drei Technologien eingeleitet, nämlich Computer-Assisted Interpreting Tools (CAI), Remote Interpreting (RI) und Machine Interpreting (MI) (vgl. Fantinuoli 2018b: 3).

Aus einer breiteren Perspektive sind die technologischen Entwicklungen im Dolmetschbereich Teil der technologischen Wende, die in der heutigen Gesellschaft in unterschiedlichen Sektoren stattfindet. Diese ist auf unterschiedliche Faktoren zurückzuführen: auf die soziale Beschleunigung des modernen Lebens, auf den Automatisierungs- und Entwicklungsprozess als

anthropologischer Antrieb und biologisches Grundprinzip, auf wirtschaftliche Bedingungen und auf Änderungen sozialpsychologischer Natur (vgl. Fantinuoli 2019). Auch das Dolmetschen ist entsprechend von diesen gesamtgesellschaftlichen Faktoren beeinflusst und wird in der Zukunft immer mehr von digitalen Veränderungen betroffen sein, die je nach Bedürfnis, Dolmetschmodus und Dolmetschsetting anders ablaufen werden (vgl. Fantinuoli 2019: 342).

In den letzten Jahrzehnten wurden also unterschiedliche Anwendungen für die Unterstützung von Dolmetscher*innen auf den Markt gebracht. Ihre Entwicklung war unzulänglich, aber gleichzeitig unaufhaltbar, und heutzutage steht das Dolmetschen erneut vor großen technologischen Veränderungen, die unser heutiges Verständnis von Dolmetschen verändern könnten.

2.1 Definition und Klassifizierung von CAI-Tools und technologischen Lösungen für Dolmetscher*innen

Das computergestützte Dolmetschen wird wie folgt definiert: "Computer-assisted or computer-aided interpreting (CAI) is defined as a form of speech translation where a human interpreter is supported by computer software developed to facilitate some aspects of the interpreting task, from preparation to information access." (Fantinuoli 2021: 512) CAI-Tools sind also solche Computersoftwares, die von Dolmetscher*innen verwendet werden, um die Dolmetschqualität und die Produktivität zu steigern. Sie dienen dazu, die Arbeitsbedingungen von Dolmetscher*innen zu optimieren, indem sie zeitaufwendige Tätigkeiten durchführen und vor allem in der Vorbereitungsphase Hilfe leisten (vgl. Fantinuoli 2018b: 4). CAI-Tools können bei der Erstellung von Glossaren in der Vorbereitung auf den Dolmetscheinsatz und bei der terminologischen Suche in der Kabine unterstützen. Sie können außerdem Anwendungen zur Verarbeitung natürlicher Sprache (auf Englisch: Natural Language Processing, abgekürzt NLP) integrieren, um Terminologie automatisch zu extrahieren, Schlüsselthemen auszufiltern, Inhalte zusammenzufassen oder automatische Spracherkennung durchzuführen (vgl. Fantinuoli 2018b: 4).

Es gibt mehreren Kriterien, die für die Klassifizierung der Technologien für Dolmetscher*innen verwendet werden können. Fantinuoli (2018a: 155) beschreibt zwei Gruppen von technologischen Lösungen: prozessorientierte und settingorientierte Technologien. Als prozessorientiert versteht man Terminologieverwaltungssysteme, Softwares zur Datenextraktion und Korpusanalyseprogramme. Diese dienen zur Unterstützung von Dolmetscher*innen während der Einzelphasen des Dolmetschprozesses, nämlich vor, während und nach der Dolmetschung.

Dazu zählen auch Computer-Assisted bzw. Computer-Aided Interpreting (CAI) Tools. Zu den settingorientierten Lösungen zählen Technologien wie die Konsolen in den Kabinen, die Technik für das Remote Dolmetschen, Weiterbildungsplattformen usw. Diese betreffen vor allem die Rahmenbedingungen des Dolmetschens und sind nicht unmittelbar mit den kognitiven Prozessen des Dolmetschens verknüpft.

Es können außerdem drei Generationen von CAI-Tools differenziert werden, die im Laufe der Jahrzehnte entwickelt worden sind. CAI-Tools der ersten Generation erfüllen hauptsächlich die Funktion von Terminologieverwaltungssystemen, deren Oberfläche für Dolmetscher*innen geeignet entwickelt wurde. Diese sind zum Beispiel Interplex[1], Terminus[2], Interpreters' Help[3], LookUp[4] und DolTerm[5]. Sie sind von einfachen Eintragsstrukturen gekennzeichnet, die während des Dolmetschens übersichtlich sind (vgl. Fantinuoli 2018a: 164). Tools der zweiten Generation können sich auf vorläufige Forschungsergebnisse stützen und bieten einen ganzheitlichen Lösungsansatz. Sie weisen zusätzliche Funktionen auf, wie zum Beispiel die Verwaltung von Unterlagen und die Extraktion von Informationen aus Korpora. Sie können bei unterschiedlichen Phasen des Dolmetschprozesses angewendet werden. Zwei Beispiele dafür sind InterpretBank[6] und Intragloss[7]. Die dritte Generation von CAI-Tools stützt sich auf künstlicher Intelligenz und automatischer Spracherkennung und dieser Kategorie können InterpretBank ASR[8], Kudo InterpreterAssist[9] und SmarTerp[10] zugeordnet werden (vgl. Frittella 2023: 51).

Braun (2019: 271) unterscheidet ebenfalls drei Arten von Technologien für das Dolmetschen: Technologien, um Dolmetschdienstleistungen zu liefern und ihre Reichweite zu erhöhen (Technology-Mediated oder Distance Interpreting); Technologien als Hilfsmittel vor, während und nach dem Dolmetscheinsatz (Technology-Supported Interpreting) und Technologien, welche Dolmetscher*innen ersetzen (Technology-Generated oder Machine Interpreting). Fantinuoli (2021: 509) beschreibt eine ähnliche Differenzierung von Tools. Darunter werden

[1] https://www.fourwillows.com/interplex.html
[2] https://www.wintringham.ch/cgi/ayawp.pl?T=terminus
[3] https://interpretershelp.com/
[4] http://www.lookup-web.de/introduction/index.html
[5] https://medien-hd.iued.uni-heidelberg.de/wordpress/digilab/dolterm.htm
[6] https://interpretbank.com/site/
[7] https://www.nimdzi.com/tbs/intragloss/
[8] https://asr.interpretbank.com/manual
[9] https://kudoway.com/articles/the-tech-behind-kudo-interpreter-assist/
[10] https://smarterp.me/

folgende Kategorien erwähnt: Tools, die für die Aus- bzw. Weiterbildung von Dolmetscher*innen gedacht sind (Computer-Assisted Interpreting Training Tools, CAIT); Tools für die Digitalisierung von Dolmetschprozessen (CAI-Tools); Technologien, die für das Remote Dolmetschen gedacht sind und Machine Interpreting, welches die vollständige Automatisierung von Dolmetschprozessen voraussetzt.

2.2 Geschichtliche Entwicklung von CAI-Tools

Dolmetscher*innen bereiten sich linguistisch und fachspezifisch vor, um Verständnis von Fachthemen zu erlangen und korrekte Terminologie bei jedem Dolmetscheinsatz zu verwenden (vgl. Fantinuoli 2016: 43). Die Notwendigkeit, sich terminologisch und kontextspezifisch für jeden Dolmetscheinsatz vorzubereiten, entsteht aus der Wissensdiskrepanz, die zwischen den Dolmetscher*innen und den Konferenzteilnehmer*innen bestehen. Diese betreffen vor allem die Ebenen des Inhalts, der Terminologie und der Phraseologie (vgl. Fantinuoli 2017: 113). Es können unterschiedliche Strategien beim Dolmetschen verwendet werden, um mit unbekannter Terminologie umzugehen, wie das Paraphrasieren, Auslassungen oder die Verwendung von Oberbegriffen. Trotzdem ist präzise Fachterminologie notwendig, um eine reibungslose Kommunikation zu erzielen und die Wahrnehmung von Dolmetscher*innen als kompetente Akteure zu gewährleisten (vgl. Fantinuoli 2016: 43).

Um die Bedürfnisse von Dolmetscher*innen bei der Vorbereitung auf Dolmetschaufträge zu erfüllen, wurden im letzten Jahrzehnt unterschiedliche technologische Lösungen auf den Markt gebracht. CAI-Tools wurden erst entwickelt, um gezielt bei der Vorbereitungsphase Dolmetscher*innen zu unterstützen und bestehende Lücken zu schließen. Glossare, die für das Simultandolmetschen als Vorbereitung auf die Aufträge verwendet wurden, waren ursprünglich einfache Vokabellisten auf Papier, die aus den Konferenzunterlagen extrahiert wurden und nach dem Dolmetschauftrag keine Verwendung mehr hatten. Während der Dolmetschung konnten sie schwer durchgeblättert werden, um die benötigte Terminologie zu finden (vgl. Will 2020: 41). In einem unveröffentlichten Beitrag unter dem Namen ‘La cabine de l’an 2000, quelques reflexion’ präsentiert Quicheron im Jahr 1995 die Ziele, Bedürfnisse und Voraussetzungen für Dolmetscher*innen im neuen Jahrhundert. Das Ziel, so viele Informationen in elektronischer Form wie möglich in der Kabine zur Verfügung zu stellen, war aufgrund der steigenden Komplexität der Konferenzinhalten notwendig. Es hätte das unnötige Drucken von Konferenzunterlagen vermieden und den Austausch zwischen Kolleg*innen mit ihren Personal Computern ermöglicht. Zu den

technischen Voraussetzungen, um die Verwendung von Informationstechnologien in den Kabinen zu ermöglichen, wurden folgende Kriterien erwähnt: geringer Platzbedarf in den Kabinen, die Mitberücksichtigung aller notwendigen Sprachen, die Benutzerfreundlichkeit, die Echtzeitreaktion, die Verfügbarkeit während der gesamten Konferenz und die Einbindung aller Systembestandteile (vgl. Mouzourakis 2000: 1f.).

Die ersten digitalisierten Glossare waren selbsterstellte Word- oder Excel-Listen, wurden lange Zeit für die Terminologieverwaltung bevorzugt (vgl. Fantinuoli 2009: 2). Diese stellten jedoch einige Schwierigkeiten bei der Verwendung in der Kabine dar und konnten keine schnelle und bedarfsorientierte Suche nach Termini gewährleisten. Laut Fantinuoli (2009: 4) mussten terminologische Ressourcen für die Kabine unterschiedliche Voraussetzungen erfüllen, darunter die Übersichtlichkeit des verwendeten Tools, eine schnelle und ergonomische Such- bzw. Eingabefunktion für die Hinzufügung neuer Termini und eine intuitive Steuerung.

Terminologiesysteme, die in der Übersetzungsbranche bereits weite Verbreitung gefunden haben, waren auch eine Möglichkeit, um komplexere Terminologiebestände zu verwalten. Die Notwendigkeit, Softwares gezielt für Dolmetscher*innen zu entwickeln, war aber schon immer ein wichtiges Anliegen. Moser-Mercer (1992: 511) hob hervor, dass Softwareentwickler*innen Anwendungen auf den Markt bringen müssen, welche die Bedürfnisse nicht nur der Übersetzer*innen, sondern auch der Dolmetscher*innen erfüllen. Eine dolmetschorientierte Terminologiearbeit unterscheidet sich maßgeblich von der übersetzungsorientierten Terminologie-arbeit, da andere Voraussetzungen erfüllt werden müssen. Der Wissenserwerb findet in der Dolmetschpraxis vor allem durch Parallelkorpora statt und muss während der Dolmetschung innerhalb weniger Sekunden abrufbar sein und in die zielsprachige Wiedergabe eingebettet werden (vgl. Will 2015: 182). Dementsprechend erwiesen sich diese Anwendungen als ungeeignet für die Bedürfnisse der Dolmetscher*innen. Simultanfähige Terminologiesysteme richten sich hingegen speziell an Simultandolmetscher*innen. Will (2015: 183) präsentiert unterschiedliche Kriterien, um dolmetschorientierte Terminologiearbeit umzusetzen. Dazu zählen Adäquatheit und Muster-bildung, um die Zuordnung und schnelle Abrufung von Informationen zu ermöglichen; Simultanfähigkeit, um diese Informationen in kurzer Zeit und ohne Ablenkungen zu visualisieren; und die phasenspezifische Verwendung, um Informationen vor, während und nach der Konferenz sinnvoll zu gestalten. Terminologiesysteme müssen entsprechend angepasst werden und drei Merkmale aufweisen. Diese bestehen aus einer flexiblen Ansicht, um die Visualisierungsoptionen

je nach Dolmetschphase bzw. je nach individuellen Bedürfnissen anzupassen; aus umfassenden Datenbank- und Hilfsfunktionen, um Filter-, Such- und Aufbereitungsmöglichkeiten zu bieten und die Erstellung terminologischer Einträge zu vereinfachen; und aus einer ergonomischen Bedienung, um intuitive Anwendungen zu gestalten (vgl. Will 2015: 186f.).

Auf der Grundlage der rudimentären Software zur dolmetschorientierten Terminologieverwaltung wurden komplexere Anwendungen entwickelt. Die zweite Generation von CAI-Tools wurde von der Einführung von zwei Softwares gekennzeichnet, nämlich InterpretBank und Intragloss. Der Fokus von Intragloss lag vor allem auf der Vorbereitungsphase, wobei Vorbereitungsunterlagen und terminologische Datenbanken auf direktem Wege interagieren können (vgl. Fantinuoli 2018a: 166). InterpretBank, das im Rahmen eines Forschungsprojekts an der Universität Mainz/Germersheim zwischen 2008 und 2012 entwickelt wurde, bot die Möglichkeit, eine linguistische und extralinguistische Vorbereitung für den Dolmetscheinsatz durchzuführen, terminologische Ressourcen zu verwalten und auf relevante Terminologie in der Kabine zuzugreifen. Es unterschied sich von anderen Tools für seine Struktur, die auf einzelne Bauteile, die sogenannte Module, bestand, um Dolmetscher*innen in den einzelnen Phasen des Dolmetschprozesses zu unterstützen (vgl. Fantinuoli 2016: 42). Diese Module waren ursprünglich ConferenceMode für den Simultaneinsatz, TermMode für die Erstellung der Terminologiebestände, CorpusMode für die Termextraktion und die Erstellung von Korpora und DrillMode für das dolmetschspezifische Lernen der Glossare (vgl. Fantinuoli 2009: 3). Die Suchfunktion in der Kabine wurde im Laufe der Jahre mehrmals verbessert. Die Fuzzy Search wurde hinzugefügt, um zu vermeiden, dass Tippfehler die Suche beeinträchtigen. Die dynamische Suche ermöglichte die Visualisierung von Ergebnissen, ohne auf Enter zu drücken, um die Treffer schon während der Eingabe anzuzeigen, während die progressive Suche eine hierarchische Suche je nach Relevanz der Glossare ermöglichte. Diese Funktionen konnten unterschiedliche Einschränkungen der vergangenen Tools beheben und damit die Wahrscheinlichkeit steigern, dass Dolmetscher*innen schnell relevante Suchergebnisse erlangen (vgl. Fantinuoli 2018a: 166).

Trotz technologischer Fortschritte haben sich CAI-Tools in der Praxis keiner großen Beliebtheit erfreut. Eine Umfrage aus dem Jahr 2016 zeigte, dass Dolmetscher*innen sich während des Simultaneinsatzes auf bilinguale Wörterbücher, auf selbsterstellte Glossare und auf die Unterstützung von Kabinenparter*innen stützten, während Datenbanken in der Vorbereitungsphase verwendet werden (vgl. Corpas Pastor & Fern 2016: 26). CAI-Tools wurden in dieser Umfrage

nicht erwähnt, und viele Befragten gaben an, dass ihnen das Vertrauen in Technologien fehlte, vor allem, weil sie diese als zeitaufwendige Ablenkungsfaktoren sahen (vgl. Corpas Pastor & Fern 2016: 34). Auch Fantinuoli (2018a: 163) argumentierte, dass die Wirkung von CAI-Tools im Dolmetschbereich nur begrenzt ist. Die Zurückhaltung bei der Anwendung von solchen Programmen kann auf Zweifeln über ihre Wirkung auf die intellektuelle Tätigkeit von Dolmetscher*innen sowie auf ihre Wahrnehmung als zeitaufwändige und ablenkende Faktoren während des Dolmetschens zurückgeführt werden.

CAI-Tools weisen außerdem viele Schwierigkeiten auf, welche ihre Popularität beschränkt haben; unter anderem ist die Verwendung von zusätzlichen Ressourcen mit der komplexen Tätigkeit des Simultandolmetschens schwer zu kombinieren. Die Technologien sind außerdem sehr heterogen, weil sie den persönlichen Bedürfnissen und Vorlieben der Entwickler*innen entsprechen, die häufig auch Dolmetscher*innen sind, und nicht die Präferenzen einer größeren Gemeinschaft vertreten (vgl. Fantinuoli 2018a: 164). Die große Anzahl der individuell gestalteten Programme, die einen hohen Grad an Heterogenität aufweisen, war ein Grund für die beschränkte Nutzung und Beliebtheit von diesen Tools (vgl. Will 2015: 180), und die mangelnde Miteinbeziehung von großen Stakeholders der Sprachindustrie führte außerdem dazu, dass Produkte häufig eingestellt oder nicht mehr weiterentwickelt wurden (vgl. Will 2020: 44).

Die mangelnde Forschung über die Rolle von Terminologie- und Wissenserwerb wird in mehreren Studien thematisiert und kritisiert (vgl. Fantinuoli 2018a: 163). Die ersten Veröffentlichungen zeigten jedoch die Wirkung dieser Tools auf die Wiedergabe von Fachtermini in den Dolmetschungen und konnten wissenschaftliche Grundlagen zur Verbesserung bestehender Anwendungen stellen. In Xu (2015) wurde eine der ersten Studien in diesem Forschungsbereich veröffentlicht. Der Schwerpunkt lag auf der korpusbasierten Vorbereitung. Ergebnisse der Dolmetschleistung wurden mit und ohne Unterstützung von automatischen Terminologie-extraktions- und Konkordanzprogrammen analysiert. Die Ergebnisse zeigten, dass die Vorbereitung mit den Tools die Genauigkeit der Terminologiewiedergabe erhöht und die Auslassungen verringert. Danach folgten eine Reihe an Masterarbeiten bzw. Doktorarbeiten zum Thema CAI-Tools und Computeranwendungen in der Kabine, darunter Biagini (2015), Gacek (2015), Prandi (2015), Monti (2016), Purkarthofer (2017) und Fossati (2021).

Prandi (2017) argumentierte, dass viele Studien sich auf die Qualität der terminologischen Wiedergabe fokussieren, um die Wirkung von CAI-Tools auf Dolmetschleistungen zu untersuchen. Mit dieser Forschungsmethode kann aber nicht erforscht werden, ob die eventuellen Verbesserungen in den Dolmetschleistungen auf den Zugang zur Terminologie oder auf Änderungen in der kognitiven Belastung zurückzuführen sind. Die erste Studie, die sich mit der Wirkung von CAI-Tools auf die kognitive Belastung beschäftigte, wurde im Rahmen einer Doktorarbeit an der Universität Mainz/Germersheim konzipiert (siehe Prandi 2022). Die Ergebnisse der experimentellen Studie zeigten, dass mit der Unterstützung von CAI bzw. ASR-gestützten Tools wenigere Auslassungen und Fehlübersetzungen als mit in PDF-Format verfügbaren Glossaren verzeichnet wurden, und die höchsten Genauigkeitsraten mit dem ASR-gestützten Tool erreicht wurden (vgl. Prandi 2022: 209). Diese Studie war der erste Versuch, die kognitiven Auswirkungen vom Simultandolmetschen mit CAI-Tools zu erforschen. Die Forschungsmethode des Eyetrackings wurde in Kombination mit etablierten, produktorientierten Methoden zur Messung der Dolmetschleistungen verwendet, wie die terminologische Genauigkeit, die Fehler und die Auslassungen in der Wiedergabe. Heutzutage ist Prandi (2022) die einzige veröffentlichte Studie, die den Dolmetschprozess mit Unterstützung von CAI-Tools in einer Eyetracking-Studie untersucht.

2.3 Automatische Spracherkennung (ASR)

Die Integration von automatischer Spracherkennung hat zu einer neuen, aktuellen Generation von ‘in-booth’ CAI-Tools geführt. Bevor diese präsentiert werden, werden in diesem Kapitel die Grundlagen und die geschichtliche Entwicklung der automatischen Spracherkennung vorgestellt.

Die automatische Spracherkennung (Automatic Speech Recognition) ist seit Jahrzehnten Gegenstand intensiver Forschung, obwohl das Interesse dafür erst in den letzten Jahren maßgeblich gestiegen ist. Die steigende Rechenleistung und Verfügbarkeit von großen Datenmengen haben zu einer immer stärkeren Verbesserung der Qualität der Ergebnisse geführt, und Geräte mit integrierter ASR (z. B. mobile oder tragbare Geräte, Smart-Home-Geräte, usw.) werden immer populärer (vgl. Yu & Deng 2015: 1). Wie Euler (2006: 15) erläutert: “Im engeren Sinn versteht man unter Spracherkennung die Aufgabe, aus einer gesprochenen Äußerung die Wörter richtig zu rekonstruieren”, wobei als Äußerung in diesem Fall sowohl ein einzelnes Wort als auch ein vollständiger Satz oder eine Reihenfolge von Sätzen verstanden werden können. Entgegen den ursprünglichen, optimistischen Erwartungen wurde festgestellt, dass die Entwicklung von automatischer

Spracherkennung viele Herausforderungen mit sich bringt (vgl. Euler 2006: 1). Diese betreffen die inhärenten Schwierigkeiten der menschlichen Sprache, die nicht nur aus dem Wortschatz und den grammatikalischen Regeln besteht (vgl. Euler 2006: 4). Von besonderer Schwierigkeit sind zunächst Variationen der Sprache. Jedes Individuum weist eine individuelle Sprechweise auf, die sowohl von anatomischen Merkmalen als auch vom erlernten und verinnerlichten Sprachgebrauch geprägt ist (vgl. Euler 2006: 7). Die Aussprache von Wörtern kann je nach Akzent oder Dialekt abweichen, und Varianten der Normlautung sind in der Aussprache mehr oder weniger üblich. Ein Beispiel ist das Wort 'sieben', das im Rahmen einer Untersuchung mit 100 Sprecher*innen zu zehn unterschiedlichen Aussprachevarianten geführt hat (vgl. Euler 2006: 11). Außerdem sind Variationen vom Akzent, Geschlecht, Redestil, Dialekt und von der Geschwindigkeit abhängig, was die Erstellung eines sprecherunabhängigen Spracherkennungssystems erschwert (vgl. Malik et al. 2021: 9412). Auch die Kontinuität der Sprache erweist sich als besonders problematisch, da im Sprachschall keine Unterbrechungen oder Differenzierungen zwischen Wörtern, Silben oder Lauten deutlich erkennbar sind, was die Erkennung der Wörter erschwert (vgl. Schukat-Talamazzini 1995: 8). Die früheren Systeme für Spracherkennung konnten zum Beispiel nur einzelne Wörter oder Ziffern erkennen und wurden als 'Command and Control' bezeichnet (vgl. Euler 2006: 18). Eine weitere Schwierigkeit stellt die Ambiguität der Sprache dar, da verschiedene Wörter bzw. Satzteile ähnlich klingen, z. B. Rad und Rat, Stau-becken und Staub-ecke (vgl. Schukat-Talamazzini 1995: 10). Außerdem ist Sprecherunabhängigkeit besonders herausfordernd, und zwar die Möglichkeit, Spracheingaben von unterschiedlichen Redner*innen zu erkennen (vgl. Malik et al. 2021: 9412). Auch die Verwendung von großen Wortschätzen stellt eine zusätzliche Aufgabe für die ASR dar. Ein umfassendes Weltwissen ist daher für das präzise Sprachverständnis nicht wegzudenken. Daher ist die Integration von erweiterten Wissensquellen in das Training von Algorithmen von erheblicher Tragweite (vgl. Euler 2006: 4). Zusätzlich kann die freie Sprechweise besonders herausfordernd sein. Während die kontinuierliche Spracheingabe bedeutet, dass Redner*innen ohne Pause zwischen den Wörtern sprechen dürfen, enthält die freie Sprechweise auch natürliche Merkmale wie Fehlstarts, Husten, Lachen oder Häsitationslaute wie ‚Ähm' (vgl. Malik et al. 2021: 9424).

Spracherkennung wird oft im Zusammenhang mit Sprachsynthese erwähnt, welche ihr Gegenstück ist, nämlich "die automatische Erzeugung eines Sprachsignals ausgehend von einem geschriebenen Text (Text To Speech, TTS)." (Euler 2006: 16f.) Wie Gottwald (2018: 1) erklärt,

erfüllt die Sprachsynthese in der Produktion von verständlicher Sprache schon hochqualitative Standards und kann weitgehend eingesetzt werden. Anders ist das aber für die automatische

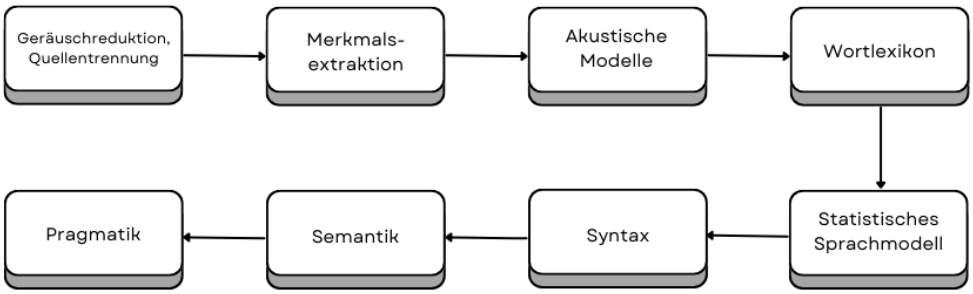

Abb. 6: Workflow für die automatische Spracherkennung adaptiert nach Euler (2006: 20)

Spracherkennung, in deren Prozessen es viele Hürden zu überwinden gibt. Die Vorgehensweise von automatischen Spracherkennungssystemen kann wie folgt beschrieben werden (Abb. 6): Hintergrundgeräusche müssen zunächst in der Vorverarbeitungsphase entfernt werden, um das Sprachsignal auszufiltern. Dazu zählt auch die Trennung der Quellen von anderen Sprecher*innen, die nicht analysiert werden sollen (vgl. Euler 2006: 20). Danach werden Merkmale extrahiert; das ist ein grundlegender Schritt für die darauffolgenden Entscheidungen über die Lauterkennung anhand von akustischen Modellen. Ein Wortlexikon wird verwendet, um die Lautfolgen der Wörter zu vergleichen, und zusätzlich wird seit der Integration von statistischen Sprachmodellen durch Wahrscheinlichkeitsrechnung versucht, die Anzahl der möglichen Wortfolgen zu verringern (vgl. Euler 2006: 20). Während vor den 1980er Jahren nur die Methode der Mustererkennung verfügbar war, um vorhandene Sprachmuster mit der Signalaufnahme zu vergleichen und Einzelwörter zu erkennen (vgl. Euler 2006: 53), können statistische Sprachmodelle die Wahrscheinlichkeiten des Vorkommens von Wortfolgen abschätzen (vgl. Malik et al. 2021: 9444). Danach prüft eine semantische Analyse die Satzstruktur. Aus der semantischen Analyse ergibt sich die Bedeutung des Satzes bzw. von Satzteilen, wobei zusätzliche Quellen berücksichtigt werden können, um weitere Informationen bzw. den Kontext abzuleiten und die Präzision zu erhöhen (vgl. Euler 2006: 20). Die Wortebene stellt also eine Zwischenschnittstelle dar, um die Bedeutung zu erschließen. Dabei gibt es eine bevorzugte Reihenfolge im Prozessablauf, die aber nicht zwingend berücksichtigt werden muss (vgl. Gottwald 2018: 1).

Auf Basis der Ergebnisse können die Wortkorrektheit und -genauigkeit der ASR gemessen werden, die zwei ausschlaggebende Faktoren sind, um die Qualität des Systems zu evaluieren. Die

Genauigkeit (‚accuracy') von ASR wird anhand von zwei Formeln berechnet, nämlich der Word Error Rate (WER), die Fehler auf Wortebene berechnet, und die entgegengesetzte Word Recognition Rate (WRR) (vgl. Malik et al. 2021: 9417). Um die WER zu berechnen, werden folgende Fehlerarten der ASR berücksichtigt: "Verwechslungen: falsche Wörter werden erkannt (NV); Auslassungen: an der Stelle gesprochener Wörter wird nichts erkannt (NA); Einfügungen: zusätzliche Wörter werden fälschlicher Weise eingefügt (NE)." (Euler 2006: 22) Die WER kann also mit dieser Formel berechnet werden, wobei N die Gesamtzahl der Wörter darstellt (vgl. Malik et al. 2021: 9417):

$$WER = \frac{NV + NA + NE}{N}$$

Ein weiterer Messwert, der für Spracherkennungssysteme besondere Relevanz hat, ist der Real-Time Factor (RTF), der die Geschwindigkeit des Modells für die Spracherkennung berechnet und auch stark von der Hardware abhängt (vgl. Malik et al. 2021: 9417). Zusammen mit Genauigkeit und Geschwindigkeit sind aber weitere wichtige Faktoren in der Entwicklung von gut funktionierenden Spracherkennungssystemen zu berücksichtigen, unter anderem die Robustheit gegen Hintergrundgeräusche, begrenzte Empfindlichkeit von Aufnahmegeräten, die Erkennung von Wörtern außerhalb des vorhandenen Wortschatzes, usw. (vgl. Euler 2006: 23).

Spracherkennung und -synthese können für zahlreiche Anwendungsmöglichkeiten eingesetzt werden. Dazu zählt die Steuerung von Geräten, die sogenannte ‚Hands free'-Anwendung, die zum Beispiel bei Geräten in Fahrzeugen Einsatz findet und die sich außerdem für Menschen mit Behinderungen als besonders vorteilhaft erwiesen hat (vgl. Euler 2006: 17). Weitere Anwendungsmöglichkeiten stellen Diktiersysteme dar, die vor allem im Gesundheits-, Rechts- und Nachrichtenwesen verwendet werden, und Sprachdialogsysteme. Sie ermöglichen, Kund*innen-Anfragen per Telefon zu bearbeiten bzw. Fragen zu beantworten, wie Auskunftssysteme von Zügen oder Filmen im Kino, Bestellungen, Telefon-Banking, Gewinnspiele, automatische Antworten von Call-Centern, usw. (vgl. Euler 2006: 17f.). Heutzutage gehören zu den beliebtesten Anwendungen in dieser Kategorie unter anderem die Sprachsuche, sprachgesteuerte digitale Assistenten und Infotainment-Portale in Autos (vgl. Yu & Deng 2015: 2). Digitale Assistenten sind vor allem seit der Einführung von Siri in Apple-Geräten populär geworden, die den Weg für viele andere ähnliche Produkte bereitet hat. Außerdem kann ASR in Videospiele integriert werden, um das Spielerlebnis

zu verbessern, oder für Geräte im Haushalt oder am Steuer verwendet werden, um zum Beispiel Musik abzuspielen, Informationen abzurufen oder Geräte zu steuern (vgl. Yu & Deng 2015: 3).

2.3.1 Entwicklung automatischer Spracherkennungssysteme

Die Geburtsstunde der ersten rudimentären Systeme für Sprachsynthese geht auf das 18. Jahrhundert zurück, als die ersten Experimente stattfanden, um die menschliche Stimme durch Maschinen nachzuahmen. 1939 wurde auf der Weltausstellung in New York das Sprachsynthesegerät VODER vorgestellt, das von Homer Dudley entwickelt wurde. Dudley zählt, zusammen mit Harvey Fletcher, zu den Vorreiter*innen der Sprachsynthese. Auf Basis ihrer Forschung wurde die moderne Methode der Spektrumanalyse für die Spracherkennung entwickelt, die bis heute mit modernen digitalen Signalverarbeitungstechniken eingesetzt wird (vgl. Juang & Rabiner 2004: 4f.). Erstmals hat aber die automatische Spracherkennung das Interesse eines großen Publikums in den 1960ern und 1970ern geweckt, als sie durch Blockbuster-Filme Bekanntheit erlangte (vgl. Juang & Rabiner 2004: 3). In der Forschung basierten die ersten Versuche zur Entwicklung von Spracherkennungssystemen auf der akustischen Theorie der Spracherzeugung und untersuchten, wie Phoneme von Menschen produziert werden. Diese rudimentären Systeme konnten aber nur eine begrenzte Anzahl an Phonemen erkennen (vgl. Juang & Rabiner 2004: 8).

Anfang der 1970er Jahre wurde von der ARPA (Advanced Research Project Agency), einer Behörde des Verteidigungsministeriums der Vereinigten Staaten, das Programm SUR (Speech Understanding Research) gefördert, das zur Herstellung von ‚Harpy' führte – einem System, das 1.011 Wörter mit angemessener Genauigkeit erkennen konnte (vgl. Juang & Rabiner 2004: 9). Entscheidend in der Geschichte der Entwicklung der automatischen Spracherkennung waren auch die Fortschritte, die von IBM und AT&T Bell Laboratories in den 1970ern gemacht wurden. IBM entwickelte das System Tangora, ein sprachabhängiges System, das mündliche Äußerungen in schriftliche Form umwandeln konnte, die auf einem Display angezeigt oder auf Papier getippt werden konnten (vgl. Juang & Rabiner 2004: 10). Der Fokus lag für AT&T Bell Laboratories auf Telekommunikationsdiensten, zum Beispiel dem sprachgesteuerten Wählen, die aber sprecherunabhängig sein mussten, um von mehreren Millionen Benutzer*innen verwendet werden zu können. Außerdem verwendete Bell Laboratories als erste den Keyword-Spotting-Ansatz, eine rudimentäre Form des Sprachverstehens, die sich auf die Bedeutung einzelner Wörter fokussiert und diejenigen ohne semantische Bedeutung auslässt (vgl. Juang & Rabiner 2004: 11).

In den 1980ern verlagerte sich der Fokus im Forschungsbereich von Mustervergleichen auf statistische Modelle, vor allem seit der Einführung von Hidden Markov Modellen (HMM), die die Variationen in den mündlichen Aussagen mittels Wahrscheinlichkeitsverteilungen berechnen. Somit wurden Trainingsdaten eingeführt, um Systeme zu trainieren und die Wahrscheinlichkeit zu berechnen, mit der unbekannte Äußerungen Wörtern entsprechen (vgl. Juang & Rabiner 2004: 12f.). Auch künstliche neuronale Netzwerke (Artificial Neural network, ANN) erlebten Ende der 1980er im Bereich Spracherkennung einen großen Aufschwung (vgl. Juang & Rabiner 2004: 15). 1990 wurde Dragon Dictate veröffentlicht, das erste Spracherkennungssystem, das für Konsument*innen verfügbar war. 1997 wurde Dragon NaturallySpeaking entwickelt, das kontinuierliche Spracheingaben und ca. 100 Wörter pro Minute erkennen konnte (vgl. Pinola 2011). In der geschichtlichen Entwicklung der künstlichen neuronalen Netze markiert die Entwicklung von Long Short-Term Memory (LSTM) einen wichtigen Fortschritt. LSTM ist eine Unterform der Recurrent Neural Networks (RNN) und wird heutzutage in Zusammenhang mit Deep Learning-Techniken verwendet (vgl. Malik et al. 2021: 9412). Deep Learning stützt sich auf die Forschungsergebnisse im Bereich Machine Learning und künstliche Intelligenz und ermöglicht es, Vorhersagemodelle zu entwickeln, die anhand von Trainingsdaten Muster erkennen und Entscheidungen treffen können (vgl. Sugomori et al. 2017: 16f.). Ein Deep Neural Network besteht aus einem Lernalgorithmus (Mehrschichtiges Perzeptron, MLP) mit einer tiefen Architektur, die mehrere versteckte Schichten enthält (vgl. Yu & Deng 2015: 57). Die Trainingsdaten stammen aus Datensätzen, die Audioaufnahmen und die entsprechenden Transkriptionen enthalten. Nach dem Training werden durch Testdaten die Ergebnisse des Modells evaluiert (vgl. Malik et al. 2021: 9418).

Am Anfang des 21. Jahrhunderts wurde die Google Voice Search App für iPhones eingeführt. Das führte dazu, dass Spracherkennung in Handys und mobile Geräte integriert werden konnte. Google hatte genug Rechenleistung und eine große Anzahl an Daten zur Verfügung, um ein potentes Spracherkennungssystem zu entwickeln. 2010 wurde für Android-Mobilgeräte eine personalisierte Spracherkennung eingeführt. 2011 wurde das Voice Search auch dem Chrome-Browser hinzugefügt (vgl. Pinola 2011). Die Entwicklungen im Bereich Spracherkennung waren nichtsdestotrotz bis 2010 laut Yu & Deng (2015: ix) langsam und wenig bemerkenswert. Das große Interesse, das danach folgte, kann auf die neuen Produkte zurückgeführt werden, die auf den Markt kamen. Dazu zählen die Sprachsuche, die Diktierfunktionen für Nachrichten und virtuelle

Assistenten wie Siri, Google Now und Cortana. Innovationen im Bereich Deep Learning waren von wesentlicher Bedeutung, um Wortfehlerraten zu verringern und somit viele Endbenutzer*innen für diese Produkte zu begeistern (vgl. Yu & Deng 2015: ix). Außerdem können heutzutage viel größere Datenmengen aufgerufen und Trainingsmodelle anhand von realen Anwendungssituationen erstellt werden (vgl. Yu & Deng 2015: 1). Ein Beispiel dafür ist Google Cloud Speech-to-Text, das 2016 unter dem Namen Cloud Speech API veröffentlicht wurde. Das ist ein Speech-to-Text-Tool, das Transkriptionen von Audio- und Videoaufnahmen ermöglicht und über die leistungsstarken und innovativen Deep Learning-Algorithmen von Google für neuronale Netzwerke verfügt (vgl. Cloudfresh 2023).

2.3.2 Die Integration von automatischer Spracherkennung in CAI-Tools

Seit den ersten Entwicklungen im Bereich der Sprachtechnologien fand die automatische Spracherkennung auch im Bereich Dolmetschen sehr rasch unterschiedliche Anwendungsmöglichkeiten. Zusammen mit Remote Simultaneous Interpreting (RSI) zählt sie zu den neuesten technologischen Entwicklungen im Bereich Simultandolmetschen und bietet sich als neue Technologie an, welche die Arbeitsweise des Dolmetschens grundlegend verändern kann. Wie Pöchhacker (2022: 49) verdeutlicht, "The use of computer software to support the interpreter's work is of course nothing new, but recent progress toward computer-assisted interpreting (CAI) enhanced by automatic speech recognition paves the way for shifting the human-machine boundary in the division of labor for accomplishing the interpreting task". Es lässt sich also klar erkennen, dass automatische Spracherkennung von entscheidender Bedeutung ist und das Potenzial in sich trägt, die Praxis des Dolmetschens radikal zu verändern (vgl. Pöchhacker 2022: 197). Sie kann zum Beispiel die Nutzbarkeit von CAI-Tools und ebenso die Präzision der Wiedergabe von Fachterminologie steigern und als eine Art 'elektronischer Kabinenpartner' fungieren (vgl. Fantinuoli 2017: 27).

Die Integration von automatischer Spracherkennung hat zu einer neuen Generation von 'in-booth' CAI-Tools geführt. Dabei geht es um technologische Lösungen, die während der Dolmetschung in Kabinen sowie in Remote Settings zum Einsatz kommen. Diese CAI-Tools sind von Automatisierung gekennzeichnet und sie unterscheiden sich von vorigen Generationen von CAI-Tools in der Möglichkeit, Eigennamen und Zahlwörter automatisch zu erkennen und anzuzeigen, damit sie von Dolmetscher*innen abgelesen werden können (vgl. Rodriguez et al.

2022). Die automatische Spracherkennung kann aber nicht nur im Bereich des Simultandolmetschens Einsatz finden, sondern auch beim Konsekutivdolmetschen. Bei Letzterem werden Transkriptionen der Ausgangsrede erstellt, die im Konsekutivmodus vom Blatt gedolmetscht werden können (vgl. Fantinuoli 2017: 26). Das könnte zu neuen Dolmetscharten führen (SimConsec und SightConsec), trotzdem stellen technologische Herausforderungen immer noch ein Hindernis für deren Entwicklung dar (vgl. Corpas Pastor 2022: 96). Diese neue Generation von Tools erfüllt ein Bedürfnis, das in CAI-Tools der zweiten Generation in Bezug auf die manuelle Suche nach Terminologie in Glossaren während der Simultandolmetschung entstanden war. Die vorangehende Generation erwies sich als problematisch, da die mangelnde Zeit während der Dolmetschung sowie die Ablenkung, die eine manuelle Suche nach Terminologie in der Kabine mit sich brachte, zu Skepsis gegenüber des CAI-Tools seitens vieler Dolmetscher*innen führte (vgl. Fantinuoli 2018a: 163). Diese Problematik kann aber behoben werden, indem eine automatisierte Suche nach Termini erfolgt und zusätzlicher kognitiver Aufwand vermieden wird (vgl. Fantinuoli 2017: 25). Wie Fantinuoli & Pisani (2021: 183) schildern, "a technological means to automate the lookup mechanism could have the potential to reduce the cognitive load, benefiting the whole interpreting process." Dieses Potenzial von ASR in Bezug auf die Reduktion der kognitiven Belastung wird in vielen Forschungsarbeiten betont (z. B. Amelina & Tarasenko 2020; Fantinuoli 2021; Fantinuoli 2017). In der Vergangenheit war außerdem eines der größten Hindernisse für die Verwendung von ASR die Tatsache, dass diese Technologien hohe Fehlerraten aufwiesen und demzufolge nur schwer in hochkomplexen und nicht kontrollierten Settings wie dem Dolmetschen einsetzbar waren (vgl. Fantinuoli 2017: 25). Die Qualität der ASR wird nämlich anhand zweier Faktoren gemessen: eine niedrige Wortfehlerrate und die Fähigkeit des Tools, relevante Informationen abzurufen bzw. zu erkennen (vgl. Fantinuoli 2017: 31). Die Einführung von Deep Learning und neuronalen Netzwerken hat aber bedeutende Fortschritte in der Genauigkeit der Ergebnisse erzielt (vgl. Fantinuoli 2017: 26), und die weitergehende Entwicklung von Big Data und Computerleistung fördert das Training von Sprachmodellen, welche schrittweise viele Einschränkungen rund um die ASR vermindern (vgl. Fantinuoli 2021: 516). Ausschlaggebend für hochqualitative Leistungen dieser Tools kann außerdem die Rolle der Vorbereitungsglossare und der Menschen sein, die eventspezifische Informationen in das ASR-System integrieren. Um die Genauigkeit der Spracherkennung zu steigern, kann das System mit unterschiedlichen Informationen über den Dolmetschauftrag trainiert und feinabgestimmt werden,

unter anderem Glossare, Listen der Teilnehmer*innen oder Namen von Produkten, wie Fantinuoli et al. (2022: 4) untersuchen. Im Rahmen eines Experiments wurde das CAI-Tool Interpret Assist verwendet, um die Nützlichkeit und Wirkung von diesen Informationen auf CAI-Tools zu untersuchen, und zwar anhand zweier Datensätze: einem Datensatz mit allgemeinen Reden und Themen, und einem mit Fachterminologie (vgl. Fantinuoli et al. 2022). Zunächst wurde der allgemeine Datensatz ohne Feinabstimmung evaluiert, während in einer zweiten Phase das ASR-System mit Termini und Eigennamen trainiert wurde, die in den Unterlagen für die Vorbereitung auf einen hypothetischen Dolmetschauftrag enthalten sind (vgl. Fantinuoli et al. 2022: 6f.). Die Ergebnisse zeigen, dass ein höherer Recall (Vollständigkeitsquote) von Glossareinträgen durch die Feinabstimmung von einem allgemeinen Korpus erzielt wird (+4,92 %), während Genauigkeit und Recall bei fachspezifischen Datensätzen höhere Werte erreichen (vgl. Fantinuoli et al. 2022: 7). Diese Studie hebt die Relevanz der Feinabstimmung bzw. der Implementierung von ad-hoc Informationen und Fachterminologie über den Dolmetschauftrag in der ASR hervor, um hochwertige Ergebnisse zu erzielen. Dieser Schritt muss nämlich von Menschen durchgeführt werden und stellt laut Fantinuoli et al. (2022: 8) die Grundlagen für eine Zukunft der ASR dar, wobei die Zusammenarbeit von Mensch und Maschine unentbehrlich ist.

Bestehende Herausforderungen in der Entwicklung von ASR für das Dolmetschen sind auf die Komplexität der Sprache zurückzuführen, wobei mehrere Elemente zum Sprachverständnis beitragen. Diese betreffen nicht nur die rein linguistische Ebene, sondern beinhalten auch zusätzliche Informationen wie Allgemeinwissen, Vorwissen über die Redner*innen, Kontext und Inhalt (vgl. Fantinuoli 2017: 28). Außerdem zählen zu den Anforderungen für die Entwicklung von automatischen Spracherkennungssystemen laut Fantinuoli (2017: 28f.) folgende Faktoren: die mündliche Sprache, die von unterschiedlichen Redestilen geprägt sein kann und vor allem in informellen Situationen von Merkmalen wie Zögern oder Wiederholungen gekennzeichnet ist; unterschiedliche Redner*innen, die unterschiedliche Stimmen, Redestile, Dialekte und Soziolekte, Akzente und Aussprachen aufweisen; die inhärente Ambiguität von Wörtern; die kontinuierliche Sprachwiedergabe ohne Pause zwischen einzelnen Wörtern, die eine Thematik von anspruchsvoller Komplexität in der ASR repräsentiert; Hintergrundgeräusche; hohe Redegeschwindigkeit, welche zu fehlerhaft ausgesprochenen Wörtern führt, und fehlender Zugang zur Körpersprache, die zum Verstehen des Inhaltes beiträgt. Diese Faktoren deuten auf die Reihe von Kriterien hin, welche die automatische Spracherkennung erfüllen muss, um in CAI-Tools integriert zu werden.

Hierzu gehören die Sprecherunabhängigkeit, das Erkennen von kontinuierlicher Sprachwiedergabe, die Erkennung großer Wortmengen, die Möglichkeit zur Anpassung für die Erkennung von Fachbegriffen, eine niedrige WER (Word Error Rate) und ein niedriger RTF (Real-Time Factor, also die Zeit für die Verarbeitung der gesprochenen Wörter) (vgl. Fantinuoli 2017: 29). Letzteres ist auf das Kaskadensystem zurückzuführen, das ASR-Systeme kennzeichnet. Sie bestehen nämlich aus unterschiedlichen Modulen, welche die automatische Spracherkennung durchführen, Interessenseinheiten wie Zahlen oder Namen erkennen und diese letztendlich anzeigen, was eine bestimmte Verarbeitungszeit voraussetzt. Diese könnte in der Zukunft von End-to-End-Systemen verkürzt werden (vgl. Fantinuoli & Montecchio 2022: 1). Die vierte Generation von CAI-Tools, die noch nicht entwickelt worden ist, wird dann mit einem natürlichen Sprachmodell ausgestattet werden und Einschränkungen der aktuellen Tools beheben, wie zum Beispiel Latenz und Fehlerfortpflanzung, die auf das Kaskadensystem zurückzuführen sind (vgl. Frittella 2023: 50; Rodriguez et al. 2022: 80). Jedenfalls müssen nicht nur ASR-Systeme verbessert werden, um die Anforderungen von Dolmetschsettings zu erfüllen. Auch CAI-Tools müssen bestimmte Voraussetzungen erfüllen, um im Einklang mit ASR-Systemen zu funktionieren. Diese Kriterien beinhalten hohe Genauigkeit und Recall von relevanten Ergebnissen im Vergleich zur Gesamtanzahl der Ergebnisse, obwohl Genauigkeit Vorrang vor dem Recall haben sollte (vgl. Fantinuoli 2017: 30). Zwei weitere Merkmale, die für die Nutzbarkeit von CAI-Tools wesentlich sind, sind der Umgang mit morphologischen Variationen der Wörter, um die Anzahl der Ergebnisse begrenzt zu halten, und eine einfache Benutzeroberfläche, welche nicht zur Ablenkung beiträgt (vgl. Fantinuoli 2017: 30). Auf Basis dieser Voraussetzungen wurde 2017 der erste Prototyp eines CAI-Tools mit ASR entwickelt, der Terminologieeinträge und Zahlen auf Basis des durch ASR erstellen Transkripts der Rede erkennt, und diese für Dolmetscher*innen auf dem Bildschirm zur Verfügung stellt. Dieser Prototyp wurde in das CAI-Tool InterpretBank integriert und wies eine benutzerfreundliche, leicht überschaubare und in drei Sektionen gegliederte Benutzeroberfläche auf, die jeweils die Transkription, Terminologie und Zahlen zeigten. Dieses Tool ermöglichte auch ein sehr hohes Maß an Individualisierung anhand von Hintergrundfarbe, Schriftgröße und weiteren Merkmalen der gezeigten Termini (vgl. Fantinuoli 2017: 31). Benutzer*innen hatten in dieser Version die Möglichkeit, die ASR-Systeme nach Belieben auszuwählen, um immer den besten Output in der Zielsprache und je nach Betriebssystem zu gewährleisten. Die Funktionsweise kann in zwei Phasen aufgeteilt werden (Abb. 7).

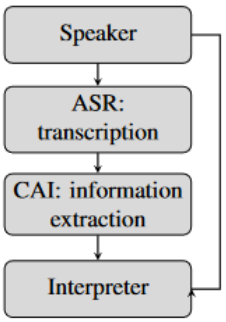

Abb. 7: Workflow des Prototyps eines CAI-Tools mit ASR nach Fantinuoli (2017: 30)

Das Tonsignal ist dasselbe, das Dolmetscher*innen durch ihre Konsole in den Kabinen hören, und wird für die Transkription verwendet. Auf Basis der Transkription und der Terminologie-datenbanken werden relevante Ergebnisse angezeigt. Die Transkription wird in Textabschnitten bearbeitet und nach Regeln normalisiert (z. B. Zahlen, Groß- und Kleinschreibung werden angepasst). Mithilfe eines N-Gramm-Matching-Ansatzes werden Terminologiedatenbanken durchsucht und mit Unterstützung von einem Fuzzy-Match-Ansatz auf relevante Ergebnisse gefiltert, um die Anzahl der Ergebnisse minimal zu halten (vgl. Fantinuoli 2017: 30f.). Trotz der Fortschritte, die in den letzten Jahren erzielt wurden, bleiben immer noch Unzulänglichkeiten bestehen, die verbessert und weiter erforscht werden müssen, damit CAI-Tools besser in die Arbeitsabläufe von Dolmetscher*innen integriert werden können. Einige technologische Hindernisse sind zum Beispiel die Plattformabhängigkeit, die geringen Präzision und Wiedererkennung, der geringe Automatisierungsgrad und der Mangel an sprachübergreifenden NLP-Sprachtechnologien (vgl. Corpas Pastor 2022: 95f.)

Ein zentrales Problem und grundlegender Nachteil ist außerdem die Benutzeroberfläche der Tools, da viele visuelle Reize gezeigt werden, was zu Ablenkung und kognitiver Überlastung für Dolmetscher*innen führen kann (vgl. Fantinuoli 2021: 516). Dieses Thema erscheint von besonderer Bedeutung in der Forschung von Frittella (2022; 2023), die sich auf die Nutzbarkeit des CAI-Tools SmarTerp fokussiert und die Beziehung zwischen kognitiven Anforderungen und der Benutzerschnittstelle präsentiert. Da der größte Vorteil von ASR eine niedrigere kognitive Belastung für Dolmetscher*innen ist, wie in diesem Kapitel schon erwähnt (z. B. Fantinuoli 2017; Frittella 2023; Fantinuoli & Pisani 2021), wird auch für die Zukunft das Ziel verfolgt, Funktionen hinzuzufügen, die die kognitive Last verringern und dementsprechend die Benutzbarkeit erhöhen. CAI-Tools könnten z. B. eine weitere Schnittstelle implementieren, um anhand von überwachten

Sequence Tags und eines trainierten Annotationssystem zu berechnen, welche Begriffe den Dolmetscher*innen Probleme bereiten können (vgl. Vogler et al. 2019: 1). Der von Vogler et al. (2019: 2) vorgeschlagene Arbeitsablauf stützt sich auf der Hypothese, dass eine Vorhersage der unübersetzten Terminologie anhand von Natural Language Processing möglich ist, und somit visuelle Reize im CAI-Tool auf ein Minimum eingeschränkt werden können (Abb. 8). Unübersetzte Termini sind Zahlen und Substantive. Sie werden anhand unterschiedlicher Kriterien berechnet und subjektiv eingestellt (vgl. Vogler et al. 2019: 2). Grundsätzlich können aber Tendenzen in Auslassungen von bestimmten Wörtern beim Dolmetschen beobachtet werden: gegen Ende der Reden oder langer Sätze, bei höherem Sprechtempo der Redner*innen oder bei seltenen oder langen Wörtern (vgl. Vogler et al. 2019: 4f.). Diese Merkmale können dann verwendet werden, um Ergebnisse der ASR zu priorisieren bzw. herauszufiltern. Aus den Ergebnissen von Vogler et al. (2019: 2) ergibt sich, dass die Genauigkeit der Wiedergabe mit dieser Funktion im Vergleich zum traditionellen Arbeitsablauf bis zu 30% gesteigert werden konnte.

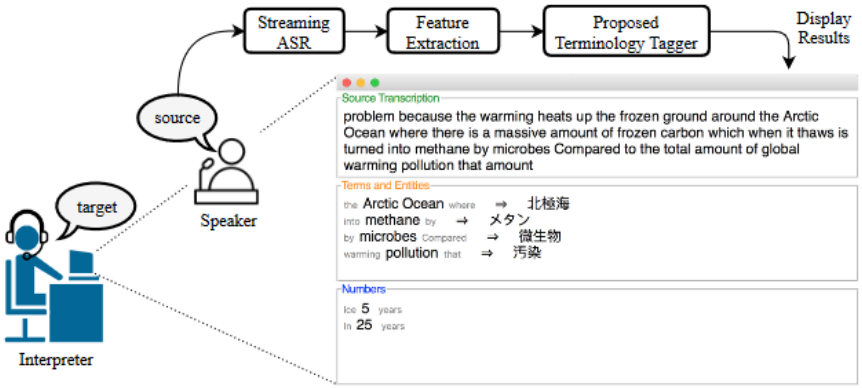

Abb. 8: Workflow des CAI-Tools mit ASR und Sequence Tags nach Vogler et al. (2019: 2)

Obwohl in den letzten Jahren ein großes Interesse auf dem Markt für ASR-Produkte entstanden ist, bleibt die Liste der CAI-Tools, die eine ASR-Funktion aufweisen, gering. Solche Tools können entweder direkt auf dem Gerät des Benutzers oder im Internet auf Servern betrieben werden und sind über einen Web-Browser abrufbar (vgl. Fantinuoli & Montecchio 2022: 1). InterpretBank war das erste CAI-Tool, das ASR integriert hat, um Terminologie, Namen und Zahlen automatisch zu erkennen. Die cloudbasierte Funktion ist derzeit durch die App 'InterpretBank Desktop' zugänglich, in der ein oder mehrere Glossare ausgewählt werden und danach eine neue Sitzung in der WebApp begonnen wird. Derzeit wird nur die Ausgangssprache Englisch unterstützt, und nach einer 14-tägigen Probezeit ist das Tool nur über eine Lizenz zugänglich (vgl. InterpretBank o.J.).

Das Center for Augmented Interpretation von Johannes Gutenberg-Universität Mainz hat außerdem ein Tool entwickelt, das mithilfe der ASR nur Zahlen erkennt und wie InterpretBank von Dr. Claudio Fantinuoli entwickelt wurde: *Augmented Interpreter - CAI v 1.1*. Das webbasierte Tool benötigt keine Lizenz und ist frei zugänglich. Neue Zahlen werden oben angezeigt und größer und rot hervorgehoben. Bis dato ist das Tool auf fünf Minuten beschränkt und muss danach neu gestartet werden (vgl. Augmented Interpreter o.J.). Zu den ASR-gestützten CAI-Tools zählt auch 'Interpreter Assist', das seit 2021 in die KUDO Konferenzplattform für mehrsprachige Meetings integriert wurde. Die zwei wesentlichen Funktionalitäten sind die automatische Erstellung von Glossaren und die Visualisierung von Vorschlägen in Echtzeit während der Dolmetschung. Das Tool ist besonders für die Integration in Remote Interpreting Settings entwickelt worden (vgl. Fantinuoli et al. 2022). Seit Mai 2023 ist auch die Funktion KUDO AI verfügbar, die den Fokus auf Speech-to-Speech Technologien legt (vgl. KUDO 2023). Ein weiteres Tool, das derzeit noch nicht kommerziell zugänglich ist, heißt SmarTerp (vgl. SmarTerp o.J.). Es wurde im Geschäftsplan 2021 im Rahmen des Programms EIT Digital von der Europäischen Union gefördert und verfolgt das Ziel, eine RSI-Plattform mit integriertem CAI-Tool und ASR zu entwickeln (vgl. Frittella 2023: 65).

2.4 Forschungsstand: Studien mit CAI-Tools und automatischer Spracherkennung

In Zeiten der Entwicklung und Verbesserung von CAI-Tools für die Unterstützung von Dolmetscher*innen und für die Überwindung von Herausforderungen während der Dolmetschprozesse hat die automatische Spracherkennung das Interesse von vielen Forscher*innen geweckt, wie die zunehmende Anzahl an Veröffentlichungen in den vergangenen Jahren verdeutlicht (z.B. Frittella 2023; Fantinuoli & Montecchio 2022; Fantinuoli et al. 2022; Fantinuoli & Pisani 2021). Seit 2017, als Fantinuoli (2017) den ersten Prototypen eines CAI-Tools mit integrierter ASR präsentierte, sind unterschiedliche Publikationen erschienen, die das Thema ‚automatische Spracherkennung' behandeln. Der Schwerpunkt liegt in diesem Forschungsbereich vor allem auf Dolmetschprodukten, wie zum Beispiel der Untersuchung von Dolmetschleistungen, oder dem Design und der Entwicklung der Tools (vgl. Rodriguez et al. 2022: 85).

Die erste Studie über die Qualität von ASR in CAI-Tools wurde mit dem in Fantinuoli (2017) beschriebenen Prototypen durchgeführt. Der Test mit 119 Termini und 11 Zahlwörtern spricht für eine hohe Genauigkeit bei der Erkennung von Zahlwörtern, die ohne Fehler transkribiert wurden, während eine Wortfehlerrate von 5,04% verzeichnet wurde. Es wurde aber angemerkt,

dass mehr Experimente unter schwierigeren Bedingungen durchgeführt werden müssen, wie zum Beispiel mit Hintergrundgeräuschen oder nicht muttersprachlichen Akzenten. Außerdem wurde angeführt, dass in der Zukunft die Miteinbeziehung von Dolmetscher*innen unverzichtbar ist, um die Dolmetschleistungen mit ASR zu messen und eine eventuelle visuelle oder kognitive Überlastung der Dolmetscher*innen zu thematisieren (vgl. Fantinuoli 2017: 32f.).

2.4.1 ASR-Mock-Up-Systeme als Forschungsgegenstand

Da ASR-gestützte CAI-Tools erst in den letzten Jahren kommerziell verfügbar geworden sind, wurden zu Beginn Mock-Up-Systeme, wie Videos oder Folien, erstellt, welche die Funktionsweisen von ASR simulieren und Forschung in diesem Bereich ermöglichen. 2018 wurde von Desmet et al. die erste Studie über die Wirkung von technologischer Unterstützung auf Dolmetschleistungen beim Simultandolmetschen veröffentlicht. Ein Tool, das die Kriterien für ASR-gestützte CAI-Tools erfüllt – ein schnelles und präzises automatisches Spracherkennungssystem, eine Software um Zahlen aus dem Text herauszufiltern, und eine ergonomische Darstellung der Zahlen – war zur Zeit der Studie nicht verfügbar (Desmet et al. 2018: 18). Daher wurde eine Simulation für das geplante Experiment konzipiert. Auf PowerPoint-Folien wurden Zahlen geschrieben, und sobald die Zahl in der Ausgangsrede vollständig ausgesprochen wurde, wurde die Folie mit der entsprechenden Zahl gezeigt und, falls vorhanden, auch die zwei Zahlen, die zuvor erwähnt wurden (Abb. 9). Das simulierte ein System mit ASR für die Erkennung von Zahlen. Die Folien wurden auf einem Bildschirm im Konferenzsaal gezeigt und manuell gesteuert, um ein ASR-System mit minimaler Latenz nachzuahmen (vgl. Desmet et al. 2018: 18). Für das Experiment wurden vier Ausgangsreden mit unterschiedlichen Themen und vergleichbarem Schwierigkeitsgrad und vergleichbarer Länge verwendet. Diese Reden bestanden aus einer 150-Wörter Einführung ohne Zahlen und einem Hauptteil mit zwanzig Zahlen pro Rede (vgl. Desmet et al. 2018: 20).

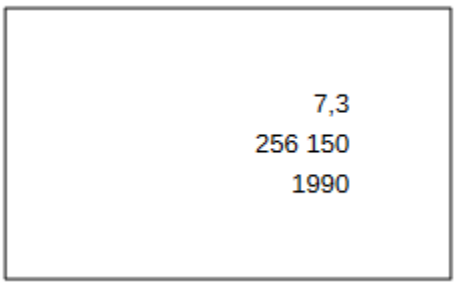

Abb. 9: Beispiel eines Mock-Up-Systems mit Folien nach Desmet et al. (2018: 19)

Die Zahlen wurden in vier Kategorien unterteilt: einfache ganze Zahlen, komplexe ganze Zahlen, Dezimalzahlen und Jahre (vgl. Desmet et al. 2018: 20). Die Komplexität der Zahlen wurde je nach Anzahl der bedeutungstragenden Einheiten beurteilt, wobei alle Zahlen ab drei bedeutungstragenden Einheiten als komplex bezeichnet wurden. Die Studienteilnehmer*innen waren zehn Dolmetschstudierende, die zur Zeit an der Universität Gent Konferenzdolmetschen studierten und Niederländisch als A-Sprache hatten. Sie dolmetschten jeweils zwei Reden mit simulierter technologischer Unterstützung und zwei Reden ohne Unterstützung (vgl. Desmet et al. 2018: 20). Die Dolmetschleistungen wurden in Anlehnung an die Zahlenkategorien von Pinochi (2009) untersucht, die sich auf die Kategorien von Braun & Clarici (1996) stützen. Aus dem Vergleich der Genauigkeit in der Wiedergabe von Zahlen mit und ohne Unterstützung durch die Folien ließ sich feststellen, dass die Unterstützung durch Folien eine beträchtliche Wirkung aufweist, mit einem Zuwachs der Genauigkeit von durchschnittlich 30% (56,5% ohne Unterstützung, 86,5% mit Unterstützung). Allerdings argumentierten die Autoren, dass diese Simulation unter optimalen Bedingungen verlaufen ist, da die Latenz minimal war und die Zahlen immer korrekt angezeigt wurden (vgl. Desmet et al. 2018: 22). Aus den Ergebnissen ließ sich ableiten, dass komplexe Zahlen und Dezimalzahlen die zwei Kategorien sind, in denen ohne technologische Unterstützung am meisten Fehler passierten. Diese Unterschiede in der Wiedergabe, je nach Art der Zahlen, verschwanden mit technologischer Unterstützung, und Verbesserungen waren in jeder Kategorie zu erkennen (vgl. Desmet et al. 2018: 24). Die Mehrheit der Fehler nach Fehlerkategorie fiel in die Kategorie Auslassung (‘omission’). Die Autoren führten als Grund entweder die Informationsdichte der Reden (Auslassungen reduzierten die kognitive Belastung oder den Timelag) oder die Unübersichtlichkeit der Folien an. Außerdem könnte die Verfügbarkeit der Zahlen eine verminderte auditorische Aufmerksamkeit zur Folge haben, und obwohl Zahlen verfügbar sind, fehlt der Kontext für die Verdolmetschung, so dass sie ausgelassen werden (vgl. Desmet et al. 2018: 24f.). Faktoren, die in dieser Forschungsarbeit unberücksichtigt bleiben, sind die Wirkung von ASR-gestützten CAI-Tools auf professionelle bzw. erfahrene Dolmetscher*innen sowie welche weitere Texteinheiten von der erhöhten Aufmerksamkeit auf Zahlen beeinflusst werden (vgl. Desmet et al. 2018: 25). Auch die Replikationsstudie, die 2022 im Rahmen einer Masterarbeit von Penta (2022) durchgeführt wurde, hat die Ergebnisse von Desmet et al. (2018) über die Dolmetschqualität und die Gesamtgenauigkeit der Zahlen, und zusätzlich auch der Eigennamen, bestätigt.

Eine weitere Möglichkeit für ein Mock-Up-System, das die Funktionsweise von ASR in den Dolmetschkabinen nachahmt, stellt die Arbeit von Canali (2019) dar. Im Rahmen einer Masterarbeit wurde ein Experiment mit einem Video gestaltet, das die im Ausgangstext erwähnten Zahlwörter zeigt, sobald sie ausgesprochen werden, um ein System mit minimaler Latenz nachzuahmen. Außerdem wurden Zahlwörter so lange wie möglich auf dem Bildschirm angezeigt, indem die neuen oben erscheinen und alle anderen nach unten rückten (Abb. 10) (vgl. Canali 2019: 111).

Abb. 10: Beispiel eines Mock-Up-Systems mit Video nach Canali (2019: 115)

Das Video wurde ähnlich wie in Desmet et al. (2018) in einem Saal auf einem großen Bildschirm gezeigt, und zwei Teilnehmer*innengruppen dolmetschten dieselbe Ausgangsrede jeweils mit und ohne Unterstützung des Videos, während die Zahlenkategorien von Desmet et al. (2018) übernommen wurden (vgl. Canali 2019: 109f.). Mit Unterstützung des Videos wurden 39% mehr richtige Zahlen gedolmetscht (36% ohne Unterstützung, 75% mit Unterstützung) und Auslassungen waren die Kategorie, die die meisten Fehler für beide Teilnehmer*innengruppen verzeichnete (vgl. Canali 2019: 139), was verglichen mit den Ergebnissen von Desmet et al. (2018) viele Ähnlichkeiten aufweist. Auch die Studie von Cheung & Li (2022) implementierte in das Forschungsdesign Videos, die die ASR nachahmen, wobei bei diesem Experiment Videos der Redner*innen mit Untertiteln versehen wurden. Der Schwerpunkt dieser Studie lag auf dem Potenzial der ASR, Dolmetscher*innen bei nicht-muttersprachlichen Akzenten (z. B. in ELF-Settings) zu unterstützen, was aufgrund der Verbreitung der englischen Sprache und der immer größeren Teilnahmemöglichkeiten bei Konferenzen via Remote Settings und Online-Meetings in der Zukunft bei Dolmetschaufträgen immer häufiger vorkommen wird (vgl. Cheung & Li 2022: 1). Die Ergebnisse zeigten, dass bei der Verfügbarkeit von Untertiteln höhere Genauigkeitsraten verzeichnet wurden und Studienteilnehmer*innen niedrigere kognitive Anstrengungen spürten. Gleichzeitig wurde aber die Flüssigkeit bei der Dolmetschung vernachlässigt (vgl. Cheung & Li 2022: 14f.).

Diese Experimente ermöglichten es, Daten zur Interaktion zwischen Dolmetscher*innen und automatischer Spracherkennung zu erheben, bevor ASR in auf dem Markt verfügbare CAI-Tools integriert wurde. Diese Ergebnisse sind daher von wesentlicher Bedeutung, indem sie relevante Rückschlüsse über die Vorteile der Anwendung von ASR in Dolmetschsettings ermöglichen und konkrete Daten über die Verbesserungen bzw. Fehler liefern, die bei Dolmetscherleistungen verzeichnet werden können. Es ist jedoch zu unterstreichen, dass diese Modelle von Menschen entworfen und optimiert wurden, was eine hohe Genauigkeit und minimale Latenz ermöglicht, und dass diese Ergebnisse von den anderen zu differenzieren sind, die mit auf dem Markt erhältlichen ASR-Tools erhoben werden könnten, bei denen viele weitere Faktoren hinzukommen und zu berücksichtigen sind.

2.4.2 InterpretBank als Forschungsgegenstand

Eine der ersten Studien über die Anwendungsmöglichkeit von InterpretBank ASR wurde von Defrancq & Fantinuoli (2021) durchgeführt. Zur Zeit des Experiments war eine Version von InterpretBank verfügbar, die eine vollständige Transkription der Ausgangsrede zeigt. Diese wurde anhand der Google Cloud Speech-to-Text API erstellt, da sich deren Qualität im Vergleich zu ähnlichen Tools sich als höher erwiesen hatte. Zahlen und Maßeinheiten sowie die Terminologie, die in der Datenbank gespeichert bzw. maschinell übersetzt wurde, wurden in der Transkription farblich markiert und groß dargestellt (vgl. Defrancq & Fantinuoli 2021: 5). Dies führte dazu, dass eine umfangreiche Menge an Informationen auf dem Bildschirm angezeigt wurde. Nichtsdestotrotz wurde diese Visualisierungsform mit der Argumentation ausgewählt, dass Informationen sinnvoller sind, wenn sie in ihren Kontext eingebettet sind (vgl. Defrancq & Fantinuoli 2021: 6). Das Tool wurde eingesetzt, um die Latenz so niedrig wie möglich zu halten. Dies wurde erzielt, indem Transkriptionen aus der Spracherkennung so schnell wie möglich auf dem Bildschirm gezeigt wurden, was dazu führte, dass Satzteile oder Zahlen für kurze Zeit eine andere Form annahmen, bevor sie in ihrer endgültigen Form aufschienen (vgl. Defrancq & Fantinuoli 2021: 5f.). Die Studienteilnehmer*innen wurden in zwei Gruppen aufgeteilt, und unterschiedliche Ausgangsreden mit und ohne Software wurden gedolmetscht. Zahlen wurden nach vier Schwierigkeitsniveaus differenziert: Niveau 1 bezieht sich auf Zahlen mit einer oder zwei bedeutungstragende Einheiten (z. B. 2 oder 40,5), Niveau 2 enthält Zahlen mit drei oder vier bedeutungstragende Einheiten (z. B. 124 oder 7,6 Millionen), Niveau 3 enthält Zahlen mit fünf oder sechs bedeutungstragende Einheiten (z. B. 1.406.000) und Niveau 4 enthält Zahlen ab sechs

bedeutungstragende Einheiten (z. B. 17.345.133) (vgl. Defrancq & Fantinuoli 2021: 10). Auch hier wie in Desmet et al. (2018) wurden bedeutungstragende Einheiten als Maßstab für die Differenzierung von Zahlen verwendet, jedoch ohne Differenzierung für Dezimalzahlen oder Jahreszahlen. Während des Experiments wurde auch anhand von Webcams in den Kabinen untersucht, ob die Studienteilnehmer*innen tatsächlich die angezeigten Zahlen auf dem Bildschirm verwendeten. Obwohl das reine Zusehen keine Sicherheit dafür ist, dass Informationen tatsächlich abgelesen und als Unterstützung verwendet wurden, zeigten die Ergebnisse der Aufzeichnungen, dass Zahlen in 55% der Fälle abgelesen wurden, unabhängig vom Schwierigkeitsgrad der Zahl oder der Rede (vgl. Defrancq & Fantinuoli 2021: 17). Während des Experiments wurden 96% der Zahlen von der ASR korrekt angezeigt (vgl. Defrancq & Fantinuoli 2021: 17), was ähnlich der Wortfehlerrate von 5,04% in Fantinuoli (2017) ist, wobei in dieser Studie kein Fehler in der Transkription von Zahlen zu verzeichnen war. Auch in dieser Studie steigt die Anzahl der korrekt wiedergegebenen Zahlen mit ASR, und zwar von 67,7% ohne ASR, auf 90,2% mit ASR (vgl. Defrancq & Fantinuoli 2021: 19). In dieser Studie haben zwei Anmerkungen eine besondere Relevanz. Sobald ASR ab einem bestimmten Punkt der Rede nicht mehr verfügbar war, sanken die Genauigkeitsraten unter den Durchschnitt, der für Reden ohne technologische Unterstützung gemessen wurde. Hier wurde die Annahme formuliert, dass diese Daten ein Zeichen für das übermäßige Vertrauen in das Tool sind (vgl. Defrancq & Fantinuoli 2021: 20). Außerdem ergibt sich aus den Ergebnissen, dass die Genauigkeit unabhängig von der tatsächlichen Verwendung von Daten vom Bildschirm ist, wenn ASR zur Verfügung steht. Das führt zur Vermutung, dass die Verfügbarkeit von ASR allgemein psychologisch positiv wirken kann, indem Stress abgebaut und Selbstvertrauen erhöht wird (vgl. Defrancq & Fantinuoli 2021: 21). Dolmetschleistungen können also wesentlich davon profitieren.

In der Folge wurden weitere Experimente durchgeführt, welche die Wirkung von ASR auf die Dolmetschleistungen untersuchten. Einigkeit herrscht über das Bedürfnis, CAI-Tools mit ASR unter realen Bedingungen zu testen – Simulationen haben den Vorteil, dass Variablen wie Fehler und Latenz kontrollierbar sind, dennoch entsprechen sie nicht den realen Rahmenbedingungen bzw. den Herausforderungen des Dolmetschens. Zu diesen Studien zählt das von Pisani & Fantinuoli (2021) durchgeführte Experiment, in dem Zahlen ohne Transkription auf dem Bildschirm angezeigt wurden. Das sollte eine höhere Benutzbarkeit des Tools ermöglichen, da nur gezielte Informationen zur Verfügung standen (vgl. Defrancq & Fantinuoli 2021: 2). Trotzdem

wird argumentiert, dass auch die Visualisierung von ungefilterten Vorschlägen, z. B. unabhängig von der Schwierigkeit der Zahlen, zu erhöhter kognitiver Belastung führen kann und in nächsten Studien mitberücksichtigt werden sollte (vgl. Fantinuoli & Pisani 2021: 186). Die Ergebnisse des Experiments verdeutlichten ein geringes Niveau an Genauigkeit in den von der ASR erkannten Zahlen (81,43%). Probleme in der Erkennung entstanden bei Auflistungen von Zahlen, die nicht als getrennt erkannt wurden, oder aufgrund der Echtzeit-Korrektur der transkribierten Zahlen, die nur kurze Zeit korrekt angezeigt wurden, um dann wieder falsch geändert vorzuliegen (vgl. Fantinuoli & Pisani 2021: 189). Trotz der hohen Fehlerraten der ASR ist bei der Teilnehmer*innengruppe, welche die technologische Unterstützung verwendete, ein Zuwachs von 25% an richtig gedolmetschten Zahlen (85,2%) im Vergleich zur Kontrollgruppe (60,2%) zu verzeichnen. Im Vergleich zu vorherigen Studien zeigen diese Daten sogar eine stärkere Verbesserung der Leistung, die z. B. in Defrancq & Fantinuoli (2021: 19) nur bei 22,5% lag.

2.4.3 Die Auswahl der Studienteilnehmer*innen in der CAI-Forschung

Eine Gemeinsamkeit aller dieser Studien ist die Auswahl der Studienteilnehmer*innen, und zwar Dolmetschstudierende. Obwohl viele diese Forschungslücke thematisieren (z. B. Defrancq & Fantinuoli 2021, Desmet et al. 2018), sind bis dato nur wenige Daten aus der Forschung über CAI-Tools und ASR mit erfahrenen Dolmetscher*innen verfügbar. Zu den Ausnahmen zählt Bereznaya (2022), wobei der Forschungsschwerpunkt auf der Benutzerfreundlichkeit und Anwendbarkeit von InterpretBank ASR in einem RSI-Kontext über die Zoom-Plattform lag. Untersucht wurde die Wirkung des Tools auf die Genauigkeit und Flüssigkeit der Dolmetschleistungen, vor allem in Bezug auf Zahlen. Die Studienteilnehmer*innen waren Dolmetscher*innen, die eine langjährige Erfahrung im Bereich Simultandolmetschen vorweisen konnten. Für das Experiment wurde das Tool *Augmented Interpreter - CAI v 1.2* verwendet, das Zahlen und Maßeinheiten in einer senkrechten Liste darstellt. Im Rahmen dieser Untersuchung wurden die kognitiven Anforderungen der Dolmetscher*innen gemessen. Sie mussten vor und nach der Dolmetschleistung das Alphabet rückwärts aufsagen, und Fehler und Dauer wurden berücksichtigt, um die kognitive Müdigkeit einzuschätzen (vgl. Bereznaya 2022: 54). Außerdem wurde die Flüssigkeit der Verdolmetschung auf Basis unterschiedlicher Kriterien untersucht. Darunter zählen Unterbrechungen im Sprechfluss länger als drei Sekunde, nonverbale Äußerungen, Fehlstarts, Zögern und Selbstkorrekturen (vgl. Bereznaya 2022: 61). Die durchschnittliche Fehlerquote betrug bei diesem Experiment 29,86% (Bereznaya 2022: 64), wobei die Genauigkeit der von der ASR gezeigten Zahlen zwischen 88,5%

und 94,4% lag (Bereznaya 2022: 58). Die Ergebnisse dieser Studie sind besonders relevant, da sie einerseits versuchen, ein RSI-Setting zu simulieren, das heißt, weitere Herausforderungen miteinzubeziehen, die mit den technologischen Schwierigkeiten eines Online-Meetings verbunden sind. Andererseits ermöglichen sie einen Einblick in die Zusammenarbeit mit erfahrenen Dolmetscher*innen bei Forschungsarbeiten und leisten somit einen großen Beitrag in diesem Forschungsfeld, das sonst weitgehend Leistungen und Meinungen von Studierenden untersucht.

Auch Frittella (2023) berücksichtigt die Wichtigkeit der Miteinbeziehung von professionellen Dolmetscher*innen in die Forschung, vor allem in Hinblick auf ihr Feedback während der Entwicklung der Schnittstellen von CAI-Tools. Die mangelnde Einbindung der Endbenutzer*innen trägt dazu bei, dass Dolmetscher*innen wenig Vertrauen in die Tools haben und Bedenken bezüglich der Wirkungen bzw. Nachwirkungen der CAI-Tools auf Dolmetschprozesse äußern (vgl. Frittella 2023: 4).

2.4.4 Das Konzept der Benutzerfreundlichkeit in der CAI-Forschung

Frittella (2023) stellt das Konzept der Benutzerfreundlichkeit der Benutzerschnittstellen von CAI-Tools in den Mittelpunkt, mit dem Ziel, sie als Unterstützung und nicht als Störfaktor für Dolmetscher*innen wahrzunehmen. Im Rahmen dieser Forschung wurden eine Pilot- und eine Hauptstudie durchgeführt, um das Feedback der Studienteilnehmer*innen in das CAI-Tool SmarTerp zu integrieren (vgl. Frittella 2023: 5).

Diese Forschungsarbeit hebt einen weiteren Aspekt hervor, welcher in vorherigen Studien mit CAI-Tools oft vernachlässigt wurde und in künftigen Forschungsdesigns vermehrt Berücksichtigung finden soll. Auswertungen von Dolmetschleistungen mit CAI-Tools fokussieren sich vor allem auf die Wiedergabe von einzelnen bedeutungstragenden Einheiten, wie Fachtermini oder Zahlwörter, und nicht auf die Satzebene bzw. auf den Kontext, in den diese Elemente eingebettet sind (vgl. Frittella 2023: 58). Die Komplexität der Wiedergabe der Zahlen liegt aber oft an der Tatsache, dass Zahlen in einer sogenannten ‘numerical information unit’ (NIU) vorkommen. Diese Einheiten bestehen aus Zahlwörtern, zusammen mit dem Referent (d.h. das Element, auf die sich die Zahl bezieht), der Maßeinheit, dem relativen Wert (z.B. Zunahme, Abnahme) sowie Informationen über die Ort und Zeit (vgl. Frittella 2023: 71). Diese Informationen können aber nicht automatisch von der ASR ausgefiltert werden, da aktuelle Algorithmen keine semantische Analyse durchführen können. Derzeit werden Zahlen also zusammen mit den nächstliegenden Satzelementen gezeigt, die in der Regel die Referenten oder die Maßeinheiten sind (vgl. Frittella

2023: 71). Das Bedürfnis, diese Elemente auch in die Analyse von Dolmetschleistungen miteinzubeziehen, wurde auch von Canali (2019: 118f.) unterstrichen. Zahlen kommen selten ohne zusätzliche Elemente wie unter anderem zeitlichen oder räumlichen Bezug vor, und die Häufigkeit der Fehler, die in die Kategorien falsche oder ungenaue numerische Informationen eingeordnet wurden, bestätigt, dass die Komplexität der Zahlen nicht nur die Ziffer betrifft, sondern auch die Elemente, die ihr vorangehen bzw. folgen (vgl. Canali 2019: 139).

Im Rahmen des Forschungsprojekts in Frittella (2023) wurde versucht, alle diese Voraussetzungen miteinzubeziehen und ein Forschungsdesign zu entwickeln, welches als Ziel die Benutzerfreundlichkeit der Benutzeroberfläche des Tools hat und untersucht, ob das Tool seinen Zweck erfüllt und in den Arbeitsalltag von Dolmetscher*innen eingebettet werden könnte (vgl. Frittella 2023: 75). In einer Fokusgruppendiskussion wurden zunächst Merkmale hervorgehoben, die in der Entwicklung des Designs von SmarTerp berücksichtigt werden müssen. Als Ergebnis wurde eine in drei Module aufgeteilte Schnittstelle entwickelt, die jeweils Namen, Terminologie, Abkürzungen und Zahlen enthält. Wenn es um die Sprache geht, wurden folgende Kriterien festgestellt: Namen wurden in der Zielsprache angezeigt, während Fachtermini und Abkürzungen auch mit ihrer Übersetzung verfügbar waren. Abkürzungen wurden sowohl in abgekürzter als auch in vollständiger Form angezeigt, während Zahlwörter als ein Mix von arabischen Ziffern und zielsprachigen Wörtern (ab der Größenordnung Tausend) gezeigt wurden (vgl. Frittella 2023: 72f.). Bei der Gestaltung der Rede wurden an unterschiedlichen Textstellen unterschiedliche Niveaus an Komplexität geschaffen, sodass Ergebnisse unter unterschiedlichen Bedingungen ausgewertet werden konnten. Um dem Anliegen von nur einzelnen bedeutungstragenden Elementen nachzukommen, wurden in der Analyse folgende Aspekte in Betracht gezogen: Terminologie und Zahlwörter wurden im Kontext evaluiert, und zwar wurde beachtet, ob sie alleine, mit Referenten, oder in hochkomplexen Textstellen vorkommen (vgl. Frittella 2023: 82f.). Außerdem wurde für die Auswertung ein kommunikativer Ansatz verwendet, in dem die Analyse auf Satzebene erfolgt und sich nicht nur auf die Auswertung der einzelnen Stimuli fokussiert; Zahlen, die richtig gedolmetscht aber falsch in Sätze eingebunden wurden, wurden als Fehler gekennzeichnet (vgl. Frittella 2023: 94).

Die Ergebnisse der Pilot- und Hauptstudie zeigen, dass von der ASR nicht erkannte Elemente in 93% der Fälle zu Auslassungen in den Dolmetschungen führen, während Fehler seitens der Dolmetscher*innen vor allem in informationsdichten Textstellen vorkommen (vgl. Frittella

2023: 139). Aus den Befragungen der Teilnehmer*innen ging das Bedürfnis hervor, Tools individuell gestalten zu können, so dass sie an individuelle Bedürfnisse angepasst werden können. Einschränkungen in der Verwendung von der ASR, die von den Befragten als Herausforderung gesehen wurden, beinhalteten die Auge-Ohr-Koordination, die Schwierigkeit, sich auf den gesamten Kontext und Sinn samt einzelnen Problemauslösern zu fokussieren, und den Umgang mit dem Tool, das unter allen Umständen verwendet wurde, ohne andere Dolmetschstrategien in Erwägung zu ziehen und wofür unrealistische Erwartungen seitens der Dolmetscher*innen bestanden (vgl. Frittella 2023: 142f.). Diese Ergebnisse werfen weiterführende Fragen für Forschung mit CAI-Tools und ASR im kognitiven und psychophysiologischen Bereich auf (vgl. Frittella 2023: 147).

Ein weiteres Forschungsprojekt, das mit einem ähnlichen Zweck entwickelt wurde, und zwar unter Miteinbeziehung der Meinungen von Dolmetscher*innen für die Erstellung eines sogenannten ‘artificial boothmate’ (ABM), wurde im Rahmen der Umfrage ‘Ergonomics for the Artificial Booth Mate’ (EABM) der Universität Gent und der Johannes Gutenberg-Universität Mainz präsentiert. Zwischen Dezember 2020 und März 2021 nahmen 525 Dolmetscher*innen an der Umfrage teil. Die Ergebnisse zeigten, dass 59,24% der Teilnehmer*innen ein vertikales Layout von CAI-Tools bevorzugen, wobei die neuen Zahlwörter unten angeführt und laut Mehrheit der Befragten (82,29%) so lange wie möglich auf dem Bildschirm angezeigt werden sollen (vgl. EABM o.J.). Außerdem wird präferiert, entweder Terminologie auf der linken und Zahlwörter auf der rechten Seite des Bildschirms zu sehen (39,62%), oder sie gemeinsam in einer Grafik auf dem Bildschirm zu visualisieren (39,43%), während neue Elemente entweder fett, mit einer anderen Farbe oder mit einer größeren Schriftgröße erscheinen sollten, um sie von den alten Elementen zu differenzieren (vgl. EABM o.J.). Diese Forschungsarbeiten liefern wichtige Angaben für die Entwicklung neuer Dolmetschtechnologien, obwohl sie sich nur auf eingeschränkte Gruppen von Studierenden oder Dolmetscher*innen mit bestimmten Sprachkombinationen beziehen und daher nicht verallgemeinert werden können. Trotzdem ist Benutzerfreundlichkeit ein grundlegender Schwerpunkt in der Forschung zu Dolmetschtechnologien, und ihr Einfluss in der tatsächlichen Verwendung von CAI-Tools scheint ausschlaggebend zu sein.

2.4.5 Die Rolle der Latenz in ASR-gestützten Tools

Ein weiterer Schwerpunkt in der Forschung liegt bei der Latenz, die zu den aktuellen Herausforderungen bei ASR-gestützten Tools zählt: Diese ist die Zeitverzögerung zwischen der Ausgangsrede und den auf dem Bildschirm angezeigten Wörtern (vgl. Frittella 2023: 49). Zu lange Zeitabstände zwischen den gehörten Inputs und den visuellen Outputs führen dazu, dass die Tools unbrauchbar werden (vgl. Fantinuoli 2021: 517).

Die Wirkung der Latenz auf die Qualität der Dolmetschung hinsichtlich Genauigkeit und Flüssigkeit wurde in Fantinuoli & Montecchio (2022) eingehend untersucht. Die Visualisierung der von automatischer Spracherkennung erkannten Zahlen oder Wörter auf dem Bildschirm erfolgt nicht gleichzeitig mit dem Audiosignal, da der Prozess, um Sprachsignale in Text umzuwandeln, komplexe Berechnungen und eine gewisse Zeit erfordert. Sollte die EVS (Ear-Voice-Span) zu lang werden, könnte sich das in den Leistungen der Dolmetscher*innen widerspiegeln (vgl. Fantinuoli & Montecchio 2022: 1f.). Für diese Studie wurde ein Video vorbereitet, das Zahlen ähnlich wie die ASR zeigt. Zahlen wurden jedoch in der Tonaufnahme von einem Hörsignal ersetzt und nur auf dem Bildschirm angezeigt. So konnte gemessen werden, je nachdem mit welchem Zeitabstand sie gezeigt wurden, ob das eine Wirkung auf die Leistung der Dolmetscher*innen gehabt hätte (vgl. Fantinuoli & Montecchio 2022: 2). Die Dolmetschungen wurden auf zwei Niveaus analysiert, nämlich Stimuli und Segmente. Für die Stimuli wurde die Genauigkeit der Wiedergabe der Zahlen gemessen; analysiert wurden die Größenordnung, die Genauigkeit der Zahl, die Aussprache und von Genauigkeit des Referenten. Als falsch wurden Rundungen, Verallgemeinerungen und Auslassungen sowie lexikalische, syntaktische, phonologische und Artikulationsfehler gesehen. Auf Ebene der Segmente wurde der Fokus auf diejenigen gelegt, in denen Zahlen vorkommen. Hier wurden folgende Aspekte analysiert: Treue, grammatikalische Korrektheit, Vollständigkeit, logischer Zusammenhalt, Konsistenz und Plausibilität. Außerdem wurde auch die Wahrnehmung von Stimme, Rhythmus, Eloquenz, Präsentation, Prosodie und kommunikativer Wirkung der Dolmetschungen evaluiert (vgl. Fantinuoli & Montecchio 2022: 3ff.). Die Ergebnisse zeigten, dass die Qualität der Verdolmetschung innerhalb von zwei Sekunden nicht von der Latenz beeinflusst wird und innerhalb von drei Sekunden ohne größere Beeinträchtigung erhalten bleibt. Längere Zeitspannen führten aber zu geringerer Genauigkeit und zu Informationsverlust (vgl. Fantinuoli & Montecchio 2022: 7).

Diese ersten Ergebnisse über die akzeptierte Latenz übernehmen eine wegweisende Funktion, um die Benutzbarkeit von ASR-Systemen beim Simultandolmetschen in der Zukunft besser evaluieren zu können. In vergangenen Studien wurde nur berücksichtigt, dass die Latenz innerhalb der durchschnittliche EVS von Dolmetscher*innen bleibt, und zwar innerhalb von 2,5 bis 3 Sekunden, wie z. B. in Defrancq & Fantinuoli (2021: 4). Wie aber Collard (2018) feststellt, ist eine gekürzte EVS eine verbreitete Dolmetschstrategie, um die Genauigkeit der Zahlen bei Dolmetschungen zu erhöhen. Dies könnte eine mögliche Erklärung sein, warum die Latenz der Tools immer noch von Dolmetscher*innen kritisiert wird, auch wenn sie minimal und innerhalb der durchschnittlichen EVS von Dolmetscher*innen gehalten wird (vgl. Frittella 2023: 107).

2.4.6 Die Rolle von akademischen Abschlussarbeiten in der CAI-Forschung

CAI-Tools sind ein relativ neues Forschungsgebiet, und diese Technologien haben sich als Forschungsgegenstand noch nicht stark etabliert (vgl. Rodriguez et al. 2022: 79). Ein beträchtlicher Teil dieser Forschung über in-booth CAI-Tools findet vor allem im Rahmen von Master- bzw. Doktorarbeiten großen Widerhalt, wie Frittella (2023: 54) anmerkt (z. B. Trenkwalder 2020; Freiberger-Geistberger 2021; Kittimongkolmar 2021; Pelizza 2021; Bereznaya 2022; Prandi 2022; Uran 2022).

Eine der ersten Masterarbeiten, die sich mit dem Einfluss von ASR auf die Wiedergabe von Zahlen und Terminologie beschäftigte, stammt aus dem Jahr 2020 und verglich die Software InterpretBank 6 mit oder ohne ASR (vgl. Trenkwalder 2020). Die Studie wurde mit 12 Masterstudierenden an der Universität Innsbruck durchgeführt und zeigte, dass mit der ASR 46,9 % mehr Termini richtig gedolmetscht wurden als ohne (46,6% vs. 68,4%) und die Qualität hinsichtlich Korrektheit, Präzision und Vollständigkeit mit ASR stieg (vgl. Trenkwalder 2020: 71f.), während die Zahlenwiedergabe mit ASR nur eine geringere Verbesserung erfuhr (+2%) (vgl. Trenkwalder 2020: 102).

Auch Pelizza (2021) untersuchte in einer Masterarbeit die Wiedergabe von Termini und Zahlen anhand der ASR-Funktion von InterpretBank 7, welche im Rahmen dieser Studie Genauigkeitsraten von 69% für Zahlen und 67,5% für Termini erreichte (vgl. Pelizza 2021: 74). Aus der Untersuchung ließ sich feststellen, dass zur Zeit des Experiments die Software technologisch fortgeschritten war. Nichtsdestotrotz gab es unterschiedliche technische Herausforderungen und Einschränkungen, unter anderem, dass Ausgangsreden auf fünf Minuten begrenzt sein mussten (vgl. Pelizza 2021: 75). Die terminologische Genauigkeit in den Dolmetschleistungen stieg mit

InterpretBank ASR im Vergleich zur Funktion BoothMode von InterpretBank, welche nicht über ASR verfügt, um bis zu 30%, während bei Zahlen eine Verbesserung von bis zu 37% mit der Unterstützung von ASR verzeichnet wurde (vgl. Pelizza 2021: 76f.).

Uran (2022) versuchte anhand von InterpretBank außerdem zwischen den unterschiedlichen Visualisierungsoptionen des Tools zu differenzieren und verglich drei Dolmetschungen: eine ohne technologische Unterstützung, eine mit InterpretBank ASR und der vollständigen Transkription und eine mit InterpretBank ASR und der punktuellen Unterstützung für Zahlen und Eigennamen. Außerdem wurde das Experiment nach zwei Wochen wiederholt, um eine potenzielle Leistungssteigerung durch die gestiegene Erfahrung mit dem Tool zu untersuchen (vgl. Uran 2022: 1). Aus den Ergebnissen geht hervor, dass die punktuelle Unterstützung mit Zahlen und Termini die bevorzugte Version im Vergleich zur vollständigen Transkription war, wobei der bewegte Text in jedem Szenario ablenkend wirkte (vgl. Uran 2022: 105). Die Hypothese, dass durch Erfahrung mit dem Tool die Dolmetschleistung eine Verbesserung erfährt, wurde außerdem in dieser Studie bestätigt (vgl. Uran 2022: 106). Die kognitive Anforderung während der Verwendung des Tools war auch eine der Forschungsfragen und wurde anhand von Interviews gemessen. Diese sollte aber laut Uran (2022: 106) in der Zukunft noch mit anderen Methoden wie Eyetracking- oder EEG-Daten erforscht werden.

Eine weitere Software, welche mittels ASR Transkriptionen in Echtzeit erstellt und sich der STT-Funktion von Google bedient, ist CASSIS, die ebenso in einer Masterarbeit für die Unterstützung während des Simultandolmetschens verwendet wurde (vgl. Szilas 2023). Für das Experiment wurde eine Rede fünf Minuten mit Unterstützung der Software und fünf Minuten ohne Unterstützung gedolmetscht, und sowohl die Genauigkeit der Wiedergabe als auch die kognitive Belastung durch die Software wurden untersucht (vgl. Szilas 2023: 4). Zahlwörter und Glossareinträge wurden in der Transkription hervorgehoben, und nur Glossareinträge waren auch separat mit ihren Übersetzungen verfügbar (Abb. 11) (vgl. Szilas 2023: 17). Wichtig ist an dieser Stelle anzumerken, dass CASSIS im Vergleich zur InterpretBank über keine Terminologieverwaltungsfunktion verfügt (vgl. Szilas 2023: 17).

Start/stop ASR	Choose input language	Choose font size	Upload term base
Transcription		Term base matches	
This is where the transcription will appear. You can adjust the font size for readability. Any terms which are in the term base are highlighted. Their target language pair appears on the right side.		term → Begriff, der	

Abb. 11: Beispiel für die Software CASSIS mit Hervorhebung von Glossareinträgen und ihrer Übersetzung nach Szilas (2023: 18)

Die Genauigkeit der Wiedergabe von Terminologie stieg mit der ASR im Vergleich zu den Verdolmetschungen mit bilingualen Glossaren um 7%; die terminologische Richtigkeit von Auflistungen stieg um 9,82% und die Genauigkeit der Zahlen um durchschnittlich 23,52% (vgl. Szilas 2023: 86). Außerdem wurden die Dolmetschstrategien analysiert, die während der Dolmetschungen eingesetzt wurden. Daten aus dieser Studie zeigen, dass Auslassungen von Zahlen und Verallgemeinerungen während der Verfügbarkeit von ASR sinken, während Interferenzen der Ausgangssprache in der Zielsprache steigen (vgl. Szilas 2023: 91).

Eine Alternative zu CAI-Tools, die gezielt für das Dolmetschen entwickelt wurden, ist auch in anderen Programmen, welche eine STT-Funktion (Speech-to-Text) aufweisen, zu finden. Das automatische Transkriptionsprogramm von Google ‘Google Voice Typing’ wird in Ricci (2020) verwendet, um anhand von zahlreichen Qualitätsmerkmalen Dolmetschleistungen zu evaluieren, darunter gefüllte und ungefüllte Pausen, unvollständige Sätze, die Sprechgeschwindigkeit, Auslassungen, Hinzufügungen und Ersetzungen (vgl. Ricci 2020: 9). Aus den Schlussfolgerungen ist abzuleiten, dass diese Funktion sowohl Vorteile als auch Nachteile hat; Vorteile betreffen vor allem die Kategorien Zahlen, Fehler und Kohärenz, während Fehler, die in der Transkription von Google gemacht wurden, keinen großen Einfluss auf die Dolmetschleistungen hatten. Dementsprechend kann diese Funktion allgemein als vorteilhaft während des Dolmetschens beurteilt werden (vgl. Ricci 2020: 129).

Eine Besonderheit unter den Masterarbeiten stellt Freiberger-Geistberger (2021) dar. Schwerpunkt dieser Arbeit ist der Einsatz von CAI-Tools in der Dolmetschlehre und ob die Integration der Tools in den Unterricht sinnvoll wäre. Berücksichtigt wurde vor allem die kognitive Belastung bei der Verwendung des Tool sowie welche Dolmetschstrategien und Störfaktoren damit verknüpft werden können. Im Rahmen einer Lehrveranstaltung an der Karl-Franzens-Universität

Graz wurde im Laufe des Semesters mehrmals das Tool verwendet, um Glossare zu erstellen und Unterstützung beim Simultandolmetschen zu bieten. Am Ende des Semesters wurden dann Meinungen zum Tool gesammelt und die Vorteile bzw. Herausforderungen des Tools in einer Gruppendiskussion besprochen. Die Ergebnisse dieser Studie heben vor, dass die Einstellung der Studierenden in Bezug auf das Tool im Dolmetschunterricht grundsätzlich positiv war. Trotzdem wurde einerseits die kognitive Belastung kritisiert, die aus den inkorrekten Ergebnissen bzw. den visuellen Reizen des Tools entsteht. Andererseits wurde argumentiert, dass technische Probleme auf die Studierenden und auf die Durchführung der Studie negativ wirkten (vgl. Freiberger-Geistberger 2021: 116f.). Die Integration von ASR in universitäre Studiengänge wurde auch von Amelina & Tarasenko (2020: 8) thematisiert. Diese Studie unterstreicht das Bedürfnis, technologische Kompetenzen während des Studiums zu entwickeln. Ein Beispiel wäre die Fähigkeit, beim Simultandolmetschen sich auf die ASR, das Publikum und den*die Redner*in gleichzeitig zu fokussieren.

2.5 Aktueller Stand der Technologie und Zukunftsperspektiven

Die Veränderungen aufgrund der Digitalisierung und Automatisierung und die steigende Anzahl von Informations- und Kommunikationstechnologien haben dazu geführt, dass eine neue technologische Wende auch im Dolmetschbereich unmittelbar bevorsteht (vgl. Fantinuoli 2018b: 6f.). Drechsel (2019: 260) beobachtet, dass es in der heutigen digitalisierten Welt von unausweichlicher Relevanz ist, Mitglied der digitalen Gemeinschaft zu sein und über digitale Fertigkeiten zu verfügen. Um an dem Digitalisierungsprozess teilzunehmen und ihn auch mitzubestimmen, müssen Dolmetscher*innen über hohe Informationskompetenzen verfügen und dabei die richtige Einstellung haben, um die menschliche Seite in die immer stärker digitalisierten und automatisierten Dolmetschprozesse einzubringen und die Zukunft des Berufsstandes zu sichern (vgl. Drechsel 2019: 266).

Auch internationale Organisationen stellen fest, dass eine Auseinandersetzung mit den neuen Technologien und mit künstlicher Intelligenz in den Sprachdienstleistungen notwendig ist. In einem Bericht der Europäischen Institutionen wird auf die transformative Wirkung von automatisierten Prozessen in der Sprachindustrie eingegangen, und Sprachtechnologien werden als integraler Bestandteil des Arbeitsalltags gesehen (vgl. EU 2019: 1f.). Das Konferenzdolmetschen wurde aber noch nicht stark davon betroffen. Verfügbare Softwares verfügen nicht über kognitive, kulturelle, intellektuelle und emotionale Fertigkeiten, die für die Gewährleistung der Qualität beim

Dolmetschen notwendig sind. Außerdem stehen viele Fragen zum Thema Sicherheit und Vertraulichkeit noch offen, die beim Dolmetschen für internationale Organisationen von großer Relevanz sind (vgl. EU 2019: 5).

Viele technologische Anwendungen, die in den vergangenen Jahrzehnten für Dolmetscher*innen entwickelt wurden, sind eingestellt oder nicht mehr weiterentwickelt worden. Laut einer Auflistung der Terminology Coordination Unit der Generaldirektion Übersetzung des Europäischen Parlaments (DG TRAD) stehen nur folgende Ressourcen derzeit zur Verfügung: Flashterm.eu[11], Glossary Assistant[12], Interplex, InterpretBank, Interpreters' Help, Intragloss, Lookup und Terminus (vgl. Terminology Coordinator o.J.).

Die neue Anwendung BoothMate[13] ist die App von Interpreters' Help, um Terminologie zu verwalten, und wurde aktualisiert, um die Suchfunktion und Glossarerstellung von Interpreters' Help zu optimieren. Zusätzlich werden einzelne Glossareinträge in einer persönlichen Datenbank gespeichert und können gleichzeitig in mehreren Glossaren hinzugefügt bzw. geändert werden. Außerdem bietet eine offline Version von BoothMate die Möglichkeit, auch offline Glossare zu erstellen und Terminologie zu verwalten (vgl. Boothmate o.J.).

Auch InterpretBank wurde vor kurzem weiterentwickelt, und InterpretBank 9 ist die neueste Version des Tools auf dem Markt. Das Tool ist in Form einer Desktop-App und einer Web-App verfügbar, um von unterschiedlichen Geräten auf die eigenen Glossaren zuzugreifen. Vier Module stehen den Benutzer*innen zur Verfügung: Edit, um Glossare zu erstellen bzw. zu ändern, Document, um Terminologie zu extrahieren und Glossare aus Vorbereitungsunterlagen automatisch zu erstellen, Memorization, um die Terminologie interaktiv zu üben und Live Search, um eine schnelle und benutzerfreundliche Suchfunktion zu ermöglichen (vgl. Interpretbank o.J.). Die Anwendung Live Search (Abb. 12) bietet den Zugriff auf Glossare während der Dolmetschung an, und die Standardeinstellungen können subjektiv angepasst werden. Es können unter anderem die Glossare ausgewählt werden, die für jeden Dolmetscheinsatz verwendet werden sollen. Zusätzliche Ressourcen, wie zum Beispiel die Datenbank IATE oder online verfügbare Übersetzungen, können auch angezeigt werden (vgl. Interpretbank o.J.).

[11] https://www.flashterm.eu/

[12] https://play.google.com/store/apps/details?id=conference.interpeter.glossaryassistant&hl=it&gl=US

[13] https://boothmate.app/dashboard

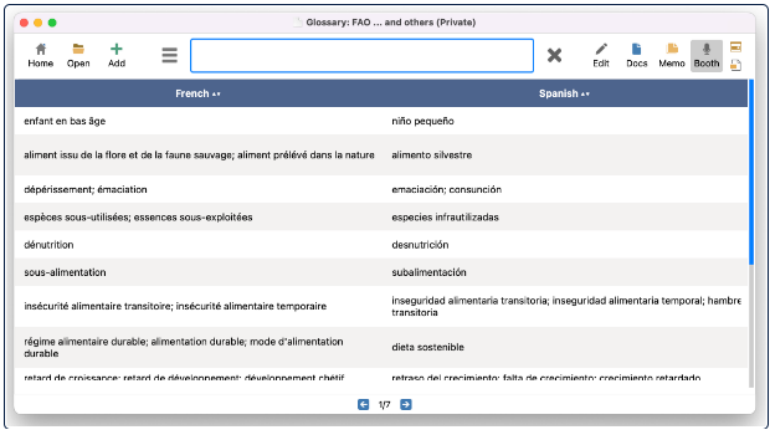

Abb. 12: Bildschirmfoto von InterpretBank Live Search nach Interpretbank (o.J.)

Künstliche Intelligenz wird für mehrere Anwendungen von InterpretBank 9 verwendet. Dazu zählen zum Beispiel Übersetzungsvorschläge während der Glossarerstellung, die automatische Erstellung von Glossaren aus einsprachigen Webseiten, aus Word- oder PDF-Dateien oder auf Basis von gegebenen Themen, die automatische Extraktion von Schlüsselthemen aus Dokumenten und Digital Boothmate. Die Anwendung InterpretBank ASR (Digital Boothmate) ermöglicht Terminologie, Eigennamen und Zahlen automatisch während der Dolmetschung anzuzeigen, und wird mit Englisch als Ausgangssprache angeboten (vgl. Interpretbank o.J.).

Auch die europäischen Institutionen haben unterschiedliche Ressourcen für die Vorbereitung von Dolmetscher*innen zur Verfügung gestellt, darunter das Interpreters' Digital Toolbox, das zur Unterlagenverwaltung, Terminologieextraktion und Glossarerstellung für Dolmetscher*innen der Generaldirektion der Europäischen Kommission (SCIC) dient, und das Knowledge Centre on Interpretation[14]. Diese ist eine digitale Plattform zum Wissens-, Best Practices und Informationsaustausch (vgl. EU 2019: 5).

Auch Aus- bzw. Weiterbildungsangebote für Dolmetscher*innen haben in den letzten Jahren Erfolge erzielt. Viele Anwendungen im Bereich Computer-Assisted Interpreting Training (CAIT) haben sich von einfachen Ressourcen zu virtuellen Lernumgebungen in 3D entwickelt, die aufgezeichnete Reden und Dolmetschübungssammlungen enthalten (vgl. Corpas Pastor 2018: 151). Eine Umfrage aus dem Jahr 2020 mit 25 Ausbildungseinrichtungen für Dolmetscher*innen wurde durchgeführt, um die Einbindung von CAI-Tools in universitären Curricula zu untersuchen. Alle Befragten gaben an, dass Technologien für das Dolmetschen in Curricula enthalten sind, darunter

[14] https://knowledge-centre-interpretation.education.ec.europa.eu/en

Fern- und Videodolmetschen, technologiegestütztes Dolmetschen, CAIT-Tools und online Ressourcen wie zum Beispiel Terminologiedatenbanken, Wörterbücher und Suchmotoren (vgl. Prandi 2020: 3). Auch das Programm vom EMCI (European Masters in Conference Interpreting) hat sich bewährt, um universitäre Curricula zu aktualisieren (vgl. Rodriguez et al. 2022: 81). Ein bemerkenswertes Beispiel unter den Ausbildungsangebote ist die Università degli Studi Internazionali di Roma (UNINT). Ab April 2024 wird eine postgraduale Ausbildung zum Thema künstliche Intelligenz und Dolmetschtechnologien angeboten. Der Studienplan enthält einen Schwerpunkt auf KI-Technologien für die Sprachindustrie, wie Natural Language Processing (NLP) und Korpora, sowie Module über das Computer-Assisted Consecutive Interpreting (CACI), Computer-Assisted Simultaneous Interpreting (CASI) und Technologien für das Dialog-dolmetschen (vgl. AltaFormazione o.J.).

Das Remote Interpreting und das Machine Interpreting sind zwei weitere Bereiche, welche die Zukunft des Dolmetschens stark prägen werden. Das Ferndolmetschen wurde am Anfang vor allem in Form von Telefondolmetschen durchgeführt, während heutzutage hochentwickelte Videokonferenzanlagen zur Verfügung stehen (vgl. Fantinuoli 2018b: 5). CAI-Tools haben immer mehr an Relevanz gewonnen, auch weil sie von internationalen RSI-Anbietern in ihre Anwendungen eingebettet wurden, wie z. B. Interpreter Assist in Kudo (vgl. Frittella 2023: 4). Das Machine Interpreting bezieht sich hingegen auf die automatische Übersetzung mündlicher Texte mittels Computersoftwares und zielt darauf ab, Dolmetscher*innen in ihrer Tätigkeit zu ersetzen. Diese Anwendung besteht aus der automatischen Spracherkennung, der automatischen Über-setzung und der Speech-to-Text Synthese, um die Übersetzung in der Zielsprache wiederzugeben (vgl. Fantinuoli 2018b: 5). Cloudbasiertes Computing und Speech-to-Speech-Technologien haben zum Aufkommen mobiler, automatisierter Dolmetschersysteme beigetragen (vgl. Corpas Pastor 2018: 153).

Die zunehmende Entwicklung der Technologie in diesem Sektor zeigt, dass die Digitalisierung und Automatisierung der Dolmetschprozesse auch in der Zukunft zwei wichtige Themen sein werden. Künstliche Intelligenz und automatische Spracherkennung spielen in diesem Zusammenhang eine grundlegende Rolle. 2010 prognostizierte Winteringham (2010: 93), dass Spracherkennungssoftwares zukünftig verwendet werden, um während der Dolmetschung eine Terminologiedatenbank zu erstellen, die in Echtzeit Fachterminologie erkennt. Seitdem stehen

ASR-gestützte CAI-Tools zur Verfügung, welche erfolgreicher sind als jene Versuche, Dolmetscher*innen mit Maschinen zu ersetzen (vgl. Braun 2019: 278).

Bevor die Fragestellung und Methode dieser Arbeit präsentiert werden, wird das Thema des Eyetrackings behandelt. Das Eyetracking wird als diagnostisches Messverfahren für die Forschung von kognitiven Prozessen eingesetzt und wurde als Forschungsmethode vorliegender Masterarbeit mit dem Ziel ausgewählt, die Verteilung der Aufmerksamkeit und die Wirkung spezifischer Problemauslöser auf Dolmetscher*innen und ihre Dolmetschleistungen zu analysieren.

3 Eyetracking und Konferenzdolmetschen

Das Eyetracking ist eine etablierte Forschungsmethode, die in der Translationswissenschaft zunächst in Forschungsbereichen wie dem Übersetzen und Post-Editing angewendet wurde. In den 1980er Jahren fanden die ersten Eyetracking-Studien auch in der Dolmetschwissenschaft Beachtung, zunächst mit exklusivem Fokus auf das Vom-Blatt-Dolmetschen bzw. Dolmetschen mit Text, bei dem ein schriftlicher Text mündlich gedolmetscht wird, während später die ersten Studien zum Thema Konsekutivdolmetschen veröffentlicht wurden, die vor allem aufgrund der schriftlichen Notizen den Einsatz von Eyetracking ermöglichen (z. B. Chen 2018; Kuang & Zheng 2023). Heutzutage wird das Eyetracking für unterschiedliche Dolmetschmodi bzw. Anwendungen im Bereich Dolmetschen eingesetzt, die z. B. in Chmiel (2021), Hu et al. (2022) und Moratto (2020) eingehend behandelt werden. Da in dieser Masterarbeit eine Studie mit Eyetracking in Bezug auf das Simultandolmetschen und das multimodale Dolmetschen durchgeführt wird, liegt der Schwerpunkt der Studien auf diesen Themen und auf den Messungsindikatoren, die in diesen Forschungsbereichen verwendet werden.

Anhand der Blickrichtung, Pupillengröße und Dauer der Fixationen vermitteln Eyetracker Daten über die visuelle Aufmerksamkeit von Studienteilnehmer*innen. In der Dolmetschwissenschaft wird das Eyetracking vor allem für die Untersuchung der Aufmerksamkeitsverteilung während des Dolmetschens verwendet, wobei Dolmetscher*innen mehreren visuellen Reizen ausgesetzt werden, wie z. B. Redner*in, Präsentationsfolien oder Texte (vgl. Chmiel 2021: 458). Obwohl akustische Reize einen Großteil der Aufmerksamkeit von Dolmetscher*innen erregen, spielen auch visuelle Inputs während des Konferenzdolmetschens eine wichtige Rolle. Diese können in zwei Kategorien unterteilt werden, nämlich bildbasierte und textbasierte Informationen (vgl. Chmiel 2021: 458). Bildbasierte Informationen bestehen aus dem*r Redner*in und dem Publikum. Ihr Verhalten, ihre Körpersprache und Gestik können relevante Informationen für Dolmetscher*innen liefern, um Bedeutung zu erschließen und die Dolmetschung entsprechend anzupassen. Außerdem gibt es Präsentationsfolien, die Bilder oder Grafiken enthalten, welche auch zu den bildbasierten Informationen zählen (vgl. Chmiel 2021: 458). Textbasierte Informationen sind hingegen entweder Materialien, die von dem*der Redner*in zur Verfügung gestellt werden (darunter das Transkript der Rede, Präsentationsfolien oder etwaige Unterlagen), oder Materialien, die von den Dolmetscher*innen erstellt werden, wie zum Beispiel Informationen zum*r Redner*in, Glossare oder Notizen, die während des Dolmetschens gemacht werden (vgl. Chmiel 2021: 458).

Die Forschung mit Eyetracking stellt die Möglichkeit dar, Antwort auf Fragen zu den von Dolmetscher*innen verwendeten visuellen Inputs zu geben (vgl. Seeber 2012: 346).

3.1 Einführung in das Eyetracking und in die Messungsindikatoren

Das Eyetracking ist eine Versuchsmethode, um Augenbewegungen und Blickpositionen zeit- und aufgabenbezogen aufzuzeichnen und die visuelle Aufmerksamkeit zu untersuchen (vgl. Carter & Luke 2020: 50). Diese Methode wurde seit Ende des 19. Jahrhunderts in der Forschung verwendet. Erst in den 1970er Jahren wurde Infrarotlicht eingesetzt, um Augenbewegungen zu erkennen, wie es in den heutigen Eyetracker üblich ist (vgl. Stachowiak-Szymczak 2019: 67). Die Methode des Eyetracking wurde ursprünglich vor allem in Forschungsbereichen wie zum Beispiel der Psychologie, der Psycholinguistik und der Kognitionswissenschaften verwendet, während es mittlerweile für eine Vielzahl von Forschungsgebieten Anwendung findet. Aus diesem Grund ist es zunächst notwendig, einen Überblick über diese Forschungsmethode zu präsentieren. Dabei ist es wichtig zu erörtern, welche Herausforderungen eine Eyetracking-Studie mit sich bringt und welche Messungsindikatoren in Studien im Bereich Konferenzdolmetschen bzw. Dolmetschen mit multimodalen Inputs von besonderer Relevanz sind. Danach wird eine Auswahl von Eyetracking-Studien in der Dolmetschwissenschaft präsentiert, um den Forschungsstand abzubilden und Forschungslücken bzw. neue Perspektiven für zukünftige Studien darzustellen.

3.1.1 Einführung in das Eyetracking als Forschungsmethode

Das Eyetracking ist eine weit verbreitete Forschungsmethode, die einen Einblick in kognitive Prozesse ermöglicht. Der Grund, warum Eyetracking in der Forschung eingeführt wurde, ist auf die sogenannte Auge-Geist-Hypothese (auf Englisch: Eye-mind assumption) zurückzuführen (vgl. Chmiel 2021: 457). Die Annahme, dass das, was man betrachtet, mit dem zusammenhängt, was kognitiv verarbeitet wird, und dass Rückschlüsse auf kognitive Prozesse anhand der Blickrichtung möglich sind, ist aber eine Fehlinterpretation. Nicht nur können Augenbewegungen oft nicht direkt auf mentale Aktivitäten zurückgeführt werden, sondern auch die kognitive und visuelle Aufmerksamkeit findet unter Umständen nicht synchron statt. Das heißt, die kognitive Informationsverarbeitung kann anfangen, bevor ein Gegenstand gesehen wird, und kann noch lange fortgesetzt werden, nachdem der Gegenstand nicht mehr sichtbar ist (vgl. Hvelplund 2017: 250).

Eyetracker verwenden Nahinfrarot, das in der Pupille und in der Hornhaut widergespiegelt und mit einer Infrarotkamera gemessen wird. Diese Form der Datenerhebung ist nicht invasiv und eine solche Bestrahlung ist nicht schädlich. Diese Daten werden dann verwendet, um Einschätzungen über die Augenbewegungen und Blickrichtungen zu machen. Dafür ist es notwendig, dass eine sorgfältige Kalibrierung stattfindet, um die Qualität der Datenerhebung mittels Eyetracking zu gewährleisten (vgl. Chmiel 2021: 459). In Studien im Bereich Konferenzdolmetschen werden vor allem drei Eyetracker verwendet (EyeLink, Tobii und SMI), wobei EyeLink das meistverbreitete ist (vgl. Chmiel 2021: 459). Auf dem Markt sind Eyetracker entweder als stationäre Eyetrackers mit Desktop-Halterung für bildschirmbasierte Messungen oder als mobiles Messsystem mit Brille verfügbar. Die Tischhalterung kann im kopfstabilisierten (auf Englisch: Head-stabilized) oder Remote-Modus verwendet werden. Im kopfstabilisierten Modus werden eine Stirn- und eine Kinnstütze verwendet. Dieser Modus wird aber in der Dolmetschforschung vermieden, da die Kinnstütze das Sprechen erschwert. Der Remote-Modus wird hingegen bevorzugt. Dafür wird ein Aufkleber auf der Stirn des*r Teilnehmers*in angebracht, um Kopfbewegungen besser zu erkennen (vgl. Chmiel 2021: 460).

Um Eyetracking-Daten zu analysieren, werden Interessensbereiche (auf Englisch: Area of Interest, abgekürzt AOI) verwendet, die in der Software EyeLink ‚Interest Areas' (abgekürzt IA) genannt werden. Diese werden fokussiert, um so analysiert werden zu können. Es kann sich dabei um einzelne Wörter, Sätze, Präsentationsfolien oder andere Bereiche handeln, die untersucht werden müssen (vgl. Chmiel 2021: 460). Gegenstand von Untersuchungen ist auch der sogenannte Interessenzeitraum (auf Englisch: Interest period). Dieser ist ein gewisses Zeitfenster, in dem Augenbewegungen analysiert werden (vgl. Chmiel 2021: 460).

Es gibt mehrere Herausforderungen, die mit der Analysemethode von Eyetracking verbunden sind. Dazu zählt die Datenqualität, die von unterschiedlichen Faktoren beeinflusst wird, wie zum Beispiel die Messrate, Genauigkeit, Präzision und Latenzzeit (vgl. Conklin et al. 2018: 22). Um diese zu gewährleisten, ist zunächst eine sorgfältige Kalibrierung wichtig. Trotzdem kann ein Eyetracker ungenau sein, und es sollten etwa 20 Prozent Datenverlust bei jedem Experiment mit eingerechnet werden (vgl. Chmiel 2021: 462). Besondere Aufmerksamkeit während der Gestaltung des Experiments sollte auch eventuellen Störungsfaktoren gewidmet werden. Vor allem für komplexe Reize wie Videos oder Präsentationsfolien können kräftige Farben oder größere Elemente eine Ablenkung sein und die Datenqualität beeinträchtigen (vgl. Chmiel 2021: 462).

3.1.2 Messungsindikatoren in Eyetracking-Studien

Es gibt unterschiedliche Messwerte, die für die Auswertung von Eyetracking-Daten verwendet werden können. In Holmqvist (2011) werden mehr als 120 Messungsindikatoren präsentiert, jedoch wird in vorliegendem Kapitel nur eine Auswahl von Messwerten aufgeführt, die im Rahmen von Studien mit Dolmetscher*innen mit Fokus auf das Konferenzdolmetschen verwendet werden.

Laut Prandi (2022) werden in der Dolmetschwissenschaft vor allem fixationsbasierte Messwerte verwendet, darunter Fixationen und Sakkaden. Conklin et al. (2018: 66) definieren diese Messungsindikatoren wie folgt: "Fixations are any point where the eye is stationary on a target, and saccades are movements from one location to another." Diese werden oft in Studien über die kognitive Belastung verwendet, da sie besonders passend in Bezug auf linguistische Faktoren verwendet werden können (vgl. Prandi 2022: 118; Conklin et al. 2018: 66). Sakkaden können zum Beispiel dazu dienen, Regressionsmuster während des Lesens zu erkennen, wenn Studienteilnehmer*innen zu bereits gelesenen Textstellen zurückkehren (vgl. Conklin et al. 2018: 66).

Anhand eines Eyetrackers kann die Anzahl und die durchschnittliche Dauer der Fixationen gemessen werden. Die Anzahl der Fixationen auf einen bestimmten Interessensbereich (auf Englisch: Area of Interest, abgekürzt AOI) gibt Auskunft über die Aufmerksamkeit, die auf eine AOI während eines gesamten Versuchs gerichtet wurde (vgl. Conklin et al. 2018: 67). Durchschnittlich dauert eine Fixation zwischen 200-300 Millisekunden, obwohl sie auch bis zu einigen Sekunden oder weniger als 30-40 Millisekunden dauern kann (vgl. Holmqvist 2011: 381). In der Mehrheit der Eyetracking-Studien wird die durchschnittliche Fixationsdauer verwendet. Eine längere Fixationsdauer deutet auf aufwändigere kognitive Verarbeitungsprozesse hin, und mehrere Studien haben eine enge Verknüpfung zwischen den Augenfixationen und der kognitiven Verarbeitung nachgewiesen (vgl. Holmqvist 2011: 382). Längere Fixationen können aber auch mit weniger Erregungsimpulsen verbunden sein, die dem Tagträumen ähneln, während kurze Fixationen in Stresssituationen üblich sind und auf eine hohe geistige Anstrengung hindeuten (vgl. Holmqvist 2011: 383). In Eyetracking-Studien muss außerdem beachtet werden, dass längere Fixationen auch auf Expertise in einem Bereich, neurologische Beeinträchtigungen oder Aufmerksamkeit auf bewegliche Elemente zurückgeführt werden können (vgl. Holmqvist 2011: 383). Es ist dementsprechend schwer, Rückschlüsse auf die Gründe für die Dauer und Anzahl der Fixationen zu ziehen.

Ein weiterer wichtiger Messungswert ist die Dauer der Erstfixation. Diese bezieht sich auf die Verweildauer, bevor es zu einem anderen Element wandert (vgl. Chmiel 2021: 461), und ist mit den kognitiven Prozessen der Erkennung und Identifizierung von Elementen verbunden (vgl. Holmqvist 2011: 385). Wenn Wörter die Stimuli sind, spiegelt die Dauer der Erstfixation lexikalische Zugangs- und Verarbeitungsschwierigkeiten wider (vgl. Chmiel 2021: 461).

Weitere Messungswerte, die in der Translationsforschung verwendet werden, sind Rückkehrbewegungen zu bereits betrachteten Elementen (auf Englisch: Regressions), Lidschläge (auf Englisch: Blinks) und die Pupillenerweiterung bzw. -verengung, die Rückschlüsse auf kognitive Verarbeitungsprozesse ermöglichen (vgl. Seubert 2019: 108).

In Studien mit visuellen Inputs wie Bildern oder Videos werden die Anzahl der Fixationen (auf Englisch: Fixation counts), die Dauer der Fixationen (auf Englisch: Dwell Time), die durchschnittliche Fixationsdauer und die Verteilung der Anzahl von Fixationen auf unterschiedliche AOI auf dem Bildschirm als Messungsindikatoren verwendet (vgl. Conklin et al. 2018: 69). Unterschiedliche Faktoren müssen in der Gestaltung dieser Studien berücksichtigt werden. Merkmale wie Kontrast, Intensität, Helligkeit, Häufigkeit und Bewegungen können die visuelle Aufmerksamkeit auf unterschiedliche Elemente steuern (vgl. Conklin et al. 2018: 115), und bei Studien mit akustischen Inputs reagieren Zuhörer*innen in der Regel auf das, worauf sich der auditive Input bezieht (vgl. Conklin et al. 2018: 119). Diese Faktoren müssen entweder im Experiment vermieden, oder als Variablen in der Analyse der Ergebnisse berücksichtigt werden (vgl. Conklin et al. 2018: 115).

3.2 Forschungsstand von Eyetracking-Studien in der Dolmetschwissenschaft

Die Vielfalt der Studien, die im Bereich Simultandolmetschen mit einem Eyetracker durchgeführt wurden, zeigt, dass diese Forschungsmethode für unterschiedliche Forschungszwecke eingesetzt werden kann. In der Dolmetschwissenschaft wurde das Eyetracking vor allem für die Forschung zum Dolmetschmodus des Vom-Blatt-Dolmetschens verwendet, welches von besonders vielen visuellen Elementen gekennzeichnet ist. In früheren Studien wurde das Leseverhalten während des Vom-Blatt-Dolmetschens mit dem von schriftlichen Übersetzungen verglichen, während in neueren Studien der Fokus vorwiegend auf der Schwierigkeit spezifischer Textstellen liegt (vgl. Chmiel 2021: 463).

Die Forschungsschwerpunkte vorliegender Masterarbeit, und zwar multimodale Inputs, CAI-Tools und die Wiedergabe von Zahlen, sind auch bereits Untersuchungsgegenstand von Eyetracking-Studien gewesen. Unterschiedliche Eyetracking-Studien beschäftigten sich mit multimodalen Inputs während des Dolmetschens, zum Beispiel Seeber (2012; 2017), Seubert (2019) und Chmiel et al. (2020). Aufgrund ihrer Natur und der Tatsache, dass sie sehr oft auch in schriftlicher Form vorhanden sind, waren auch Zahlen der Schwerpunkt von mehreren Studien (z. B. Seeber 2012, Korpal & Stachowiak-Szymczak 2018 und Korpal & Stachowiak-Szymczak 2020). Bis dato wurde nur eine Eyetracking-Studie über CAI-Tools veröffentlicht (siehe Prandi 2022). Im folgenden Kapitel wird eine Auswahl von Eyetracking-Studien angeführt, die in chronologischer Folge präsentiert werden, um einen Überblick über die unterschiedlichen Anwendungsbereiche des Eyetrackers und die Ergebnisse, die bisher in diesem Forschungsbereich erzielt wurden, darzustellen. Diese Veröffentlichungen wurden als Grundlage herangezogen, um Schlussfolgerungen aus den Ergebnissen des Experiments zu ziehen, das im Rahmen dieser Arbeit durchgeführt wurde.

Die erste Studie, die im Bereich Simultandolmetschen mit einem Eyetracker durchgeführt wurde, fand im Jahr 1981 statt. In McDonald & Carpenter (1981) liegt der Forschungsschwerpunkt auf die Dolmetschstrategien bzw. -prozesse bei der Dolmetschung mehrdeutiger idiomatischer Ausdrücke. Dafür wurde ein schriftlicher Text auf Englisch ins Deutsche vom Blatt gedolmetscht. Die Fixationen während der Dolmetschleistungen wurden verwendet, um zu beobachten, ob bei der Dolmetschung die Satzteile ohne Unterbrechungen im Lesemuster gelesen wurden, und ob es Vorwärtsfixationen oder Rückschritten während des Lesens vom Text auftraten (vgl. McDonald & Carpenter 1981: 236). Diese Studie kam zu dem Ergebnis, dass Entscheidungen über die Dolmetschstrategie getroffen wurden, während die neuen Satzteile gelesen wurden. Dabei wurde versucht, die vorhergehenden Informationen mit den neuen zu vergleichen bzw. zu ergänzen. Dieser Prozess spiegelte sich dann in der Blickrichtung wider (vgl. McDonald & Carpenter 1981: 246).

In Hyönä et al. (1995) wurde untersucht, ob die Pupillengröße im Laufe verschiedener mentaler Aktivitäten Veränderungen unterlag. Das Experiment zeigte, dass während des Simultandolmetschens eine größere Pupillendilatation als während des Shadowings verzeichnet wurde, und während des Shadowings war sie wiederum größer als während des reinen Zuhörens.

Das konnte also die Hypothese bestätigen, dass Aktivitäten mit unterschiedlichen Graden von Schwierigkeit auf die Pupillengröße wirken (vgl. Hyönä et al. 1995: 603).

Die erste Studie, die die Eyetracking-Methode verwendete, um zu untersuchen, worauf sich Dolmetscher*innen während des Simultandolmetschens fokussierten, wurde erst 2012 durchgeführt. In Seeber (2012) lag der Forschungsschwerpunkt auf Zahlen, die in drei unterschiedlichen Modalitäten präsentiert wurden: akustisch-verbal (in der Ausgangsrede mündlich erwähnt), visuell-räumlich (durch Gestik mit den Händen gezeigt) und visuell-verbal (als Wörter bzw. Zahlwörter gezeigt). Für die Studie wurde entschieden, den Dolmetscher*innen Zahlen gleichzeitig in akustischer und visueller Form zur Verfügung zu stellen, wie es in der Arbeitspraxis auch gängig ist. Das Experiment wurde mit professionellen Konferenzdolmetscher*innen durchgeführt, um zu untersuchen, inwiefern sie visuelle Reize in Form von Gestik oder Präsentationsfolien verwendeten (vgl. Seeber 2012: 343). Zahlen von eins bis zehn wurden durch Gestik gezeigt, während höhere Zahlen auf Präsentationsfolien angezeigt wurden (vgl. Seeber 2012: 344). Für das Experiment wurden drei Areas of Interest festgelegt, nämlich das Gesicht des Redners für die Mimik, sein Körper für die Hände und Gestik und die Zahlwörter auf den Präsentationsfolien. Die gesamte Dauer der Fixationen wurde bemessen. Durchschnittlich schauten die Teilnehmer*innen bei Sätzen, die kleine Zahlen beinhalteten, deutlich länger auf das Gesicht des Redners als bei Sätzen, die große Zahlen beinhalteten. Die Dauer der Fixationen war länger bei Folien, die große Zahlen enthielten, als bei Folien, die kleine Zahlen enthielten (vgl. Seeber 2012: 345). Aus diesen Daten lässt sich feststellen, dass Dolmetscher*innen dazu neigten, auf das Gesicht des Redners zu schauen. Präsentationsfolien wurden verwendet, um akustische Informationen zu ergänzen, und Zahlen wurden unabhängig von ihrer Größe auf den Präsentationsfolien gesucht. Das Experiment deutete darauf hin, dass Simultandolmetscher*innen aktiv nach Informationen auf den Folien suchten, um das Gehörte zu ergänzen (vgl. Seeber 2012: 345).

In Chmiel & Mazur (2013: 189f.) wurde die Hypothese formuliert, dass Dolmetschstudierende in fortgeschrittenen Phasen des Studiums eine niedrigere kognitive Belastung aufwiesen als Studierende, die sich am Anfang des Studiums befanden. Diese Hypothese basierte auf der Annahme, dass das Dolmetschen ein Prozess ist, der mit der Erfahrung immer mehr automatisiert werden kann und dementsprechend eine immer niedrigere kognitive Belastung erfordert (vgl. Chmiel & Mazur 2013: 192). Für diese Studie wurden als AOI Subjekt-Verb-Objekt (SVO) und nicht-SVO Sätze ausgewählt, weil nicht-SVO Sätze in der Sprachkombination

Polnisch-Englisch eine besondere Schwierigkeit darstellen (vgl. Chmiel & Mazur 2013: 194). Die Daten widerlegten die Annahme, dass große Unterschiede in den Eyetracking-Daten von Studierenden in früheren und fortgeschrittenen Phasen des Studiums zu erkennen sind, und dass die Eyetracking-Daten von Studierenden mit weniger Erfahrung auf eine erhöhte kognitive Belastung hinweisen. Studierenden in fortgeschrittenen Phasen des Studiums zeigten weiters eine effizientere Technik des Scanning, und die Unterschiede der zwei Teilnehmer*innen-Gruppen in Bezug auf die Gesamtdauer der Aufgabe (auf Englisch: Total task time) wiesen keine statistische Signifikanz auf. Es ließ sich daraus ableiten, dass in der Ausbildung Unterschiede von zwei Semestern nicht ausreichend sind, um Abweichungen in der Dolmetschtechnik zu erkennen (vgl. Chmiel & Mazur 2013: 201).

In Korpal & Stachowiak-Szymczak (2018: 340f.) wurde ein EyeLink 1000+ Remote Eyetracker verwendet, um die kognitive Belastung in Bezug auf Zahlen und ihrem Kontext zu untersuchen. Das gleiche Experiment wurde mit zwei unterschiedlichen Gruppen von Teilnehmer*innen mit unterschiedlichem Kompetenz- und Erfahrungshintergrund durchgeführt, und zwar mit 25 Dolmetschstudierenden und 30 professionellen Dolmetscher*innen. Im Rahmen des Experiments wurde eine englischsprachige Ausgangsrede von vier Minuten verwendet, und gleichzeitig wurde eine Präsentation gezeigt, die aus fünf PowerPoint-Folien mit den wichtigsten Inhalten der Rede bestand (vgl. Korpal & Stachowiak-Szymczak 2018: 342). Um die Augenbewegungen zu messen, wurden Areas of Interest (AOI) für Zahlen und Kontextinformationen erstellt und die durchschnittliche Dauer der Fixationen jeweils untersucht (vgl. Korpal & Stachowiak-Szymczak 2018: 345). Die Ergebnisse zeigten, dass die Genauigkeitsraten bei professionellen Dolmetscher*innen höher waren als bei Dolmetschstudierenden (vgl. Korpal & Stachowiak-Szymczak 2018: 348). Außerdem wurde festgestellt, dass hohe Genauigkeitswerte in der Wiedergabe von Zahlen mit hohen Genauigkeitswerten in der Wiedergabe von Kontextinformationen verknüpft waren. Die Dauer der Fixationen war außerdem höher für Zahlen als für andere Elemente auf den Folien, was zeigte, dass diese kognitiv herausfordernder als ihr Kontext waren (vgl. Korpal & Stachowiak-Szymczak 2018: 349).

In Seubert (2019: 19) wurde eine Studie mit professionellen Konferenzdolmetscher*innen durchgeführt, und das Experiment wurde im Rahmen einer realitätsnahen Konferenz vor Ort mit Präsentationsfolien durchgeführt. Der Schwerpunkt lag auf den visuellen Elementen, die als Unterstützung oder Störung wahrgenommen werden können, und auf der Relevanz visueller

Informationen (vgl. Seubert 2019: 120f.). Das Blickfeld wurde anhand der Dauer und Anzahl der Fixationen vor, während und nach dem Dolmetschen untersucht, und die AOI, die dafür ausgewählt wurden, waren folgende: der*die Redner*in, die Präsentationsfolien, die Leinwand auf der der*die Redner*in angezeigt wurde, die Sitzungsleitung und das Publikum (vgl. Seubert 2019: 152). Die Studie zeigte, dass Studienteilnehmer*innen in der Lage waren, relevante Informationen zu erkennen und Störungsfaktoren auszublenden (vgl. Seubert 2019: 293). Von großer Relevanz waren vor allem Folieninhalte und Erklärungen, welche den Gehörten entsprachen, obwohl der Zeitpunkt der Verwendung der visuellen Informationen sehr individuell war (vgl. Seubert 2019: 259).

In Chmiel et al. (2020) ging es um das Simultandolmetschen mit Text. Schwerpunkt dieser Studie war der Umgang von Dolmetscher*innen mit visuellen Inputs, die mit dem Gehörten entweder übereinstimmen oder nicht. Laut professionellen Dolmetschstandards sollte das Gehörte im Fall einer Diskrepanz zwischen dem Gehörten und den schriftlich zur Verfügung gestellten Informationen immer Vorrang haben. Dies steht im Gegensatz zu der menschlichen kognitiven Tendenz, im Fall einer Diskrepanz die schriftliche Version anstelle des Gehörten zu priorisieren (vgl. Chmiel et al. 2020: 38f.). Für das Experiment wurden den Dolmetscher*innen schriftliche Texte zur Verfügung gestellt, die teilweise vom Ausgangstext abwichen. Die Ergebnisse zeigten, dass übereinstimmende Elemente zu höheren Genauigkeitsraten als nicht übereinstimmende Elemente führten. Wenn Informationen schriftlich und akustisch abwichen, hatte in vielen Fällen der schriftlichen Input Vorrang, und viele Fehler waren auf visuelle Interferenzen zurückzuführen (vgl. Chmiel et al. 2020: 49f.). Daten aus dem Eyetracker zeigten, dass nicht übereinstimmende Elemente länger betrachtet wurden. Das galt als Beweis, dass Dolmetscher*innen im Falle von Diskrepanz nicht sofort versuchten, sich nur auf den akustischen Kanal zu fokussieren, sondern stattdessen versuchten, die Abweichung zu klären. Elemente, die länger beobachtet wurden, wurden aber in der Mehrheit der Fälle auch richtig gedolmetscht, was impliziert, dass eine längere Reflexion zu richtigen Entscheidungen führte (vgl. Chmiel et al. 2020: 52).

In Korpal & Stachowiak-Szymczak (2020) lag der Schwerpunkt der Studie auf der Wirkung des Redetempos des Ausgangstextes auf die Genauigkeit der Wiedergabe von Zahlen und auf der Frage, ob Änderungen im Sprechtempo einen Einfluss auf die Augenbewegungen haben. Die Ergebnisse der Studie bestätigten die Hypothese, dass die Redegeschwindigkeit einen Einfluss auf

die Genauigkeitsraten sowohl von professionellen Dolmetscher*innen als auch von Dolmetsch-studierenden hat, obwohl die durchschnittliche Genauigkeitsrate von professionellen Dolmetscher*innen höher war (vgl. Korpal & Stachowiak-Szymczak 2020: 137). Die Ergebnisse verdeutlichten Änderungen im Blickverhalten in Bezug auf die Anzahl der Fixationen für die Zahlen und die gesamten Folien. Diese waren während der schnell vorgetragenen Reden viel höher, was auf ein sogenanntes Scanning der Informationen hindeutet, das sich von einem sorgfältigen Lesemuster unterscheidet. Korpal & Stachowiak-Szymczak (2020: 138) kamen zur Schlussfolgerung, dass Scanning in diesem Fall eine Strategie sein kann, um akustische Inhalten zu antizipieren bzw. um sie mit dem Text auf den Präsentationsfolien zu überprüfen. Somit kann auch das Arbeitsgedächtnis entlastet werden (vgl. Korpal & Stachowiak-Szymczak 2020: 138).

In Seeber et al. (2020) wurde der SR Research EyeLink 1000 Eyetracker verwendet, um einen Einblick in die kognitiven Prozesse während des Simultandolmetschens mit Text zu erlangen. Der Forschungsschwerpunkt lag auf dem während des Simultandolmetschens erkennbaren Lesemuster. Unterschiedliche Satzteile wurden als Kontrollsätze ausgewählt und mit den Satzteilen davor und danach verglichen, um Daten über die visuelle Aufmerksamkeit der Dolmetscher*innen während der entsprechenden akustischen Inputs (der Ausgangsrede) auszuwerten. Dafür wurden fünf AOI erstellt: der vorherige Satz, der Kontrollsatz, das Ende des Satzes, der darauffolgende Satz und das Video des*r Redner*s (vgl. Seeber et al. 2020: 1487). Die Ergebnisse legten nahe, dass während ein Satz von dem*der Redner*in ausgesprochen wird, mehr Aufmerksamkeit auf den vorherigen Satz im Text gelegt wurde. Der Satz, der nach dem Kontrollsatz kam, bekam hingegen weniger visuelle Aufmerksamkeit. Das widerlegte die Vermutung, dass Dolmetscher*innen den zur Verfügung gestellten Text verwenden, um Inhalte zu antizipieren (vgl. Seeber et al. 2020: 1488). Außerdem zeigte die Studie, dass die Kontrollsätze selbst wenig visuelle Aufmerksamkeit von den Dolmetscher*innen bekamen, während sie vom Redner ausgesprochen wurden (vgl. Seeber et al. 2020: 1488). Die Aufmerksamkeit der Dolmetscher*innen lag auf den Sätzen, die sie gerade dolmetschten und die sie mit einem Timelag wiedergeben mussten.

Im Rahmen einer Doktorarbeit wurde in Prandi (2022) die kognitive Belastung von Simultandolmetscher*innen bei der Verwendung von CAI-Tools untersucht, um Unterschiede in der manuellen Suchfunktion von Terminologie und in der automatischen Erkennung von Termini zu analysieren. Dabei liegt der Fokus auf der Wirkung unterschiedlicher technologischer Lösungen

und auf den kognitiven Prozessen von Dolmetscher*innen. Ein Eyetracker SIM red250 wurde dabei im Remote-Modus verwendet (vgl. Prandi 2022: 158). Mit dem ASR-gestützten Tool wurden zwei AOI ausgewählt, nämlich eine für das Video des*r Redners*in und eine für das ASR-Tool; außerdem wurden einzelne AOI für alle einzelnen Vorschläge der ASR festgelegt (vgl. Prandi 2022: 169). Das Experiment, das in Prandi (2022) beschrieben wird, war das erste, das die Eyetracking-Methode verwendete, um die Anwendungen von CAI-Tools bzw. ASR-gestützten Tools zu erforschen. Dabei wurden aber nicht nur Eyetracking-Daten, sondern auch Dolmetschleistungen in Bezug auf Genauigkeit und Auslassungen analysiert, um die Unterschiede zwischen bestehenden Anwendungen zu untersuchen.

Dieser Überblick der bisher veröffentlichten Eyetracking-Studien zeigt, dass im Laufe der Jahre das Eyetracking als Forschungsmethode in unterschiedlichen Studien verwendet wurde. Ausgehend von den Experimenten, die in der Vergangenheit zum Thema der visuellen Aufmerksamkeit durchgeführt wurden, wurde der experimentelle Teil dieser Masterarbeit gestaltet.

4 Fragestellung und Methode

Wie bereits erwähnt, ist das Simultandolmetschen mit Unterstützung durch automatische Spracherkennung ein neues Forschungsgebiet, das erst seit 2017 in der wissenschaftlichen Literatur Erwähnung, und in der Arbeitspraxis von Dolmetscher*innen Einsatz findet. Diese bahnbrechende Technologie, die mit künstlicher Intelligenz ausgestattet wurde, spiegelt den modernsten Stand der Technik wider. Aus diesem Grund wurde eine Studie über die Wirkung von ASR-gestützten Tools auf die visuelle Aufmerksamkeit von Dolmetscher*innen entworfen, um einen Beitrag zur Forschung über die Anwendungsmöglichkeit dieser Tools zu leisten. In diesem Kapitel werden die Zielsetzung und die Forschungsfrage sowie das Forschungsdesign, die Planung des Versuchs, die Auswahl einer Software für die ASR, die Miteinbeziehung von Studienteilnehmer*innen, die Vorbereitung einer Ausgangsrede und eines Fragebogens präsentiert. Danach wird auf die Durchführung des Experiments und die Vorgehensweise bei der Analyse der Dolmetschleistungen eingegangen.

4.1 Zielsetzung und Forschungsfrage

Die Erforschung von kognitiven Prozessen bzw. Dolmetschprozessen ist von großer Bedeutung, um CAI-Tools mit ASR an die Bedürfnisse und Arbeitsabläufe von Dolmetscher*innen anzupassen und damit die Funktionalität dieser Tools und folglich ihre Akzeptanz in der Arbeitspraxis zu erhöhen. Um kognitive bzw. neurokognitive Prozesse in der Dolmetschwissenschaft zu untersuchen, stehen heutzutage unterschiedliche Methoden, darunter Eyetracking, PET-Untersuchungen, fMRI und EKP, zur Verfügung (vgl. Hu et al. 2022: 2). Aus diesem Grund wurde in dieser Studie ein Eyetracking-Experiment durchgeführt, um die visuelle Aufmerksamkeit von Dolmetschstudierenden während des Simultandolmetschens zu untersuchen.

Die Forschungsfrage, die im Mittelpunkt dieser Arbeit steht, lautet: “Welche Wirkung haben visuelle Inputs in Form von durch automatische Spracherkennung erkannten Zahlen im CAI-Tool InterpretBank auf die Verteilung der visuellen Aufmerksamkeit beim Simultandolmetschen und auf die Richtigkeit der gedolmetschten Zahlen?”. Um diese Frage zu beantworten, soll zunächst die Wirkung von visuellen Inputs in Form von Vorschlägen eines ASR-Tools auf die visuelle Aufmerksamkeit von Simultandolmetscher*innen untersucht werden. Zwei Areas of Interest (AOI), nämlich ein Video der Rednerin und eine Software mit der ASR von Zahlen, wurden festgelegt, um das Blickverhalten von Teilnehmer*innen während des Simultandolmetschens zu untersuchen. Im zweiten Teil der Analyse soll die Wirkung der ASR auf die

Genauigkeit der Wiedergabe der Zahlen untersucht werden. Zahlen sollen in ihrem Kontext (Referenten) berücksichtigt werden, um eine ausführliche Analyse der Dolmetschleistungen durchzuführen. Danach sollen die Daten des Eyetrackers und der Dolmetschleistungen miteinander in Bezieht gesetzt werden.

Zusätzlich zu der aufgeführten Fragestellung wurden folgende Hypothesen aufgestellt: dass seit Beginn der Rede die visuelle Aufmerksamkeit auf die Rednerin und auf ihr Gesicht gerichtet wird. Das wird erwartet, da das Vorhandensein vom Video der Ausgangsrede viele Informationen, wie Mimik und Gestik, zur Verfügung stellt, die beim Dolmetschen auch zum besseren Verständnis der Inhalte führen können. In Bezug auf das Blickverhalten während des Dolmetschens wird auch vermutet, dass die Unterbrechung der ASR-Anwendung zu Fixationen auf die Software führen wird, obwohl Zahlen nicht mehr angezeigt werden, da Studienteilnehmerinnen nicht im Vorhinein informiert werden, dass die Software im zweiten Teil des Videos nicht mehr verfügbar sein wird. Um die Hypothese zu überprüfen, wird die Anzahl und Dauer der Fixationen verwendet.

In Bezug auf die Genauigkeit bei der Wiedergabe von Zahlwörtern kann vermutet werden, dass ein hoher Prozentsatz der Zahlwörter richtig wiedergegeben wird, da sie von der ASR-Software angezeigt werden. Gleichzeitig kann aber erwartet werden, dass die Referenten häufiger als die Zahlwörter ausgelassen werden und ihre Genauigkeitsrate niedriger als die Genauigkeitsrate der Zahlwörter sein wird, da Referenten nicht von der Software erkannt werden. Weiters wird die Hypothese aufgestellt, dass die visuelle Aufmerksamkeit auf die von der ASR erkannten Zahlen und die korrekte Wiedergabe der Zahlen verglichen werden können.

Der Fragebogen, der hauptsächlich zur Klärung von individuellen Schwierigkeiten bzw. Meinungen der einzelnen Teilnehmerinnen dient, wird verwendet, um Informationen über die Herausforderungen des Experiments und über den Umgang mit dem Tool zu sammeln. Hierbei kann davon ausgegangen werden, dass die Dichte der Zahlen als eine der Schwierigkeiten der Rede erwähnt wird und dass die Software als unzuverlässig eingestuft wird, da Teilnehmerinnen nicht über die geplante Unterbrechung informiert wurden. Zuletzt kann vermutet werden, dass das experimentelle Setting als ablenkend bzw. als ein Stressfaktor interpretiert wird, der einen Einfluss auf die Dolmetschleistung hat.

4.2 Forschungsdesign

Im Rahmen dieser Masterarbeit wurde ein Experiment mit neun Studentinnen des Masterstudiums Translation an der Universität Wien durchgeführt, um ihr Blickverhalten während einer Simultandolmetschung mit Unterstützung durch ASR für die Erkennung der Zahlen zu untersuchen. Zunächst musste das Versuchssetting im Eyetracker-Labor organisiert werden, danach wurde eine passende Software für die ASR ausgewählt. Eine Ausgangsrede wurde gezielt für dieses Experiment vorbereitet. Zur Erhebung der subjektiven Wahrnehmung der Teilnehmer*innen in Bezug auf das Versuchssetting, die Software und den Umgang mit visuellen Inputs wurde ein Fragebogen verwendet.

4.2.1 Versuchssetting

Das Experiment fand am Zentrum für Translationswissenschaft der Universität Wien statt, das in seinen Räumlichkeiten über einen Eyetracker *EyeLink® Portable Duo* von *SR Research* verfügt. In den Monaten vor dem Experiment wurde Kontakt mit dem Forschungsteam HAITrans an der Universität Wien aufgenommen, um das Experiment und seine Ausführbarkeit hinsichtlich der technischen Bedingungen bzw. Ausstattung zu besprechen. Die Ausstattung bestand aus einer Eyetracking-Einheit, die eine High-Speed-USB-Kamera und einen Infrarotstrahler enthält, um das Blickverhalten zu erkennen. Die Kamera wurde auf einem Stativ montiert und mit dem sogenannten ‚head free-to-move'-Modus verwendet, das leichte Kopfbewegungen während des Experiments ermöglicht. Dieser Modus wurde einem ‚head-stabilised'-Modus vorgezogen, um ein natürlicheres Setting für die Dolmetscher*innen zu simulieren. Hinter der Eyetracking Unit stand der sogenannte Display PC Monitor, den Studienteilnehmer*innen während des Experiments verwendeten. In einem Nebenraum (Kontrollraum) standen der Host PC, der das Eyetracking durchführt, und der Display PC, der die visuellen Reize während des Experiments anzeigt (vgl. SR Research 2017: 3f.). Tab. 1 fasst die Hauptmerkmale des Eyetrackers *EyeLink® Portable Duo* von *SR Research* zusammen (vgl. SR Research 2017: 7).

Tab. 1: Hauptmerkmale des Eyetrackers *EyeLink® Portable Duo* von *SR Research*

Merkmale	
Eyetracking-Methode	Dark-Pupil, Cornea-Reflex-Methode
Abtastrate (Sampling Rate)	500 Hertz
Durchschnittsgenauigkeit (Average Accuracy)	Bis zu 0,15° (0,25° bis 0,5° typisch)

Aufnahmetyp (Type of recording)	Binokular
Kalibrierungstyp	13-Punkt-Kalibrierung

Am 21. November 2023 wurden die technischen Bedingungen für das Experiment im Rahmen eines Vorversuches überprüft; unter anderem wurde festgestellt, dass im Experimentraum kein Ton hörbar war. Zu einem späteren Zeitpunkt wurden Kopfhörer zur Verfügung gestellt, trotzdem waren keine Audioaufnahmen der Dolmetschleistungen über das Mikrofon möglich. Aus diesem Grund wurde während des Experiments ein externes Aufnahmegerät in den Raum gestellt, um die Dolmetschungen aufzuzeichnen, die später für die Transkriptionen und die Analyse der Leistungen verwendet wurden. Im Rahmen des Vorversuchs wurden außerdem die Konfiguration, Kalibrierung und Validierung des Eyetrackers getestet. Diese wurden mit einer Versuchsperson durchgeführt, die nicht an dem Experiment teilnahm. Jede Versuchsperson muss ca. 50-55 cm von der Kamera entfernt sein (vgl. SR Research 2017: 46), was dazu führt, dass für jeden*r Teilnehmer*in, je nach Größe, das Stativ umgestellt werden muss, um die genannte Distanz zu gewährleisten. Außerdem müssen sich Studienteilnehmer*innen im Remote-Modus einen Sticker auf die Stirn kleben, um dem System die Kopfposition zu signalisieren (vgl. SR Research 2017: 48). Nachdem diese Schritte mit der Versuchsperson erfüllt wurden, wurde die Kalibrierung und Validierung durchgeführt. Die Validierung wurde erfolgreich abgeschlossen und die durchschnittlichen und maximalen Fehler des Eyetrackings wurden als angemessen angezeigt. Ein zweiter Vorversuch wurde am 23. November 2023 mit einer brillentragenden Person durchgeführt, da für diese eine genauere Anpassung der Einstellungen (Licht, Position der Kamera, Sitzplatz) notwendig ist (vgl. SR Research 2017: 44). Zu diesem Zeitpunkt war das Tonproblem bewältigt. Trotzdem konnte die Lautstärke mit den zur Verfügung gestellten Kopfhörern nicht angepasst werden. Der Ton wurde trotzdem als angemessen für das Dolmetschen beurteilt, da er laut hörbar war. Nachdem auch dieser Vorversuch erfolgreich abgeschlossen wurde, wurden die Studienteilnehmer*innen für die Durchführung der Studie kontaktiert.

4.2.2 Software für die ASR

Da zur Zeit der Durchführung der Studie nicht viele CAI-Tools mit integrierter ASR auf dem Markt verfügbar waren, waren die Auswahlmöglichkeiten sehr begrenzt. KUDO Interpreter Assist befand sich noch in der Entwicklungsphase, genauso wie SmarTerp, wofür nur Infomaterial zur Verfügung gestellt werden konnte. Aus diesem Grund wurde InterpretBank für die Durchführung des

Experiments ausgewählt. Zunächst wurde die 14-tägige Testlizenz der Software heruntergeladen, um ihre Anwendung zu überprüfen. Für die InterpretBank ASR-Funktion musste zuerst *Stereomix* heruntergeladen werden, sodass die Software mit der richtigen Audioquelle verbunden werden konnte (vgl. InterpretBank ASR o.J.). Dieses Programm war im für das Experiment verwendeten Computer noch nicht installiert (Fujitsu Lifebook A3511 i13, Betriebssystem Windows 11 Education). Es war aber für das Experiment unabdingbar, da eine Audioquelle aus dem Computer (das Video, das für das Experiment aufgezeichnet wurde) von der ASR erkannt werden musste. Zu diesem Zeitpunkt waren aber zwei Herausforderungen entstanden, die dazu führten, dass das Experiment angepasst werden musste. Zunächst wurde beobachtet, dass InterpretBank ASR zu diesem Zeitpunkt nur die Ausgangssprache Englisch für die automatische Spracherkennung anbot, obwohl davor weitere Ausgangssprachen verfügbar waren, zum Beispiel Italienisch (wie in Pelizza 2021). Außerdem waren Zahlen in der Anwendungsmöglichkeit Digital Boothmate von InterpretBank nur sehr klein abgebildet, da viel Platz für die von der ASR erkannte Terminologie eingeräumt wird. Da Fachtermini im Rahmen dieses Experiments nicht untersucht werden sollten, hätte die Benutzerschnittstelle von InterpretBank zu Ablenkung führen können und somit die Ergebnisse des Experiments zu stark beeinflusst. Aus diesem Grund wurde *Augmented Interpreter - CAI v 1.1* als Software für die Erkennung der Zahlen ausgewählt (vgl. Augmented Interpreter o.J.). Auch diese Software wurde von dem Entwickler von InterpretBank, Dr. Claudio Fantinuoli, entwickelt, und teilt viele Ähnlichkeiten mit InterpretBank. Zahlen waren hier größer und auch hervorgehoben. Die einfache Benutzeroberfläche sollte sich positiv auf die kognitiven Leistungen von Dolmetscher*innen auswirken, da diese zu den grundlegenden Merkmalen von CAI-Tools zählt (vgl. Fantinuoli 2017: 30). *Stereomix* musste auch für diese Software aktiviert werden. *Augmented Interpreter* ist ohne Lizenz online verfügbar und kann ohne Einschränkungen verwendet werden. Trotzdem ist die Erkennung von Zahlen nur auf fünf Minuten beschränkt; danach muss die Seite neu geladen werden (vgl. Augmented Interpreter o.J.). Wie in den Anleitungen von InterpretBank erklärt wird Wichtig funktioniert *Augmented Interpreter* nur mit dem Browser Google Chrome (vgl. InterpretBank o.J.).

4.2.3 Studienteilnehmer*innen

Neun Studentinnen des Masterstudiums Translation am Zentrum für Translationswissenschaft der Universität Wien wurden für die Durchführung des Experiments kontaktiert. Voraussetzungen waren eine längere Erfahrung mit dem Simultandolmetschen (mind. 2 Semester) und Italienisch

und Englisch in der Sprachkombination. Sechs Studentinnen hatten die Sprachkombination Italienisch (A-Sprache), Deutsch (B-Sprache) und Englisch (C-Sprache), und eine Studentin Deutsch (A-Sprache), Italienisch (B-Sprache) und Englisch (C-Sprache). In die Studie wurden außerdem zwei Studentinnen involviert, die die oben genannten Voraussetzungen nicht erfüllen, das heißt kein Englisch in der Sprachkombination an der Universität Wien hatten. Sie wurden trotzdem eingeladen, da sie schon Erfahrung im Rahmen von universitären Lehrveranstaltungen mit der Sprache Englisch gesammelt hatten und hohe Niveaus in der italienischen Sprache vorweisen konnten. Ihre Sprachkombinationen waren jeweils Deutsch (A-Sprache), Italienisch (B-Sprache) und Spanisch (C-Sprache) und Italienisch (A-Sprache), Deutsch (B-Sprache) und Russisch (C-Sprache). Alle Studentinnen waren zur Zeit der Studie zwischen 23 und 27 Jahre alt (Durchschnittsalter: 24,5 Jahre alt). Sie befanden sich zwischen dem 3. und 7. Semester (Durchschnitt: 4,2) und hatten zwischen zwei und sechs Semester Erfahrung mit Simultandolmetschen an der Universität (Durchschnitt: 3,1). Es wurde versucht, ähnliche Bedingungen zu schaffen wie für die Studienteilnehmer*innen-Auswahl der vorherigen Studien mit ASR-gestützten CAI-Tools (z. B. Desmet et al. 2018; Fantinuoli & Montecchio 2022; Trenkwalder 2020), um eventuelle Vergleiche bei der Analyse der Ergebnisse zu ermöglichen.

Wie in anderen Studien wurden den Teilnehmer*innen einige Informationen über die Studie zur Verfügung gestellt, um die Teilnehmer*innen an die Bedingungen des Experiments zu gewöhnen (z. B. Defrancq & Fantinuoli; Fantinuoli & Montecchio 2022). Als mit den Teilnehmerinnen Kontakt aufgenommen wurde, wurden ihnen die Sprachrichtung, die ungefähre Länge und das allgemeine Thema der Rede (Mobilität) übermittelt. Außerdem wurde ihnen erklärt, dass es sich um eine Studie mit automatischer Spracherkennung und Eyetracking handeln sollte und dass keine terminologische Vorbereitung vor dem Experiment notwendig wäre. Um sie an das Setting zu gewöhnen, wurde ihnen außerdem ein Video geschickt. Es war ähnlich wie das Video, das während des Experiments verwende wurde, und zeigte auf der linken Hälfte des Bildschirms die Rednerin, und auf der rechten Hälfte das Programm *Augmented Interpreter CAI v 1.1* (Abb. 13).

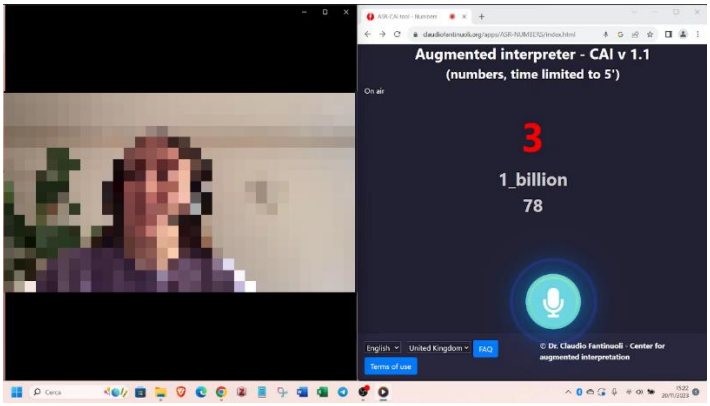

Abb. 13: Screenshot aus dem Beispielvideo mit Redner (links) und Augmented Interpreter (rechts)

4.2.4 Vorbereitung der Ausgangsrede

Nachdem die Software für die Durchführung des Experiments ausgewählt worden war, wurde eine Ausgangsrede vorbereitet. Wie Frittella (2022: 38) erklärt, muss das Design der Ausgangsrede für die Analyse der Dolmetschleistung gezielt sein und an die Forschungsfrage angepasst werden. Variablen in der Rede müssen zum Beispiel reflektiert eingesetzt werden. Es müssen Elemente vermieden werden, die vom Forschungsinteresse ablenken. Wichtig ist auch eine genaue Beschreibung der Spracheinheiten, in denen Zahlen vorkommen, um die Replikation der Studie und die Vergleichbarkeit der Ergebnisse gewährleisten zu können. Aus diesen Gründen wurden unterschiedliche Voraussetzungen in der Erstellung der Rede berücksichtigt. Als Thema wurde ‚Mobilität' ausgewählt, da es ein allgemeines Thema ist, das keiner fachlichen Vorbereitung bedarf und das die Erstellung einer Rede mit vielen Zahlwörtern ermöglicht. Wie in vorangehenden Studien, die sich mit ASR-gestützten CAI-Tools beschäftigen, wurde die Dauer der Rede auf ca. sieben Minuten beschränkt (z. B. Fantinuoli & Montecchio 2022; Ricci 2020; Trenkwalder 2020). Obwohl die Software *Augmented Interpreter* die Spracherkennung nur bis maximal fünf Minuten ermöglicht, bevor sie neu gestartet werden muss, wurde trotzdem diese Länge für die Ausgangsrede festgelegt. Es wurde außerdem entschieden, dass die letzten Minuten ohne Software gedolmetscht werden mussten. Einerseits ist die plötzliche Unterbrechung eines Online-Dienstes ein Szenario, das im Alltag vorkommen könnte. Andererseits konnten damit das Verhalten und die Dolmetsch-strategien und -leistungen der Studienteilnehmer*innen bei einer Unterbrechung der Software untersucht werden. Die Struktur der Rede wurde anhand der Anweisungen von Frittella (2023: 85) entworfen, nämlich mit einer Einleitung ohne Problemauslöser, die als Aufwärmzeit diente.

Danach wurde der Hauptteil der Rede vorbereitet, der aus unterschiedlichen Abschnitten bestand. Zwischen den Sätzen, die Zahlen enthielten, wurden sogenannte Kontrollsätze eingefügt. Diese bestanden aus 30-50 Wörtern und enthielten keine Problemauslöser oder besondere Schwierigkeiten (vgl. Frittella 2022: 85). Die Ausgangsrede bestand aus 732 Wörtern, darunter 29 Zahlen. Der Titel der Rede lautete ‚Mobility now and in the future', und die Informationen für die Erstellung der Rede wurden unterschiedlichen Webseiten entnommen und teilweise geändert, um die für die Forschung erforderlichen Voraussetzungen zu erfüllen oder um die Verständlichkeit des Textes zu erhöhen (vgl. MeinBezirk 2023, National Geographic 2013; Statista 2023; Statistik 2023) (siehe Anhang I). Außerdem wurden zwei weitere Faktoren der Rede hinzugefügt, um Unterschiede im Verhalten der Studienteilnehmer*innen zu untersuchen. Es wurde entschieden, eine Zahl zusätzlich bei der Videoaufnahme mit den Fingern zu signalisieren und eine Zahl falsch auszusprechen.

Nach Erstellung der Ausgangsrede wurde das Video vorbereitet, das von Studienteilnehmer*innen gedolmetscht werden sollte. Die Länge des Videos betrug insgesamt 7 Minuten und 20 Sekunden, was bedeutet, dass die Rede mit einem durchschnittlichen Sprechtempo von 99,7 Wörtern pro Minute vorgetragen wurde. Dieses Sprechtempo ähnelte dem in vorherigen Studien (z. B. Fantinuoli & Montecchio 2022; Frittella 2022). Das Video wurde in 1080p-Auflösung mit einem iPad Air und mit einem Headset Sennheiser PC 5.2 CHAT in einem ruhigen Raum aufgenommen, um eine möglichst hohe Audio- und Videoqualität zu erzielen.

Vor der Dolmetschung wurden der Titel der Ausgangsrede und acht Vokabeln mit Übersetzung zur Verfügung gestellt, die während der Dolmetschung vorkamen. Die zwei Präsentationsfolien mit einer Zusammenfassung der Bedingungen des Experiments und mit dem Glossar wurden 1,5 Minuten vor dem Experiment auf dem Bildschirm gezeigt (siehe Anhang II). Um vergleichbare Vorbereitungszeit und -bedingungen für alle Teilnehmer*innen zu gewährleisten, wurden diese Präsentationsfolien in das Video eingegliedert. Danach wurde die Software *Augmented Interpreter* mit dem Video der Ausgangsrede getestet, um die Erkennung der Zahlen zu überprüfen. Es wurde festgestellt, dass die Software manchmal mehrmals neu gestartet bzw. die Webseite neu geladen werden musste, bevor die Spracherkennung funktionierte. Aus diesem Grund wurde entschieden, eine Aufnahme des Bildschirms mit dem Video der Ausgangsrede und mit der ASR zu machen, die dann für die Dolmetschungen verwendet werden konnte. Die Bildschirmaufnahme wies viele Vorteile für die Durchführung und Analyse des Experiments auf.

Damit konnte sichergestellt werden, dass die ASR allen Teilnehmerinnen die gleichen Ergebnisse anzeigt. Außerdem musste bei den Experimenten nur ein Video (statt eines Videos und einer Webapp) abgespielt werden, was die Wahrscheinlichkeit des Auftretens technischer Probleme verringerte. Der Bildschirm wurde mit dem Programm OBS Studio (vgl. OBS Project o.J.) aufgenommen und die Länge des gesamten Videos, inklusive einleitender Folien, betrug insgesamt 9 Minuten.

4.2.5 Vorbereitung des Fragebogens

Es wurde ein Fragebogen mit Google Formulare erstellt, um die subjektiven Wahrnehmungen von Studienteilnehmer*innen zu beurteilen. Ziel des Fragebogens war vor allem zu erfahren, welche Faktoren laut den Studienteilnehmer*innen ihre Aufmerksamkeit beeinflussen und ob es nennenswerte Schwierigkeiten oder Störfaktoren gab. Der Fragebogen bestand aus 23 Fragen. Im ersten Teil wurden Fragen zum Experiment gestellt. Es ging um die Wahrnehmung des Themas der Rede, der Geschwindigkeit, der Fachlichkeit des Wortschatzes, den allgemeinen Eindruck von der Funktion *Augmented Interpreter CAI v 1.1*, die Richtigkeit der Erkennung der Zahlen, welche Aspekte der Software ihre Dolmetschleistung verbessert bzw. verschlechtert haben und was besonders herausfordernd (in Bezug auf Setting, Rede, Software) war. Im zweiten Abschnitt wurden Fragen zur Software gestellt, nämlich zu ihren Kenntnissen über CAI-Tools, InterpretBank und ihrer Erfahrung damit. Im letzten Absatz wurden Fragen zu den Zahlen gestellt; wie sie normalerweise mit Zahlen in der Kabine umgehen und welche Strategien sie verwenden, um Zahlen zu dolmetschen. Der Fragebogen bestand aus Multiple-Choice, Single-Choice und offenen Fragen, während subjektive Wahrnehmungen anhand von fünfstelligen Skalen ausgedrückt werden konnten. Die Studienteilnehmerinnen wurden gebeten, den Fragebogen gleich nach der Dolmetschung auszufüllen.

4.3 Durchführung des Experiments

Für die Durchführung des Experiments wurden den Studienteilnehmerinnen unterschiedliche Zeitfenster zur Verfügung gestellt, um sich für das gewünschte Datum und Uhrzeit anzumelden. Es wurde entschieden, dass für jede Teilnehmerin 45 Minuten für die Durchführung des Experiments einberechnet werden mussten. Jede Teilnehmerin wurde zunächst über die Bedingungen des Experiments informiert. Anschließend wurde ihr die technische Ausstattung gezeigt. Danach erfolgten die Kalibrierung und Validierung des Eyetrackers, deren Dauer individuell verschieden war. Danach wurde das Video gestartet und die Dolmetschungen

aufgenommen. Nach dem Experiment wurden die Teilnehmerinnen gebeten, den Fragebogen umgehend auszufüllen und alle Anmerkungen schriftlich anzugeben. Sie hatten die Möglichkeit, die Antworten des Fragebogens auf Deutsch, Italienisch oder Englisch zu schreiben. Das Experiment fand an insgesamt vier Tagen statt: am 30. November (P1), am 4. Dezember (P2, P3), am 5. Dezember (P4-P8) und am 11. Dezember 2023 (P9). Am ersten Tag traten bei P1 besonders viele technische Probleme auf. Die Teilnehmerin trug eine Blaulichtfilterbrille, und die Kalibrierung des Eyetrackers konnte nicht durchgeführt werden. Obwohl der Eyetracker mehrmals angepasst wurde und unterschiedliche Positionen ausprobiert wurden, konnten keine hohen Genauigkeitswerte der Messungen erzielt werden. Es wurde trotzdem entschieden, das Experiment durchzuführen. Danach wurde das Experiment mit P1 auch ohne Brille wiederholt, obwohl dadurch die Fähigkeit von P1, die Zahlen auf dem Bildschirm zu lesen, stark eingeschränkt wurde. Beide Dolmetschleistungen wurden aufgenommen und beide Sets von Daten des Eyetrackers gespeichert. Nach diesem Experiment wurde festgestellt, dass Brillen mit Blaulichtfilter die Messung von Augenbewegungen sehr stark beeinträchtigen, wie in SR Research (2017: 55) erwähnt wird: “The corneal reflection may not be stable with all participants, particularly those wearing glasses with a heavy anti-reflection coating.” Um dieses Problem zu beheben, wurden die anderen Teilnehmerinnen darum gebeten, keine Brillen während des Experiments zu tragen. Nur P9 hat eine Brille mit entspiegelten Brillengläsern getragen. Da die Kalibrierung und Validierung mit dieser einwandfrei erfolgten, wurde das Experiment mit Brille durchgeführt. Bei den anderen Experimenten traten keine weiteren technischen Schwierigkeiten diesbezüglich auf. Zwei Ausnahmen galten für P7 und P8. Obwohl für beide die Genauigkeitswerte des Eyetrackers zufriedenstellend waren, wurde während des Experiments die Augenerkennung unterbrochen. Trotz eines Neustarts des Experiments von P7 wurden die oben erwähnten technischen Probleme nicht behoben und es wurde entschieden, die Aufnahmen beider Versuche des Eyetrackers zu speichern und das Experiment für beendet zu erklären. Ähnliches galt für P8, bei der von Beginn des Experiments an keine Augenbewegungen erkannt wurden. Mit P8 wurde das Experiment nicht wiederholt. Es sind also sechs erfolgreiche Sitzungen mit dem Eyetracker zu vermerken, während bei drei (P1, P7, P8) Probleme mit der Erkennung der Augenbewegungen entstanden. Insgesamt wurden elf Dolmetschaufnahmen gesammelt, da P1 und P7 das Experiment einmal wiederholt haben.

4.4 Vorgehensweise bei der Analyse der Dolmetschleistungen

Bevor die Ergebnisse dargestellt werden, werden nun die Kriterien präsentiert, die für die Analyse der Ergebnisse verwendet wurden. Es wurden zunächst Schwerpunkte in der Auswertung der Daten gesetzt, die für die Beantwortung der Forschungsfragen ausschlaggebend sind. Die Verarbeitung visueller Informationen aus ASR-Tools während des Simultandolmetschens wird in den nächsten Kapiteln mittels Eyetracking-Daten analysiert.

Im zweiten Teil der Analyse werden die Dolmetschleistungen der Studienteilnehmer*innen in Bezug auf die Genauigkeit der wiedergegebenen Zahlen untersucht und mit der Genauigkeit der vom ASR-Tool erkannten Zahlen und den Ergebnissen der Eyetracking-Daten zum Vergleich herangezogen. Um die Wiedergabe der Zahlen auszuwerten, müssen sie zunächst in Kategorien unterteilt werden. Anschließend werden Fehlerkategorien festgelegt, um die Dolmetschleistungen zu bewerten. Wie in Kapitel 1.3 ausführlich erklärt wurde, gibt es in der Literatur keine Einigkeit über die Methode, die für die Auswertung der Daten von Dolmetschleistungen verwendet werden soll. Aus diesem Grund wurden unterschiedliche Methoden kombiniert bzw. angepasst, um den Forschungszweck zu erfüllen.

Für die Differenzierung der Zahlen wurde die Entscheidung getroffen, sich auf die Analysemethode von Desmet et al. (2018) und von Defrancq & Fantinuoli (2021) zu stützen. Zahlen wurden je nach Anzahl der bedeutungstragenden Einheiten differenziert, und nicht nach der Anzahl der Ziffern (vgl. Desmet et al. 2018: 20). Zum Beispiel enthält 7.000 nur zwei bedeutungstragende Einheiten (sieben und Tausend), während 565 drei bedeutungstragende Einheiten enthält (fünf, sechs, fünf), und das, obwohl 565 eine kleinere Zahl als 7.000 ist. Keine Differenzierung wurde dagegen zwischen ganzen Zahlen, Jahreszahlen oder Dezimalzahlen vorgenommen. Elemente wie Tausend, Million und Prozent wurden als eine bedeutungstragende Einheit behandelt. Die insgesamt 30 Zahlen wurden also wie folgt in vier Kategorien unterteilt (Tab. 2).

Tab. 2: Kategorien für die Auswertung der Zahlen

Kategorien der Zahlen	**Im Text enthaltene Zahlen**
Kategorie 1	4 twice
Kategorie 2	82 110 305 1.000(x3) 2050(x2) 7.000 10.000 half a million

Kategorie 3	0,30% 68% 131 565 621 639 654 655 659 679 2019 2021 2022 2025
Kategorie 4	12.045 31st December 2022 5.51 million

Für die Fehlerkategorien der Zahlen wurde die Unterteilung von Desmet et al. (2018) verwendet, die in Anlehnung an Pinochi (2009) und Braun & Clarici (1996) erstellt wurde:

- Auslassung: Eine Zahl wurde ausgelassen oder durch allgemeine Ausdrücke ersetzt (z. B. ‚32' wird ‚mehrere').
- Annäherung: Die Größenordnung wurde korrekt wiedergegeben, aber die Zahl wurde auf- oder abgerundet (z. B. ‚32' wird ‚ca. 30').
- Lexikalischer Fehler: Die Größenordnung wurde korrekt wiedergegeben, aber einige Bestandteile wurden geändert (z. B. ‚32' wird ‚34').
- Transposition: Alle Bestandteile wurden korrekt wiedergegeben, aber ihre Reihenfolge ist falsch (z. B. ‚32' wird ‚23'). Das ist vor allem für Sprachkombinationen üblich, wo Einer und Zehner in unterschiedlicher Reihenfolge gesprochen werden.
- Syntaktischer Fehler: Die Größenordnung ist falsch, obwohl alle Bestandteile korrekt sind (z. B. ‚32' wird ‚320').
- Phonologischer Fehler: Der Fehler kann auf eine phonologische Verwechslung zurückgeführt werden (z. B. ‚15' und ‚50').
- Andere Fehler: Zu dieser Kategorie gehören Fehler, die keiner anderen Kategorie zuzuordnen sind oder die aus der Summe verschiedener Fehler bestehen (z. B. ‚32' wird ‚230') (vgl. Desmet et al. 2018: 21).

Für die Analyse wurden diese Kategorien wie folgt abgekürzt: Auslassung (AU), Annäherung (AN), Lexikalischer Fehler (LF), Transposition (TR), Syntaktischer Fehler (SF), Phonologischer Fehler (PF), Andere Fehler (AF). Als korrekt wurden genaue Wiedergaben der Zahlen sowie auch Dolmetschstrategien, die zu keinem Informationsverlust führen, betrachtet. Das gilt zum Beispiel für Jahreszahlen, die durch Ausdrücke wie ‚voriges Jahr' ersetzt werden konnten (vgl. Frittella 2023: 96). Korrekte Wiedergaben wurden als KW abgekürzt.

Es wurde für diese Studie außerdem die Entscheidung getroffen, nicht nur die Genauigkeit der wiedergegebenen Zahlen nach Fehlerkategorie zu messen, sondern auch zu berücksichtigen, ob sie korrekt in den Satzteil eingebettet wurden. Obwohl in einigen Studien erwähnt wurde, dass Zahlen korrekt wiedergegeben, aber falsch im Kontext verwendet wurden (z. B. Canali 2019; Fantinuoli & Pisani 2021), wurde nur in Frittella (2023) versucht, einen sogenannten kommunikativen Ansatz zu systematisieren. Die Problemauslöser wurden also in ihrem gesamten Kontext (auf Wort-, Satz-, Text-, Kontextebene und mit ihrer Funktion) betrachtet (vgl. Frittella 2023: 95). Da in der vorliegenden Studie nur Zahlen untersucht wurden, wurden die Kategorien der Analyse von Frittella (2023: 84) nicht übernommen. Für jede Zahl wurde aber berücksichtigt, ob ihr Referent korrekt wiedergegeben wurde. Für den Zweck der Analyse wurden folgende Referenztabellen erstellt. Ein Punktesystem wurde verwendet, um die Genauigkeit der Wiedergabe von Zahlwort und Referent auszuwerten (Tab. 3 und 4).

Tab. 3: Punktevergabe je nach Fehlerkategorie der Zahlen

Fehlerkategorien der Zahlen	**Punktevergabe**
Korrekte Wiedergabe (KW)	1
Annäherung (AN)	0,5
Auslassung (AU), Lexikalischer Fehler (LF), Transposition (TR), Syntaktischer Fehler (SF), Phonologischer Fehler (PF), Andere Fehler (AF)	0

Tab. 4: Punktevergabe je nach Wiedergabe des Referenten

Referent	**Punktevergabe**
Referent richtig wiedergegeben, oder Synonym mit selber Bedeutung	1
Referent ausgelassen oder falsch wiedergegeben	0

Alle Fehlerkategorien werden also als falsch bewertet. Es wurde aber entschieden, Annäherungen als teilweise richtig zu betrachten. Für die Referenten wurde nur berücksichtigt, ob sie richtig wiedergegeben wurden. Da es 30 Zahlen mit 30 Referenten gab, konnten pro Teilnehmerin maximal bis zu 60 Punkte erreicht werden.

5 Ergebnisse

Im vorliegenden Kapitel werden die Ergebnisse des durchgeführten Experiments dargestellt. Ziel der Untersuchung war der Einfluss von automatisch erkannten Zahlen auf die Wiedergabe von Zahlen und ihre Referenten und wie sich ein ASR-gestütztes CAI-Tool auf die visuelle Aufmerksamkeit während des Simultandolmetschens auswirken kann. Im ersten Teil des vorliegenden Kapitels werden die Ergebnisse für die Leistungen des Tools *Augmented Interpreter* in Bezug auf die Genauigkeit der Erkennung von Zahlen dargestellt und die Visualisierungsmerkmale der Zahlen im Tool präsentiert. Danach werden die einzelnen Dolmetschleistungen in Bezug auf die Wiedergabe der Zahlen und Referenten ausgewertet. Im zweiten Teil des Kapitels werden die Eyetracking-Daten verwendet, um die visuelle Aufmerksamkeit der Teilnehmer*innen während des Experiments in Bezug auf die erstellten Areas of Interest zu untersuchen. Zum Schluss werden die Ergebnisse der Befragung dargestellt und mit den Dolmetschleistungen und Eyetracking-Daten verglichen.

5.1 Ergebnisse der Dolmetschungen: Wiedergabe und Genauigkeit der Zahlen

Um den zweiten Teil der Forschungsfrage zu beantworten, nämlich welche Wirkung visuelle Inputs von automatisch erkannten Zahlen auf die Richtigkeit der gedolmetschten Zahlen haben, muss zunächst die Leistung von *Augmented Interpreter* bewertet werden. Während des Experiments wurde allen Teilnehmerinnen dasselbe Video von der Bildschirmaufnahme der vorgelesenen Ausgangsrede und der Software gezeigt. Das ermöglichte sowohl vergleichbare Versuchsbedingungen und vergleichbare Ergebnisse der Wiedergabe unter allen Teilnehmerinnen als auch die Verringerung der Wahrscheinlichkeit des Auftretens technischer Probleme. Für die Erstellung der Bildschirmaufnahme musste zum Beispiel das Video mehrmals neu gestartet werden, bevor die Zahlen vom Programm erkannt werden konnten. Während der Durchführung des Experiments mit Studienteilnehmerinnen traten hingegen keine technischen Probleme auf.

5.1.1 Auswertung der automatischen Erkennung von Zahlen

Um die Ergebnisse der Leistungen von *Augmented Interpreter* auszuwerten, wurden zunächst die Zahlen in der Ausgangsrede mit der Bildschirmaufnahme verglichen, um die Genauigkeitsrate des Erkennungssystems zu messen. Von den insgesamt 30 Zahlen, die im Ausgangstext enthalten waren, wurden nur 21 Zahlen ausgesprochen, bevor die Software aufhörte zu funktionieren. Wie im vorigen Kapitel erwähnt, muss die Software nach fünf Minuten neu gestartet werden, um die

automatische Erkennung fortzuführen, was absichtlich vermieden wurde. Die Software konnte also nur bis zu Minute 5′20″ Zahlen erkennen (Tab. 5).

Tab. 5: Vergleich zwischen den im Text enthaltenen Zahlen und den von der ASR angezeigten Zahlen

Zahlen im Ausgangstext	**ASR**		
2050	2.050	565	5 // 00 >> 560 >> 565
2022	2.022	1.000	1000 >> 1 // 000
110	110	1.000	1 // 000
305	305	679	679
7.000	7 // 000	655	655
	1	654	654
	1	639	639
	2	621	
4	4	569	
2019	2.019	68%	
10.000	10 // 000	twice	
1.000	1 // 000	2021	
half a million		2025	
31st December 2022	31 // 2.022	2050	
5.51 million	5.50 // 1_million	82	
0,30%	3% // 0	131	
12.045	12 // 054		

Es wurden 24 Zahlen von der Software gezeigt. Alle Zahlen ab Minute 5′20″ werden aus der Auswertung der Leistungen der Software ausgeschlossen, da sie erst nach der Unterbrechung von *Augmented Interpreter* vorgekamen. Außerdem wurden drei Zahlen, die nicht für die Auswertung der Leistungen mitberücksichtigt wurden, von der Software erkannt. Diese waren eins, eins und zwei und wurden aus folgendem Satz erkannt: “Car sharing is an alternative to enjoy the advantages of driving a car, without having to own *one*. It can be more cost-effective than owning a car, because users share the price of fuel, and it reduces the number of cars on the road. *One* of the most renowned companies in Austria is Car*2*go.” (vgl. Anhang I). Da es sich in diesem Fall um Indefinitpronomen bzw. um Teile von Eigennamen handelt, werden auch diese Zahlen aus der Analyse der Ergebnisse ausgeschlossen. Es werden daher insgesamt 21 Zahlen analysiert.

Aus der Auswertung ist ersichtlich, dass 20 von 21 Zahlen von der Software richtig erkannt wurden. Die einzige Zahl, die nicht erkannt wurde, war ‚half a million’. Dies entspricht einem Prozentsatz von ca. 95 %. Das Video, das den Studienteilnehmerinnen während des Experiments

zur Verfügung gestellt wurde, enthielt 20 von den insgesamt 30 Zahlen, die im Ausgangstext erwähnt wurden, sprich zwei Drittel der gesamten Zahlen.

An dieser Stelle ist es wichtig, folgendes anzumerken: obwohl alle Zahlen richtig erkannt wurden, wurden einige auf dem Bildschirm abweichend von der traditionellen Schreibweise angezeigt. Mehrere Zahlen wurden getrennt und die jeweils letzten Ziffern wurden in der nächsten Zeile abgebildet. Die Verrückung der Ziffern wird in Tab. 5 mit dem Symbol // dargestellt. Diese Visualisierungsform betraf Zahlen ab eintausend, darunter 1.000, 7.000, 10.000 und 12.054, aber auch 500, das als 5 // 00 angezeigt wurde. Auch die Zahl 0,3 % wurde als 3 % // 0 angezeigt, während 5,51 Millionen als 5.50 // 1_million angezeigt wurde. Die letzten zwei Beispiele zeigen, dass obwohl alle Ziffern richtig erkannt und angezeigt wurden, die Visualisierungsform teilweise sehr stark von der traditionellen Rechtschreibkonvention abwich. Außerdem kann ein Mangel an Einheitlichkeit beobachtet werden, da die Ziffern, die in der unteren Zeile abgebildet werden, beide die Anfangs- und Endziffern sein können, wie die Beispiele 3 % // 0 und 5.50 // 1_million zeigen (Abb. 14).

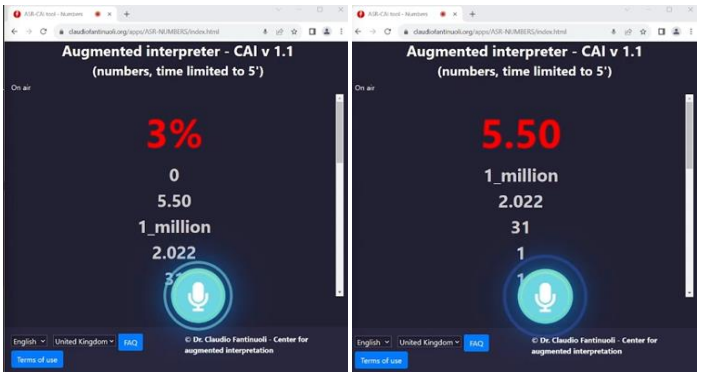

Abb. 14: Bildschirmaufnahme von Augmented Interpreter bei der Verrückung von Ziffern

Einige Zahlen haben sich außerdem geändert, nachdem sie angezeigt wurden. Das betraf nur zwei Zahlen, nämlich 565 und 1.000, und die Veränderung wird in Tab. 5 mit dem Symbol >> dargestellt. Die Zahl 1.000 wurde zunächst als 1000 und dann als 1 // 000 angezeigt. Die Zahl 565 wurde mehrmals auf dem Bildschirm geändert, und zwar zuerst 5 // 00, dann 560 und schließlich 565. Dieses Phänomen ist auf den Workflow der automatischen Spracherkennung zurückzuführen, die vorläufige Ergebnisse mit geringer Latenz zeigt, während sich der Satz noch entfaltet (vgl. Defrancq & Fantinuoli 2021: 6). Hier ist es wichtig anzumerken, dass die Zahl 565 außerdem absichtlich von der Rednerin falsch ausgesprochen und mit Unterbrechungen vorgelesen wurde.

Die letzte Abweichung von der konventionellen Rechtschreibung ist für Jahreszahlen zu beobachten. Diese wurden von der ASR nicht als Jahreszahlen erkannt und entsprechend mit einem Punkt als Tausendertrennzeichen dargestellt. Das betraf die Jahreszahlen 2050, 2022 und 2019, die als 2.050, 2.022 und 2.019 angezeigt wurden.

Zuletzt wurde auch die Latenz der Software berechnet. Die Aufnahme des Videos und Audios der Ausgangsrede wurden manuell mit Zeitstempeln markiert. Es wurde ein Zeitstempel sowohl nach vollständig ausgesprochener Zahl als auch nach vollständigem Transkribieren der Software hinzugefügt. Die durchschnittliche Zeitspanne lag bei ca. 0,70 Sekunden (siehe Anhang III).

5.1.2 Auswertung der Dolmetschleistungen

Nach der Darstellung der Ergebnisse der Zahlen, die von der ASR-Software erkannt und angezeigt wurden, werden in diesem Kapitel die Ergebnisse der Dolmetschleistungen ausgewertet. Diese wurden hinsichtlich der Wiedergabe von Zahlen und ihrer Referenten analysiert. Die Audioaufnahmen, die für jede Studienteilnehmerin realisiert wurden, wurden zunächst transkribiert. Die Kontrollsätze, die Zahlen und ihre Referenten enthielten, wurden in Excel-Listen transkribiert, um eine Übersicht über die wiedergegebenen Zahlen und Referenten zu ermöglichen. Zahlen wurden außerdem in die vier Kategorien aufgeteilt, die für die Auswertung der Zahlen beschrieben wurden (s. Kapitel 5.4). Abschließend wurden sie nach Fehlerkategorien aufgeteilt und sowohl für Zahlen als auch für ihre Referenten wurde die Punktevergabe vorgenommen. Eine Excel-Liste, die alle Merkmale enthält, wurde angelegt: Punkte und Referenten im Ausgangstext und pro Studienteilnehmerin, Kategorie der Zahlen, Fehlerkategorie der Zahlen und Punktevergabe der Zahlen und Referenten (siehe Anhang IV). Außerdem wurde für den Zweck der Übersichtlichkeit der Ergebnisse eine zweite Excel-Liste erstellt, welche nur die Fehlerkategorie und Punktevergabe für jede Studienteilnehmerin enthält (siehe Anhang V).

Es wurden die Ergebnisse von neun Teilnehmerinnen ausgewertet. Es wurde entschieden, auch die Ergebnisse von P1b und P7b miteinzubeziehen, die das Experiment aufgrund von technischen Problemen mit dem Eyetracker wiederholen mussten. Für die Auswertung der Richtigkeit der Wiedergabe wurden aber vorwiegend die Daten des ersten Versuchs (P1a und P7a) berücksichtigt, während die Ergebnisse des zweiten Versuchs nur dort berücksichtigt wurden, wo es ausdrücklich vermerkt ist. Im Allgemeinen wurden 30 Zahlen bei neun Teilnehmerinnen ausgewertet. Für die insgesamt 270 Zahlen, die während des Experiments gedolmetscht wurden,

konnten maximal 270 Punkte erreicht werden. Insgesamt wurden von den Teilnehmerinnen 179,5 von 270 Punkten erreicht. Diese bestehen aus korrekten Wiedergaben, die einen Punkt wert waren, und Annäherungen, die einen halben Punkt wert waren. Dies entspricht einem Prozentsatz von 66,5 % der Zahlen, die korrekt wiedergegeben wurden. Insgesamt wurden von den Teilnehmerinnen 174 Zahlen korrekt wiedergegeben, was einen Prozentanteil von 64,4 % entspricht, während elf Zahlen zur Kategorie Annäherung zugeordnet werden konnten. 26 % der Zahlen wurden ausgelassen, und insgesamt kam es zu sechs lexikalischen und acht syntaktischen Fehlern. Zwei Fehler wurden der Kategorie Anderer Fehler zugeordnet, während keine Fehler in den Kategorien Transposition und Phonologischer Fehler vorkamen (Abb. 15).

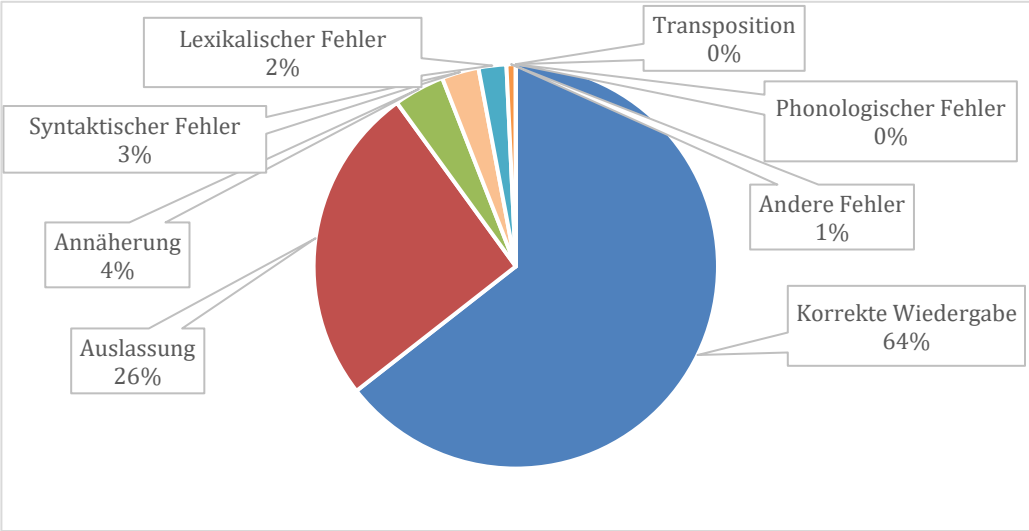

Abb. 15: Verteilung der Fehlerkategorien der Zahlen

Die Ergebnisse schwanken jedoch stark je nach Teilnehmerin (Abb. 16). Die höchste Punktezahl wurde von P5 erreicht, die 28 Zahlen und 30 Referenten richtig wiedergabt, also 96,7 % der maximalen Punktezahl, während die geringste Punktezahl von P3 verzeichnet wurde, die 14 Punkte in der Wiedergabe der Zahlen und 12 für die Referenten verzeichnete, was 41,7 % der maximalen Punktezahl entspricht. Der Mittelwert der gesamten Dolmetschleistungen liegt bei ca. 20 Zahlen und 22 Referenten.

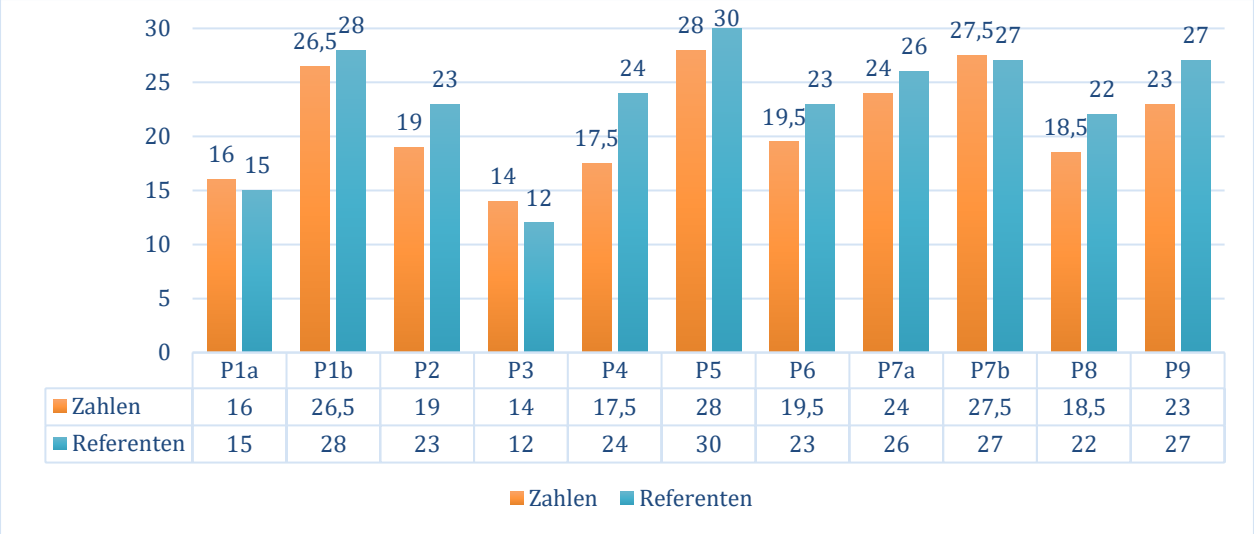

Abb. 16: Punktevergabe für Zahlen und Referenten pro Teilnehmerin

Die Verteilung der Fehlerkategorien pro Teilnehmerin hinsichtlich der Zahlen zeigt, dass die Ergebnisse je nach Teilnehmerin stark voneinander abweichen (Abb. 17). Die Anzahl der korrekten Wiedergaben liegt zwischen 13 (P3) und 27 (P5), und der Mittelwert liegt bei 19,3. Die Auslassungen liegen zwischen null (P5) und 15 (P3), und der Mittelwert liegt bei 7,7. Als Auslassungen wurden nicht nur die Fälle berechnet, in denen keine Erwähnung der Zahlen stattfand, sondern auch sehr allgemeine Ausdrücke. Diese waren zum Beispiel “mehr” (P2; P3; P4), “superano dei numeri incredibili” [überschreiten unglaubliche Zahlen] (P2), “un altro numero” [eine andere Zahl] (P6), “ancora di più” [noch mehr] (P7a), oder “un numero più grande” [eine höhere Zahl] (P8). Allgemein wurden elf Annäherungen verzeichnet, und der Mittelwert liegt bei 1,2. Beispiele für Dolmetschungen, die als Annäherung gekennzeichnet wurden, sind ‚mehr als 12.000’ (P2; P9) und 12.000 (P6) statt 12.045, 60 % (P2) oder mehr als 60 % (P5) statt 68 %, oder mehr als 300 (P4) statt 305. Sechs lexikalische Fehler wurden verzeichnet, wobei die Hälfte auf P2 zurückzuführen war, während acht semantische Fehler erfasst wurden, wobei P7a drei verzeichnete. Zwei Fehler, die keiner der oben genannten Kategorien zugeordnet werden konnten, wurden der Kategorie Andere Fehler zugeordnet. Diese waren 576 (P1a), die statt 621 gedolmetscht wurde, und “0,5 (...) 3%” (P2), die statt 0,3 % gedolmetscht wurde.

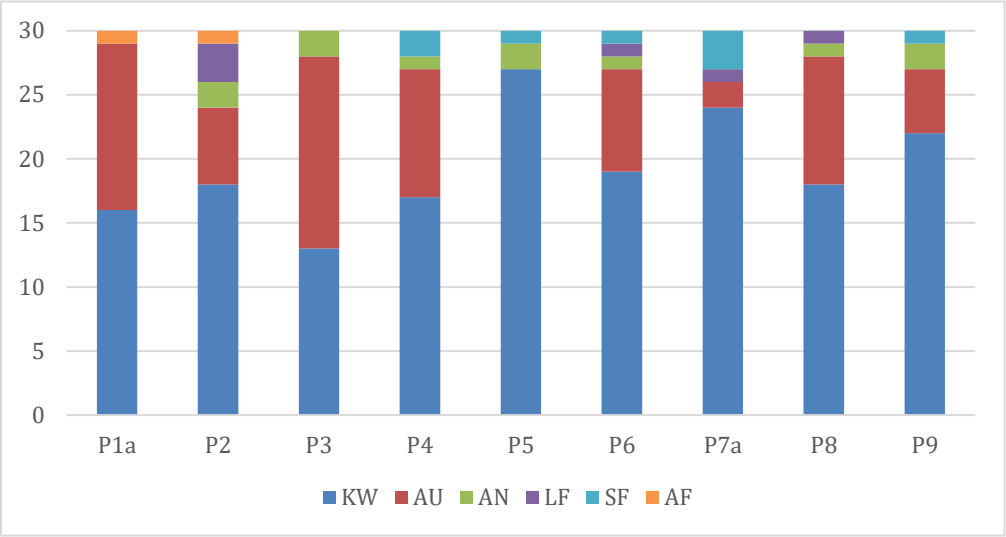

Abb. 17: Verteilung der Fehlerkategorien der Zahlen pro Teilnehmerin

Für die Referenten wurden hingegen keine Fehlerkategorien festgesetzt. Es wurden nur solche Referenzen als richtig bezeichnet, welche den Referenten im Ausgangstext genau entsprachen, und alle Varianten, die teilweise als korrekt angesehen werden konnten (z. B. Ober- und Unterbegriffe) wurden als falsch gewertet. Für ‚locations' wurden zum Beispiel folgende Übersetzungen akzeptiert: "posti" [Stellen] (P1a), "car sharing" (P4), "luoghi" [Orte] (P5, P6, P9), "locations" (P7a), aber nicht "aziende" [Unternehmen] (P2), "compagnie" [Unternehmen] (P3). Ein weiteres Beispiel ist ‚regular users': dafür wurden unterschiedliche Lösungen akzeptiert, wie zum Beispiel "utenti" [Benutzer] (P1a, P2, P5, P7a), "persone che la usano" [Menschen, die es verwenden] (P4), "persone che hanno usato" [Menschen, die es verwendet haben] (P8). Insgesamt wurden 202 Referenten als richtig gewertet.

Es ist erwähnenswert, dass aus den Ergebnissen eine weitere Kategorie von Fehlern erkannt werden kann, die zu keiner der vorher bestimmten Kategorien gehört. Für die englische Zahl ‚5.51 million' wurde von drei (P2, P4, P9) der neun Teilnehmerinnen und zusätzlich bei dem zweiten Versuch von P1b und P7b der Punkt als Trennzeichen für Dezimalzahlen übernommen, obwohl es auf Italienisch das Komma als Trennzeichen für Dezimalzahlen gibt. Dieses Phänomen kann auf eine Interferenz der Ausgangssprache zurückgeführt werden. Weitere Interferenzen, die während der Dolmetschungen vorkamen und gleich danach von den Teilnehmerinnen korrigiert wurden, waren "Siria, Stiria" [Syrien-Steiermark] (P1a), "Salzburg, Salisburgo" (P2) und "eighty, ottantadue" [*eighty*, zweiundachtzig] (P4). Da diese unmittelbar ausgebessert wurden, wurden sie für die Auswertung als richtig gewertet.

Danach wurde auch untersucht, ob es eine Beziehung zwischen den korrekt wiedergegebenen Zahlen und Referenten gibt. In 61 % der Fälle wurden beide, Zahl und Referent, richtig gedolmetscht, während in 18 % der Fälle beide falsch gedolmetscht oder ausgelassen wurden. In 21,1 % der Fälle wurden entweder die Zahl oder der Referent richtig gedolmetscht, während jeweils der andere falsch wiedergegeben bzw. ausgelassen wurde. Bei den insgesamt 179,5 Zahlen und 202 Referenten, die als korrekt eingestuft wurden, können außerdem starke Unterschiede in den Ergebnissen nach Zahlen beobachtet werden (Abb. 18).

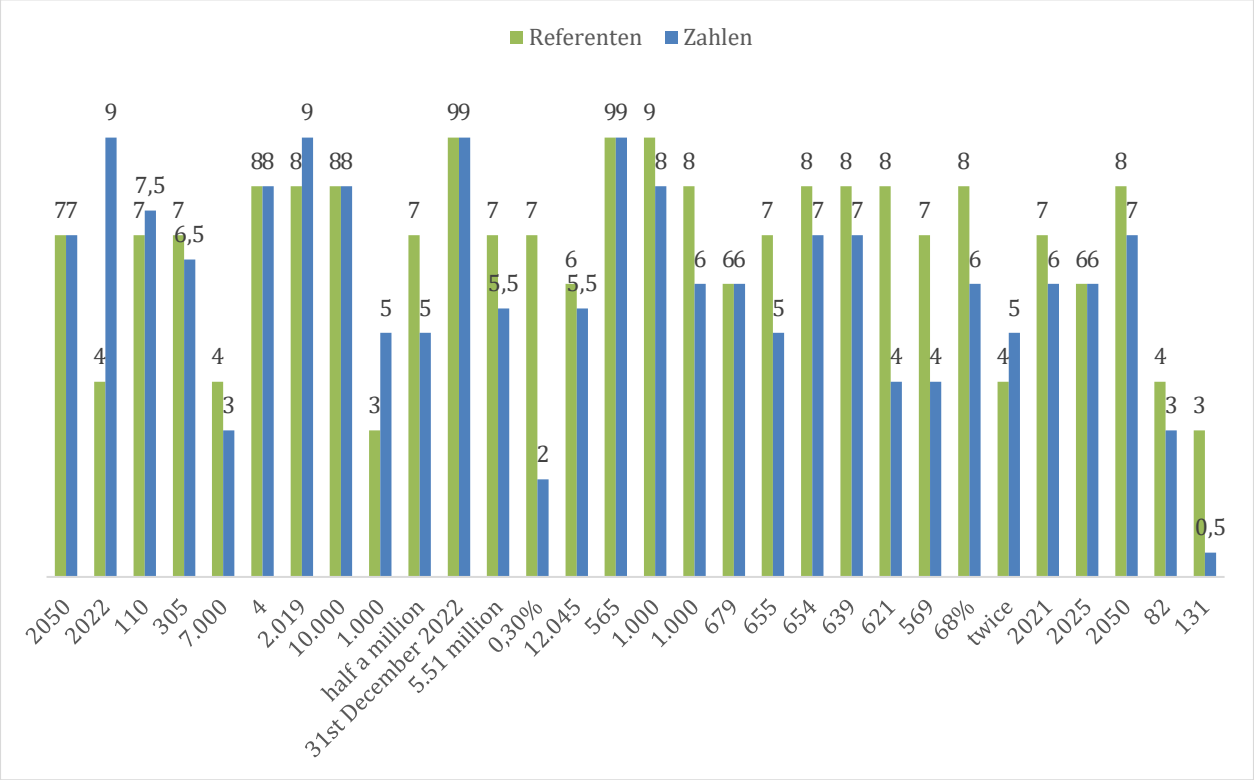

Abb. 18: Punktevergabe nach Zahl hinsichtlich der Wiedergabe von Zahlen und Referenten

Die Zahl mit der niedrigsten Punktezahl ist 131, die nur 0,5 Punkte erreichte. Danach folgen 0,03 %, die zwei Punkte, und 7.000, die drei Punkte verzeichnete. Hingegen wurden vier Zahlen von jeder Teilnehmerin korrekt wiedergegeben. Diese waren 2022, 2019, 31. Dezember 2022 und 565. Der Mittelwert der korrekt wiedergegebenen Zahlen liegt bei ca. 6. Die Referenten, die am häufigsten ausgelassen bzw. falsch wiedergegeben wurden, waren “e-smarts” und “bus lines” mit drei Punkten und “in Austria”, “vehicles” “bikes as cars” und “tramway lines” mit vier Punkten. Drei Referenten wurden hingegen von allen neun Teilnehmerinnen richtig gedolmetscht; diese waren “registered vehicles”, “cars” und “inhabitants”.

Die Ergebnisse für die Dolmetschleistungen nach Zahlen ermöglichen außerdem den Vergleich zwischen den Dolmetschleistungen vor und nach der Unterbrechung der ASR. Beim Aufstellen der Hypothesen wurde bereits vermutet, dass die Genauigkeitsrate der Dolmetschleistungen ab der Unterbrechung der Software sinkt, während gegen Ende der Rede mehr Zahlen bzw. Referenten erkannt werden. Die Ergebnisse zeigen, dass von den 189 Zahlen mit Referenten, die während des laufenden Betriebs der Software angezeigt wurden, 138 Zahlen und 147 Referenten richtig gedolmetscht wurden. Dies entspricht einem Anteil von 73 % der Zahlen und 78 % der Referenten. Die zwei Zahlen (621 und 569), die unmittelbar nach der Unterbrechung der Software erwähnt wurden, wurden in 44,4 % der Fälle richtig gedolmetscht, während ihre Referenten in 83,3 % der Fälle richtig gedolmetscht wurden. Ab der Unterbrechung der Software werden bis zum Ende des Videos neun Zahlen mit Referenten erwähnt, die nicht angezeigt wurden. Diese Zahlen wurden in 51,2 % der Fälle richtig gedolmetscht, während ihre Referenten einen Prozentanteil von 67,9 % erreichten.

Abschließend kann noch untersucht werden, ob es Unterschiede in der Wiedergabe nach Zahlenkategorien gibt. Zahlen wurden, je nach Anzahl der bedeutungstragenden Einheiten, in vier Kategorien unterteilt (s. Kapitel 5.4). Die Fehler in jeder Kategorie in Bezug auf die Gesamtanzahl der enthaltenen Zahlen zeigen, dass Kategorie 3 anfälliger für Fehler während der Dolmetschung ist. Für Kategorie 1 wurden die Zahlen in 72,2 % der Fälle richtig gedolmetscht. Für Kategorie 2, die zehn Zahlen enthielt, wurde ein Prozentanteil der richtig wiedergegebenen Zahlen von 73,3 % verzeichnet, während es für Kategorie 3, die 14 Zahlen enthielt, 63,9 % waren. Für Kategorie 4, die drei Zahlen erfasste, wurden diese in 74,1 % der Fälle richtig gedolmetscht. Da die Anzahl der Zahlen in jeder Kategorie unterschiedlich war, wurde außerdem die Fehlerquote im Verhältnis zur Anzahl der Elemente in jeder Kategorie berechnet. Diese verdeutlicht, welche Kategorien anfälliger für Fehler sind (Abb. 19).

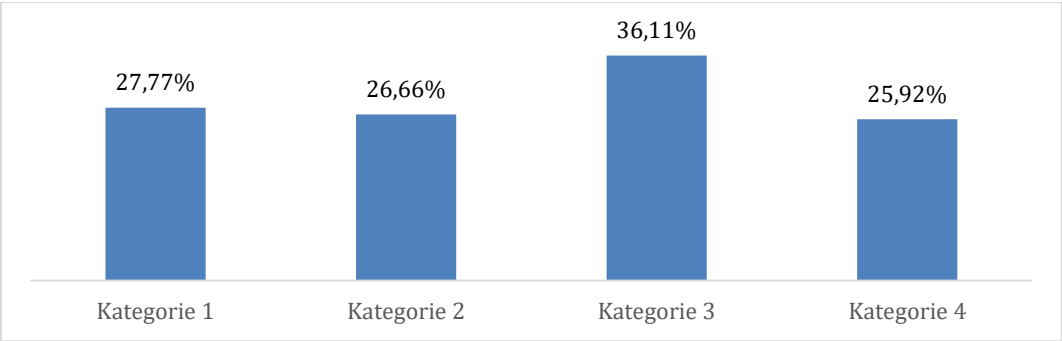

Abb. 19: Fehlerquote im Verhältnis zur Anzahl der Elemente pro Kategorie

5.2 Ergebnisse des Eyetrackers

Ziel der vorliegenden Arbeit war es, die visuelle Aufmerksamkeit von Simultandolmetscher*innen mit einem ASR-gestützten CAI-Tool zu messen. Multimodale Inputs wie die Ausgangsrede, das Video eines*r Redners*in und zusätzliche visuelle Reize aus einer Software spielen alle eine wichtige Rolle in der Verteilung der Aufmerksamkeit von Dolmetscher*innen. Obwohl Rückschlüsse auf die kognitive Aufmerksamkeit bzw. auf die Informationen, die in dem Moment verarbeitet werden, schwer zu ziehen sind, kann die Eyetracking-Methode Informationen über kognitive Prozesse liefern. Allgemein kann zum Beispiel behauptet werden, dass Problemauslöser während des Dolmetschens zu längeren Fixationen führen. Längere Fixationen sind also mit einer erhöhten kognitiven Belastung assoziiert, während unnatürlich kurze Fixationen ein Zeichen von Stress sein können (vgl. Stachowiak-Szymczak 2019: 68). Es ist jedoch wichtig, an dieser Stelle hervorzuheben, dass Eyetracking-Ergebnisse sehr individuell sein können. Augenbewegungen sind eine Verhaltensäußerung der visuellen Aufmerksamkeit und können von unterschiedlichen visuellen Merkmalen wie Kontrast, Farbe, Intensität, Helligkeit, Häufigkeit und Bewegung beeinflusst werden (vgl. Conklin et al. 2018: 115). Aus diesem Grund werden im folgenden Kapitel unterschiedliche Messungsindikatoren vorgestellt, die für die Auswertung der Eyetracking-Daten ausgewählt wurden. Diese dienen jedoch nur zur Veranschaulichung unterschiedlicher Phänomene wie zum Beispiel der Blickrichtung und des Blickverhaltens während des Dolmetschens und sind nicht als Indikatoren kognitiver Aufmerksamkeit bzw. kognitiver Belastung zu verstehen.

Für die Auswertung der Eyetracking-Daten wurde die Software *EyeLink Data Viewer* verwendet. Dieses Programm ermöglicht die Veranschaulichung, Ausfilterung und Verarbeitung von EDF-Dateien, die mit Eyetrackers der Marke EyeLink aufgezeichnet wurden (vgl. SR Research 2023: 1). Zunächst wurde der EyeLink Data Viewer verwendet, um Daten zu visualisieren. Die Animation Playback View zeigt die Blickdaten der Versuchsteilnehmer*innen

während der gesamten Aufzeichnung, um Einblicke in das visuelle Verhalten der Versuchspersonen zu ermöglichen. Wie schon während der Durchführung des Experiments festgestellt werden konnte, hat die Kalibrierung des Eyetrackers nicht bei allen Teilnehmerinnen funktioniert. Von den ursprünglichen neun Studienteilnehmerinnen wurden drei Teilnehmerinnen aus der Auswertungen ausgeschlossen (P1, P7 und P8). Die manuelle Inspektion dieser Daten hat gezeigt, dass die Eyetracking-Genauigkeit nicht hoch genug war. Hohe Unterschiede konnten zwischen Testreizen auf dem Bildschirm und der gemessenen Blickposition beobachtet werden.

5.2.1 Analyse der Blickhäufigkeit anhand einer Heatmap

Bevor die Eyetracking-Daten anhand von Zeitangaben und Areas of Interest (AOI) analysiert wurden, wurde eine sogenannte Heatmap erstellt. Diese Form der Datendarstellung zeigt die Verteilung der Blickdaten auf einer bestimmten Oberfläche. Wärmekarten sind eine vereinfachte Visualisierungsform von Eyetracking-Daten und müssen mit Vorsicht interpretiert werden, da sie zur Vereinfachung der Interpretation der Ergebnisse führen können. Sie zeigen nämlich die Blickposition aller Versuchsteilnehmer*innen während des gesamten Versuchs und bestehen aus einer Mischung unterschiedlicher Informationen über einen längeren Zeitraum (vgl. Holmqvist 2011: 238). In diesem Fall kann eine Heatmap verwendet werden, um das allgemeine Blickverhalten zu beobachten, ohne Rückschlüsse auf seine Ursachen zu ziehen. Eine Heatmap des gesamten Ablaufs des Experiments mit sechs Teilnehmerinnen wurde daher entsprechend erstellt und eine Bildschirmaufnahme wurde als Hintergrund ausgewählt, um anzuzeigen, welche Bereiche des Bildschirms der Karte entsprechen (Abb. 20).

Die Legende auf der rechten Seite des Bildes zeigt anhand einer Farbskala die Häufigkeit der Blicke für unterschiedliche Bereiche auf dem Bildschirm (Abb. 20). Die Häufigkeit wird anhand der Zeit gemessen, die die Versuchsteilnehmer*innen auf einen Bereich des Bildschirmes geblickt haben, und wird entsprechend in Millisekunden gemessen. Der rote Bereich zeigt die am häufigsten betrachteten Elemente. Auf diesem Bereich verweilte der Blick im Rahmen dieses Experiments bis zu 101490,93 Millisekunden (1,69 Minuten).

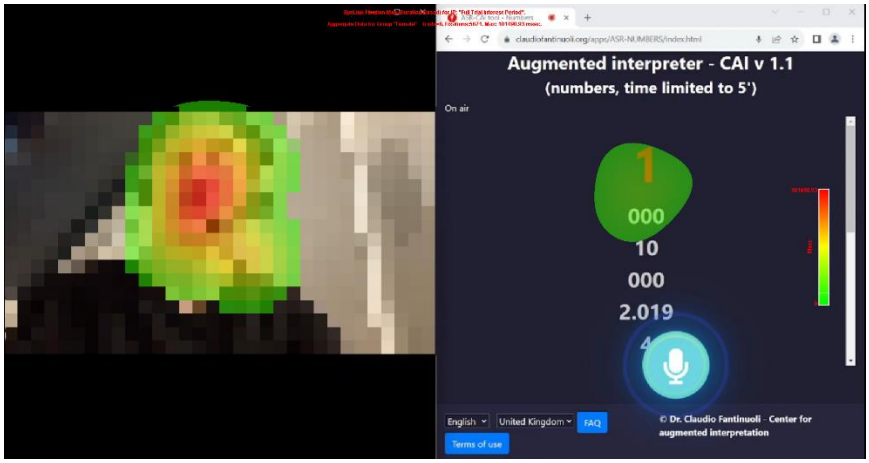

Abb. 20: Wärmekarte des gesamten Ablaufs des Experiments

In der ausgewerteten Heatmap sind die zwei auffälligsten Bereiche farblich markiert (Abb. 20). Bei der Interpretation der Heatmap gilt zu beachten, dass die Elemente auf dem Bildschirm während des Videos weitgehend statisch waren. Auf der linken Seite wurde das Video der Rednerin gezeigt, die sich während des Vortrags kaum bewegte. Auf der rechten Seite wurden Zahlen bis zur Unterbrechung der Software, also bis zu Minute 5′20″, angezeigt. Die zwei Bereiche, die auf der Karte markiert sind, sind das Gesicht der Rednerin, d.h. der obere Teil des Videobildes, und die letzten erschienen Zahlen. Der gekennzeichnete Bereich des Gesichts ist der einzige Bereich, der rot und gelb markiert wurde. Das zeigt, dass die Blicke der Studienteilnehmerinnen am häufigsten darauf verweilten. Auf der rechten Seite ist auch der Bereich grün markiert, in dem die neuen Zahlen erschien. Die grüne Markierung stellt den Bildschirmbereich dar, in dem die letzte Zahl erschienen ist, die auch rot markiert wurde und größer angezeigt wurde. Außerdem ist auch der Bereich markiert, wo die vorletzte Zahl angezeigt wurde. An dieser Stelle ist es ebenfalls wichtig anzumerken, dass einige Zahlen auf zwei Zeilen verteilt gezeigt wurden und die Hälfte der Zahl nach unten rutschte.

Heatmaps können auch als Orientierungshilfe für die Erstellung von AOI verwendet werden (vgl. Holmqvist 2011: 243). Sie zeigen, welche Bereiche für den*die Forscher*in von Interesse sein können, insbesondere wenn bisher nicht berücksichtigte Bereiche farbig markiert sind. In diesem Fall entsprechen die in der Heatmap markierten Bereiche den Elementen, die von vornherein als AOI ausgewählt wurden. Die Auswahl der AOI wird im nächsten Abschnitt ausführlich präsentiert und begründet, um die Grundlagen für die Analyse der Eyetracking-Daten zu schaffen.

5.2.2 Vorstellung der Areas of Interest

Die Bestimmung der Interessengebiete (AOI) wurde anhand der Hypothesen zur Verteilung der Aufmerksamkeit festgelegt. Es wurde entschieden, die rechte und linke Seite des Bildschirmes zu vergleichen, indem als AOI das Gesicht der Rednerin und die Zahlen ausgewählt wurden. Die AOI, welche im EyeLink Data Viewer als Interest Areas (abgekürzt IA) bezeichnet werden, wurden händisch erstellt. Es wurde entschieden, unterschiedliche AOI für die Zahlen zu zeichnen. Die AOI wurde, ähnlich wie in vorherigen Forschungsarbeiten, aktiviert, sobald eine Zahl auf dem Bildschirm erschien, und deaktiviert, wenn sie verschwand (vgl. Prandi 2022: 170). Das ermöglichte es, genauere Daten über die einzelnen visuellen Reize zu erhalten. Es wurde auch eine zusätzliche AOI für den letzten Teil ab Minute 5′20″ gezeichnet, welche die Zahlenerkennungssoftware enthält. Dies wurde gemacht, um überprüfen zu können, ob der Blick sich auch nach der Unterbrechung noch auf die Software richtete. Alle diese AOI wurden in Form eines Rechtecks gezeichnet (Abb. 21). Gleichzeitig wurden unterschiedliche AOI auch auf der linken Seite des Bildschirms festgelegt. Diese wurden auf das Gesicht der Rednerin und auf ihr Handzeichen der Zahl vier gezeichnet. Da die AOI auf das Gesicht von der AOI des Handzeichens unterbrochen wurde, da das Handzeichen vor dem Gesicht gezeigt wurde, mussten zwei AOI für das Gesicht (vor und nach dem Handzeichen) gezeichnet werden. Als letzte wurde eine AOI auf das Gesicht der Rednerin ab Minute 5′20″ gezeichnet, um einen besseren Vergleich der Verteilung der visuellen Aufmerksamkeit zwischen der Software und des Gesichts der Rednerin zu ermöglichen. Die AOI auf dem Gesicht der Rednerin wurde in Form einer Ellipse gezeichnet, um den Gesichtskonturen der Rednerin so nahe wie möglich zu kommen (Abb. 21).

Im Allgemeinen wurden also 29 AOI festgelegt. Diese enthalten alle Zahlen, die erwähnt wurden, bis die Software aufhörte zu funktionieren. Nur die Zahl ‚half a million' wurde nicht erkannt und dementsprechend nicht analysiert. Außerdem wurden die Zahlen eins und zwei auch als zwei AOI berücksichtigt. Diese wurden von der Software erkannt, obwohl sie nicht ursprünglich in der Gestaltung des Experiments als Zahlen aufgelistet wurden, da sie jeweils ein Indefinitpronomen und Teil von Eigennamen (Car2go) sind (s. Tabelle 4). Diese Areas of Interest wurden verwendet, um mit der EyeLink Data Viewer Software Berichte über die Dauer und Anzahl der Fixationen auf die AOI zu erstellen. Es wurde zunächst ein Gesamtbericht für alle AOI erstellt, welcher aus der Summe aller Eyetracking-Daten der sechs Versuchsteilnehmerinnen besteht.

Danach wurde ein Ergebnisbericht für jede Versuchsteilnehmerin erstellt. Er ermöglicht genaue Vergleiche zwischen den einzelnen Teilnehmerinnen.

Abb. 21: Areas of Interest (AOI) während der automatischen Erkennung von Zahlen (links) und nach der Unterbrechung der Software (rechts)

5.2.3 Blickverhalten während des Simultandolmetschens mit ASR

Anhand der erstellten AOI konnten unterschiedliche Messungsindikatoren verwendet werden, um das Blickverhalten während des Simultandolmetschens mit automatischer Erkennung der Zahlen zu untersuchen (siehe Anhang VI). Zunächst werden die Ergebnisse aus dem Gesamtbericht der Areas of Interest vorgestellt. Dieser besteht aus der Gesamtheit der erhobenen Daten und liefert einen Überblick über alle Ergebnisse der Versuchsteilnehmerinnen in Bezug auf die festgelegten AOI. Es wurde entschieden, zwei Messungsindikatoren für diese Untersuchung auszuwählen: die Dauer der Fixationen (auf Englisch: Dwell time) und die Anzahl der Fixationen (auf Englisch: Fixation count).

Zunächst wurde die Dauer der Fixationen analysiert. Diese besteht aus der Zeit (in Millisekunden), in den jede AOI betrachtet wurde (vgl. SR Research 2023: 78). Die durchschnittliche Zeit, in der die Teilnehmerinnen ihren Blick auf die AOI gerichtet hielten, wurde also berechnet. Die durchschnittliche Dauer der Fixationen (in Millisekunden) zeigt, dass unterschiedliche Zahlen zu unterschiedlichen Ergebnissen geführt haben (Abb. 22). Die Zahl, die durchschnittlich am längsten angesehen wurde, war ‚5,51 million' (durchschnittlich 2846 Millisekunden). Danach folgen 565 (2818,33 Millisekunden), 12.054 (2431,33 Millisekunden), 1.000 in dem Satz „*In Austria, there are 565 cars per 1.000 inhabitants*" (1796,67 Millisekunden) und 679 (1584,33 Millisekunden). Die niedrigsten Werte wurden für die Zahlen 1.000 in dem Satz

„*Burgenland has the most cars per 1,000 inhabitants*“ (138,33 Millisekunden), 2 von dem Namen Car2go (170,33 Millisekunden) und 1.000 in dem Satz „*of which 1.000 were e-smarts*“ (258 Millisekunden) verzeichnet.

Die AOI auf dem Handzeichen ermöglicht einen direkten Vergleich der durchschnittlichen Dauer der Fixationen auf die Zahl 4 und auf das entsprechende Handzeichen. Die Zahl verzeichnete eine durchschnittliche Dauer von 374,33 Millisekunden, während das Handzeichen eine durchschnittliche Dauer von 187,67 Millisekunden verzeichnete(Abb. 22). Das bedeutet, dass die AOI der Zahl ca. drei Mal länger als die AOI des Handzeichens betrachtet wurde.

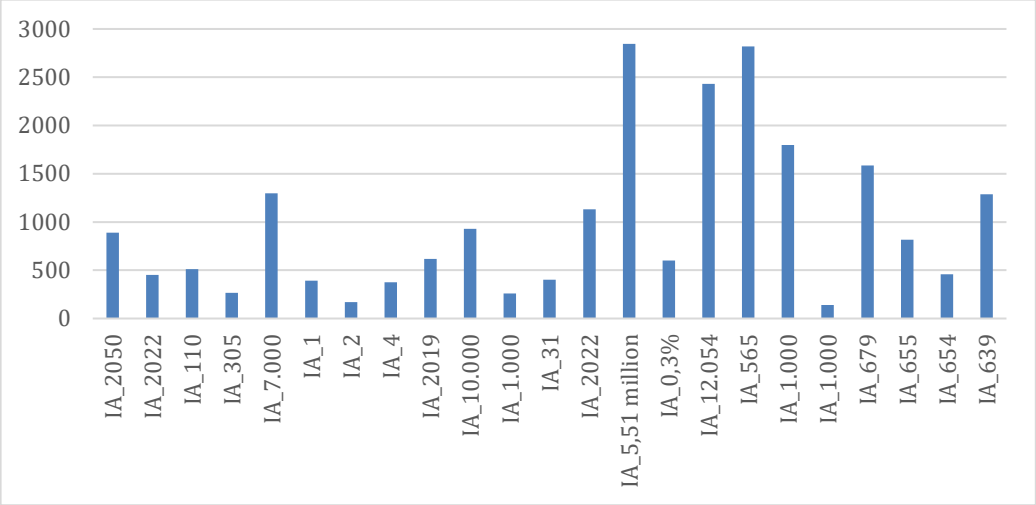

Abb. 22: Durchschnittliche Dauer der Fixationen (in Millisekunden)

Der letzte Teil des Experiments, nämlich jener ab dem Zeitpunkt der Unterbrechung der Software, erweist sich als besonders interessant. Zwei AOI wurden jeweils für das Gesicht der Rednerin und die Software erstellt, um die Verteilung der visuellen Aufmerksamkeit für diesen Teil des Experiments spezifisch zu untersuchen. Die durchschnittliche Dauer der zwei Bereiche zeigt, dass die AOI der Rednerin mit 104259,66 Millisekunden viel länger als die AOI der Software mit 8699,67 Millisekunden betrachtet wurde. Das bedeutet, dass die AOI der Rednerin ca. 12-mal länger betrachtet wurde als die AOI der Software. Obwohl diese Werte im Vergleich einen großen Unterschied aufweisen, zeigen diese Daten, dass Versuchsteilnehmerinnen auch ab der Unterbrechung der Software erwarteten, dass neue Zahlen erkannt werden. Im Zusammenhang mit diesem Aspekt ist es durchaus sinnvoll, eine weitere Erkenntnis in Bezug auf die Dauer der Erstfixation zu erwähnen. Diese entspricht der Dauer des ersten Blicks auf eine AOI, bevor es zum nächsten Element weitergeht. Diese wurde für alle Zahlen berechnet und belief sich

durchschnittlich auf 565,22 Millisekunden. Die durchschnittliche Dauer der Erstfixation auf die AOI der Software ab dem Zeitpunkt der Unterbrechung lag bei 871,67 Millisekunden. Diese war also in Vergleich zu den Zahlen viel höher. Das bedeutet, dass die Versuchsteilnehmerinnen zunächst länger auf die Zahl gewartet haben, bevor sie den Blick auf ein anderes Element richteten.

Die drei AOI, die für das Gesicht der Rednerin während der gesamten Durchführung des Experiments erstellt wurden, verzeichneten den größten Anteil der Dauer der Fixationen. Der Prozentsatz der Dauer der Fixationen auf die unterschiedlichen AOI zeigt, dass die Teilnehmerinnen durchschnittlich länger auf die AOI des Gesichts (vor und nach der Unterbrechung der Software) geschaut haben (Abb. 23). Der Prozentanteil der durchschnittlichen Dauer der Fixationen wird als die Summe der Dauer der Fixationen in den AOI berechnet, geteilt durch die gesamte Fixationsdauer der Versuche (vgl. SR Research 2023: 78). Der Bereich des Kreisdiagramms, der als „Sonstiges“ bezeichnet wurde, stellt hingegen jene Bereiche dar, die nicht als AOI markiert wurden. Diese enthalten unter anderem die Benutzerschnittstelle der Software, die viele visuelle Elemente enthält, so wie auch die Windows-Taskleiste am unteren Rand des Bildschirms und die Tab-Leiste mit dem offenen Browserfenster der ASR-Webseite. Diese wurden während des Experiments nicht ausgeblendet, um eine reale Situation zu simulieren. Diese Kategorie betrug ca. 29,5 % der gesamten Dauer der Fixationen und stellt den zweitgrößten Anteil dar. 70,52 % der durchschnittlichen Dauer der Fixationen liegt hingegen innerhalb der festgelegten AOI. Die Zahlen, die von der ASR erkannt wurden, betragen insgesamt ca. 5 % der gesamten Blickverteilung. Ab dem Zeitpunkt der Unterbrechung sinkt der Anteil der Dauer der Fixationen in der Software auf ca. 2 %. Wichtig ist an dieser Stelle anzumerken, dass die von der ASR erkannten Zahlen nur bei bestimmten Textstellen der Ausgangsrede vorkamen, während das Video der Rednerin während des gesamten Experiments angezeigt wurde. Die große Unterschiede in der prozentuellen Blickverteilung der Dauer der Fixationen können auch auf diese Tatsache zurückgeführt werden.

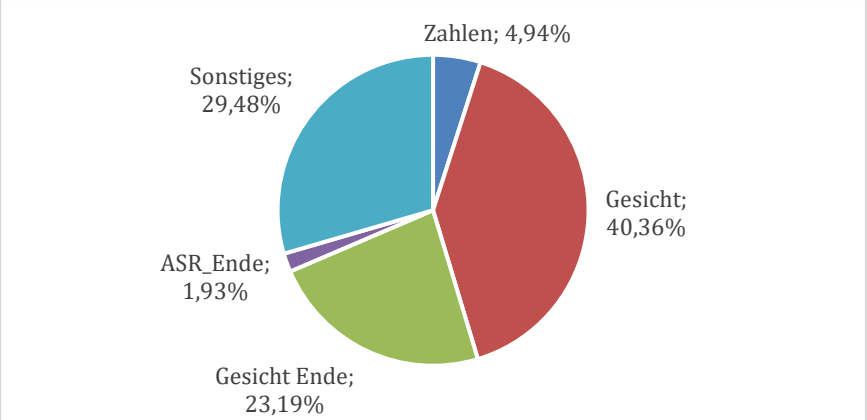

Abb. 23: Prozentuelle Blickverteilung der Dauer der Fixationen auf die AOI

Neben der Dauer der Fixationen kann auch die Anzahl der Fixationen untersucht werden (Abb. 24). Diese bietet eine andere Möglichkeit, die visuelle Aufmerksamkeit, die einer AOI geschenkt wird, zu betrachten (vgl. Conklin et al. 2018: 67). Dieser Messungsindikator bezieht sich auf die durchschnittliche Anzahl der Fixationen, die innerhalb aller Versuche von allen Versuchsteilnehmerinnen auf einer bestimmten AOI gemacht werden (vgl. SR Research 2023: 77). Die durchschnittliche Anzahl der Fixationen zeigt, dass die Zahlen, welche durchschnittlich mehr Fixationen verzeichneten, 12.054 (mit durchschnittlich 7,17 Fixationen), ‚5,51 million' (mit durchschnittlich 6 Fixationen), 565 (mit durchschnittlich 5,67 Fixationen) und 1.000 in dem Satz „*In Austria, there are 565 cars per 1.000 inhabitants*" (mit durchschnittlich 5,17 Fixationen) waren (Abb. 24). Die geringste durchschnittliche Anzahl der Fixationen wurde von der Zahl 1.000 in dem Satz „*Burgenland has the most cars per 1,000 inhabitants*" (mit durchschnittlich 0,33 Fixationen) und in dem Satz „*of which 1.000 were e-smarts*" (mit durchschnittlich 0,5 Fixationen) verzeichnet. Da die durchschnittlichen Fixationen für diese Zahlen geringer als 1 sind, lässt sich daraus ausschließen, dass nicht jede Teilnehmerin diese Zahlen direkt betrachtet hat, oder dass ihre Blicke nicht vom Eyetracker aufgezeichnet wurden. Die anderen Zahlen, die eine durchschnittliche Anzahl der Fixationen unter 1 ergaben, waren 31 (mit durchschnittlich 0,67 Fixationen), 2 (mit durchschnittlich 0,67 Fixationen), 645 (mit durchschnittlich 0,83 Fixationen) und 305 (mit durchschnittlich 0,83 Fixationen).

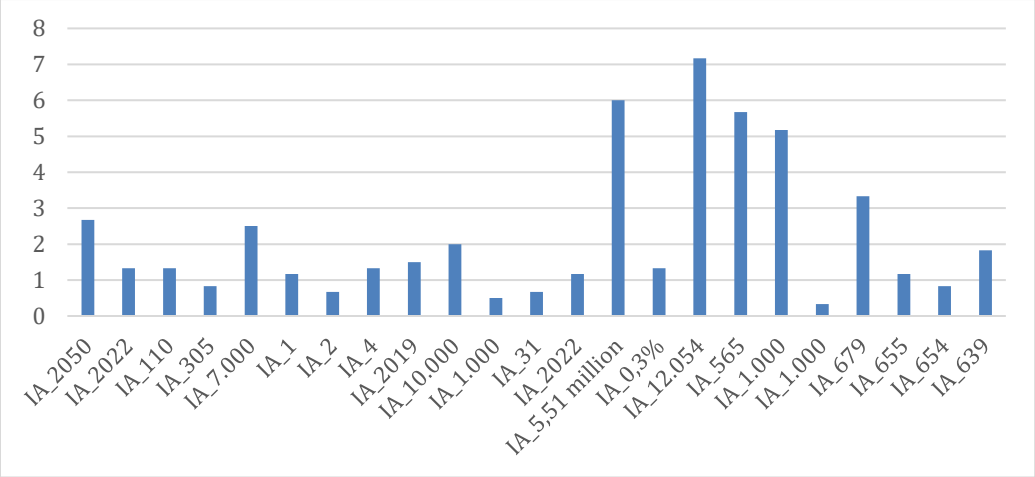

Abb. 24: Durchschnittliche Anzahl der Fixationen

Die AOI, die auf dem Handzeichen der Zahl 4 gestellt wurde, hat eine durchschnittliche Anzahl der Fixationen von 0,33 verzeichnet, währen die Zahl 4 auf der ASR-Software hat eine durchschnittliche Anzahl der Fixationen von 1,33 ergeben. Das bedeutet, dass die ASR-Software ca. viermal mehr Fixationen als das Handzeichen verzeichnet hat, während das Handzeichen nicht von allen Teilnehmer*innen betrachtet wurde.

Die Anzahl der Fixationen ermöglicht auch eine andere Perspektive auf das Blickverhalten während des letzten Teils des Vortrags. Ab dem Zeitpunkt der Unterbrechung wurden durchschnittlich 141 Fixationen für die AOI der Rednerin und 25,5 Fixationen für die AOI der Software verzeichnet. Wichtig ist an dieser Stelle anzumerken, dass insgesamt neun Zahlen ab diesem Zeitpunkt noch erwähnt wurden, ohne dass sie von der ASR-Software angezeigt wurden. Aus diesen Daten geht deutlich hervor, dass während der letzten Minuten des Videos die Teilnehmerinnen ihren Blick sehr oft bewegt haben, obwohl die Situation auf dem Bildschirm unverändert geblieben ist, und nur noch die Rednerin vorgetragen hat. Obwohl die AOI der Rednerin durchschnittlich fünfmal mehr als die AOI der Software betrachtet wurde, ist die Zahl der Fixationen für die AOI der Software nicht zu unterschätzen. Diese zeigen, dass die Versuchsteilnehmerinnen mehrmals ihren Blick auf die rechte Seite der Software gerichtet haben, während neue Zahlen nicht mehr angezeigt wurden. In diesem Zeitraum waren nur noch die letzten Zahlen (639, 654, 655, 679 und 1//000) zu sehen.

Die AOI auf dem Gesicht der Rednerin vor und nach dem Handzeichen haben jeweils durchschnittlich 118,67 und 137,67 Fixationen verzeichnet. Der durchschnittliche Prozentsatz der Anzahl der Fixationen liefert einen Überblick über die Verteilung der visuellen Aufmerksamkeit

anhand der Anzahl der Fixationen (Abb. 25). Der größte Prozentanteil ist die Kategorie „Sonstiges“. Das lässt schlussfolgern, dass die größte Anzahl der Fixationen nicht innerhalb der festgelegten AOI lag. Der zweitgrößte Prozentanteil ist auf der AOI des Gesichts mit 27,1 % zu verzeichnen. Die Summe des Prozentanteils aller AOI auf dem Gesicht der Rednerin ergibt einen gesamten Prozentanteil von 42 %. Auch in Bezug auf die durchschnittliche Anzahl der Fixationen ergeben sich die AOI des Gesichts der Rednerin als die am häufigsten betrachteten AOI. Danach folgen die AOI der Zahlen mit 5,3 % der durchschnittlichen Anzahl der Fixationen und der ASR-Software ab der Unterbrechung mit 2,7 %. Das führt dazu, dass die ASR-Software insgesamt einen durchschnittlichen Anteil der Fixationen von 8,0 % erreicht hat.

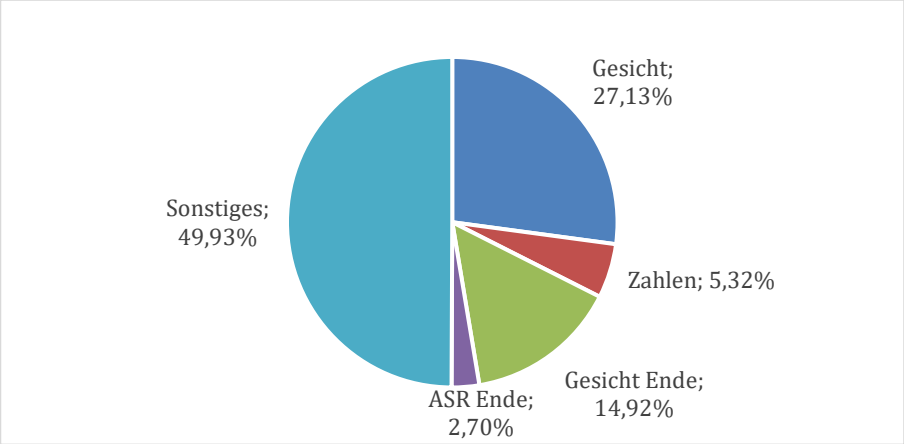

Abb. 25: Prozentuelle Blickverteilung der Anzahl der Fixationen

Nach der Auswertung der durchschnittlichen Dauer und Anzahl der Fixationen erweist es sich von besonderer Relevanz, die Eyetracking-Daten von zwei Versuchsteilnehmerinnen genauer zu untersuchen. Anhand der bis jetzt präsentierten Daten kann ein Überblick über das Blickverhalten aller Teilnehmerinnen geliefert werden. Um genauer die Wirkung einzelner Faktoren während der Durchführung des Experiments auf das Blickverhalten bzw. auf die Dolmetschleistung einzelner Studienteilnehmerinnen zu untersuchen, wurde der Entschluss gefasst, einen Ergebnisbericht für P3 und P4 zu erstellen und ihre Eyetracking-Daten mit ihren Dolmetschleistungen zu vergleichen.

5.2.3 Vergleich zweier Studienteilnehmerinnen anhand des Blickverhaltens und der Dolmetschleistungen während des Simultandolmetschens mit ASR

Da der Umfang der Eyetracking-Daten sehr groß ist, wurde eine ausführliche Analyse dieser Daten auf zwei Teilnehmerinnen beschränkt. Der Ergebnisbericht der AOI (auf Englisch: Output Report) wurde daher für P3 und P4 erstellt. Die Auswahl der Teilnehmerinnen war zufällig.

Zunächst wurde überprüft, ob die Kalibrierung des Eyetrackers für P3 und P4 erfolgreich war und ob die Validierungsergebnisse unter den Fehlergrenzen lagen. Ein Versuchsbericht (auf Englisch: Trial Report) wurde erstellt, um die Validierungsergebnisse für das linke und rechte Auge zu prüfen (Tab. 6). Für jedes Auge wurden die durchschnittlichen und maximalen Fehler berechnet, welche unter einem Grad liegen müssen (vgl. SR Research 2017: 61). In diesem Fall wurde dieser Wert nur von P3 für das linke Auge überschritten. Da die Kalibrierung vom System positiv beurteilt wurde, wurde die Auswertung der Daten weitergeführt.

Tab. 6: Validierungsergebnisse des Eyetrackers für P3 und P4

P3	HV13 LR LEFT: GOOD - ERROR: 0.47 avg. 1.33 max
	HV13 LR RIGHT: GOOD - ERROR: 0.33 avg. 0.99 max
P4	HV13 LR LEFT: GOOD - ERROR: 0.31 avg. 0.69 max
	HV13 LR RIGHT: GOOD - ERROR: 0.28 avg. 0.53 max

Der gesamte Ergebnisbericht der AOI ermöglicht den Zugang zu den unterschiedlichen Messungsindikatoren (siehe Anhang VII). Für die erstellten AOI wurden die Dauer und Anzahl der Fixationen berechnet. Diese wurden dann anhand der Ergebnisse der Dolmetschleistungen evaluiert, um zu überprüfen, ob ein Zusammenhang zwischen dem Blickverhalten und den Dolmetschungen besteht.

5.2.3.1 Analyse von P3 anhand des Blickverhaltens und der Dolmetschleistungen während des Simultandolmetschens mit ASR

Zunächst wurden die Eyetracking-Daten von P3 analysiert. Das Blickverhalten während des gesamten Experiments wurde anhand der prozentuellen Blickverteilung der Dauer der Fixationen auf die AOI untersucht (Abb. 26). Der höchste Wert wurde für die AOI auf dem Gesicht der Rednerin (vor und nach dem Handzeichen) verzeichnet. Danach folgte die AOI des Gesichts während der Unterbrechung der Software (17,41 %). Die von der ASR erkannten Zahlen zeigten einen Prozentanteil von 10,99 %, während die ASR-Software ab dem Zeitpunkt der Unterbrechung einen Prozentanteil von 5,54 % verzeichnete. 32,90 % der gesamten Dauer der Fixationen wurde in Bereichen außerhalb der AOI verzeichnet. Diese Daten entsprechen den Ergebnissen der durchschnittlichen Dauer der Fixationen aller Teilnehmerinnen (siehe Abb. 22).

Auch die prozentuelle Blickverteilung der Anzahl der Fixationen wurde für P3 berechnet (Abb. 27). Diese zeigt, dass knapp die Hälfte der Fixationen innerhalb der AOI des Gesichts der Rednerin lagen. Die AOI „Sonstiges“ verzeichnete 50,54 % der Gesamtdauer der Fixationen. 9,89 % der Fixationen wurden auf von der ASR erkannten Zahlen verzeichnet, während 7,18 % auf der ASR-Software nach der Unterbrechung stattfanden.

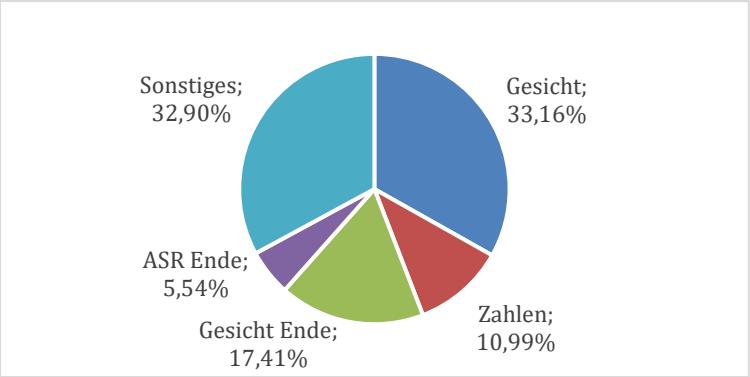

Abb. 26: Prozentuelle Blickverteilung der Dauer der Fixationen auf die AOI für P3

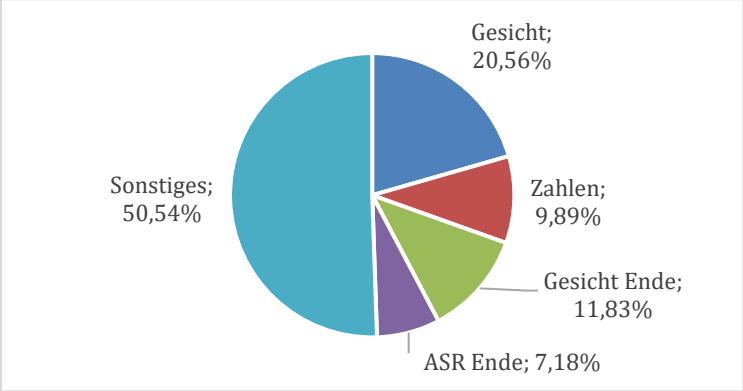

Abb. 27: Prozentuelle Blickverteilung der Anzahl der Fixationen für P3

Die Dauer der gesamten Fixationen und der Erstfixation zeigen die Zeit, die jeweils insgesamt und bei der Erstfixation auf einer AOI verbracht wurde (Abb. 28). Die AOI der Zahlen, die die längste Dauer der Fixationen verzeichnet haben, waren 12.045 (5582 Millisekunden), 1.000 im Satz „*In Austria, there are 565 cars per 1.000 inhabitants*“ (5228 Millisekunden), ‚5,51 million‘ (4716 Millisekunden), 565 (3806 Millisekunden) und 2022 (3100 Millisekunden). Hingegen verzeichneten die AOI folgender Zahlen die geringste Dauer der Fixationen: 1 (232 Millisekunden), 655 (290 Millisekunden), 2 (394 Millisekunden) und 305 (584 Millisekunden). Drei AOI verzeichnetenhingegen keine Fixation: 1.000 in den Sätzen „*of which 1.000 were e-smarts*“ und „*Burgenland has the most cars per 1,000 inhabitants*“ und das Handzeichen der Zahl

4. In Bezug auf die Dauer der Erstfixation sind folgende AOI mit einer höheren Dauer zu verzeichnen: 2022 (3100 Millisekunden), 639 (2558 Millisekunden) und 7.000 (1320 Millisekunden). Die niedrigste Dauer der Erstfixationen wurde für folgende AOI verzeichnet: 2 (160 Millisekunden), 1 (232 Millisekunden) und 110 (240 Millisekunden). Es muss jedoch berücksichtigt werde, dass die Dauer und Anzahl der Fixationen nur innerhalb des Versuchs von P3 evaluiert wurden: Infolgedessen wurde festgestellt, welche AOI zu höheren bzw. niedrigen Werten geführt haben. Diese Werte sind jedoch nicht als allgemein hoch bzw. niedrig zu bezeichnen, da für diese Differenzierung nur der Versuch von P3 ausgewertet wurde und kein Vergleich mit allgemeinen Eyetracking-Werten angestellt wurde.

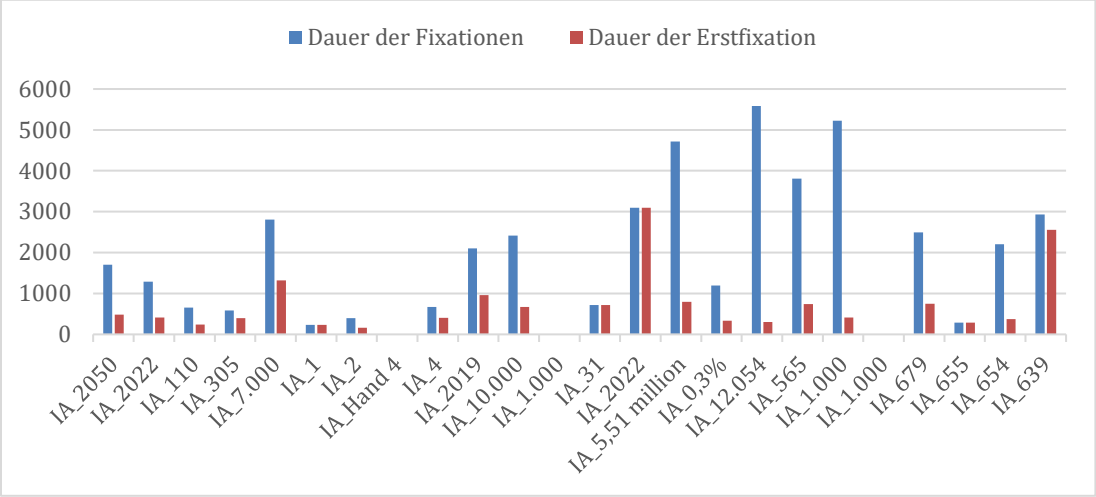

Abb. 28: Dauer der gesamten Fixationen und der Erstfixation (in Millisekunden) für P3

Wenn die Dauer der Erstfixationen bzw. der gesamten Fixationen auf einer AOI verglichen werden, lässt sich beobachten, dass bei einigen AOI diese Werte gleich sind, d.h. nur eine Fixation wurde verzeichnet. Für die Mehrheit der AOI wurden aber mehrere Fixationen verzeichnet. Wenn die drei AOI auf der Rednerin (mit insgesamt 334 Fixationen) und auf der Software ab dem Zeitpunkt der Unterbrechung (mit insgesamt 74 Fixationen) ausgeschlossen werden, da sie im Vergleich zu den AOI der Zahlen länger angezeigt wurden, lassen sich folgende Ergebnisse feststellen. Die AOI, die die höchste Zahl an Fixationen verzeichneten, waren 12.054 (18 Fixationen), 1.000 im Satz „*In Austria, there are 565 cars per 1.000 inhabitants*“ (17 Fixationen), 565 (10 Fixationen) und ‚5.51 million‘ (9 Fixationen). Die AOI der Zahlen, die nur eine Fixation verzeichneten, waren 1, 655, 31 und 2022.

Anhand dieser Daten über die Dauer und Anzahl der Fixationen kann ein Vergleich mit der Wiedergabe dieser Zahlen und ihren Referenten gezogen werden. Zur Veranschaulichung der Ergebnisse wurde eine Tabelle erstellt, welche jeweils die Zahlen mit der höchsten Dauer und Anzahl der Fixationen (und Erstfixationen) und mit den niedrigsten Werten darstellen. Zahlen, die die höchste Dauer der Fixationen bzw. Erstfixationen verzeichneten, stimmen teilweise mit den Zahlen mit der höchsten Anzahl der Fixationen überein, und umgekehrt gilt dies auch für die Zahlen mit den niedrigsten Werten. Nur die Zahl 2022 verzeichnete die höchste Dauer der Erstfixationen (3100 Millisekunden), obwohl nur eine Fixation auf der AOI verzeichnet wurde. Um den Vergleich zwischen den Eyetracking-Daten und den Dolmetschleistungen darzustellen, wurden in die Tabelle die Ergebnisse der Dolmetschleistungen in Form von Fehlerkategorien der Zahlen und korrekte bzw. falsche Wiedergabe des Kontexts hinzugefügt (Tab. 7).

Tab. 7: Vergleich zwischen den Eyetracking-Daten und der Dolmetschleistung von P3

Hohe Dauer und Anzahl der Fixationen	**Wiedergabe der Zahl**	**Wiedergabe des Kontexts**
12.045	KW	0
5,51 million	AN (5 milioni)	0
565	KW	1
2022*	KW	1
639	KW	1
7.000	AU	0
1.000 (in Austria)	AU	0
Niedrige Dauer und Anzahl der Fixationen	**Wiedergabe der Zahl**	**Wiedergabe des Kontexts**
1	-	-
2	-	-
655	KW	1
305	AU	1
1.000 (e-smarts)	AU	0
1.000 (Burgenland)	AU	1
110	AU	0
31/2022	KW	1

Allgemein kann festgestellt werden, dass die AOI der Zahlen mit der höchsten Dauer und Anzahl der Fixationen zu genaueren Wiedergaben der Zahlen geführt haben. Während in dieser Kategorie Zahlen eine gesamte Punktezahl von 4,5 erzielten, wurde in der Kategorie der Zahlen mit der

niedrigsten Dauer und Anzahl der Fixationen eine Punktezahl von 2 erreicht. Zu dieser Kategorie gehören jedoch auch die Zahlen 1 und 2, die von der ASR erkannt wurden, obwohl sie ursprünglich in der Gestaltung des Experiments nicht als Zahlen berücksichtigt wurden. Die Analyse der Transkription der Dolmetschleistung von P3 zeigt, dass der Satz, der diese zwei Zahlen enthielt, nicht gedolmetscht wurde: „Ci sono molte compagnie in Austria, per questo obiettivo. E molte." [Es gibt viele Unternehmen in Österreich, für dieses Ziel. Und viele.] [sic] (P3) (siehe Anhang VIII). Wenn die Zahlen eingebettet in ihren Kontext analysiert werden, kann beobachtet werden, dass vor allem Zahlen aus zwei Sätzen zur höheren Dauer und Anzahl der Fixationen geführt haben. Der erste Satz lautet: „Dagli ultimi dati, entro il 31 dicembre 2022, ci sono stati più di 5 milioni di registrazioni in Austria. E 12.054 in più dell'anno precedente" [Laut den neuesten Daten gab es bis zum 31. Dezember 2022 mehr als 5 Millionen Registrierungen in Österreich. Und 12.054 mehr als im Vorjahr] (P3). In diesem Satz bezogen sich die Zahlen 5 Millionen und 12.045 auf die Autos, die verzeichnet wurden. Der Kontext wurde aus diesem Grund für die zwei Zahlen als falsch eingestuft. Alle anderen Zahlen wurden korrekt mit ihrem Kontext wiedergegeben. Der zweite Satz lautet: „Ci sono 565 auto in Austria per abitanti" [In Österreich gibt es 565 Autos pro Einwohner*innen] [sic] (P3). In diesem Fall wurde die Zahl 565 richtig gedolmetscht, während 1.000 trotz der längeren Dauer der Fixation ausgelassen wurde. Das führte auch zu einem grammatikalischen Fehler im Satz, da die richtigen Satzstrukturen im Italienischen folgende sind: „per abitante" [pro Einwohner] und „per (Zahl) abitanti" [je (Zahl) Einwohner*innen]. Auch 7.000 wurde, trotz der längeren Dauer der Fixation, ausgelassen.

Die Zahlen, für die keine Fixationen verzeichnet wurden, waren 1.000 in den Sätzen „*of which 1.000 were e-smarts*" und „*Burgenland has the most cars per 1,000 inhabitants*". Diese wurden in der Dolmetschung ausgelassen. Der erste Satz wurde wie folgt gedolmetscht: „E dopo il 2019, queste compagnie hanno registrato più di 10.000 utilizzi di car sharing" [Und nach 2019 verzeichneten diese Unternehmen mehr als 10.000 Carsharing-Nutzungen] (P3). Der zweite Satz wurde hingegen wie folgt wiedergegeben: „Il Burgenland ha il numero più importante, il numero più alto di auto, ovvero 679 per abitanti" [Burgenland hat die größte Zahl, die höchste Zahl an Autos, also 679 pro Einwohner*innen] [sic] (P3). Aus der Analyse dieser Sätze kann beobachtet werden, dass im ersten Fall der Satzteil mit der Zahl 10.000 ausgelassen wurde, während im zweiten Satz nur die Zahl ausgelassen wurde, was zu einem grammatikalischen Fehler führte. Ein weiterer Satz enthielt mehrere Auslassungen. Der Satz „In 2022 in Austria there were already 110

locations in 305 cities with more than 7000 vehicles" wurde wie folgt gedolmetscht: "E entro il 2022 ci dovranno essere più compagnie per questo e molte più città" [Und bis 2022 müssen es dafür mehr Unternehmen und mehr Städte geben] (P3). Zwei Zahlen, die auch eine geringe Dauer der Fixationen verzeichneten, wurden ausgelassen: 305 (Gesamtdauer der Fixationen: 584 Millisekunden) und 110 (Dauer der Erstfixation: 240 Millisekunden). Auch die Zahl 7.000 wurde ausgelassen, obwohl sie eine höhere Dauer der Erstfixation verzeichnete (1320 Millisekunden). Die Zahl 655 und das Datum 31. Dezember 2022 wurden trotz der geringen Anzahl und Dauer der Fixationen richtig gedolmetscht und in den Kontext eingebettet.

5.2.3.2 Analyse von P4 anhand des Blickverhaltens und der Dolmetschleistungen während des Simultandolmetschens mit ASR

Nach der Analyse der Eyetracking-Daten von P3 wurden auch die Ergebnisse von P4 untersucht (siehe Anhang IX). Zunächst wurde die prozentuelle Verteilung der Dauer der Fixationen auf die AOI für P4 untersucht (Abb. 29). Diese Daten, die mit denen von P3 und dem Durchschnitt der allgemeinen Ergebnisse aller Studienteilnehmerinnen vergleichbar sind, zeigen, dass der größte Anteil der AOI des Gesichts (vor und nach der Unterbrechung der Software) entspricht (60,3 %). Für P4 ist der Prozentsatz der Dauer der Fixationen auf die AOI der Zahlen (7,9 %) höher als der Durchschnitt (4,9 %), sowie auch der Prozentsatz der AOI der ASR-Software ab der Unterbrechung (P4: 2,6 %, Durchschnitt: 1,93%). Die AOI „Sonstiges" verzeichnete 29,1 % der Gesamtdauer der Fixationen. Das bedeutet, dass 70,9 % der Gesamtdauer der Fixationen innerhalb der erstellten AOI verzeichnet wurden.

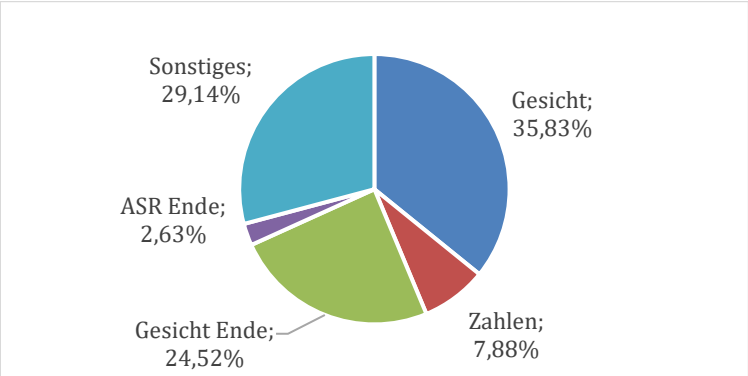

Abb. 29: Prozentuelle Blickverteilung der Dauer der Fixationen auf die AOI für P4

Die prozentuelle Blickverteilung der Anzahl der Fixationen veranschaulicht hingegen, dass, obwohl das Gesicht vor und nach der Unterbrechung der Software die AOI mit der höchsten Anzahl an Fixationen blieb (38,4 %), beinahe die Hälfte der aufgezeichneten Fixationen außerhalb der AOI stattfand (Sonstiges: 49,7 %) (Abb. 30). Auch die AOI der Zahlen ergaben eine hohe Anzahl an Fixationen (8,7 %), vor allem im Vergleich mit den durchschnittlichen Daten aller Studienteilnehmerinnen (5,3 %). Die Eyetracking-Daten zur Dauer und Anzahl der Fixationen von P4 weichen also nicht viel von den durchschnittlichen Ergebnissen aller Versuche ab.

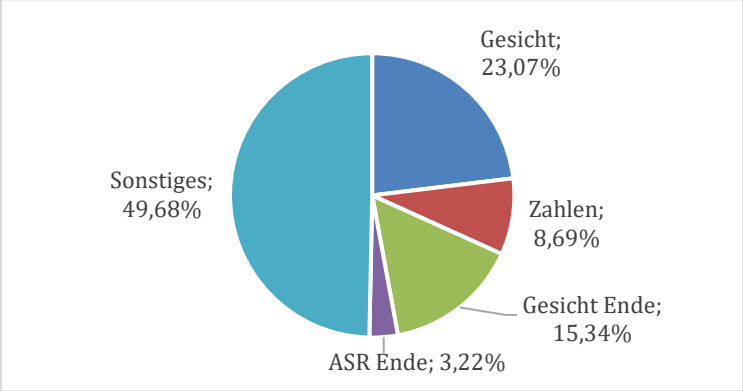

Abb. 30: Prozentuelle Blickverteilung der Anzahl der Fixationen für P4

Da die AOI der Zahlen in Bezug auf die Dauer und Anzahl der Fixationen einen geringen Anteil darstellen – und das nicht zuletzt aufgrund der Tatsache, dass sie im Verhältnis zu den anderen AOI viel weniger auf dem Bildschirm angezeigt wurden – erscheint es von großer Relevanz, eine gezielte Analyse der Eyetracking-Daten in Bezug auf die AOI der Zahlen durchzuführen. Diese ermöglicht einen direkten Vergleich zwischen den unterschiedlichen Zahlen, die mit ihrer Wiedergabe in den Dolmetschleistungen verglichen werden. Ein Diagramm wurde für die Dauer der gesamten Fixationen und der Erstfixation auf die AOI erstellt (Abb. 31).

Die grafische Darstellung der Eyetracking-Daten zeigt große Unterschiede zwischen den einzelnen AOI. Für P4 betrug die Dauer der Fixationen durchschnittlich 1.565,5 Millisekunden und die Dauer der Erstfixation durchschnittlich 499,1 Millisekunden. Anhand der Grafik kann jedoch festgestellt werden, dass die Werte für die einzelnen Zahlen stark voneinander abweichen. Während einige AOI längere Fixationen zeigen, wurde für andere AOI keine Fixation verzeichnet. Sechs AOI verzeichneten keine Fixationen. Diese waren 110, 305, 7.000, 1, 654 und das Handzeichen der Zahl 4. In Bezug auf die Dauer der Fixationen zeigten folgende Zahlen die geringste Dauer der Fixationen: 4 (424 Millisekunden), 31 (442 Millisekunden) und 0,3 % (486

Millisekunden). Die kürzesten Erstfixationen wurden für folgende Zahlen verzeichnet: 639 (222 Millisekunden), 679 (232 Millisekunden) und 0,3 % (270 Millisekunden).

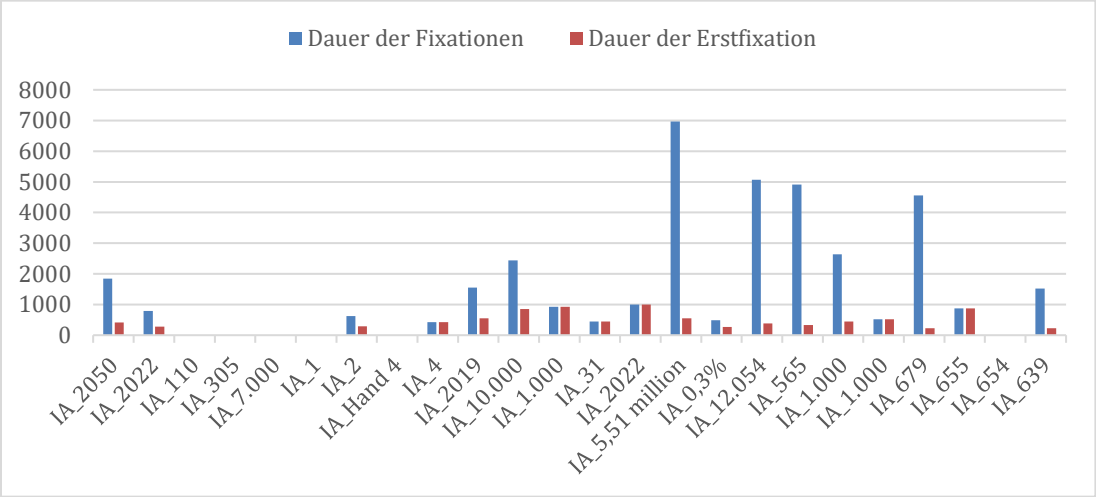

Abb. 31: Dauer der gesamten Fixationen und der Erstfixation (in Millisekunden) für P4

Die Zahl, welche die längste Dauer der Fixationen verzeichnete, war ‚5.51 million' mit 6.966 Millisekunden. Danach folgten 12.054 mit 5064 Millisekunden, 565 mit 4.908 Millisekunden und 679 mit 4.562 Millisekunden. Diese Zahlen verzeichneten also Fixationen, die deutlich länger als der Durchschnittswert in der Leistung von P4 waren. Über dem Durchschnitt liegen auch 1.000 im Satz „*In Austria, there are 565 cars per 1.000 inhabitants*" (2.634 Millisekunden), 10.000 (2.440 Millisekunden) und 2050 (1.844 Millisekunden). Wenn es um die Dauer der Erstfixation geht, verzeichneten folgende Zahlen die höchste Dauer: 2022 (994 Millisekunden), 1.000 im Satz „*of which 1.000 were e-smarts*" (926 Millisekunden) und 655 (876 Millisekunden). Diese drei Werte entsprechen auch der Gesamtdauer der Fixationen für diese AOI der Zahlen.

Auch die Anzahl der Fixationen kann Auskunft über das Blickverhalten geben. Sechs AOI verzeichneten nur eine Fixation: 4, 31, 1.000 im Satz „*of which 1.000 were e-smarts*" und „*Burgenland has the most cars per 1,000 inhabitants*", 655 und 2022. Durchschnittlich wurden für P4 4,5 Fixationen auf jeder AOI verzeichnet. Die AOI, welche die größte Anzahl der Fixationen aufwies, war ‚5.51 million' (14 Fixationen) und danach folgten 12.054 (13 Fixationen), 679 (12 Fixationen) und 565 (9 Fixationen).

Die Eyetracking-Daten zur Dauer und Anzahl der Fixationen können mit der Wiedergabe der Zahlen bzw. ihrem Kontext verglichen werden. Zu diesem Zweck wurde eine Tabelle erstellt, welche die Zahlen mit der höchsten Dauer und Anzahl der Fixationen (und Erstfixationen) und mit

den niedrigsten Werten vergleicht (Tab. 8). Zur Kategorie der AOI mit hoher Dauer und Anzahl der Fixationen wurden drei Zahlen hinzugefügt, welche nur eine Fixation verzeichneten. Da diese drei Zahlen die längste Dauer der Erstfixation verzeichneten, wurden sie zur dieser Kategorie in die Tabelle hinzugefügt und mit einem Sternzeichen (*) markiert.

Tab. 8: Vergleich zwischen den Eyetracking-Daten und der Dolmetschleistung von P4

Hohe Dauer und Anzahl der Fixationen	**Wiedergabe der Zahl**	**Wiedergabe des Kontexts**
5.51 million	KW	0
12.054	KW	1
565	KW	1
2022*	KW	0
1.000* e-smarts	KW	1
655*	AU	1
679	AU	0
Niedrige Dauer und Anzahl der Fixationen	**Wiedergabe der Zahl**	**Wiedergabe des Kontexts**
110	KW	1
305	AN	1
7.000	AU	0
1	*KW*	*1*
654	AU	1
4	KW	1
31	KW	1
0,3 %	SF (3 %)	1
639	AU	1

Allgemein kann beobachtet werden, dass die sieben Zahlen der ersten Kategorie (hohe Dauer und Anzahl der Fixationen) eine Punktezahl von 5 in Bezug auf die Genauigkeit ihrer Wiedergabe erreichten, während die neun Zahlen der zweiten Kategorie (keine Fixation oder niedrige Dauer und Anzahl der Fixationen) eine Punktezahl von 4,5 erreichten, obwohl zwei Zahlen mehr in dieser Kategorie sind.

Die Zahlen, die der ersten Kategorie zugeordnet wurden, wurden in den meisten Fällen richtig mit ihrem Kontext wiedergegeben. Die Zahl ‚5.51 million', welche die längste Dauer der Fixationen und die größte Anzahl der Fixationen erhielt, wurde wie folgt gedolmetscht: „I dati dicono che il 31 dicembre 2022 cinque punto cinque uno milioni di persone" [Die Zahlen zeigen, dass am 31. Dezember 2022 fünf punkt fünf ein Millionen Menschen] [sic] (P4) (siehe Anhang X).

In diesem Fall wurde der Kontext als falsch ausgewertet, da es in der Originalrede um Autos statt Menschen ging: „On the 31st December 2022, 5.51 million vehicles were registered in Austria“. Außerdem kann in der Transkription der Dolmetschung beobachtet werden, dass die Wiedergabe der Zahl, die als korrekt gewertet wurde, Schwierigkeiten für P4 bereitete. Der Punkt wurde als Trennzeichen für Dezimalzahlen übernommen, obwohl im Italienischen das Komma als Trennzeichen für Dezimalzahlen gilt. Außerdem wurde die Zahl nicht als „cinquantuno“ [einundfünfzig] ausgesprochen, sondern als einzelne Ziffern 5 und 1. Die Zahl 655 verzeichnete nur eine Fixation, die aber eine der längsten Erstfixationen war. Die Zahl 679 verzeichnete 12 Fixationen und eine der längsten Gesamtdauer der Fixationen. Diese zwei Zahlen, die im selben Satz enthalten sind („Burgenland has the most cars per 1,000 inhabitants (679) and the highest level of motorization, followed by Lower Austria (655)”) wurden beide von P4 ausgelassen. Wichtig ist an dieser Stelle anzumerken, dass diese Zahlen in der Auflistung der Zahlen und der Bundesländer enthalten waren. P4 hat die Auflistung wie folgt gedolmetscht: „Ci sono tantissime macchine per esempio in Carinzia, in Austria Bassa, in Austria Alta in Stiria e Salisburgo sono sempre circa 600 per mille abitanti” [Es gibt viele Autos, z.B. in Kärnten, Niederösterreich, Oberösterreich, in der Steiermark und Salzburg sind es immer um die 600 pro tausend Einwohner] (P4). Die Transkription zeigt, dass alle Zahlen ausgelassen wurden, und nur die Namen der Bundesländer wiedergegeben wurden.

Von den fünf Zahlen, die keine Fixation verzeichneten und in der Tabelle (Tab. 8) grau markiert wurden, wurden drei ausgelassen und zwei korrekt wiedergegeben. Im Satz „In 2022 in Austria there were already 110 locations in 305 cities with more than 7000 vehicles” wurde die Zahl 110 korrekt wiedergegeben, während 300 abgerundet und 7.000 ausgelassen wurde: „Nel 2022 c'erano già 110 car sharing in tantiss- più di 300 città“ [2022 gab es bereits 110 Car-Sharing-Angebote in sehr viel- in mehr als 300 Städten] (P4). Die Zahl 1 erhielt keine Fixation. Trotzdem wurde sie im Satz korrekt wiedergegeben. Sie wurde in der Tabelle berücksichtigt, obwohl sie als Indefinitpronomen als Ausnahme gilt. Von den restlichen Zahlen, die eine niedrige Dauer oder Anzahl der Fixationen verzeichneten, wurden zwei korrekt in ihrem Kontext wiedergegeben (4 und 31). Die Zahl 0,3 % wurde syntaktisch falsch wiedergegeben, nämlich als 3 %. Die Zahl 639, welche die kürzeste Erstfixation verzeichnete, wurde in der Auflistung der Bundesländer ausgelassen, so wie auch die anderen Zahlen in diesem Satz.

Die Ergebnisse des Experiments mit dem Eyetracker stellen die Verteilung des Blickverhaltens während des Simultandolmetschens auf die unterschiedlichen AOI dar. Die grafischen Darstellungen ermöglichen eine präzise Veranschaulichung der Verteilung der visuellen Aufmerksamkeit auf das Gesicht der Rednerin und auf die von der ASR erkannten Zahlen, die auf dem Bildschirm angezeigt wurden. Eine ausführliche Analyse zweier Dolmetschleistungen mit den entsprechenden Eyetracking-Daten wurde außerdem verwendet, um mögliche Zusammenhänge zwischen der Wiedergabe der Zahlen und dem Blickverhalten der Studienteilnehmerinnen herzustellen. Da mehrere Faktoren während der Durchführung des Experiments das Blickverhalten der Teilnehmerinnen und ihre Dolmetschleistungen beeinflussen konnten, wurde jede Teilnehmerin am Ende des Versuchs gebeten, einen Fragebogen auszufüllen. Damit konnten die persönlichen Meinungen und Einschätzungen der Teilnehmerinnen erhoben werden, um zu untersuchen, ob die Ausgangsrede, die Sprachkombination, das Setting des Experiments oder andere Faktoren besondere Schwierigkeiten für die Teilnehmerinnen darstellten.

5.3 Ergebnisse des Fragebogens

Im folgenden Kapitel werden die Ergebnisse der Befragung präsentiert. Zur Erstellung des Online-Fragebogens wurde die Funktion Google Formulare verwendet, und der dafür erstellte Link wurde den Studienteilnehmerinnen im Anschluss an das Experiment zugeschickt. Es wurden 23 Fragen in Form von offenen Fragen, Bewertungen mit einer Skala von eins bis fünf, Single- und Multiple-Choice Fragen verwendet (siehe Anhang XI). Antworten wurden auf Italienisch und Deutsch verfasst. Relevante Ergebnisse des Fragebogens werden in Folge vorgestellt.

Die ersten Fragen des Fragebogens befassten sich mit allgemeinen Eindrücken in Bezug auf die gedolmetschte Rede: das ausgewählte Thema, die Redegeschwindigkeit und die Fachlichkeit des Fachwortschatzes. Diese Fragen dienten zur Auswertung der Wahrnehmungen der Studienteilnehmerinnen. Das Thema der Rede wurde von der Mehrheit der Studienteilnehmerinnen als „Sehr bekannt“ bzw. „Bekannt“ bewertet. Was die Geschwindigkeit der Rede betrifft, so stimmten die meisten Studienteilnehmerinnen in einer Skala von “Sehr schnell” bis “Sehr langsam” für die mittlere Option. Neben der Geschwindigkeit der Rede und der Bekanntheit des Themas wurde auch die Fachlichkeit des Wortschatzes evaluiert, da diese auch ein potenzieller Problemauslöser sein kann, der sich auf die gesamte Dolmetschleistung auswirkt. Die Fachlichkeit wurde anhand einer Skala von 1 „Sehr fachspezifisch“ bis 5 „Nicht fachspezifisch“ evaluiert. Ein

Mittelwert von 3,66 wurde verzeichnet. Das bedeutet, dass die Rede als eher nicht fachspezifisch empfunden wurde (Abb. 32).

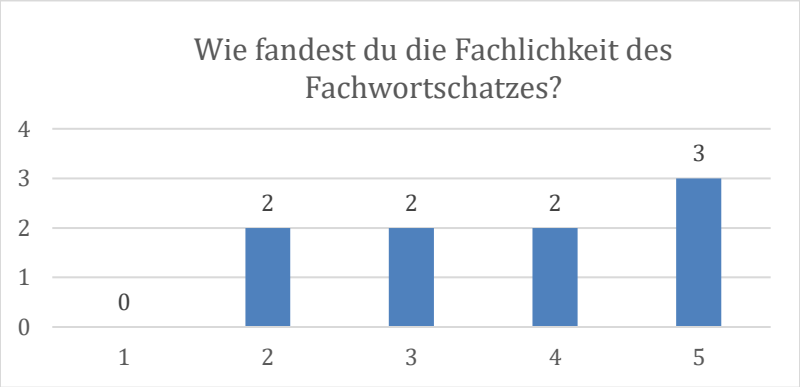

Abb. 32: Frage 3

Für die Frage, welche Wirkung *Augmented Interpreter* auf die Dolmetschleistungen hatte, wurde eine Skala von 1 „Viel verbessert“ bis 5 „Viel verschlechtert“ verwendet. Ein Mittelwert von 2,5 wurde verzeichnet (Abb. 33). Die offenen Antworten zur Erklärung der Aspekte, welche die Leistungen verbessert bzw. verschlechtert haben, waren verschieden. Es wurde erwähnt, dass die Software für dreistellige Zahlen (P5) bzw. zwei- bis vierstellige Zahlen (P4) hilfreich war. Außerdem gaben zwei Teilnehmerinnen an, dass die geringe Latenz ein positiver Aspekt war (P2 und P9). Zu den Aspekten, welche die Dolmetschleistungen verschlechtert haben, zählen die Unterbrechung der Software (P2, P6, P7) und die Trennung der Ziffern ab ein Tausend (P2, P3, P5, P7). Falsch angezeigte Zahlen wurden von P1 und P9 auch erwähnt. Weitere visuelle Störfaktoren, die erwähnt wurden, waren die Farbe der Zahlen und die vielen visuellen Inputs (P1) und der Punkt als Tausendertrennzeichen auch bei Jahreszahlen (P8).

Im folgenden Abschnitt wurden Fragen zu den Zahlen gestellt, zu ihrer Genauigkeit und ihrem Visualisierungsformat. Die Genauigkeit der Erkennung der Zahlen wurde auf einer Skala von 1 „Sehr genau“ bis 5 „Sehr ungenau“ evaluiert. Ein Mittelwert von 2,77 wurde verzeichnet (Abb. 34). Alle Studienteilnehmerinnen waren sich darüber einig, dass Zahlen nicht immer richtig erkannt wurden. Wenn Zahlen falsch angezeigt wurden, gaben Studienteilnehmerinnen an, dass sie die Zahlen auch falsch gedolmetscht (P1), sich auf das Video der Rednerin fokussiert (P2, P9), oder die Vorschläge ignoriert (P8) und sich auf das eigene Gedächtnis gestützt haben (P4, P8).

Die Zufriedenheit mit dem Visualisierungsformat wurde auf einer Skala von 1 „Ja, sehr zufrieden“ bis 5 „Nein, sehr unzufrieden“ mit einem Mittelwert von 2,5 bewertet (Abb. 35). Sieben Teilnehmerinnen (77,8 %) hätten sich kein anderes Format gewünscht, während zwei angaben, dass die Ziffern nicht abgetrennt werden sollten (P6, P7).

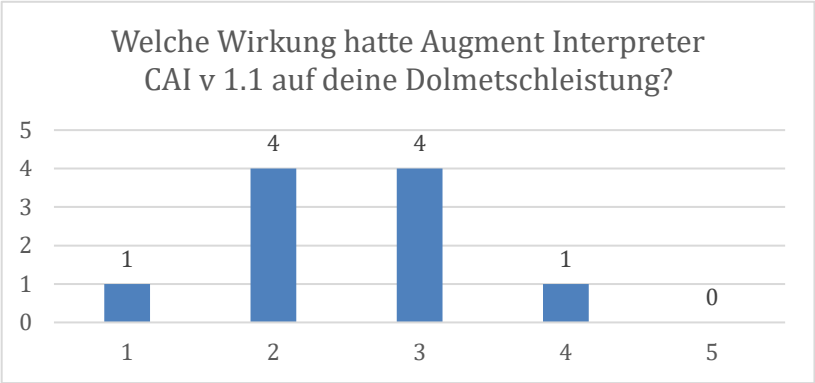

Abb. 33: Frage 5

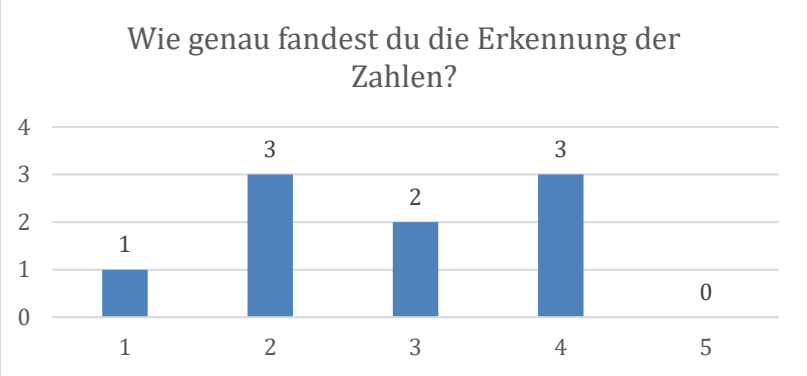

Abb. 34: Frage 8

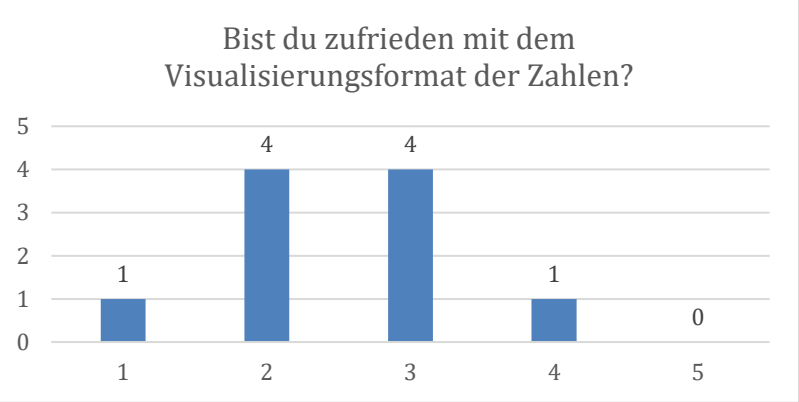

Abb. 35: Frage 11

Für die Frage über die besonderen Herausforderungen bei den Dolmetschungen wurden die ungewohnte Sprachkombination bzw. Italienisch als Fremdsprache (P4, P7, P8), die eingeschränkte Bewegungen aufgrund des Eyetrackers (P1, P2, P5) und das Fehlen von schriftlichen Notizen während des Dolmetschens (P4, P9) erwähnt.

Im Anschluss wurden Fragen zu CAI-Tools und zur Software *Augmented Interpreter CAI v 1.1* gestellt. Sechs Teilnehmerinnen (75 %) hatten schon vor dem Experiment über CAI-Tools gehört, darunter vier während des Studiums (P2, P4, P7, P9). Nur eine Teilnehmerin (P7) gab an, dass sie vor dem Experiment schon mit CAI-Tools gedolmetscht hatte, und zwar mit BoothMate, das bei universitären Lehrveranstaltungen verwendet wurde. Sieben Teilnehmerinnen (87,5 %) hatten schon vor dem Experiment von InterpretBank gehört.

Im letzten Abschnitt wurden Fragen zu den Zahlen gestellt. Auf die Frage, wie sie normalerweise mit Zahlen in der Kabine umgehen, gaben alle Teilnehmerinnen an, dass sie Zahlen auf einem Blatt Papier notieren. Weitere Strategien sind außerdem, dass Kabinenpartner*innen Zahlen für sie notieren (66,7 %) und dass sie sich Zahlen merken (55,6 %) (Abb. 36). Zu den häufigen Dolmetschstrategien für die Wiedergabe von Zahlen wurde von mehreren Teilnehmerinnen erwähnt, dass sie Annäherungen bzw. Verallgemeinerung verwenden (P2, P3, P6, P8, P9). P2 gab außerdem an, dass sie bei der Wiedergabe von Zahlen ein langsameres Sprechtempo verwendet und den Inhalt vereinfacht. P6 erwähnte auch, dass sie Zahlen sofort wiedergibt und danach versucht, den restlichen Satz daran anzupassen.

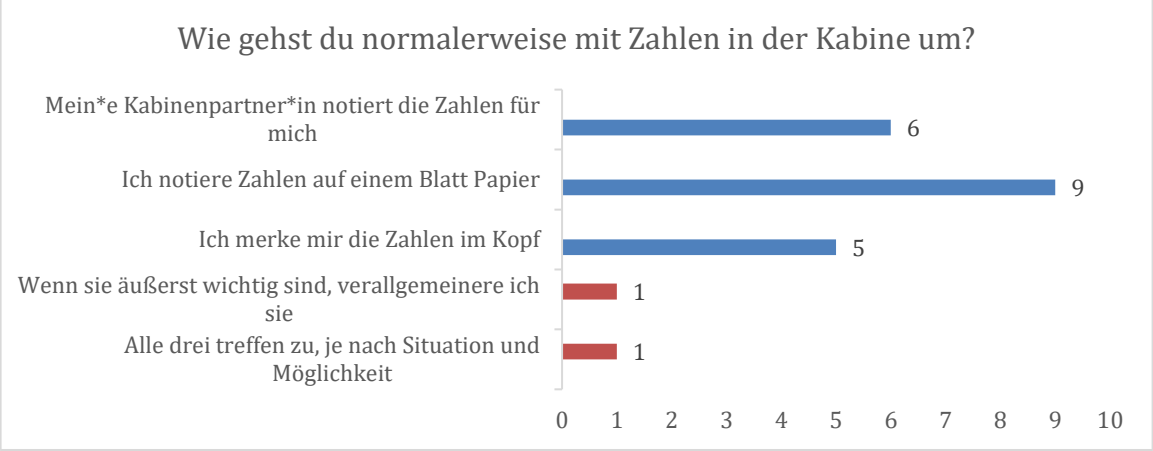

Abb. 36: Frage 21

6 Diskussion und Schlussfolgerungen

Die Fortschritte, die in den letzten Jahren im Bereich automatische Spracherkennung erreicht wurden, haben dazu geführt, dass Spracherkennungssysteme Anwendung in verschiedenen Sektoren gefunden haben, unter anderem in Computer Assisted Interpreting (CAI) Tools. Ein wachsendes Interesse seitens der Dolmetscher*innen und Akteur*innen im Dolmetschsektor wurde verzeichnet, wie die hohe Anzahl an Veröffentlichungen über dieses Thema in den letzten Jahren zeigt. Auf Basis dieser Erkenntnisse entstand die Entscheidung, visuelle Aufmerksamkeit beim Simultandolmetschen mit ASR-gestützten Anwendungen zu erforschen. Anhand der Eyetracking-Methode wurde ein Experiment über die visuelle Aufmerksamkeit von Simultandolmetscher*innen durchgeführt, wobei der Fokus auf der automatischen Erkennung von Zahlen lag.

Im Rahmen dieser Arbeit sollte folgende Forschungsfrage beantwortet werden: „Welche Wirkung haben visuelle Inputs in Form von durch automatische Spracherkennung erkannten Zahlen im CAI-Tool InterpretBank auf die Verteilung der visuellen Aufmerksamkeit beim Simultandolmetschen und auf die Richtigkeit der gedolmetschten Zahlen?“. Diese Forschungsfrage zielte einerseits darauf ab, die Dolmetschleistungen im Hinblick auf die von der ASR erkannten und angezeigten Zahlen auszuwerten. Andererseits wurde ein Eyetracker verwendet, um das Blickverhalten und die visuelle Aufmerksamkeit der Studienteilnehmer*innen auf unterschiedlichen Bereichen des Bildschirmes während des Simultandolmetschens zu untersuchen. In diesem Kapitel werden nun die wichtigsten Erkenntnisse zu den Leistungen der Software *Augmented Interpreter - CAI v 1.1* bei der automatischen Erkennung von Zahlen, zu den Dolmetschleistungen der Studienteilnehmer*innen und zur Verteilung der visuellen Aufmerksamkeit beim Simultandolmetschen mit der Eyetracking-Methode zusammengefasst.

Vor der Durchführung des Experiments wurde zunächst die englische Ausgangsrede vorbereitet. Die Vorteile der Erstellung einer Ausgangsrede ad-hoc für die Durchführung des Experiments lagen darin, dass diese an die Forschungsfrage angepasst werden konnte. Außerdem wurde diese in Anlehnung an zuvor durchgeführte Studien mit Schwerpunkt auf der Wiedergabe von Zahlen bzw. auf dem Dolmetschen mit CAI-Tools gestaltet. Die Rede enthielt 29 Zahlen und dauerte insgesamt 7 Minuten und 20 Sekunden. Das Sprechtempo von 99,7 WPM ähnelte dem in vorherigen Studien (z. B. Fantinuoli & Montecchio 2022; Frittella 2022). Die Rede wurde aufgezeichnet und es wurde ein Video erstellt, das auf einem geteilten Bildschirm das Video der Rednerin und die Software zur automatischen Erkennung der Zahlen, also *Augmented Interpreter*,

zeigte. Das Video wurde im Rahmen des Experiments den neun Studienteilnehmerinnen gezeigt, um ähnliche Bedingungen für alle Teilnehmerinnen zu gewährleisten. Da die Erkennung von Zahlen bei der Software nur auf fünf Minuten beschränkt war, bevor die Webseite neu geladen werden musste, wurden Zahlen nur bis zu Minute 5′20″ erkannt. Für die letzten neun Zahlen wurde entschieden, die Webseite nicht neu zu laden, um eine Unterbrechung der ASR-Software nachzuahmen. Studienteilnehmerinnen wurden nicht im Vorhinein informiert, dass die Software im zweiten Teil des Videos aufhören würde zu funktionieren. Das Video, das den Studienteilnehmerinnen während des Experiments zur Verfügung gestellt wurde, enthielt 20 von den insgesamt 30 Zahlen, die im Ausgangstext vorkamen, also zwei Drittel der gesamten Zahlen.

Die Untersuchung der Softwareleistung von *Augmented Interpreter - CAI v 1.1* bei der Erkennung von Zahlen ergab einen durchschnittlichen Genauigkeitswert von ca. 95 %. Nur eine Zahl wurde nicht von der Software erkannt. Dieser Wert ist vergleichbar mit den Ergebnissen von zuvor durchgeführten Studien (z. B. Defrancq & Fantinuoli 2021; Fantinuoli 2017). Obwohl Zahlen richtig erkannt wurden, wurden sie in der Software oft anders dargestellt als in der traditionellen Schreibweise. Die letzten Ziffern wurden zum Beispiel in der nächsten Zeile gezeigt, und Jahreszahlen wurden mit einem Punkt als Tausendertrennzeichen ausgegeben.

Um die Teilfrage der Forschungsfrage „Welche Wirkung haben visuelle Inputs in Form von durch automatische Spracherkennung erkannten Zahlen im CAI-Tool InterpretBank auf die Richtigkeit der gedolmetschten Zahlen?“ zu beantworten, wurden zunächst die Dolmetschleistungen der einzelnen Studienteilnehmerinnen ausgewertet (s. Kapitel 5.1.2). Die Analyse ergab, dass 64,4 % der in der Ausgangsrede erwähnten Zahlen korrekt wiedergegeben wurden. Wenn die Annäherungen dazugezählt werden, die einen halben Punkt wert waren, stieg dieser Wert auf bis zu 66,5 %. Diese Ergebnisse schwankten stark, je nach Teilnehmerin (von 41,67 % bis zu 96,67 %). Obwohl die ASR-Software nur Zahlen ohne Referenten anzeigte, wurden 257 Referenten von den Teilnehmerinnen richtig wiedergegeben, nämlich 95,2 %. Damit ist die Hypothese widerlegt, dass ihre Genauigkeitsrate niedriger als die Genauigkeitsrate der Zahlwörter ist.

In den Hypothesen wurde auch vermutet, dass die Genauigkeitsrate der Dolmetschleistungen ab der Unterbrechung der Software gesunken wäre. Die Analyse der Dolmetschleistungen vor und nach der Unterbrechung der ASR ergab, dass 73,02 % der Zahlen und 77,78 % der Referenten vor der Unterbrechung der Anwendung richtig gedolmetscht wurden. Die zwei Zahlen, die unmittelbar nach der Unterbrechung der Software erwähnt wurden, wurden in 44,44 % der

Fälle richtig gedolmetscht, während ab der Unterbrechung 51,23 % der Zahlen und 67,90 % der Referenten richtig gedolmetscht wurden. Das bestätigt die Hypothese, dass ohne Software niedrigere Genauigkeitswerte erreicht wurden.

Im zweiten Teil der Datenanalyse (s. Kapitel 5.2) wurden die Ergebnisse des Eyetrackers ausgewertet, um die Wirkung von visuellen Inputs auf die Verteilung der visuellen Aufmerksamkeit zu untersuchen. Eine Heatmap wurde erstellt, um das allgemeine Blickverhalten aller Teilnehmerinnen während der gesamtem Durchführung der Versuche zu beobachten. Hierbei waren zwei Bereiche farblich markiert. Diese Karte zeigte, dass die Blicke der Studienteilnehmerinnen am häufigsten auf dem Bereich des Bildschirms verweilten, in dem sich das Gesicht der Rednerin befand. Außerdem wurde auch der Bereich markiert, wo die neuen Zahlen erschienen.

Anhand dieser Heatmap und der Hypothesen zur Verteilung der Aufmerksamkeit wurden Interessengebiete (AOI) festgelegt. Als AOI wurden das Gesicht der Rednerin und die Zahlen ausgewählt. Um Unterschiede beim Blickverhalten für die einzelnen Zahlen zu analysieren, wurde jede AOI aktiviert, sobald eine Zahl auf dem Bildschirm aufschien, und folglich deaktiviert, wenn sie verschwand. Da die Zahl 4 von der Rednerin nicht nur vorgelesen, sondern auch mit einem Handzeichen gezeigt wurde, wurde eine AOI dafür erstellt, um auch die Wirkung des Handzeichens auf die visuelle Aufmerksamkeit zu untersuchen. Ab der Minute 5′20″ wurden nur zwei AOI verwendet, nämlich eine auf dem Gesicht der Rednerin und eine auf der Software. Ziel war es zu untersuchen, ob auch ab der Unterbrechung der ASR-Anwendung Fixationen verzeichnet wurden, obwohl Zahlen nicht mehr angezeigt wurden.

Um die visuelle Aufmerksamkeit der Studienteilnehmer*innen während des Experiments zu analysieren, wurden zwei Messungsindikatoren für die Analyse der Eyetracking-Daten ausgewählt: die Dauer und die Anzahl der Fixationen. Die Ergebnisse der Datenauswertung zeigten, dass die AOI der Rednerin (beides vor und nach der Unterbrechung der Software) die höchsten Werte in Bezug auf Dauer und Anzahl der Fixationen verzeichnete. Dabei muss aber berücksichtigt werden, dass die AOI des Videos der Rednerin länger als alle anderen AOI sichtbar war. Darüber hinaus nahm sie eine größere Fläche auf dem Bildschirm als die anderen AOI ein. Alle diese Faktoren können einen Einfluss auf die Eyetracking-Ergebnisse haben.

Auch ab dem Zeitpunkt der Unterbrechung der ASR-Software wurde die AOI der Rednerin länger als die AOI der Software betrachtet, und zwar ca. 12-mal länger. Außerdem war die Dauer der Erstfixation auf die AOI der Software ab dem Zeitpunkt der Unterbrechung viel höher als die

durchschnittlichen Erstfixationen auf die AOI der Zahlen. Ab dem Zeitpunkt der Unterbrechung wurden durchschnittlich 141 Fixationen für die AOI der Rednerin und 25,5 Fixationen für die AOI der Software verzeichnet. Das zeigt, dass Teilnehmerinnen während der letzten Minuten des Videos ihren Blick sehr oft bewegt haben, obwohl auf dem Bildschirm keine neuen Zahlen erschienen. Das lässt sich darauf zurückführen, dass Versuchsteilnehmerinnen auch nach der Unterbrechung der Software erwarteten, dass neue Zahlen angezeigt werden.

Obwohl die Hypothese bestätigt wird, dass von Anfang der Rede an die visuelle Aufmerksamkeit auf die Rednerin und auf ihr Gesicht gerichtet wird, erwies es sich als sinnvoll, auch die Dauer und die Anzahl der Fixationen für die einzelnen Zahlen zu untersuchen. Es stellte sich heraus, dass große Unterschiede unter den Zahlen erkennbar sind. Es wurde festgestellt, dass die Zahl, die durchschnittlich am längsten angesehen wurde, ‚5,51 million' war. Danach folgten 565, 12.054, 1.000 in dem Satz „*In Austria, there are 565 cars per 1.000 inhabitants*" und 679. Die niedrigsten Werte wurden hingegen für die Zahlen 1.000 in dem Satz „*Burgenland has the most cars per 1,000 inhabitants*", 2 in dem Namen Car2go und 1.000 in dem Satz „*of which 1.000 were e-smarts*" verzeichnet. Die AOI der Zahl 4 wurde ca. drei Mal länger als die AOI des Handzeichens betrachtet. Auch die Anzahl der Fixationen zeigte ähnliche Ergebnisse. Die Zahlen, deren AOI durchschnittlich mehr Fixationen verzeichneten, waren 12.054, ‚5,51 million', 565 und 1.000 in dem Satz „*In Austria, there are 565 cars per 1.000 inhabitants*". Die geringste durchschnittliche Anzahl der Fixationen wurde für die Zahl 1.000 in den Sätzen „*Burgenland has the most cars per 1,000 inhabitants*" und „*of which 1.000 were e-smarts*" verzeichnet.

Wenn die prozentuelle Verteilung der Anzahl und Dauer der Fixationen betrachtet wird, kann außerdem festgestellt werden, dass viele Fixationen außerhalb der festgelegten AOI verzeichnet wurden. Diese Kategorie betrug ca. 29,5 % der gesamten Dauer der Fixationen und 50,5 % der Anzahl der Fixationen. Auf dem Bildschirm gab es zahlreiche zusätzliche Elemente, wie zum Beispiel die Benutzerschnittstelle der Software, die Windows-Taskleiste am unteren Rand des Bildschirms und die Tab-Leiste mit dem offenen Browserfenster der ASR-Webseite. Diese wurden nicht ausgeblendet, um das Setting realitätsnah zu halten. Es ist also davon auszugehen, dass auch auf diese Elemente visuelle Aufmerksamkeit gerichtet war.

Diese Ergebnisse ermöglichten einen Überblick über das durchschnittliche Blickverhalten aller Studienteilnehmerinnen. Da Eyetracking-Daten sehr individuell sind, wurde entschieden, die Ergebnisse von zwei Versuchsteilnehmerinnen genauer zu untersuchen. Damit konnte das

Blickverhalten einzelner Teilnehmerinnen ausführlich beobachtet werden. Das wurde auch anhand der Ergebnisse ihrer Dolmetschleistungen evaluiert, um die Hypothese zu überprüfen, ob es einen Zusammenhang zwischen dem Blickverhalten und ihren Dolmetschungen gibt. Die Auswahl der Teilnehmerinnen P3 und P4 war in diesem Fall zufällig.

Die Eyetracking-Daten von P3 und P4 in Bezug auf der Dauer und Anzahl der Fixationen weichen nicht stark von den durchschnittlichen Ergebnissen aller Teilnehmerinnen ab. Zwei Tabellen wurden erstellt, in denen die Zahlen mit der höchsten Dauer und Anzahl der Fixationen (und Erstfixationen) und die Zahlen mit den niedrigsten Werten dargestellt wurden (s. Tabelle 8 und 9). Das war möglich, da Zahlen, die die höchste Dauer der Fixationen bzw. Erstfixationen verzeichneten, in den meisten Fällen mit den Zahlen mit der höchsten Anzahl der Fixationen übereinstimmten, und umgekehrt galt dies auch für die Zahlen mit den niedrigsten verzeichneten Werten. Anhand der Tabellen wurde geprüft, ob die Eyetracking-Daten Rückschlüsse auf die Dolmetschleistungen ermöglichen.

Für P3 konnte festgestellt werden, dass die AOI der Zahlen mit der höchsten Dauer und Anzahl der Fixationen zu genaueren Wiedergaben der Zahlen führten. Zahlen, für welche keine Fixationen verzeichnet wurden, wurden ausgelassen, was auch zu grammatikalischen Fehlern führte. Es wurden aber nicht nur Zahlen ausgelassen, die eine geringe Dauer der Fixationen verzeichneten, sondern auch die Zahl 7.000, obwohl sie eine vergleichsweise höhere Dauer der Erstfixation verzeichnete. Auch die Zahl 655 und das Datum 31. Dezember 2022 wurden trotz der geringen Anzahl und der geringen Dauer der Fixationen richtig gedolmetscht und in den Kontext eingebettet.

Die grafische Darstellung der Eyetracking-Daten von P4 zeigte hingegen große Unterschiede zwischen den einzelnen AOI der Zahlen (s. Abbildung 31). Während einige AOI längere Fixationen verzeichneten, wurde für andere AOI keine Fixation verzeichnet. Die Zahlen der ersten Kategorie (hohe Dauer und Anzahl der Fixationen) erreichten für P4 im Allgemeinen höhere Genauigkeitsraten als Zahlen der zweiten Kategorie (keine Fixation oder niedrige Dauer und Anzahl der Fixationen). Zahlen mit hoher Dauer und Anzahl der Fixationen wurden in den meisten Fällen von P4 richtig mit dem Kontext wiedergegeben. Die Zahl ‚5.51 million', welche die längste Dauer der Fixationen und die größte Anzahl der Fixationen verzeichnete, stellte Schwierigkeiten für P4 dar. Der Punkt wurde als Trennzeichen für Dezimalzahlen verwendet, und die Zahl wurde nicht als „cinquantuno" [einundfünfzig] ausgesprochen, sondern als einzelne Ziffern 5 und 1. Von

den fünf Zahlen, die keine Fixation verzeichneten, wurden drei ausgelassen und zwei korrekt wiedergegeben.

Zusammenfassend lässt sich feststellen, dass ein Muster zwischen der visuellen Aufmerksamkeit auf die von der ASR erkannten Zahlen und der korrekten Wiedergabe der Zahlen erkennbar ist. Zahlen, die länger betrachtet wurden bzw. eine höhere Anzahl an Fixationen verzeichneten, wurde in der Mehrheit der Fälle von P3 und P4 richtig wiedergegeben und in den Kontext eingebettet. Die Zahlen 7.000 für P3 und ‚5.51 million' für P4 waren aber zwei Ausnahmen. Obwohl sie hohe Werte in Bezug auf die Dauer und Anzahl der Fixationen erreichten, wurden sie ausgelassen bzw. nicht einwandfrei wiedergegeben. Da diese aber nur einzelne Beispiele sind, und nur zwei Teilnehmerinnen für die ausführliche Analyse der Eyetracking-Daten ausgewählt wurden, können diese Erkenntnisse nicht verallgemeinert werden. Es wäre in diesem Zusammenhang lohnenswert, eine systematisierte Datenauswertung der Eyetracking-Daten und der Dolmetschleistungen zu entwerfen, um zu untersuchen, ob diese Hypothese bei einem größeren Umfang an Daten verifiziert werden kann.

Zuletzt wurden die Antworten in den Fragebögen ausgewertet, da mehrere Faktoren während der Durchführung des Experiments das Blickverhalten der Teilnehmerinnen und ihre Dolmetschleistungen beeinflussen konnten. Das Thema der Rede, *Mobilität*, wurde von der Mehrheit der Studienteilnehmerinnen als „Sehr bekannt" bzw. „Bekannt" bewertet, während die Geschwindigkeit der Rede in einer Skala von "Sehr schnell" bis "Sehr langsam" von der Mehrheit der Befragten als mittelmäßig eingestuft wurde. Die Rede wurde eher als nicht fachspezifisch empfunden. Zusammenfassend lässt sich feststellen, dass diese Faktoren von den Teilnehmerinnen nicht als störend empfunden wurden.

Wenn es um die Beurteilung der Software *Augmented Interpreter CAI v 1.1* ging, waren die Antworten über die Wirkung der Software auf die eigenen Leistungen unterschiedlich. Es wurde angegeben, dass die automatische Erkennung vor allem für Zahlen unter Eintausend (P5) bzw. zwei- bis vierstellige Zahlen (P4) hilfreich war. Die geringe Latenz wurde von zwei Teilnehmerinnen als ein positiver Aspekt vermerkt. Zu den negativen Aspekten der Software zählen folgende Faktoren: die Unterbrechung der Software (P2, P6, P7), die Trennung der Ziffern ab Eintausend (P2, P3, P5, P7) und falsch angezeigte Zahlen (P1, P9). Als visuelle Störfaktoren wurden die Farbe der Zahlen, die vielen visuellen Inputs (P1) und der Punkt als Tausendertrennzeichen auch bei Jahreszahlen (P8) genannt. Die große Menge an Zahlen in der

Rede wurde aber nicht als Störfaktor empfunden. Hingegen konnte die Hypothese bestätigt werden, dass die Software aufgrund der Unterbrechung als unzuverlässig eingestuft wurde.

Das experimentelle Setting stellte wie erwartet einige Stressfaktoren dar. Zahlen durften während des Experiments nicht aufgeschrieben werden, ansonsten hätte der Eyetracker die Positionierung des Blicks verloren. Aus diesem Grund wurde es den Teilnehmerinnen nicht gestattet, Zahlen zu notieren. Laut der Ergebnisse der Befragung wäre das Notieren jedoch eine beliebte Methode, um mit Zahlen beim Dolmetschen umzugehen. Aus demselben Grund wurden Teilnehmerinnen aufgefordert, sich nicht allzu viel zu bewegen. Die beschränkte Bewegungsfreiheit aufgrund des Eyetrackers wurde im Fragebogen unter den Herausforderungen beim Dolmetschen erwähnt (P1, P2, P5).

Zusammenfassend kann gesagt werden, dass die vorliegende Arbeit eine der ersten Studien über die visuelle Aufmerksamkeit im Zusammenhang mit ASR-gestützten Tools für Dolmetscher*innen ist. Diese Eyetracking-Studie fügt sich in einen brandaktuellen Forschungskontext ein, und mithilfe des experimentellen Teils dieser Arbeit wurde versucht, einen Beitrag zur prozessorientierten Dolmetschforschung zu leisten. Während die produktorientierte Forschung im Bereich CAI-Tools etabliert ist und mehrere Forschungsarbeiten zum Thema veröffentlicht wurden (s. Kapitel 2.4), zeigte sich eine Lücke in Bezug auf die Wirkung, die CAI-Tools auf den Dolmetschprozess bzw. auf die kognitiven Prozesse beim Dolmetschen haben (vgl. Prandi 2022; Rodriguez et al. 2022). Diese Arbeit ebnet den Weg für zukünftige Forschung im Bereich der visuellen Aufmerksamkeit und der Integration von automatischer Spracherkennung in CAI-Tools. Die vorgestellten Ergebnisse werfen aber auch weiterführende Fragen zu weiteren Aspekten auf, die in dieser Arbeit nicht erforscht wurden, jedoch von großem Interesse wären.

Was die ASR-Software betrifft, wurde im Rahmen dieser Studie die Anwendung *Augmented Interpreter CAI v 1.1*. verwendet. Dieses Tool bietet eine gezielte Hilfestellung für Dolmetscher*innen an, indem Zahlen automatisch erkannt werden. Allerdings ist diese Software kein CAI-Tool im weiteren Sinne. Eine Anwendung, welche auch über ein Terminologieverwaltungssystem verfügt und Fachtermini automatisch erkennt, könnte verwendet werden, um den Rahmen der Forschungsarbeit zu erweitern. Außerdem erweisen sich auch Fachterminologie und Eigennamen als Problemauslöser beim Simultandolmetschen und müssten dementsprechend in Forschungsarbeiten berücksichtigt werden. Eine ASR-Software, welche Zahlen mit ihren Referenten erkennt und anzeigt, wäre auch bei zukünftigen Forschungen zu untersuchen, jedoch

wurde zum Zeitpunkt der Durchführung der folgenden Arbeit keine Software gefunden, die diese Möglichkeit anbietet, gezielt für Dolmetscher*innen entwickelt wurde und gleichzeitig auch auf dem Markt verfügbar ist.

Die Forschung über visuelle Aufmerksamkeit könnte sich des Weiteren als lohnenswert erweisen, um das Thema Benutzerfreundlichkeit der Benutzerschnittstelle von CAI-Tools zu untersuchen. Die Wirkung, welche unterschiedliche Elemente auf dem Bildschirm auf die visuelle Aufmerksamkeit von Dolmetscher*innen haben, könnte mit der Eyetracking-Methode erforscht werden.

Im Zuge der Gestaltung des Experiments wurden lediglich Dolmetschstudierende als Studienteilnehmer*innen herangezogen. Es wäre für die Forschung von bedeutendem Interesse, ob professionell ausgebildete Dolmetscher*innen unterschiedliche Ergebnisse liefern würden. Diese Frage wird in mehreren Studien thematisiert und erweist sich als eine der häufigsten Hürden, was die Untersuchungen im Dolmetschbereich betrifft.

Eine weitere Variable, die in zukünftigen Studien berücksichtigt werden könnte, ist die Sprachkombination. Ursprünglich war für dieses Experiment geplant worden, die Sprachkombination Deutsch-Italienisch zu verwenden. Diese hätte einerseits den Vorteil gehabt, genug Studienteilnehmer*innen mit Italienisch als A-Sprache und Deutsch als B-Sprache zur Verfügung zu haben. Alle anderen Sprachkombinationen hätten ausgeschlossen werden können, um mehr Einheitlichkeit zwischen den Teilnehmer*innen zu haben. Andererseits werden im Italienischen und im Deutschen zwei unterschiedliche Zahlensysteme verwendet, wobei die Zehner und Einser jeweils an der Stelle des Anderen stehen. Dabei hätte die Hilfestellung des ASR-Tools gezielt auch in Bezug auf dieses Thema erforscht werden können. Das war aber aufgrund der Einschränkung der Software in der Auswahl der Ausgangssprache Englisch im Rahmen dieser Studie nicht möglich.

Bei der Forschungsmethode des Eyetrackings wäre es interessant gewesen, auch andere Messungsindikatoren, wie zum Beispiel die Pupillengröße, zu untersuchen. Hierbei handelt es sich um einen verbreiteten Messungsindikator, um kognitiv herausfordernde Aufgaben zu analysieren. Diese Daten waren im Ergebnisbericht der einzelnen Teilnehmerinnen verfügbar, werden aber von der Datenauswertung ausgeschlossen. Außerdem sollte bei zukünftigen Forschungsarbeiten die Untersuchung der Zusammenhang zwischen den Eyetracking-Daten und den Dolmetschleistungen nicht unbeachtet bleiben und für alle Studienteilnehmer*innen durchgeführt werden.

Diese Studie stellt eine erste Forschungsarbeit zum Thema der Auswirkungen von automatischer Spracherkennung auf das Simultandolmetschen dar. Da die Integration von automatischer Spracherkennung in unterschiedliche Anwendungen nicht mehr wegzudenken, und die Zusammenarbeit zwischen Mensch und Maschine zunehmend notwendig, ist, um die Herausforderungen der aktuellen und zukünftigen technologischen Entwicklung zu bewältigen, erweist sich dieses Thema für die Dolmetschforschung und für alle Akteur*innen im Dolmetschbereich als äußerst bedeutsam. Die präsentierten Ergebnisse sind somit eine solide Grundlage für weitere Analysen in diesem Forschungsbereich, der für zukünftige Forschungsvorhaben großes Potenzial birgt.

Bibliografie

Ahrens, Barbara (2017). Interpretation and Cognition. In: Schwieter, John W. & Ferreira, Aline (Hg.) *The handbook of translation and cognition.* Hoboken: John Wiley & Sons, 445-460.

AltaFormazione (o.J.). Master di II livello in "Intelligenza artificiale e tecnologie per l'interpretariato – (IATI)" https://altaformazione.unint.eu/corso/master-universitario-annuale-di-ii-livello-in-intelligenza-artificiale-e-tecnologie-per-linterpretariato-iati/ (Stand: 06.04.2024).

Amelina, Svitlana & Tarasenko, Rostyslav (2020). Using modern simultaneous interpretation tools in the training of interpreters at universities. *CEUR workshop proceedings.* https://ceur-ws.org/Vol-2740/20200188.pdf (Stand: 23.01.2024).

Arzik Erzurumlu, Özüm & Perihan Demir (2022). Taking stock of the COVID-19 working conditions on the performance of the interpreters: Rendering the numbers in the 2020 American presidential debates. *Çeviribilim ve uygulamaları dergisi*/ *Journal of translation studies* 2022 (32), 25-54.

Augmented Interpreter (o.J.). Augmented interpreter - CAI v 1.1. https://www.claudiofantinuoli.org/apps/ASR-NUMBERS/index.html (Stand: 06.04.2024).

Bereznaya, Violetta (2022). *Can Computer-Assisted Interpreting (CAI) Tools be reliable artificial boothmates? Testing the usability of InterpretBank's Automatic Speech Recognition feature in a remote interpreting setting*. Masterarbeit, Universität Wien.

Biagini, Giulio (2015). *Glossario cartaceo e glossario elettronico durante l'interpretazione simultanea: uno studio comparativo*. Masterarbeit, Università di Trieste.

Boothmate (o.J.). Boothmate (v1.1.19). https://boothmate.app/dashboard (Stand: 06.04.2024).

Braun, Sabine (2019). Technology and Interpreting. In: O'Hagan, Minako (Hg.) *The Routledge handbook of translation and technology*.1. Auflage. New York: Routledge, 271-88.

Braun, Susanne & Clarici, Andrea (1996). Inaccuracy for numerals in simultaneous interpretation: neurolinguistic and neuropsychological perspectives. *The Interpreters' Newsletter* 7, 85-102.

Canali, Sara (2019). *Technologie und Zahlen beim Simultandolmetschen: Utilizzo del riconoscimento vocale come supporto durante l'interpretazione simultanea dei numeri.* Masterarbeit, Universität UNINT Rom.

Carter, Benjamin T. & Luke, Steven G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology* (155), 49-62.

Chen, Sijia (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives* 25 (4), 640-657.

Chen, Sijia (2018). *Exploring the process of note-taking and consecutive interpreting: A pen-eye-voice approach towards cognitive load*. PhD Thesis, Macquarie University.

Cheung, Andrew & Li, Tianyun (2022). Machine Aided Interpreting: An experiment of automatic speech recognition in simultaneous interpreting. *Translation Quarterly* 104, 1-20.

Chmiel, Agnieszka (2021). Eye-tracking studies in conference interpreting. In: Albl-Mikasa, Michaela & Tiselius, Elisabet (Hg.) *The Routledge handbook of conference interpreting*. London: Routledge, 457-70.

Chmiel, Agnieszka; Janikowski, Przemysław & Lijewska, Agnieszka (2020). Multimodal processing in simultaneous interpreting with text: Interpreters focus more on the visual than the auditory modality. *Target* 32 (1), 37-58.

Chmiel, Agnieszka & Mazur, Iwona (2013). Eye tracking sight translation performed by trainee interpreters. In: Way, Catherine; Vandepitte, Sonia; Meylaerts, Reine & Bartłomiejczyk, Magdalena (Hg.) *Tracks and treks in translation studies.* Amsterdam: John Benjamins, 189-205.

Cloudfresh (2023). Speech-to-text from google cloud: What are the reasons to use it. https://cloudfresh.com/en/blog/google-speech-to-text-why-to-use/ (Stand: 06.04.2024).

Collard, Camille (2018). *A corpus-based study of simultaneous interpreting with special reference to sex.* Masterarbeit, Universität Gent.

Conklin, Kathy; Pellicer-Sánchez, Ana & Carrol, Gareth (2018) *Eye-Tracking: A guide for applied linguistics research*. Cambridge: Cambridge University Press.

Corpas Pastor, Gloria (2022). Technology solutions for interpreters: the VIP system. *Hermēneus* 23, 91-123.

Corpas Pastor, Gloria & Durán-Muñoz, Isabel (Hg.) (2018). *Trends in e-tools and resources for translators and interpreters.* Leiden/Boston: Brill Rodopi.

Corpas Pastor, Gloria & Fern, Lily May (2016). A survey of interpreters' needs and practices related to language technology. https://www.researchgate.net/publication/303685153_A_survey_of_interpreters'_needs_and_practices_related_to_language_technology (Stand: 06.04.2024).

Defrancq, Bart & Fantinuoli, Claudio (2021). Automatic Speech Recognition in the booth: Assessment of system performance, interpreters' performances and interactions in the context of numbers. *Target* 33 (1), 73-102.

Desmet, Bart & Vandierendonck, Mieke & Defrancq, Bart (2018). Simultaneous interpretation of numbers and the impact of technological support. In: Fantinuoli, Claudio (Hg). *Interpreting and technology*. Berlin: Language Science Press, 13-28.

Drechsel, Alexander (2019). Technology literacy for the interpreter. In: Sawyer, David B.; Austermühl, Frank & Enríquez Raído, Vanessa (Hg.) *The Evolving Curriculum in Interpreter and Translator Education : Stakeholder Perspectives and Voices*. Amsterdam: John Benjamins, 259-268.

EABM (o.J.). Ergonomics for the artificial booth mate (EABM). https://www.eabm.ugent.be/survey/ (Stand: 06.04.2024).

EU (2019). New Technologies and Artificial Intelligence in the field of language and conference services. https://www.knowledge-centre-interpretation.education.ec.europa.eu/en/news/eu-host-paper-new-technologies-and-artificial-intelligence-field-language-and-conference (Stand: 06.04.2024).

Euler, Stephen (2006). *Grundkurs Spracherkennung: Vom Sprachsignal zum Dialog. Grundlagen und Anwendung verstehen*. Wiesbaden: Vieweg.

Fantinuoli, Claudio (2009). InterpretBank: Ein Tool zum Wissens- und Terminologiemanagement für Simultandolmetscher. https://www.staff.uni-mainz.de/fantinuo/download/publications/Fantinuoli_InterpretBank.pdf (Stand: 06.04.2024).

Fantinuoli, Claudio (2016). InterpretBank. Redefining computer-assisted interpreting tools. *Proceedings Translating and the Computer 38 Conference.* London: Tradulex, 42-52.

Fantinuoli, Claudio (2017). Speech Recognition in the interpreter workstation. *Proceedings of the Translating and the Computer 39 Conference*. Genf: Editions Tradulex, 25-34.

Fantinuoli, Claudio (2018a). Computer-assisted Interpreting: Challenges and future perspectives. In: Corpas Pastor, Gloria & Durán-Muñoz, Isabel (Hg.). *Trends in e-tools and resources for translators and interpreters*. Leiden: Brill/Rodopi, 153-176.

Fantinuoli, Claudio (2018b). Interpreting and technology: The upcoming technological turn. In: Fantinuoli, Claudio (Hg.) *Interpreting and technology*. Berlin: Language Science Press, 1-12.

Fantinuoli, Claudio (2019). The technological turn in interpreting: The challenges that lie ahead. In: Baur, Wolfram & Mayer, Felix (Hg.) *Übersetzen und Dolmetschen 4.0: Neue Wege im Digitalen Zeitalter*. Berlin: BDÜ Fachverlag, 334-354.

Fantinuoli, Claudio (2021). Conference interpreting and new technologies. In: Albl-Mikasa, Michaela & Tiselius, Elisabet (Hg.) *The Routledge handbook of conference interpreting.* London/New York: Routledge, 508-522.

Fantinuoli, Claudio; Marchesini, Giulia; Landan, David & Horak, Lukas (2022). KUDO Interpreter Assist: Automated real-time support for remote interpretation. https://doi.org/10.48550/arXiv.2201.01800 (Stand: 06.04.2024).

Fantinuoli, Claudio & Montecchio, Maddalena (2022). Defining maximum acceptable latency of AI-enhanced CAI tools. https://arxiv.org/abs/2201.02792 (Stand: 06.04.2024).

Fantinuoli, Claudio & Pisani, Elisabetta (2021). Measuring the impact of automatic speech recognition on number rendition in simultaneous interpreting. In: Wang, Caiwen & Zheng, Bingham (Hg.) *Empirical studies of translation and interpreting: The post-structuralist approach.* New York: Routledge, 181-197.

Fossati, Giona (2021). *SmarTerp: Applying the user-centered design process in a Computer-Assisted Interpreting (CAI) Tool.* Masterarbeit, Universidad Politécnica de Madrid.

Freiberger-Geistberger, Christian (2021). *CAI-Tools mit automatischer Spracherkennung – ein neuer Bestandteil der Dolmetschlehre?* Masterarbeit, Universität Graz.

Frittella, Francesca Maria (2019). "70.6 billion world citizens": Investigating the difficulty of interpreting numbers. *Translation & Interpreting* 11 (1), 79-99.

Frittella, Francesca Maria (2022). CAI Tool-Supported SI of numbers: A theoretical and CAI Tool-Supported SI of numbers: A theoretical and methodological contribution. *International Journal of Interpreter Education* 14 (1), 32-56.

Frittella, Francesca Maria (2023). *Usability research for interpreter-centred technology: The case study of SmarTerp.* Berlin: Language Science Press.

Gacek, Michael (2015). *Softwarelösungen für DolmetscherInnen.* Masterarbeit, Universität Wien.

Gile, Daniel (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum* 6 (2), 59-77.

Gile, Daniel (2009). *Basic concepts and models for interpreter and translator training: Revised edition.* Amsterdam: John Benjamins.

Gottwald, Alfons (2018). *Automatische Spracherkennung: Methoden der Klassifikation und Merkmalsextraktion.* 2. Auflage. Berlin/Boston: Oldenbourg.

Hansen-Schirra, Silvia (2012). Nutzbarkeit von Sprachtechnologien für die Translation. *Trans-kom* 5 (2), 211-226.

Holmqvist, Kenneth (2011). *Eye tracking: A comprehensive guide to methods and measures.* 1. Auflage. Oxford: Oxford University Press.

Hu, Ting; Wang, Xinyu & Xu, Haiming (2022). Eye-Tracking in interpreting studies: A review of four decades of empirical studies. *Frontiers in psychology* 13. https://doi.org/10.3389/fpsyg.2022.872247 (Stand: 06.04.2024).

Hvelplund, Kristian Tangsgaard (2017). Eye Tracking in translation process research. In: Schwieter, John W. & Ferreira, Aline (Hg.) *The handbook of translation and cognition*. Hoboken: John Wiley & Sons, 248-264.

Hyönä, Jukka; Tommola, Jorma & Alaja, Anna-Mari (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology* 48 (3), 598-612.

InterpretBank (o.J.). AI-based tool for interpreters. https://www.interpretbank.com/site/ (Stand: 06.04.2024).

InterpretBank ASR (o.J.). Artificial Boothmate: User manual. https://asr.interpretbank.com/manual (Stand: 06.04.2024).

Juang, Biing Hwang & Rabiner, Lawrence (2004). Automatic Speech Recognition - A brief history of the technology development. https://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALIASRHistory-final-10-8.pdf (Stand: 06.04.2024).

Kajzer-Wietrzny, Marta; Ilmari, Ivaska & Ferraresi, Adriano (2021). 'Lost' in interpreting and 'found' in translation: Using an intermodal, multidirectional parallel corpus to investigate the rendition of numbers. *Perspectives* 29 (4), 469-88.

Kajzer-Wietrzny, Marta; Ilmari, Ivaska & Ferraresi, Adriano (2023). Fluency in rendering numbers in simultaneous interpreting. *Interpreting*, 26 (1), 1-23.

Kittimongkolmar, Chirattikarn (2021). *Investigation of an automatic speech recognition software for numbers trigger management in remote simultaneous interpretation from English to Thai*. Masterarbeit, Chulalongkorn University.

Korpal, Paweł & Katarzyna Stachowiak-Szymczak (2018). The whole picture: Processing of numbers and their context in simultaneous interpreting. *Poznan Studies in Contemporary Linguistics* 54 (3), 335-354.

Korpal, Paweł & Katarzyna Stachowiak-Szymczak (2020). Combined problem triggers in simultaneous interpreting: Exploring the effect of delivery rate on processing and rendering numbers. *Perspectives* 28 (1), 126-143.

Kuang, Huolingxiao & Zheng, Binghan (2023). Note-taking effort in video remote interpreting: effects of source speech difficulty and interpreter work experience. *Perspectives* 31 (4), 724-744.

KUDO (2023). KUDO Releases KUDO AI engine v.2.0 for improved speech translation accuracy. https://kudoway.com/product-release/kudo-releases-kudo-ai-engine-v-2-0-for-improved-speech-translation-accuracy/ (Stand: 06.04.2024).

Lambert, Sylvie (2004). Shared attention during sight translation, sight interpretation and simultaneous interpretation. *Meta* 49 (2), 294-306.

Malik, Mishaim; Malik, Muhammad Kamran; Mehmood, Khawar & Makhdoom, Imran (2021). Automatic speech recognition: A survey. *Multimedia Tools and Applications* 80, 9411-9457.

Mazza, Cristina (2001). Numbers in simultaneous interpretation. *The Interpreters' Newsletter* 11, 87-104.

McDonald, Janet L. & Carpenter, Patricia A. (1981). Simultaneous Translation: Idiom interpretation and parsing heuristics. *Journal of Verbal Learning and Verbal Behavior* 20, 231-247.

MeinBezirk (2023). Tag des Fahrrads am 3. Juni. https://www.meinbezirk.at/c-freizeit/tag-des-fahrrads-am-3-juni_a6070914 (Stand: 06.04.2024).

Monti, Lucia (2016). *Die Verwendung von Terminologieprogrammen in der Dolmetschkabine: Wertvolle Unterstützung oder zusätzliche Belastung?* Masterarbeit, Karl-Franzens-Universität Graz.

Moratto, Riccardo (2020). A preliminary review of eye tracking research in interpreting studies: Retrospect and prospects. *Studies in Linguistics and Literature* 4 (2), 19-32.

Moser-Mercer, Barbara (1992). Simultaneous interpreting. Cognitive potential and limitations. *Interpreting* 5 (2), 83-94.

Mouzourakis, Panayotis (2000). Interpretation Booths for the Third Millennium. https://academia.edu/25845632/Interpretation_Booths_for_the_Third_Millennium (Stand: 06.04.2024).

National Geographic (2013). Die Zukunft der Mobilität. https://www.nationalgeographic.de/umwelt/die-zukunft-der-mobilitaet (Stand: 06.04.2024).

OBS Project (o.J.). OBS Open Broadcaster Software. https://obsproject.com/ (Stand: 06.04.2024).

Pelizza, Francesco Antonio (2021). *Automatische Spracherkennung und Simultandolmetschen: Eine Untersuchung der ASR-Funktion von InterpretBank 7*. Masterarbeit, Universität Wien.

Penta, Martina (2022). *Eigennamen und Zahlen beim Simultandolmetschen: Schwierigkeiten und Strategien*. Masterarbeit, Leopold-Franzens-Universität Innsbruck.

Pinochi, Diletta (2009). Simultaneous interpretation of numbers: Comparing German and English to Italian. An experimental study. *The Interpreters' Newsletter* 14, 33-57.

Pinola, Melanie (2011). Speech Recognition through the decades: How we ended up with Siri. https://www.pcworld.com/article/477914/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html (Stand: 06.04.2024).

Pöchhacker, Franz (2022). *Introducing interpreting studies*. 3. Auflage. Abingdon, Oxon & New York: Routledge.

Prandi, Bianca (2015). The Use of CAI Tools in interpreters' training: A pilot study. *Proceedings of the 37th conference translating and the computer*. Genf: Editions Tradulex, 48-57.

Prandi, Bianca (2017). Designing a multimethod study on the use of CAI tools during simultaneous interpreting. In: Esteves-Ferreira, João; Macan, Juliet; Mitko, Ruslan & Stefanov, Olaf-Michael (Hg.) *Proceedings of the 39th conference Translating and the Computer*. Genf: Editions Tradulex, 76-88.

Prandi, Bianca (2022). *Computer-Assisted simultaneous interpreting: A cognitive-experimental study on terminology*. Berlin: Language Science Press.

Purkarthofer, Mirjam (2017). *Technologiegestütztes Dolmetschen. Der Einsatz von Softwareapplikationen im Dolmetschprozess*. Masterarbeit, Karl-Franzens-Universität Graz.

Ricci, Beatrice (2020). *Simultandolmetschen mit Speech-to-Text-Funktion: Ein Experiment im Sprachenpaar Deutsch-Italienisch*. Masterarbeit, Universität Wien.

Rodriguez, Susana; Frittella, Francesca Maria & Okoniewska, Alicja (2022). A paper on the conference panel “In-booth CAI tool support in conference interpreter training and education”. *AsLing TC43 conference proceedings*, 78-88.

Schukat-Talamazzini, Ernst Günter (1995). *Automatische Spracherkennung: Grundlagen, statistische Modelle und effiziente Algorithmen*. Braunschweig: Vieweg.

Seeber, Kilian G. (2007). Thinking outside the cube: Modeling language processing tasks in a multiple resource paradigm. *Proc. Interspeech* 2007, 1382-1385.

Seeber, Kilian G. (2011). Cognitive load in simultaneous interpreting Existing theories - new models. *Interpreting* 13 (2), 176-204.

Seeber, Kilian G. (2012). Multimodal input in simultaneous interpreting: An eye-tracking experiment. In: Zybatov, L.N.; Petrova, A. & Ustaszewski, M. (Hg.) *Proceedings of the 1st international conference TRANSLATA, translation & interpreting research: yesterday - today – tomorrow*. Frankfurt am Mein: Peter Lang, 341-347.

Seeber, Kilian G. (2017). Multimodal processing in simultaneous interpreting. In: Schwieter, John W. & Ferreira, Aline (Hg.) *The handbook of translation and cognition.* Hoboken: John Wiley & Sons, 461-475.

Seeber, Kilian G. & Delgado Luchner, Carmen (2020). Simulating simultaneous interpreting with text: From training model to training module. In: Dolores Rodríguez Melchor, María; Horváth, Ildikó & Ferguson, Kate (Hg.) *The role of technology in conference interpreter training.* New York: Peter Lang, 129-151.

Seeber, Kilian G. & Kerzel, Dirk (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism* 16 (2), 228-242.

Seubert, Sabine (2019). *Visuelle Informationen beim Simultandolmetschen: Eine Eyetracking-Studie.* Berlin: Frank & Timme.

SmarTerp (o.J.). Smarterp&me. https://smarterp.me/ (Stand: 06.04.2024).

SR Research (2017). EyeLink® Portable Duo user manual. https://www.manualslib.com/manual/2114770/Sr-Research-Eyelink-Portable-Duo.html (Stand: 06.04.2024).

SR Research (2023). *EyeLink® Data Viewer user’s manual*. Oakville/Ontario: SR Research.

Stachowiak-Szymczak, Katarzyna (2019). *Eye movements and gestures in simultaneous and consecutive interpreting*. Warschau: Springer.

Statista (2023). Statistiken zu Fahrrädern in Österreich. https://de.statista.com/themen/2775/fahrraeder-in-oesterreich/#topicOverview (Stand: 06.04.2024).

Statistik (2023). Insgesamt 5,15 Mio. Pkw in Österreich zugelassen. https://www.statistik.at/fileadmin/announcement/2023/02/20230224KfzBestand2022.pdf (Stand: 06.04.2024).

Sugomori, Yusuke; Kaluza, Bostjan; Soares, Fabio M. & Souza, Alan M. F. (2017). *Deep earning: Practical neural networks with Java*. Birmingham: Packt.

Szilas, Benjamin (2023). *Computer-assisted interpreting and automatic speech recognition: Effects of the CASSIS software on simultaneous interpreting.* Masterarbeit, Universität Wien.

Terminology Coordinator (o.J.). Terminology management tools for interpreters. https://termcoord.eu/interpretation/terminology-management-tools-for-interpreters/ (Stand: 06.04.2024).

Trenkwalder, Barbara (2020). *Die Bedeutung von CAI-tools in der Dolmetschdisziplin. Eine Untersuchung des ASR-features von InterpretBank 6 und der Situation Bezüglich CAI an der Universität Innsbruck*. Masterarbeit, Universität Innsbruck.

Uran, Simone (2022). *Multimodale Verarbeitung beim Simultandolmetschen mit (bewegtem) visuellen Text-Input: Ein Experiment im Sprachenpaar Englisch-Deutsch.* Masterarbeit, Universität Wien.

Vogler, Nikolai & Stewart, Craig & Neubig, Graham (2019). Lost in Interpretation: Predicting untranslated terminology in simultaneous interpretation. https://ArXiv:1904.00930 (Stand: 06.04.2024).

Wickens, Christopher D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3 (2), 159-177.

Will, Martin (2015). Zur Eignung simultanfähiger Terminologiesysteme für das Konferenz-dolmetschen. *Trans-kom* 8 (1): 179-201.

Will, Martin (2020). Computer Aided Interpreting (CAI) for conference interpreters. Concepts, content and prospects. *Essachess* 25 (1), 37-71.

Winteringham, Sarah Tripepi (2010). The usefulness of ICTs in interpreting practice. *The Interpreters' Newsletter* 15, 87-99.

Xu, Ran (2015). *Terminology preparation for simultaneous interpreters.* PhD Thesis, University of Leeds.

Yu, Dong & Deng, Li (2015). *Automatic speech recognition: A Deep Learning approach*. London: Springer.

Anhang

Anhang I: Ausgangsrede

Dear participant, I would like to welcome you kindly to our experiment today and I hope you are familiar with the topic of mobility, because today we will talk about mobility in the present and in the future. I will talk about some recent trends referring to cars and bikes in Austria, and how these are connected to rising environmental awareness.

Means of transportation play an important role in climate policy.

By **2050** a regulation of the European Commission will ban cars with combustion engines from the city centers, in an effort to reach the goal of zero-emissions.

Climate neutrality has become a high priority in the legislation of many European countries, and new regulations and initiatives on this topic are being implemented to reach the goals. The proposal of banning cars from the city centers during the weekend is an example that has already been brought into force in many European cities.
Another possibility is represented by car sharing companies:

In **2022** in Austria there were already **110** locations in **305** cities with more than **7000** vehicles.

Car sharing is an alternative to enjoy the advantages of driving a car, without having to own one. It can be more cost-effective than owning a car, because users share the price of fuel, and it reduces the number of cars on the road.
One of the most renowned companies in Austria is Car2go.

At the end of this year, only **4** years after it was founded in **2019**, the company registered worldwide more than **10.000** smarts on the streets, of which **1.000** were e-smarts, and almost **half a million** regular users.

Clients describe car sharing apps as convenient and flexible instruments, particularly for those who do not need a car on a daily basis and may only need one to come home at night when public transportation is not available anymore.
Private cars remain still one of the most used means of transportation, as can be observed from recent data.

On the **31st** December **2022**, **5.51 million** vehicles were registered in Austria - these are **0,3%**, or **12.054** cars more than the previous year.

This shows a rising tendency among car owners, contrary to the belief that Austrians prefer bikes. Austria, as a country, has made many efforts in implementing its infrastructure to promote the use of bikes. Still, many prefer to own a private car for the daily commute. But would you guess how many cars there are per inhabitant?

In Austria, there are **565** cars per **1.000** inhabitants.

Austria, like many other countries, has a diverse population, and while some bigger cities like Graz or Vienna may offer many alternative means of transportation, there are also country areas where cars are still necessary for many everyday activities.

Burgenland has the most cars per **1,000** inhabitants (**679**) and the highest level of motorization, followed by Lower Austria (**655**), Carinthia (**654**), Upper Austria (**639**), Styria (**621**) and Salzburg (**569**).

Still, Austria is worldwide renowned as the country of bikes, and statistics about bicycle users are impressive. The rising climate awareness and the call for sustainability are having a considerable impact on the mobility sector, and bikes are a sustainable alternative to travel in the city.

Around **68** percent of the Austrian population owns a bike.

This number is very high compared to other European countries, which can be seen as proof that bikes, in fact, enjoy great popularity in Austria. Cycling is not only environmentally sustainable, but also healthy, particularly for those that spend many hours in front of the computer during working hours.

Twice as many new bicycles were purchased as cars in **2021**.

A keyword for the future will be sharing. Whether car sharing or bike sharing, this trend is efficient and practical, and Vienna is already playing a pioneering role in this area.

Whether the European goals for **2025** and **2050** will be met, this question cannot be answered now. For sure the metro lines, the **82** tramway lines and the **131** bus lines are a big step towards public transportation for the mobility of the future.

How the capital city of Austria will look like in the next few years, this is a question that remains unanswered, and pivotal will be the role that every single citizen plays. Do not forget about this the next time you will choose your bike or a bus instead of a car.

[732 Wörter]

Anhang II: Einleitung des Videos

Slide 1

Benvenuta all'esercitazione pratica!

Stai per interpretare un video con la combinazione linguistica EN>IT

La durata del video è di circa 5 minuti

A sinistra dello schermo vedrai l'oratore

A destra vedrai i numeri riconosciuti automaticamente dal programma *Augmented Interpreter CAI v. 1.1*

Puoi scegliere se far riferimento al riconoscimento automatico dei numeri sullo schermo o meno, ma non potrai appuntare nulla su carta o in altro modo.

Il titolo del video e la terminologia con traduzione ti saranno mostrati nella prossima slide per la durata di 1 minuto, e potrai far riferimento solo alle informazioni sulla slide. La terminologia è da vedere come un suggerimento e non è in alcun modo vincolante.

Slide 2

Queste informazioni sul tema e la terminologia saranno disponibili per 1 minuto:

Titolo: **Mobility now and in the future**

Terminologia:

Combustion engine	Motore a combustione[1]
Car2go	(servizio di car sharing che offre la possibilità di condividere un'auto, noleggiandola e pagandola al minuto)[2]
Level of motorisation	Grado di motorizzazione[3]
Regulation	Regolamento[4]
Zero-emission	Emissioni zero[5]
Fuel	Carburante[6]
Commute	Traffico pendolare[7]
Environmental awareness	Consapevolezza ambientale[8]

[1] Fonte: IATE
[2] Fonte: webnews.it
[3] Fonte: IATE
[4-8] Fonte: IATE

Anhang III: Zeitstempel-Tabelle

Zahlen im Ausgangstext	ASR	Zeitstempel – Audio	Zeitstempel – Video	Timelag
2050	2.050	02:11.20	02:11.60	00:00.80
2022	2.022	03:01.07	03:01.83	00:00.76
110	110	03:05.07	03:05.91	00:00.84
305	305	03:07.55	03:08.54	00:00.99
7.000	7 // 000	03:11.04	03.11.98	00:00.94
	1	03:21.57	03:22.23	00:00.66
	1	03:35.79	03:36.41	00:00.62
	2	03:40.05	03:40.75	00:00.70
4	4	04:45.15	03:45.73	00:00.58
2019	2.019	03:49.81	03:50.36	00:00.55
10.000	10 // 000	03:55.00	03:55.66	00:00.66
1.000	1 // 000	03:58.90	04:00.28	00:01.38
31st December 2022	31 // 2.022	04:40.63	04:41.23	00:00.60
5.51 million	5.50 // 1_million	04:44.60	04:45.36	00:00.76
0,30%	3% // 0	04:50.85	04:51.21	00:00.36
12.045	12 // 054	04:54.79	04:55.23	00:00.44
565	5 // 00 >> 560 >> 565	05:39.19	05:39.33	00:00.14
1.000	1000 >> 1 // 000	05:41.88	05:42.66	00:00.78
1.000	1 // 000	06:10.17	06:11.01	00:00.84
679	679	06:13.17	06:13.46	00:00.29
655	655	06:22.04	06:22.39	00:00.35
654	654	06:26.28	06:26.59	00:00.31
639	639	06:29.89	06:30.26	00:00.37
				Durchschnitt = 0,705 Sekunde

Anhang IV: Ergebnisse der Dolmetschleistungen

AUSGANGSTEXT					P1a			P1b			P2			P3		
Z	K	R	Z	K	P	R	P	Z	K	P	R	Z	K	P	R	P
		regulation of the European Commission will ban		AU	0		0	2050	KW	1	un regolamento dell'Unione Europea ha	2050	KW	1	la Commissione europea prevede di	AU		0		0
2050	2	cars		KW	1		0	2022	KW	0	vietato	2022	KW	1	bannare	2022	KW	1		0
2022	3	in Austria	2022	KW			1			1				1	automobili-stiche aziende					
110	2	locations	110	KW	1	posti	1	110	KW	1	posti	110	KW	1		0	più	AU	0	compagnie	0
305	2	cities		AU	0		0	305	KW	1	città	305	KW	1		0	più	AU	0	città	1
7.000	2	vehicles	7.000	KW	1	veicoli	1	7.000	KW	1	veicoli		AU	1		0		AU	0		0
4	1	years	4	KW	1	anni	1	4	KW	1	anni	4	KW	1	anni	1		AU	0		0
2.019	3	founded in	2.019	KW	1	invenzione	1	2.019	KW	1	fondazione	2.019	KW	1	nascita	1	2.019	KW	1	utilizzi di car	0
10.000	2	smarts on the streets	10.000	KW	1	macchine smart	1	10.000	KW	1	smart macchine in strade	10.000	KW	1	auto	1	10.000	KW	1	sharing	0
1.000	2	e-smarts		AU	0		0	1.000	KW	1	e-smarts	1.000	KW	1		0		AU	0		0
half a million	2	regular users	500 mila	KW	1	utenti	1	500 mila	KW	1	utenti	mezzo milione	KW	1	utenti	1		AU	0		0
31st December 2022	4	registered vehicles	31-dic-22	KW	1	registrato	1	31-dic-22	KW	1	registrati	nel 31 dicembre del 2022	KW	1	registrati	1	entro il 31 dicembre 2022	KW	1	ci sono stati	1
5,51 million	4	vehicles	5,51 milioni	KW	1	macchine private	1	**5.51 milioni**	KW	1	veicoli	**5.1 milioni**	LF	1	veicoli	0	più di 5 milioni	AN	0	registrazioni	0
0,30%	3	cars		AU	0	più delle	1	0,30%	KW	1		più di 12.000	AF	1	auto	0		AU	0		0
12.045	4	cars	più	AU	0		0	12.044	AN	0,5		auto	AN	0,5	auto	1	12.045	KW	1		0
565	3	cars	565	KW	1	macchine	1	565	KW	1	automobili	565	KW	1	auto	1	565	KW	1		0
1.000	2	inhabitants	1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1.000	KW	1	abitanti	1		AU	0	abitanti	1
1.000	2	inhabitants		AU	0		0	1.000	KW	1	abitanti	1.000	KW	1	abitanti	1		AU	0	abitanti	1
679	3	Burgenland		AU	0		0	679	KW	1	Burgenland	679	KW	1	particolare una città	0	679	KW	1	Burgenland	1
655	3	Lower Austria		AU	0		0	655	KW	1	Bassa Austria		AU	0		0	655	KW	1	Bassa Austria	1
654	3	Carinthia		AU	0		0	654	KW	1	Carinzia	654	KW	1	Carinzia	1	654	KW	1	Carinzia	1
639	3	Upper Austria	639	KW	1	Alta Austria	1	639	KW	1	Alta Austria	639	KW	1	Austria superiore	1	639	KW	1	Alta Austria	1
621	3	Styria	579	AF	1		0	**Stiria-Stiria**	1	631	LF	1	Stiria	621	KW	1	Stiria	1			0
569	3	Salzburg		AU	0		0	659	TR	0	Salisburgo popolazione austriaca	1	569	KW	1	**Salzburg, Salisburgo** popolazione austriaca	1	500	AN	0,5	Salzburg	1
68%	3	Austrian population	68%	KW	1		1	68%	KW	1		il 6, il 60%	AN	0,5		1	68%	KW	1	popolazione	1
twice	1	bikes as cars	doppio	KW	0		1	doppio	KW	1	biciclette rispetto alle auto	più	AU	1	biciclette rispetto alle automobili	1		AU	0		0
2021	3	purchased	2021	KW	1	biciclette vendute	1	2021	KW	1	acquistate	2022	LF	1	acquistate	0		AU	0		0
2025	3	European goals	2025	KW	0		1	2025	KW	1	obiettivi	2024	LF	1	obiettivi	0	nel 2025 anche nel 2050	KW	1		1
2050	2	European goals		AU	0		0	2050	KW	1			AU	0		0					
												superano dei numeri incredibili	AU	0		0		AU	0		0
82	2	tramway lines		AU	0		0			0				0	linee tram	0		AU	0		0
131	3	bus lines	131	KW	0		0		AU	1	tratte dei bus		AU	1	i bus	0		AU	0		0
SUMME					**16**		**15**			**26,5**				**28**		**19**			**23**		**14**		**12**

P4					P5					P6					P7a				
Z	K	P	R	P	Z	K	P	R	P	Z	K	P	R	P	Z	K	P	R	P
2050	KW	1	non si potranno più usare le macchine	1	2050	KW	1	un regolamento della Commissione Europea impedirà	1	2050	KW	1	un regolamento dell'Unione Europea eliminerà	1	2050	KW	1	c'è un regolamente che proibirà	1
2022	KW	1		0	2022	KW	1	in Austria	1	2022	KW	1	in Austria	1	2022	KW	1	in Austria	1
110	KW	1	car sharing	1	110	KW	1	luoghi	1	110	KW	1	luoghi	1	110	KW	1	locations	1
tantiss- più di 300	AN	0,5	cittá	1	305	KW	1	cittá	1	305	KW	1	cittá	1	305	KW	1	cittá	1
	AU	0		0	7.000	KW	1	veicoli di car sharing	1	7.000	KW	1	veicoli	1	7	SF	0	possibilitá	0
4	KW	1	anni	1	4	KW	1	anni	1	4	KW	1	anni	1	4	KW	1	anni	1
2.019	KW	1	fondata	1	2.019	KW	1	fondazione	1	2.019	KW	1	fondazione	1	2.019	KW	1	fondato	1
10.000	KW	1	macchine	1	10.000	KW	1	macchine	1	10.000	KW	1	macchine	1	10.000	KW	1	macchine	1
1.000	KW	1	di e-mobilitá	1	1.000	KW	1	smart elettriche	1	1.000	KW	1		0	1.000	KW	1	e-smarts	1
più di 500	SF	0	persone le usano	1	più di mezzo milione	KW	1	utenti	1		AU	0		0	mezzo milione	KW	1	utenti	1
31-dic-22	KW	1	I dati dicono	1	31-dic-22	KW	1	stati registrati	1	il 31 dicembre del 2022	KW	1	stati registrati	1	nel 31 dicembre del 2022	KW	1	utilizzati	1
5.51 milioni	KW	1	persone	0	5 virgola, 5 milioni virgola 50	KW	1	veicoli	1	5,5	SF	0	veicoli	1	550 milioni	SF	0	macchine	1
3,00%	SF	0	macchine	1	3,00%	SF	0	questo corrisponde a	1	0,30%	KW	1	macchine	1	3,00%	SF	0		1
12.045	KW	1	macchine	1	12.045	KW	1	macchine	1	12 mila	AN	0,5		1	12.045	KW	1	macchine	1
565	KW	1	macchine	1	565	KW	1	macchine	1	565	KW	1	macchine	1	500 e... 65	KW	1	macchine	1
1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1
1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1	1.000	KW	1	abitanti	1
tantissime	AU	0		0	679	KW	1	Burgenland	1	679	KW	1	Burgenland	1	679	KW	1	Burgenland	1
sempre circa 600	AU	0	Austria Bassa	1	655	KW	1	Bassa Austria	1	655	KW	1	Bassa Austria	1	655	KW	1	Bassa Austria	1
	AU	0	Carinzia	1	654	KW	1	Carinzia	1	654	KW	1	Carinzia	1	654	KW	1	Carinzia	1
	AU	0	Austria Alta	1	639	KW	1	Alta Austria	1	639	KW	1	Alta Austria	1	639	KW	1	Alta Austria	1
	AU	0	Stiria	1	621	KW	1	Stiria	1	un altro numero	AU	0	Stiria	1		AU	0	Stiria	1
	AU	0	Salisburgo	1	569	KW	1	Salisburgo	1		AU	0	Salzburg	1	ancora di piú	AU	0		0
68%	KW	1	abitanti dell'Austria	1	Più del 60%	AN	0,5	austriaci	1		AU	0		0	68%	KW	1	popolazione austriaca	1
piú	AU	0	bici che macchine	1	doppio	KW	1	biciclette rispettp al numero di macchine	1		AU	0		0	doppio	KW	1	delle biciclette	0
2021	KW	1	comprate	1	2021	KW	1	acquistate	1		AU	0		0	2021	KW	1	acquisiti	1
	AU	0		0	2025	KW	1	obiettivi	1	2024	LF	0	obiettivo europeo	1	2025	KW	1	obiettivi dell'UE	1
2050	KW	1	obiettivi	1	2050	KW	1	obiettivi europei	1	2050	KW	1		1	2050	KW	1		1
82 (eighty-ottantadue)	KW	1	tram	1	82	KW	1	linee del tram	1		AU	0		0	82	KW	1	possibilitá di tram	1
	AU	0		0	piú di 100	AN	0,5	linee degli autobus	1		AU	0		0	139	LF	0		0
		17,5		**24**			**28**		**30**			**19,5**		**23**			**24**		**26**

P7b					P8					P9					TOT Zahle	TOT Refer
Z	K	P	R	P	Z	K	P	R	P	Z	K	P	R	P		
2050	KW	1	un regolamento della Commissione europea proibirà	1	2050	KW	1	un regolamento della Commissione europea vorrà eliminare	1	2050	KW	1	un regolamento della Commissione europea	1	7	7
2022	KW	1	in Austria	1	2022	KW	1	in Austria	0	2022	KW	1	in Austria	1	9	4
110	KW	1	locations	1	120	AN	0	punti	1	110	KW	1	luoghi	1	7	7
305	KW	1	cittá	1	305	KW	1	cittá	1	305	KW	1	cittá	1	6,5	7
7.000	KW	1	veicoli	1		AU	0	macchine a noleggio	1	700	SF	0	postazioni	0	3	4
4	KW	1	anni	1	4	KW	1	anni	1	4	KW	1	anni	1	8	8
2.019	KW	1	fondata	1	2.019	KW	1	fondata	1	2.019	KW	1	nascita	1	9	8
10.000	KW	1	smart	1	10.000	KW	1	macchine	1	numero importante di	AU	0	macchine	1	8	8
1.000	KW	1	e-smarts	1		AU	0		0		AU	0		0	5	3
mezzo milione	KW	1	user	1	mezzo milione	KW	1	partecipanti, persone che hanno usato	1		AU	0	utenti	1	5	7
31-dic-22	KW	1	registrati	1	il 31 dicembre 2022	KW	1	state registrate	1	Il 31 dicembre 2022	KW	1	stati registrati	1	9	9
5.50 million	KW	1	veicoli	1	5 milioni, 51 milioni	KW	1	auto	1	**5.51 (punto) milioni**	KW	1	veicoli	1	5,5	7
0,30%	KW	1	macchine	1		AU	0		0	0,30%	KW	1	auto	1	2	7
12.055	AN	0,5		1	numero piú grande	AU	0		0	12.000	AN	0,5		1	5,5	6
565	KW	1	**macchina**	1	565	KW	1	macchine	1	500...565	KW	1	auto	1	9	9
1.000	KW	1	abitanti	1	1.000	KW	1	persone	1	1.000	KW	1	abitanti	1	8	9
1.000	KW	1	abitanti	1		AU	0	quantitá di abitanti	1	1.000	KW	1	abitanti	1	6	8
679	KW	1	Burgenland	1		AU	0	Burgenland	1	679	KW	1	Burgenland	1	6	6
655	KW	1	Bassa Austria	1	665	LF	0	Bassa Austria	1	655	KW	1	Bassa Austria	1	5	7
654	KW	1	Carinzia	1	654	KW	1	Carinzia	1	654	KW	1	Carinzia	1	7	8
639	KW	1	Austria	0		AU	0		0	639	KW	1	Alta Austria	1	7	8
621	KW	1	Stiria	1	621	KW	1	Stiria	1	621	KW	1	Stiria	1	4	8
569	KW	1	Salzburgo	1	569	KW	1	Salisburgo	1	circa 600	AN	0,5	Salisburgo	1	5	7
68%	KW	1	popolazione austriaca	1	praticamente ogni persona	AU	0	In Austria	1	68%	KW	1	la popolazione austriaca	1	6	8
doppio	KW	1	nuove biciclette rispetto alle macchine	1	doppio	KW	1	biciclette rispetto alle macchine	0	doppio	KW	1	delle biciclette rispetto alle macchine	1	5	4
2021	KW	1	acquisite	1	2021	KW	1	acquistate	1	2021	KW	1	acquistate	1	6	7
2025	KW	1	obiettivi dell'UE	1	2025	KW	1	obiettivi dell'Unione Europea	1	2025	KW	1	obiettivi	1	6	6
2050	KW	1		1	2050	KW	1		1	2050	KW	1		1	7	8
	AU	0		0		AU	0	rete dei trasporti	0		AU	0		0	3	4
	AU	0		0		AU	0		0		AU	0	dei pullman	1	0,5	3
		27,5		**27**			**18**		**22**			**23**		**27**		

Anhang V: Zusammengefasste Ergebnisse der Dolmetschleistungen

	AUSGANGSTEXT		P1a		P2		P3		P4	
	Zahlen	Kategorie	Kategorie	Punkte	Punkte	Kategorie	Punkte	Kategorie	Punkte	Punkte
1	2050	2	AU	0	0	KW	1	AU	0	KW	1	1
2	2022	3	KW	1	0	KW	1	KW	1	KW	1	0
3	110	2	KW	1	1	KW	1	AU	0	KW	1	1
4	305	2	AU	0	0	KW	1	AU	0	AN	0,5	1
5	7.000	2	KW	1	1	AU	0	AU	0	AU	0	0
6	4	1	KW	1	1	KW	1	AU	0	KW	1	1
7	2.019	3	KW	1	1	KW	1	KW	1	KW	1	1
8	10.000	2	KW	1	1	KW	1	KW	1	KW	1	1
9	1.000	2	AU	0	0	KW	1	AU	0	KW	1	1
10	half a million	2	KW	1	1	KW	1	AU	0	SF	0	1
11	31st December 2022	4	KW	1	1	KW	1	KW	1	KW	1	1
12	5.51 million	4	KW	1	1	LF	0	AN	1	KW	0	0
13	0,30%	3	AU	0	1	AF	0	AU	1	SF	0	1
14	12.045	4	AU	0	0	AN	0,5	KW	1	KW	1	1
15	565	3	KW	1	1	KW	1	KW	1	KW	1	1
16	1.000	2	KW	1	1	KW	1	AU	1	KW	1	1
17	1.000	2	AU	0	0	KW	1	AU	0	KW	1	1
18	679	3	AU	0	0	KW	1	KW	1	AU	0	0
19	655	3	AU	0	0	AU	0	KW	1	AU	0	1
20	654	3	AU	0	0	KW	1	KW	1	AU	0	1
21	639	3	KW	1	1	KW	1	KW	1	AU	0	1
22	621	3	AF	0	1	KW	1	AU	0	AU	0	1
23	569	3	AU	0	0	KW	1	AN	0,5	AU	0	1
24	68%	3	KW	1	1	AN	0,5	KW	1	KW	1	1
25	twice	1	KW	1	0	AU	0	AU	0	AU	0	1
26	2021	3	KW	1	1	LF	0	AU	0	KW	1	1
27	2025	3	KW	1	0	LF	0	KW	1	AU	0	0
28	2050	2	AU	0	0	AU	0	KW	1	KW	1	1
29	82	2	AU	0	0	AU	0	AU	1	KW	1	1
30	131	3	AU	0	0	AU	0	AU	1	AU	0	0
SUMME				**16**	**15**		**19**		**23**		**12**		**14**		**17,5**	**24**

P5			P6			P7a			P8			P9		
Kategorie	Punkte	Punkte	Kategorie	Punkte	Punkte	Kategorie	Punkte	Punkte	Kategorie	Punkte	Punkte	Kategorie	Punkte	Punkte
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	0	KW	1	1
KW	1	1	KW	1	1	KW	1	1	AN	0	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	SF	0	0	AU	0	1	SF	0	0
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	AU	0	1
KW	1	1	KW	1	0	KW	1	1	AU	0	0	AU	0	0
KW	1	1	AU	0	0	KW	1	1	KW	1	1	AU	0	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	SF	0	1	SF	0	1	KW	1	1	KW	1	1
SF	0	1	KW	1	1	SF	0	1	AU	0	0	KW	1	1
KW	1	1	AN	0,5	1	KW	1	1	AU	0	0	AN	0,5	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	AU	0	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	AU	0	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	LF	0	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	AU	0	0	KW	1	1
KW	1	1	AU	0	1	AU	0	1	KW	1	1	KW	1	1
KW	1	1	AU	0	1	AU	0	0	KW	1	1	AN	0,5	1
AN	0,5	1	AU	0	0	KW	1	1	AU	0	1	KW	1	1
KW	1	1	AU	0	0	KW	1	0	KW	1	0	KW	1	1
KW	1	1	AU	0	0	KW	1	1	KW	1	1	KW	1	1
KW	1	1	LF	0	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	KW	1	1	KW	1	1	KW	1	1	KW	1	1
KW	1	1	AU	0	0	KW	1	1	AU	0	0	AU	0	0
AN	0,5	1	AU	0	0	LF	0	0	AU	0	0	AU	0	1
	28	**30**		**19,5**	**23**		**24**	**26**		**18**	**22**		**23**	**27**

Anhang VI: Eyetracker Aggregate IA Report

IA_LABEL	IA_TYPE	IA_ID	IA_AREA	IA_DWELL_TIME	IA_DWELL_TIME_%	IA_FIXATION_COUNT	IA_FIX_COUNT_%	IA_FIRST_FIXATION_DURATION	IA_VISITED_TRIAL_%	TRIALS_IN_GROUP	TRIALS_USING_THIS_IA
IA_Gesicht 1	ELLIPSE	3	111919,2	85053,34	18,92	118,67	12,56	1383,33	100	6	6
IA_Gesicht 2	ELLIPSE	4	111919,2	96419,66	21,44	137,67	14,57	730,33	100	6	6
IA_2050	RECTANGLE	6	31007	888,33	0,2	2,67	0,28	379,6	83,33	6	6
IA_2022	RECTANGLE	7	28809	450,33	0,1	1,33	0,14	330	66,67	6	6
IA_110	RECTANGLE	8	23256	511,67	0,11	1,33	0,14	388	50	6	6
IA_305	RECTANGLE	9	28304	265,33	0,06	0,83	0,09	383,33	50	6	6
IA_7000	RECTANGLE	10	43338	1296,33	0,29	2,5	0,26	1145	66,67	6	6
IA_1	RECTANGLE	11	14388	391	0,09	1,17	0,12	270	50	6	6
IA_2	RECTANGLE	12	11187	170,33	0,04	0,67	0,07	226	33,33	6	6
IA_Hand 4	RECTANGLE	1	108240	187,67	0,04	0,33	0,04	563	33,33	6	6
IA_4	RECTANGLE	13	16884	374,33	0,08	1,33	0,14	381	66,67	6	6
IA_2019	RECTANGLE	14	23318	616,67	0,14	1,5	0,16	515,33	50	6	6
IA_10000	RECTANGLE	15	33734	930,67	0,21	2	0,21	585,33	50	6	6
IA_1.000	RECTANGLE	16	12257	258	0,06	0,5	0,05	643	33,33	6	6
IA_31	RECTANGLE	17	15855	402,67	0,09	0,67	0,07	604	66,67	6	6
IA_2022	RECTANGLE	18	33335	1132,33	0,25	1,17	0,12	1489	66,67	6	6
IA_5,5	RECTANGLE	19	53448	2846	0,63	6	0,63	470,4	83,33	6	6
IA_3,0	RECTANGLE	20	37370	599,33	0,13	1,33	0,14	450	66,67	6	6
IA_12054	RECTANGLE	21	39140	2431,33	0,54	7,17	0,76	323,6	83,33	6	6
IA_565	RECTANGLE	22	39996	2818,33	0,63	5,67	0,6	616,4	83,33	6	6
IA_1000	RECTANGLE	23	24208	1796,67	0,4	5,17	0,55	364,5	66,67	6	6
IA_1000	RECTANGLE	24	30380	138,33	0,03	0,33	0,04	415	33,33	6	6
IA_679	RECTANGLE	25	30150	1584,33	0,35	3,33	0,35	415,5	66,67	6	6
IA_655	RECTANGLE	26	26082	816,67	0,18	1,17	0,12	483,5	66,67	6	6
IA_654	RECTANGLE	27	26270	458	0,1	0,83	0,09	308	33,33	6	6
IA_639	RECTANGLE	28	27010	1286,33	0,29	1,83	0,19	856	66,67	6	6
IA_Gesicht_Ende	ELLIPSE	5	108875	104259,7	23,19	141	14,92	286,67	100	6	6
IA_Ende_ASR	RECTANGLE	29	107590	8699,67	1,93	25,5	2,7	871,67	100	6	6

Anhang VIII

P3: Per raggiungere l'obiettivo di emissioni zero. La politica ambientale è diventata molto importante per l'Unione Europea e ci sono molte iniziative per raggiungere gli obiettivi prefissati. Per esempio, si può evitare di usare le auto nel weekend, specialmente nelle città più importanti dell'Unione Europea.

Un altro esempio è il car sharing.

E entro il **2022** ci dovranno essere più compagnie per questo e molte più città.
Il car sharing consiste nel usare soltanto un'auto, ma da più persone. Ed è molto più conveniente perché gli utilizzatori possono dividere il prezzo per quanti sono.

Ci sono molte compagnie in Austria, per questo obiettivo. E molte.

E dopo il **2019**, queste compagnie hanno registrato più di **10.000** utilizzi di car sharing.

E si può diventare anche dei utilizzatori frequenti di questo servizio. Possono essere degli strumenti molto utili e anche flessibili. E possono richiedere soltanto... Possono richiedere anche meno carburante di tutte le auto.

Dagli ultimi dati, entro il **31** dicembre **2022**, ci sono stati più di **5 milioni** di registrazioni in Austria. E **12.054** in più dell'anno precedente.

Questo vuol dire che ci sono più proprietari di auto rispetto agli anni passati. L'Austria ha fatto molti sforzi, anche dal punto di vista infrastrutturale, sotto questo punto di vista.
Ma le persone continuano a volere un'auto di proprietà. Ma adesso, quante auto hanno ogni abitante dell'Austria?

Ci sono **565** auto in Austria per abitanti.

Ma ci sono anche delle città, o magari come Vienna, che offrono più possibilità alternative rispetto all'auto di proprietà.

Il Burgenland ha il numero più importante, il numero più alto di auto, ovvero **679** per abitanti.
Mentre nella bassa Austria **655**. In Carinzia, **654**. Nell'alta Austria, **639**.
In Salzburg, **500** per abitanti.

Inoltre l'Austria è famosa anche per le biciclette. I numeri di biciclette sono impressionanti. Stanno crescendo sempre di più e hanno un impatto importante sul settore della mobilità.

Circa il **68%** della popolazione ha una bicicletta

e questo numero è molto alto rispetto ad altri paesi dell'Unione Europea. Questo vuol dire che le biciclette sono molto usate in Austria e piacciono molto alle persone. Specialmente per le persone

che magari lavorano in ufficio, spendono molte ore davanti a uno schermo, apprezzano molto muoversi in bicicletta.

Una parola chiave per il futuro dovrebbe essere la condivisione, che sia per auto o biciclette. Condividere è efficiente e pratico e può avere un ruolo molto importante in questo settore.

Dal **2000**… che possa essere raggiunto nel **2025** e anche nel **2050** ci sono molti obiettivi da raggiungere riguardo il settore della mobilità per il futuro.

Come sembrerà la capitale dell'Austria nel futuro, nei prossimi anni? Questo è una domanda che può essere risposta e può essere utile anche per il futuro. Vi auguro di usare una bicicletta invece della macchina.

Anhang VII: Eyetracker IA Output Report P3

IA_LABEL	IA_AVERAGE_FIX_PUPIL_SIZE	IA_DWELL_TIME	IA_DWELL_TIME_%	IA_FIRST_FIXATION_DURATION	IA_FIXATION_%	IA_FIXATION_COUNT	IA_REGRESSION_IN	IA_REGRESSION_IN_COUNT	TRIAL_DWELL_TIME	TRIAL_FIXATION_COUNT	TRIAL_IA_COUNT
IA_Gesicht 1	1633,13	75360	0,1836	3012	0,1086	112	1	8	410540	1031	28
IA_Gesicht 2	1479,86	60746	0,148	732	0,097	100	1	19	410540	1031	28
IA_Gesicht Ende	1464,02	71470	0,1741	248	0,1183	122	1	20	410540	1031	28
IA_2050	1620,17	1700	0,0041	482	0,0058	6	0	0	410540	1031	28
IA_2022	1821	1286	0,0031	412	0,0039	4	0	0	410540	1031	28
IA_110	1878,5	652	0,0016	240	0,0019	2	0	0	410540	1031	28
IA_305	1833	584	0,0014	398	0,0019	2	0	0	410540	1031	28
IA_7.000	1869,43	2808	0,0068	1320	0,0068	7	0	0	410540	1031	28
IA_1	1721	232	0,0006	232	0,001	1	0	0	410540	1031	28
IA_2	1731,5	394	0,001	160	0,0019	2	0	0	410540	1031	28
IA_Hand 4	.	0	0	.	0	0	.	.	410540	1031	28
IA_4	1645,67	668	0,0016	404	0,0029	3	0	0	410540	1031	28
IA_2019	1623,75	2106	0,0051	958	0,0039	4	0	0	410540	1031	28
IA_10.000	1637	2416	0,0059	670	0,0029	3	0	0	410540	1031	28
IA_1.000	.	0	0	.	0	0	.	.	410540	1031	28
IA_31	1521	720	0,0018	720	0,001	1	0	0	410540	1031	28
IA_2022	1559	3100	0,0076	3100	0,001	1	0	0	410540	1031	28
IA_5,51 million	1426,78	4716	0,0115	798	0,0087	9	0	0	410540	1031	28
IA_0,3%	1465	1196	0,0029	334	0,0019	2	0	0	410540	1031	28
IA_12.054	1421,72	5582	0,0136	302	0,0175	18	0	0	410540	1031	28
IA_565	1352,5	3806	0,0093	738	0,0097	10	0	0	410540	1031	28
IA_1.000	1429,88	5228	0,0127	410	0,0165	17	0	0	410540	1031	28
IA_1.000	.	0	0	.	0	0	.	.	410540	1031	28
IA_679	1545,5	2494	0,0061	750	0,0039	4	0	0	410540	1031	28
IA_655	1428	290	0,0007	290	0,001	1	0	0	410540	1031	28
IA_654	1386,67	2202	0,0054	376	0,0029	3	0	0	410540	1031	28
IA_639	1379	2928	0,0071	2558	0,0019	2	0	0	410540	1031	28
IA_ASR Ende	1421,64	22732	0,0554	458	0,0718	74	0	0	410540	1031	28

Anhang X

P4: Cari colleghi, voglio darvi un benvenuto a questo esperimento di oggi e spero che vi piaccia la mobilità. Oggi parleremo della mobilità, adesso e nel futuro, parlo di macchine e biciclette in Austria e come sono connessi all'ambientale, alla responsabilità.

È molto importante per l'ambiente e entro il **2050** non si potranno più usare le macchine al centro per avere zero emissioni.

L'ambiente, la responsabilità per l'ambiente è diventata una cosa molto importante per tanti paesi dell'Unione Europea. Si investe molto per questo. Durante i fine settimana, per esempio, non si possono più usare le macchine in tante città dell'Unione... dell'Europa.
Un'altra possibilità sono le compagnie di car sharing.

Nel **2022** c'erano già **110** car sharing in tantiss- più di **300** città.

Carsharing è la possibilità di usare una macchina senza averla ed è più economico perché le persone usano una macchina insieme, cioè.
Una compagnia molto conosciuta è Car2Go.

È stata fondata nel **2019** e ora solo **quattro** anni dopo ci sono **10 mila** macchine e **mille** di questi sono di e-mobilità e più di 500 persone le usano

ed è una possibilità per quelli che non hanno bisogno di una macchina ogni giorno però hanno bisogno di una macchina solo quando non c'è il traffico, hanno la possibilità di usare altri medi di mobilità.

I dati dicono che il **31** dicembre **2022 5.51 (cinque punto cinque uno) milioni** di persone, cioè **3%** o **12.054** macchine sono state usate in questi anni,

cioè sempre più persone hanno delle macchine. In Austria tante persone vogliono usare la bicicletta e infatti in Austria ci sono tante possibilità per quelli che guidano o vanno in bici. Però sai quante macchine esistono per persona?

In Austria ci sono **565** macchine per **mille** abitanti.

Ci sono città più grandi come Graz e Vienna dove c'è il traffico pubblico, però ci sono anche delle aree rurali dove le macchine sono necessarie per quello che si fa ogni giorno.

Ci sono tantissime macchine per esempio in Carinzia, in Austria Bassa, in Austria Alta in Stiria e Salisburgo sono **sempre circa 600 per** mille abitanti.

L'Austria è conosciuta come il paese delle bici e le statistiche sono impressionanti. Sempre più persone vogliono essere responsabili e sostenibili e questo ha un impatto sulla sostenibilità

Più o meno **68%** degli abitanti dell'Austria hanno una bici

e questo è un numero molto alto in cooperazione con altri paesi. E infatti la gente va volentieri in bici e andare in bici non è solo sostenibile ma anche buono per la salute, perché spesso ci mettiamo lì seduti davanti al computer, non ci muoviamo.

Infatti più bici sono state comprate in **2021** che macchine.

Non dipende… sia bike sharing che car sharing sono efficienti, efficaci e l'Austria ha già fatto tanti passi.

Vedremo se possiamo raggiungere gli obiettivi del **2050** non lo sappiamo ancora però abbiamo (eighty-ottantadue) **82** tram il trasporto pubblico è veramente molto elevato qua e questo è importante.

Vedremo anche come sarà Vienna nei prossimi anni, però al momento non lo sappiamo ancora. E è importante quello che fanno tutti, ognuno, ogni individuolo, e pensaci la prossima volta che usi la bici al posto della macchina. Grazie

Anhang XI: Fragebogen

***Vielen Dank* für die Teilnahme an meinem Experiment!**

Hier findest du einige Fragen in Bezug auf das Experiment heute.

Bitte fülle alle Fragen aus. Alle offenen Fragen können auch stichwortartig beantwortet werden.

Fragen zum Experiment

1. War das Thema der Rede für dich bekannt?

	1	2	3	4	5	
Ja, sehr bekannt	○	○	○	○	○	Nein, überhaupt nicht

2. Wie fandest du die Redegeschwindigkeit der Ausgangsrede?

	1	2	3	4	5	
Sehr schnell	○	○	○	○	○	Sehr langsam

3. Wie fandest du die Fachlichkeit des Fachwortschatzes?

	1	2	3	4	5	
Sehr fachspezifisch	○	○	○	○	○	Nicht fachspezifisch

4. Wie ist dein allgemeiner Eindruck der Funktion *Augmented Interpreter CAI v 1.1*?

	1	2	3	4	5	
Sehr positiv	○	○	○	○	○	Sehr negativ

5. Welche Wirkung hatte *Augmented Interpreter CAI v 1.1* auf deine Dolmetschleistung?

	1	2	3	4	5	
Viel verbessert	○	○	○	○	○	Viel verschlechtert

6. Welche Aspekte von *Augmented Interpreter CAI v 1.1* haben deine Dolmetschleistung verbessert?

7. Welche Aspekte von *Augmented Interpreter CAI v 1.1* haben deine Dolmetschleistung verschlechtert?

8. Wie genau fandest du die Erkennung der Zahlen?

	1	2	3	4	5	
Sehr genau	○	○	○	○	○	Sehr ungenau

9. Wurden Zahlen immer richtig erkannt?

- ☐ Ja
- ☐ Nein

10. Falls *nein*, wie hast du darauf reagiert?

11. Bist du zufrieden mit dem Visualisierungsformat der Zahlen?

	1	2	3	4	5	
Ja, sehr zufrieden	○	○	○	○	○	Nein, sehr unzufrieden

12. Hättest du dir ein anderes Format der Zahlen gewünscht?

- ☐ Ja
- ☐ Nein

13. Wenn *ja*, was hättest du geändert?

14. Was war besonders herausfordernd für dich bei der Dolmetschung (in Bezug auf Setting, Rede, Software)?

Fragen zur Software

15. Hattest du vor dem Experiment schon über CAI-Tools gehört?

- ☐ Ja
- ☐ Nein

16. Wenn *ja*, wann und in welchem Ausmaß?

17. Hattest du vor dem Experiment schon über InterpretBank gehört?

- ☐ Ja
- ☐ Nein

18. Wenn *ja*, wann und in welchem Ausmaß?

19. Hattest du vor dem Experiment schon mit CAI-Tools gedolmetscht?

- ☐ Ja
- ☐ Nein

20. Wenn *ja*, mit welchen Tools und in welchem Ausmaß?

Fragen zu den Zahlen

21. Wie gehst du normalerweise mit Zahlen in der Kabine um? (mehrere Antworte möglich)

- ☐ Mein*e Kabinenpartner*in notiert Zahlen für mich
- ☐ Ich notiere Zahlen auf einem Blatt Papier
- ☐ Ich merke mir die Zahlen im Kopf
- ☐ Sonstiges:

22. Welche Dolmetschstrategie(n) nutzt du häufig, um Zahlen zu dolmetschen?

23. Hast du Anmerkungen bzw. Feedbacks über das Experiment und die Software?

Vielen Dank für das Ausfüllen des Fragebogens!

Zusammenfassung

Diese Arbeit befasst sich mit dem Simultandolmetschen mit ASR-gestützten Anwendungen. Das setzt ein multimodales Dolmetschsetting mit visuellen Inputs der ASR und mündlichen Inputs der Ausgangsrede voraus. Ziel der Arbeit ist es, die Wirkung der Anwendung *Augmented Interpreter - CAI v 1.1* auf die Verteilung der visuellen Aufmerksamkeit beim Simultandolmetschen und auf die Richtigkeit der gedolmetschten Zahlen zu untersuchen.

Im ersten Teil werden technologische Lösungen für Dolmetscher*innen vorgestellt und die Entwicklung der automatischen Spracherkennung bis zur Integration von ASR in CAI-Tools thematisiert. Für das Experiment wurde eine Rede von ca. 7 Minuten erstellt, wobei Zahlen automatisch erkannt und mit einem Genauigkeitswert von ca. 95,24 % von der Software *Augmented Interpreter* angezeigt wurden. Die Dolmetschleistungen von neun Dolmetschstudierenden wurden anhand der Wiedergabe der Zahlen ausgewertet. Durchschnittlich 64,4 % der in der Ausgangsrede erwähnten Zahlen und 95,2 % ihrer Referenten wurden korrekt wiedergegeben, obwohl die Ergebnisse je nach Teilnehmerin stark variierten (von 41,67% bis zu 96,7 % für Zahlen).

Die Auswertung der Eyetracking-Daten der Studienteilnehmerinnen in einer Heatmap zeigte, dass die am meisten beobachteten Bereichen auf dem Bildschirm das Gesicht der Rednerin und die letzten angezeigten Zahlen waren. Anhand der einzelnen Interessensgebiete (Areas of Interest, abgekürzt AOI) auf dem Gesicht der Rednerin und auf den einzelnen Zahlen wurden dann die Dauer und die Anzahl der Fixationen gemessen. Die Auswertung der Eyetracking-Daten zeigt, dass die höchsten Werte in Bezug auf die Dauer und Anzahl der Fixationen für die AOI der Rednerin verzeichnet wurden. Viele Fixationen wurden auch außerhalb der festgelegten AOI, d. h. auf anderen Elementen der Benutzerschnittstelle, verzeichnet, während große Unterschiede unter den verschiedenen AOI der Zahlen erkennbar sind.

Zwischen den Eyetracking-Daten und der Widergabe der Zahlen, die für zwei Studienteilnehmerinnen ausführlich analysiert wurden, konnte ein Muster erkannt werden. Zahlen, die länger betrachtet wurden bzw. eine höhere Anzahl an Fixationen verzeichneten, wurden in der Mehrheit der Fälle richtig wiedergegeben und in ihren Kontext eingebettet. Diese waren jedoch einzelne Beispiele, die nicht verallgemeinert werden können. Diese Studie legt die Grundlagen für eine eingehendere Untersuchung mit mehr Teilnehmer*innen und Messungsindikatoren, um aussagekräftigere Daten zu diesem Thema zu erhalten.

Abstract

This paper investigates the topic of Automatic Speech Recognition (ASR) tools for interpreters, which implies a multimodal interpreting setting based on visual inputs of the ASR and oral inputs of the source speech. The objective of this research is to assess the impact of the *Augmented Interpreter* tool on visual attention during interpreting and on the accuracy of the interpretation of numbers.

The first part presents technological solutions for interpreters and discusses the development of automatic speech recognition up to the integration of ASR in CAI tools. For the experiment, a speech of approx. 7 minutes was created, displaying numbers automatically and with an accuracy rate of approx. 95% by the software. The interpreting accuracy of the nine study participants was analyzed based on the numbers displayed. On average, 95.2% of the speakers and 64.44% of the numbers mentioned in the initial speech were reproduced correctly, although the results varied significantly depending on the participant (from 41.67% to 96.67% for numbers).

The analysis of the eye-tracking data of the study participants in form of a heat map showed that the most fixated areas on the screen were the speaker's face and the last numbers. The single areas of interest (AOI) on the speaker's face and the individual numbers were then used to measure the dwell time and fixation count. The analysis of the eye tracking data revealed that the highest values in terms of dwell time and the fixation count were recorded for the speaker's AOI. Many fixations were also recorded outside the defined AOI, i.e. on other elements of the user interface, while large differences are also recognizable between the various AOIs of the numbers.

A pattern could be identified between the eye-tracking data and the interpretation of the numbers, which were analyzed in detail for two study participants. Numbers that were looked at for longer or had a higher number of fixations were accurately reproduced in the majority of cases and embedded in their context. However, these were individual examples that cannot be generalized. Even so, this study lays the foundations for more in-depth research with more participants and measurement indicators to obtain more meaningful data on the topic.