

Informationsintegration in mehrsprachigen Textchats

Der Skype Translator im
Sprachenpaar Katalanisch-Deutsch

Felix Hoberg

Translation and Multilingual Natural
Language Processing 17



Translation and Multilingual Natural Language Processing

Editors: Oliver Czulo (Universität Leipzig), Silvia Hansen-Schirra (Johannes Gutenberg-Universität Mainz), Reinhard Rapp (Hochschule Magdeburg-Stendal), Mario Bisiada (Universitat Pompeu Fabra)

In this series (see the complete series history at <https://langsci-press.org/catalog/series/tmnlp>):

6. Menzel, Katrin, Ekaterina Lapshinova-Koltunski & Kerstin Anna Kunz (eds.). New perspectives on cohesion and coherence: Implications for translation.
7. Hansen-Schirra, Silvia, Oliver Czulo & Sascha Hofmann (eds). Empirical modelling of translation and interpreting.
8. Svoboda, Tomáš, Łucja Biel & Krzysztof Łoboda (eds.). Quality aspects in institutional translation.
9. Fox, Wendy. Can integrated titles improve the viewing experience? Investigating the impact of subtitling on the reception and enjoyment of film using eye tracking and questionnaire data.
10. Moran, Steven & Michael Cysouw. The Unicode cookbook for linguists: Managing writing systems using orthography profiles.
11. Fantinuoli, Claudio (ed.). Interpreting and technology.
12. Nitzke, Jean. Problem solving activities in post-editing and translation from scratch: A multi-method study.
13. Vandevoorde, Lore. Semantic differences in translation.
14. Bisiada, Mario (ed.). Empirical studies in translation and discourse.
15. Tra&Co Group (ed.). Translation, interpreting, cognition: The way out of the box.
16. Nitzke, Jean & Silvia Hansen-Schirra. A short guide to post-editing.
17. Hoberg, Felix. Informationsintegration in mehrsprachigen Textchats: Der Skype Translator im Sprachenpaar Katalanisch-Deutsch.

Informationsintegration in mehrsprachigen Textchats

Der Skype Translator im
Sprachenpaar Katalanisch-Deutsch

Felix Hoberg



Felix Hoberg. 2022. *Informationsintegration in mehrsprachigen Textchats: Der Skype Translator im Sprachenpaar Katalanisch-Deutsch* (Translation and Multilingual Natural Language Processing 17). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/340>

© 2022, Felix Hoberg

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

ISBN: 978-3-96110-337-9 (Digital)

978-3-98554-033-4 (Hardcover)

ISSN: 2364-8899

DOI: 10.5281/zenodo.5902971

Source code available from www.github.com/langsci/340

Errata: paperhive.org/documents/remote?type=langsci&id=340

Cover and concept of design: Ulrike Harbort

Typesetting: Felix Hoberg, Felix Kopecky

Proofreading: Andreas Hözl

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: X_EL_AT_EX

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

<http://langsci-press.org>

Storage and cataloguing done by Zenodo



Inhaltsverzeichnis

Danksagung	vii
Abkürzungsverzeichnis	ix
1 Einführung	1
1.1 Ausgangslage	1
1.2 Hintergrund	2
1.3 Kapitelübersicht	5
1.4 Formelle Hinweise	6
2 Linguistischer Aspekt	7
2.1 Computervermittelte Kommunikation	7
2.1.1 Charakteristika der Chatkommunikation	10
2.1.1.1 Definitorische Grenzziehung	10
2.1.1.2 Nähekommunikation	11
2.1.1.3 Distanzkommunikation	11
2.1.1.4 Ablauf der Chatkommunikation	12
2.1.2 Rollenverteilung in der Chatkommunikation	13
2.2 Medientheoretische Verortung	15
2.2.1 Medientheoretische Verbindung von Internet und Kommunikation	16
2.2.2 Modelle und Theorien der computervermittelten Kommunikation	18
2.2.2.1 Medien- und kanalbezogene Modelle	18
2.2.2.1.1 Kanalreduktionsmodell	18
2.2.2.1.2 Filter-Modell	19
2.2.2.1.3 Mediale Reichhaltigkeit	19
2.2.2.2 Medienwahlmodelle	19
2.2.2.2.1 Rationale Medienwahl	19
2.2.2.2.2 Normative Medienwahl	20
2.2.2.2.3 Interpersonale Medienwahl	20

Inhaltsverzeichnis

2.2.2.3	Individuumsbezogene Modelle	20
2.2.2.3.1	Theorie der sozialen Informationsverarbeitung	20
2.2.2.3.2	SIDE-Modell	21
2.2.2.3.3	Hyperpersonale Kommunikation	21
2.3	Rückanbindung an konkrete Chat-Anwendungen	22
2.4	Der Textbegriff im Umfeld von Chatkommunikation	24
2.5	Wissensrepräsentation in digitalen Räumen	26
2.6	Der soziale Status einer Sprache im digitalen Raum	30
2.6.1	Das Katalanische im digitalen Raum	32
2.6.2	Das Katalanische im internationalen Kontext	35
2.7	Zusammenfassung	36
3	Technologischer Aspekt	37
3.1	Geschichtlicher Überblick der Technologie	37
3.1.1	Bis zum 2. Weltkrieg	37
3.1.2	Ab 1946	38
3.1.3	Der ALPAC-Bericht	39
3.1.4	Technologische Weiterentwicklung	41
3.1.5	Ausweitung des Produktpportfolios	42
3.1.6	Das neue Jahrtausend	43
3.2	Maschinelle Übersetzung	43
3.2.1	Regelbasierte Maschinelle Übersetzung	43
3.2.2	Statistische Maschinelle Übersetzung	45
3.2.3	Neuronale Maschinelle Übersetzung	46
3.2.3.1	Künstliche neuronale Netze	47
3.2.3.2	Ein einfaches Beispiel	48
3.2.3.3	Backpropagation	49
3.2.3.4	Vektoren	50
3.2.3.5	Rekurrenz	51
3.2.4	Leistungsmerkmale der Systeme	51
3.3	CAT und CAI	52
3.4	Dienste für Voice- und Video-Chat	54
3.4.1	Skype	56
3.4.2	Skype Translator	57
3.4.3	Differenzierung der Technologie	57
3.5	Zum Qualitätsbegriff	60
3.5.1	Qualitätsbewertung	61
3.5.2	Qualitätsnormen	62

3.5.3	Evaluation von menschlichen Übersetzungen	63
3.5.4	Evaluation von maschineller Übersetzung	65
3.6	Postediting	67
3.7	Usability	68
3.8	Zusammenfassung	70
4	Arbeitshypothesen	71
5	Empirischer Teil	75
5.1	Begrifflicher Hinweis	75
5.2	Methodik	76
5.2.1	Online-Befragung	76
5.2.2	Eye-Tracking	78
5.2.2.1	Dynamische Areas of Interest	82
5.2.2.2	Einsatz von Eye-Tracking zur MÜ-Evaluation	83
5.2.3	Konzept der Online-Umfrage	85
5.2.4	Datenschutz und Anonymität	88
5.2.5	Konzept der Feldstudie	88
5.3	Die Feldstudie	91
5.3.1	Abhängige Variablen der statistischen Analyse	92
5.3.2	Unabhängige Variablen der statistischen Analyse	95
6	Analyse	97
6.1	Online-Umfrage zur allgemeinen Nutzung von Skype	97
6.1.1	Ergebnisse der Umfrage	97
6.1.2	Generelle Nutzungsangaben	100
6.1.3	Nutzung von Alternativen zu Skype	103
6.1.4	Situierung der Ergebnisse	103
6.2	Die Proband·innen im Setting Katalanisch-Deutsch	105
6.2.1	Angaben der Proband·innen im Eingangsfragebogen . .	108
6.2.2	Nutzung von Alternativen zu Skype	110
6.2.3	Angaben der Proband·innen im Ausgangsfragebogen .	111
6.3	Der Datensatz im Setting Katalanisch-Deutsch	121
6.3.1	Visuelle Inspektion der Bildschirmmitschnitte	122
6.3.2	Fixationen der Probanden	126
6.3.2.1	Fixationsanzahl	126
6.3.2.2	Dauer der ersten Fixation	129
6.3.2.3	Regressionen	133
6.3.2.3.1	Eingehende Regressionen	133

Inhaltsverzeichnis

6.3.2.3.2	Ausgehende Regressionen	136
6.3.2.4	Dauer des ersten Durchlaufs	139
6.3.2.5	Gesamtverweildauer	142
6.3.2.6	Regressive Durchlaufdauer	146
6.3.2.7	Selektive regressive Durchlaufdauer	148
6.3.2.8	Pupillengröße	152
6.3.3	Das sakkadische Blickverhalten	157
6.3.3.1	Sakkadenamplitude	159
6.3.3.2	Sakkadendauer	161
6.3.4	Zusammenfassung der Ergebnisse im Setting Katalanisch-Deutsch	164
6.4	Die Proband·innen im Setting Deutsch-Deutsch	166
6.4.1	Angaben der Proband·innen im Eingangsfragebogen . .	169
6.4.2	Nutzung von Alternativen zu Skype	169
6.5	Der Datensatz im Setting Deutsch-Deutsch	170
6.5.1	Visuelle Inspektion der Bildschirmmitschnitte	170
6.5.2	Fixationen der Probanden	173
6.5.2.1	Fixationsanzahl	173
6.5.2.2	Dauer der ersten Fixation	176
6.5.2.3	Regressionen	179
6.5.2.3.1	Eingehende Regressionen	179
6.5.2.3.2	Ausgehende Regressionen	182
6.5.2.4	Dauer des ersten Durchlaufs	183
6.5.2.5	Gesamtverweildauer	186
6.5.2.6	Regressive Durchlaufdauer	189
6.5.2.7	Selektive Regressive Durchlaufdauer	191
6.5.2.8	Pupillengröße	193
6.5.3	Das sakkadische Blickverhalten	196
6.5.3.1	Sakkadenamplitude	197
6.5.3.2	Sakkadendauer	200
6.5.4	Zusammenfassung der Ergebnisse der einsprachigen Vergleichsstudie	202
7	Diskussion	205
7.1	Merkmale der CvK an Skype und dem Translator	205
7.2	Visuelle Inspektion	208
7.3	Fixatorische Augenbewegung	210
7.3.1	Fixationsanzahl	210
7.3.2	Dauer der ersten Fixation	211

7.3.3	Regressionen	213
7.3.4	Dauer des ersten Durchlaufs	216
7.3.5	Gesamtverweildauer	218
7.3.6	Regressive Durchlaufdauer	219
7.3.7	Pupillengröße	221
7.4	Sakkadische Augenbewegungen	223
7.4.1	Sakkadenanzahl	223
7.4.2	Sakkadenamplitude	225
7.4.3	Sakkadendauer	226
7.5	Blinde Flecken	228
7.5.1	Controlled World	228
7.5.2	Visual World Paradigm	229
7.5.3	Sichtbarkeit der Übersetzer-innen	229
7.5.4	Pilotierung	230
7.6	Zusammenfassung	232
8	Fazit und Ausblick	235
Anhang A:	Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype	237
Anhang B:	Fragebogen 2: Fragebogen Pre-Experiment	251
Anhang C:	Fragebogen 3: Fragebogen Post-Experiment	263
Literaturverzeichnis		271
Register		285
Autorenregister	285
Sachregister	289

Danksagung

Für Unterstützung, Motivation, Revision und Kritik, für Planung, Durchführung, Kaffee in jeglicher Couleur und Mittagessen, kurzum: für alle Phasen und Momente eines solchen Promotionsprojektes möchte ich mich ganz offenherzig und aufrichtig bedanken bei:

Uta Leier	Anastasiia Vaskova
Klaus Leier	Adam Weyell
Nikolas Hoberg	Christine Paasch-Kaiser
Leandra Käsler	Zwetelina Steinbach
Oliver Czulo	Alexandra Bartelt
Silvia Hansen-Schirra	Malina Schriek
Axel Hoberg	Gesa-Yasmin Rosebrock
Sven Kotlarow	Tim Feldmüller
Monique Theile	Christine Paasch-Kaiser
Miriam Heike Schroers	David Imgrund

Außerdem den Katalan-innen, die in verschiedener Funktion bei der Durchführung des empirischen Teils behilflich waren:

Paloma Abad Vilar	Albert Martí Ferrer
Ferran Pérez	Meritxell Feliu Ribas
Àlex Mir Romero	Alba Romero
Angels Adell	Núria Esther Monzonís Carda

Abkürzungsverzeichnis

ANN	Artificial Neural Network	LSP	Language Service Provider
BLEU	Bi-Lingual Evaluation Understudy	KNN	Künstliches Neuronales Netz
CAI	Computer Aided Interpretation	MAHT	Machine Assisted Human Translation
CAT	Computer Aided Translation	MÜ	Maschinelle Übersetzung
CMC	Computer Mediated Communication	MT	Machine Translation
CvK	Computervermittelte Kommunikation	NIST	National Institute of Standards and Technology
ESK	Eingebettete Systeme (und) Kommunikationstechnik	NMT	Neural Machine Translation
ESN	European Student Network	NLP	Natural Language Processing
GTM	General Text Matcher	PE	Post Editing
HAMT	Human Assisted Machine Translation	RBMT	Rule Based Machine Translation
HTWK	Hochschule für Technik, Wirtschaft und Kommunikation	RNN	Recurrent Neural Network
IRC	Internet Relay Chat	SMT	Statistical Machine Translation
		ST	Skype Translator
		TQA	Translation Quality Assessment

AOI-Tags (Englisch)

GerO	German Original	GerMT	German Machine Translation
CatO	Catalan Original	CatMT	Catalan Machine Translation

1 Einführung

Auf den folgenden Seiten wird die Motivation für dieses Promotionsprojekt dargelegt und das Themengebiet anfänglich situiert. Weiterhin ist eine Kapitelübersicht enthalten sowie einzelne Hinweise zur stilistischen und formellen Ausgestaltung dieser Schrift.

1.1 Ausgangslage

Seit ungefähr 2010 lassen sich innerhalb des ohnehin exponentiell wachsenden Forschungsgebietes der digitalen Datenverarbeitung noch stärker herausragende Innovationen und Leistungszugewinne beobachten. Das autonome Fahren scheint plötzlich in greifbare Nähe gerückt zu sein, Technik und Technologie sind so kostengünstig und vielfältig wie nie zuvor, oder kosten für den Endnutzer gar nichts mehr dank engagierter Organisationen, die qualitativ hochwertige Open Data- und Open Source-Projekte betreuen. Dank Smartphones und vergleichbarer mobiler Endgeräte sind Milliarden Menschen mit dem Internet verbunden und können hierüber Informationen zu nahezu jedem Lebensbereich abrufen und beitragen. Kurz gefasst: Allein innerhalb der vergangenen zehn Jahre hat sich auf diesem Gebiet sowohl strukturell, finanziell, ideologisch, innovativ als auch konzeptuell derart viel getan, dass diese Aufzählung nur die größte aller denkbaren Einbettungen liefern kann.

All dies ist nur auf Grundlage der Akkumulation von immensen Datenmengen, internationalen Wissensverbünden und dem weltweiten Technologie- sowie Informationstransfer möglich. Die hier vorgestellte Monographie mit dem Titel *Informationsintegration in mehrsprachigen Textchats* greift dabei einen Teilaspekt aus dem Blickwinkel der Sprachtechnologieforschung, Übersetzungswissenschaft sowie Kommunikationswissenschaft auf. Anhand des Skype Translators als Fallbeispiel wird der Frage nachgegangen, welche Auswirkung die Verbindung von computervermittelter Kommunikation und Maschineller Übersetzung (MÜ) auf das Nutzungsverhalten derartiger Technologien hat. Nicht erst innerhalb des genannten Zehnjahreszeitraums hat sich, auch geleitet vom exponentiellen Wachstum der Rechenfähigkeit und der immer effizienter werdenden Computertechnologie, das Feld der Mensch-Maschine-Interaktion ausgeweitet.

1 Einführung

Eine Unterform davon ist die sog. Usability, sprich: die Nutzerfreundlichkeit, bei der die Interaktion von Mensch und Maschine in Hinblick auf den erfolgreichen Austausch von Informationen hin untersucht wird. Hieran anschließend lautet die Frage, wie eine Kommunikationssituation unter Beteiligung der MÜ strukturiert wird.

Diese zwei Leitfragen lassen sich ferner unterteilen in kleinere Einheiten auf den Gebieten der Soziolinguistik und der Sprachtechnologie. Welche Eigenschaften weist ein Chat mit Beteiligung eines Echtzeit-Übersetzungsdienstes auf? Inwiefern kann in diesem Kontext von einem Gespräch die Rede sein? Welche Konsequenzen hat es zuletzt für vergleichsweise kleine Sprachen im europäischen Kontext wie dem Katalanischen, wenn eine solche Technologie zum Einsatz kommt? Wie formt – auf der vorausgehenden Frage aufbauend – der Einsatz von solchen Technologien wie dem Skype Translator die Informationsextraktion aus Textchats?

Über all diesen Fragen schwebt zudem unausgesprochen die allgemeine Sorge der gesamten Sprachmittlerbranche, in wenigen Jahren nicht mehr benötigt zu werden. Bereits häufig bemüht und sicherlich auch mittlerweile als Verweis abgenutzt ist dabei der Babelfish aus Douglas Adams' *Per Anhalter durch die Galaxis*. Der Grund, weshalb dieser kleine, fiktive Fisch, den man sich zwecks Universalübersetzungswerkzeug aller Sprachen in alle Sprachen ins Ohr stecken kann, auch in dieser Arbeit genannt wird, ist das nach wie vor sehnsgütig betrachtete Menschheitsziel, womöglich irgendwann Sprachgrenzen mithilfe von Maschinen komplett überwindbar und somit obsolet zu machen. Diese Sehnsucht steht in einem Widerspruch zur exotischen Anziehung von Fremdsprachen, die wiederum auch einen Kontrast bildet zur Hegemonie des Englischen in der globalisierten Welt, und dort besonders im Internet.

1.2 Hintergrund

Das übergeordnete Themengebiet des Forschungsvorhabens ist die Translations-technologie, konkreter: die Maschinelle Übersetzung (MÜ) sowie die automatische Verarbeitung natürlicher Sprache. Da mit Katalanisch jedoch auch – im europäischen Kontext – eine vergleichsweise „kleine“ Sprache im Fokus der Ausarbeitung steht, gewinnt die Monographie einen soziolinguistischen Anteil hinzu. Zwar ist es nicht Ziel dieser Arbeit, auf Grundlage der Betrachtung des Skype Translators eine Abwägung der Vor- und Nachteile sowie die Chancen und Risiken für kleinere Sprachen zu diskutieren. Allerdings besteht der Anspruch, einen möglichen Einstieg für die Auseinandersetzung aufzuzeigen und einige Denkanstöße auch für weitere Forschung auf diesem Gebiet zu liefern.

Die Kommunikationssoftware Skype ist bereits seit 2003 auf dem Markt. Ursprünglich als eigenständiges Unternehmen unter gleichem Namen gegründet, wurde es 2011 von Microsoft aufgekauft. Weltweit bekannt ist das Programm vor allem für die Möglichkeit, Echtzeit-Videochats zu führen. Auch wenn es eine explizit beworbene Business-Version für Unternehmen und Dienstleister gibt, wird Skype hauptsächlich von Privatanwendern genutzt. Vereinzelt lassen sich jedoch auch gerade Dienstleister verzeichnen, die die Videofunktion zur Fernberatung ihrer Kunden einsetzen (s. z. B. Leipziger Wohnungsbaugesellschaft, LWB). Gegenwärtig ist Skype neben dem ebenfalls aus den USA stammenden Konkurrenten *Zoom* auch medial präsent. Im Zuge der durch die weltweite Corona-Pandemie bestehenden Kontakt- und Reisebeschränkungen und dem entsprechend sich verändernden Kommunikationsverhalten kommt es vor, dass ein offizielles Interview im Fernsehen per Videochat durchgeführt wird. Skype ist dabei eine häufig genutzte Anwendung.

Die hier zur Diskussion gestellte Funktion des Skype Translators wurde 2015 implementiert und befindet sich im stetigen Wandel. Bereits die Version, die dieser Arbeit zugrunde liegt, ist eine andere, als die aktuell für die Endnutzer·innen verfügbare. Zu Beginn der Ausarbeitung war sie nur für Windows 7 und 10 verfügbar und auch dort nur in der jeweils aktuellsten Version der Software. Gegenwärtig jedoch ist der Skype Translator bei Skype auf Windows, Linux und MacOS sowie in der Smartphone-App enthalten. Die Version *Skype for Business* unterstützt den Skype Translator bislang hingegen nicht. Aktuell sind mehr als 60 Sprachen im Textchat sowie 11 Sprachen für die Übersetzung von gesprochener Sprache in Voice- und Videochats bereitgestellt. Dieser Bereich befindet sich in einem steten Wandel und wird im Jahresrhythmus von Neuerungen geprägt. Die Darstellung bestehender MÜ-Ansätze, von regelbasierter über statistischer bis hin zur jüngsten neuronalen MÜ, ist in diesem Zusammenhang notwendig, um die Funktionalität des Skype Translators einschätzen zu können und belastbare Hypothesen für diese Arbeit aufzustellen.

Da es mittlerweile eine Vielzahl weltweit agierender Dienste gibt, die ihren Nutzer·innen eine maschinelle Übersetzungsausgabe zur Verfügung stellen (z. B. Facebook, Instagram), wird der Skype Translator zunächst in den gegenwärtigen Forschungsstand der MÜ eingebettet. Hierzu ist in der Arbeit ein Abriss über die derzeit verfügbaren Technologien enthalten, bevor auch Computer Assisted Translation (CAT) und Computer Assisted Interpretation (CAI) Tools betrachtet werden, um den Skype Translator dazwischen anzusiedeln. Weiterhin wird im theoretischen Teil dieses Buches auf Evaluierungsmöglichkeiten der MÜ eingegangen.

1 Einführung

Eine besondere Berücksichtigung erfährt dabei das Post-Editing (PE), da die traditionsreiche, jedoch zugleich sich stetig weiterentwickelnde Technologie wie die MÜ bei Weitem nicht fehlerfrei arbeitet und daher der Nachbesserung durch den Menschen bedarf. Innerhalb des Forschungsstrangs zum Post-Editing findet sich eine Vielzahl an Publikationen zu Evaluations- und Qualitätssicherungsmethoden, die in diesem Zusammenhang abstrahiert und auf das Modell angepasst werden können. Diese Punkte stecken die Ecken des Untersuchungsfeldes ab, in dem die Informationsverarbeitung in mehrsprachigen Textchats praktisch untersucht wird.

Der o. g. soziolinguistische Anteil wird parallel dazu ausgearbeitet. Das Katalanische als untersuchte Sprache ist vor allem durch die Unabhängigkeitsbewegung in Katalonien bekannt. Mit ca. 9 Mio. aktiven Sprechern ist das Katalanische, das in seinen Varietäten, die in Katalonien, der Valencianischen Gemeinschaft, Südfrankreich, den Balearen, Andorra und einem kleinen Teil von Sardinien angesiedelt sind, annähernd so groß wie die Gesamtbevölkerung Belgiens oder Tschechiens. Das Katalanische ist jedoch keine offizielle Amtssprache der EU. Außerdem ist es einzig in Andorra alleinige offizielle Sprache. In Spanien haben die Regionalsprachen laut Verfassung den Status einer offiziellen Sprache innerhalb der jeweiligen autonomen Gemeinschaften inne, wobei das Katalanische durch den Regionalstatut der autonomen Gemeinschaft gestützt wird (Bochmann u. a. 2011: 414). Aufgrund des zentralistischen Staatsaufbaus in Frankreich gestaltet sich die Situation sprachpolitisch dort anders.

Diese knappe Skizze stellt die Notwendigkeit der soziolinguistischen Betrachtung innerhalb des Forschungsvorhabens dar, wobei der Schwerpunkt hier besonders auf die Stellung der Sprache im digitalen Raum und damit wieder auf dem Informationsaustausch in mehrsprachigen Konstellationen liegt. Von dort aus lässt sich wiederum die Brücke schlagen zur Sprachtechnologie und dem Einsatz der maschinellen Übersetzung.

Der empirisch-experimentelle Teil ist in drei größere Bereiche gegliedert. Neben einer grundlegenden Skizzierung des Nutzungsverhaltens von Skype und dem Skype Translator auf Grundlage eines Online-Fragebogens, sieht der zweite Teil eine Eye-Tracking-Studie vor, in der deutschsprachige Studierende während einer Chatkommunikation mit katalanischen Muttersprachler·innen vom Skype Translator unterstützt werden. Der dritte Teil, ebenfalls eine Eye-Tracking-Studie, liefert hierzu Vergleichsmaterial. Studierende werden während der Chatkommunikation mit anderen deutschsprachigen Muttersprachler·innen erfasst, ohne dass der Skype Translator aktiviert ist.

Zuletzt werden, ausgehend von den gewonnenen Erkenntnissen am Skype Translator, weiterführende Überlegungen für die Forschung abstrahiert.

1.3 Kapitelübersicht

Wie bereits in der Einführung deutlich wird, setzt diese Arbeit an mehreren unterschiedlichen Themengebieten und Teilespekten an. Um hieraus eine logische Ordnung mit angemessener Argumentationsstruktur aufzubauen, setzt das folgende Kapitel 2 zunächst ganz allgemein in der Linguistik an. So wird eine Definition der computervermittelten Kommunikation herausgearbeitet (Abschnitt 2.1, S. 7), die sowohl der Wissensrepräsentation (Abschnitt 2.5, S. 26) als auch einen Textbegriff im Rahmen der digitalen Lebenswelt (Abschnitt 2.4, S. 24) des Menschen, und damit besonders dem Chat, Rechnung trägt. Die Rückanbindung an die in dieser Arbeit beteiligten Sprachen findet ebenfalls in diesem Kapitel ab Abschnitt 2.6 (S. 30) statt.

Kapitel 3 greift die theoretischen Grundlagen auf und verknüpft sie mit einem geschichtlichen Überblick über die Entwicklung von Sprachtechnologien (Abschnitt 3.1, S. 37). Die historische Übersicht endet mit dem gegenwärtigen Stand der Forschung zur MÜ (Abschnitt 3.2, S. 43), in dessen Rahmen eine Differenzierung der Technologien (Abschnitt 3.4, S. 54) stattfindet. Skype als Herzstück dieser Arbeit erhält dabei eine besondere Zuwendung (Abschnitt 3.4.1, S. 56). Weiterhin umfasst das Kapitel die Qualitätsbewertung, sowohl mit Blick auf das Leistungsvermögen der Technologie (Abschnitt 3.5, S. 60) als auch aus Sicht der Endnutzer-innen in Form der *Usability* (Abschnitt 3.7, S. 68). Kapitel 4 (S. 71) leitet mit der Aufstellung der konkreten Forschungsfragen vom Theorienteil zum praktischen Teil dieser Arbeit über.

Kapitel 5 dient dann der Darstellung der methodologischen Grundlagen. Nach einer generellen Vorstellung der Gepflogenheiten bei der Arbeit mit Online-Umfragen und mit Eye-Tracking-Studien (Abschnitt 5.2, S. 76) erfolgt die Konzeption beider Erhebungsformate: Für die Online-Umfrage in Abschnitt 5.2.5 (S. 88) und für die Feldstudie mit dem Skype Translator in Abschnitt 5.3 (S. 91).

Die beiden Kapitel 6 und 7 umfassen dann die Präsentation, Analyse und Diskussion der gewonnenen Daten. Zunächst werden die Ergebnisse der Online-Umfrage (Abschnitt 6.1, S. 97) vorgestellt und situiert. Anschließend erfolgt die Beschreibung der beiden Studienvarianten *Katalanisch-Deutsch* (Abschnitt 6.2, S. 105) und *Deutsch-Deutsch* (Abschnitt 6.4, S. 166), bevor die Daten dieser beiden Teile ebenfalls analysiert werden (Abschnitte 6.3, S. 121 und 6.5, S. 170). Kapitel 7 dient der zusammenführenden Diskussion der Ergebnisse sowohl aus visueller als auch statistischer Sicht (Abschnitte 7.2, S. 208 und 7.3, S. 210) und verweist darüber hinaus auf Bereiche dieser Arbeit, die lohnenswert für eine eingehende Folgeuntersuchung erscheinen (7.5, S. 228).

Mit Kapitel 8 (S. 235) wird ein Fazit gezogen und bestehende Desiderata ausgesprochen.

1.4 Formelle Hinweise

Zum Abschluss der Einleitung seien an dieser Stelle noch einzelne Hinweise formeller Natur genannt. Die Monographie ist gemäß der Stileitfäden von Language Science Press formatiert und ausgestaltet worden. Die sog. *Generic Style Rules for Linguistics*¹ können ebenso online eingesehen werden wie die *Language Science Press Guidelines*². Beispiele, erstmalige Erwähnungen, fremdsprachliche Ausdrücke sowie nicht konventionalisierte Abkürzungen werden kursiv geschrieben. Zitate, die kürzer als drei Zeilen sind, werden in den folgenden Anführungszeichen gesetzt: „...“

Über die Arbeit hinweg werden textinterne Referenzen wie Tabellen, Beispiele und Abbildungen immer mit der fortlaufenden Nummerierung sowie der Seitenanzahl angegeben. Eine Ausnahme wird nur gemacht, wenn eine unmittelbar anliegende Textpassage direkt auf das Element verweist, dann erscheint nur die jeweilige fortlaufende Nummer.

Diese Arbeit umfasst zudem zwei Indices. Der erste Index ist ein allgemeines Stichwortverzeichnis mit wichtigen Begriffen dieser Arbeit. Der zweite Index ist ein erweitertes Namensregister. Er führt die Namen wichtiger Personen im Umfeld dieses Forschungsbereichs auf.

¹S. hierzu <https://langsci.github.io/gsr/GenericStyleRulesLangsci.pdf>. Letzter Zugriff am 31. Januar 2022.

²S. hierzu <https://langsci.github.io/guidelines/latexguidelines/LangSci-guidelines.pdf>. Letzter Zugriff am 31. Januar 2022.

2 Linguistischer Aspekt

Dieses Kapitel wendet sich der computervermittelten Kommunikation sowie der Wissensrepräsentation in digitalen Räumen, genauer gesagt: in Chatsituationen, zu. Hierfür werden zunächst die grundlegenden Elemente der computervermittelten Kommunikation (folgender Abschnitt 2.1) aufgezeigt und damit einhergehend die Wesenszüge der Chatkommunikation dargestellt (Abschnitt 2.1.1, S. 10). Es folgt ein Abriss geläufiger Analysemodelle von computervermittelter Kommunikation, bevor über eine Betrachtung des Textbegriffs im Chat die Brücke zur Wissensrepräsentation geschlagen wird. Mit Blick auf die in dieser Arbeit behandelte Sprache Katalanisch enthält dieses Kapitel ebenfalls eine Exkursion in die soziale, linguistische und digitale Lebenswelt der Sprache (Abschnitt 2.6, S. 30).

2.1 Computervermittelte Kommunikation

Unter Kommunikation verstehen wir die Informationsübertragung zwischen Menschen. Kommunikation kann mithilfe von Sprache, Mimik, Gestik oder durch Schrift stattfinden. Dabei werden verschiedene Kommunikationskanäle bedient: akustisch, optisch, über Individualmedien (z. B. Telefon oder Computer) oder Massenmedien (z. B. Fernsehen oder Internet). (Trepte & Reinecke 2012: 158)

Dieses Zitat liefert in wenigen Zeilen eine grobe Definition von Kommunikation. Darauf soll im Folgenden aufgebaut werden. Um zusätzlich noch die Brücke zur sog. computervermittelten (*computer-mediated*) Kommunikation (eng. *CMC*, dt. *CvK*) zu schlagen, erweitern Trepte & Reinecke die obige Aussage. Computervermittelte Kommunikation ist der „zwischen zwei oder mehr Personen stattfindende, interaktive Prozess des Erstellens, Austauschens und Empfangens von Informationen mithilfe von Computern“ (Trepte & Reinecke 2012: 158). Hierfür wurden über die Jahrzehnte mehrere Anwendungen und Plattformen entwickelt, die in der Forschung gleichermaßen betrachtet werden: Neben dem für diese Arbeit zentralen Chat (z. B. Internet Relay Chat, IRC) sind dies E-Mail, Newsgroups bzw. Foren, Blogs, Videospiele, Videoplattformen, SMS, Internettelefonie uvm.

2 Linguistischer Aspekt

Eine umfassende und differenzierte Aufzählung liefern beispielsweise (beispielsweise Beck 2006, Misoch 2006).

Nun ist die Kommunikationswissenschaft ein ausgesprochen breites Forschungsfeld und zugleich stetigen Umwälzungen unterworfen, wie die Ausführungen verschiedener Autor·innen aus unterschiedlichen Jahrzehnten belegen (z. B. Döring 2003, Beißwenger 2007, Dürscheid 2018, Storrer 2001b, Kienle u. a. 2017, Baechler u. a. 2016). Beide eingangs genannten Zitate bedürfen daher einer eingehenden Situierung. Es ist zu bedenken, dass der Textchat von Skype, also eine Form von textbasierter Kommunikation, im Mittelpunkt dieser Untersuchung steht. Die anderen, zumeist audio-visuellen, Formen (Voice- und Videochat) werden deshalb nur angeschnitten. Mit dieser Abgrenzung wird ein grundlegendes Spannungsfeld bei der Betrachtung von (Chat- bzw. computervermittelter) Kommunikation aufgeworfen, das schon häufig Gegenstand von Studien war: die dichotomische Unterscheidung von Mündlichkeit und Schriftlichkeit. An diesem Punkt setzen Koch & Oesterreicher (2011: 13) mit ihrem Modell an und bieten somit einen soliden Ausgangspunkt für die intendierte Kartographierung der CvK (s. Abbildung 2.1).

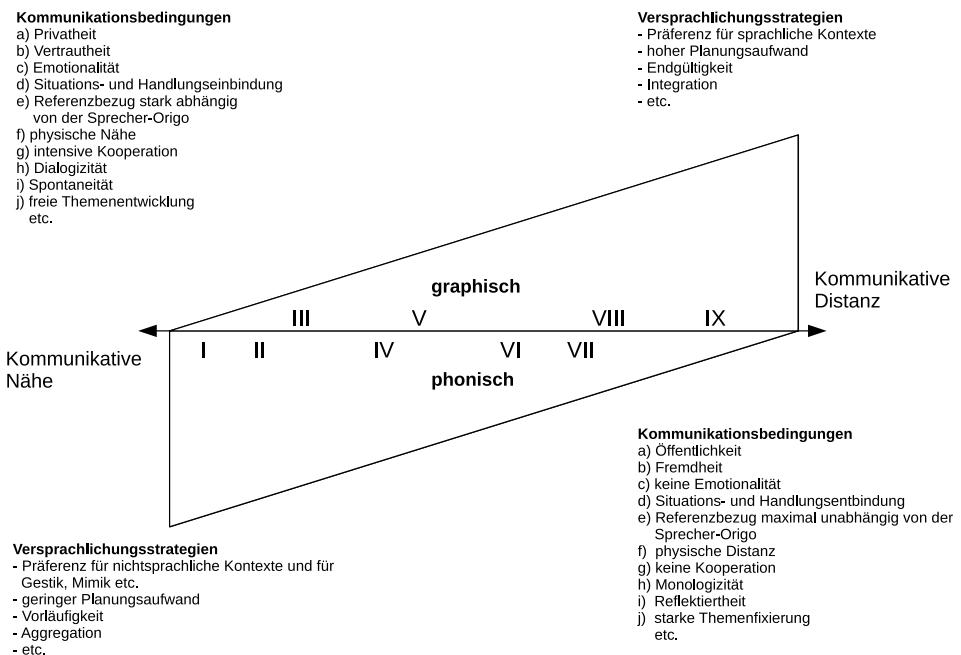


Abbildung 2.1: Das Modell von Koch und Oesterreicher als von F.H. nachgebildete Darstellung

2.1 Computervermittelte Kommunikation

Nähe- und Distanzkommunikation bilden im Modell die Pole eines Kontinuums, das sich entlang einer weiteren Achse in eine phonische und in eine graphische Dimension teilt. Die Bezeichnung als phonisch und graphisch ist auf eine terminologische Präzisierung zurückzuführen: Koch & Oesterreicher weisen darauf hin, dass das Begriffspaar *mündlich* und *schriftlich* unterschiedlich verstanden werden kann. Einerseits sei damit das „*Medium* der Realisierung sprachlicher Äußerungen“ (Koch & Oesterreicher 1994: 587, Kursivierung im Original) gemeint, weshalb hierfür von den Autoren fortan *phonisch* für *mündlich* und *graphisch* für *schriftlich* verwendet wird. Andererseits „meinen die beiden Termini oft den Duktus, die Modalität der Äußerungen sowie die verwendeten Varietäten, kurz: die *Konzeption*“ (Koch & Oesterreicher 1994: 587, Kursivierung im Original). Koch & Oesterreicher halten weiterhin fest, dass einerseits

zwischen dem phonischen Medium und konzeptionell mündlichen Äußerungsformen, andererseits zwischen dem graphischen Medium und konzeptionell schriftlichen Äußerungsformen eine ausgeprägte Affinität besteht.
(Koch & Oesterreicher 1994: 587)

Die Eigenschaften der Pole können unter Kommunikationsbedingungen und Versprachlichungsstrategien subsummiert werden. Auf der Seite der Nähekommunikation stehen die Parameter Privatheit, Vertrautheit, Emotionalität, Situations- und Handlungseinbindung, Referenzbezug abhängig vom Sprecher, physische Nähe, intensive Kooperation, Dialogizität, Spontaneität und freie Themenentwicklung. Die Sprache ist im Rahmen der Nähekommunikation nur ein Faktor (Koch & Oesterreicher 1994: 591). Dem entgegen stehen weitestgehend die Antonyme der genannten Begriffe als Parameter der kommunikativen Distanz: Öffentlichkeit, Fremdheit, keine Emotionalität, Situations- und Handlungsentbindung, Referenzbezug unabhängig vom Sprecher, physische Distanz, keine Kooperation, Monologizität, Reflektiertheit und starke Themenfixierung. In diesem Fall ist die Sprache nicht nur ein Faktor unter vielen. Als zentraler Bestandteil der Kommunikation erfährt das lexikalische Material eine Erweiterung, um auf außersprachliche Kontexte und Bezüge verweisen oder zugreifen zu können (Koch & Oesterreicher 1994: 591).

Mit Ausnahme von Referenzbezug und physischer Nähe bzw. Distanz handelt es sich um graduelle Parameter (Koch & Oesterreicher 2011: 7). Weiterhin unterscheiden sich die Pole nach den Versprachlichungsstrategien. Auch diese Aspekte weisen eine graduelle Orientierung auf: (nicht-)sprachlicher Kontextbezug, Planungsaufwand, Vorläufigkeit bzw. Endgültigkeit sowie Integration bzw. Aggregation.

2 Linguistischer Aspekt

Das Modell beschreibt mit der Polarität von Nähe- und Distanzkommunikation einen Parameter, der so in dem Eingangszitat von Trepte & Reinecke zur CvK keine – allenfalls implizite – Beachtung findet. Dabei ist dieser Aspekt allein schon von einem rein menschlichen-intuitiven Standpunkt aus sinnvoll: Einerseits erscheint es auf den ersten Blick wenig sinnvoll, mit einer Person im unmittelbaren, physischen Umfeld computergestützt zu kommunizieren. Andererseits gibt es durchaus Konstellationen, in denen diese Möglichkeit zur Kommunikation genutzt wird (bspw. Schüler·innen, die im Unterricht heimlich miteinander kommunizieren).

Aus diesem denkbar knappen Umriss des Modells von Koch & Oesterreicher sollen nun im nachstehenden Abschnitt die wesentlichen Merkmale der Chatkommunikation aufgezeigt werden.

2.1.1 Charakteristika der Chatkommunikation

2.1.1.1 Definitorische Grenzziehung

Bevor sich den eigentlichen Charakteristika der Chatkommunikation gewidmet wird, sei an dieser Stelle noch eine definitorische Grenze gezogen. Die folgenden Ausführungen beschreiben den Textchat als durch eine entsprechende Anwendung ermöglichte Kommunikation zwischen mindestens zwei Personen. Diese Personen befinden meist an unterschiedlichen Orten, nutzen unterschiedliche Endgeräte und tauschen sich schriftlich über das Chatfenster der gewählten Anwendung (sei es Skype, WhatsApp, Telegram, Facebook Messenger uvm.) aus. Chats mit mehreren beteiligten Personen, die etwa *Chat-Rooms* nutzen, werden in der Forschung seit mehreren Jahrzehnten untersucht (vgl. z. B. Kienle u. a. 2017, Schweiger & Beck 2019), an dieser Stelle kommt ihnen allerdings nur Beachtung im minimal nötigen Umfang zu. Der Skype Textchat sowie der Skype Translator können zwar auch in Konstellationen mit mehreren beteiligten Personen genutzt werden. Dies sei an dieser Stelle jedoch zu vernachlässigen. Es ist davon auszugehen, dass die Untersuchung des Skype Translators in Textchats mit mehr als zwei beteiligten Personen ganz andere Anforderungen an die theoretische Ausarbeitung sowie den Versuchsaufbau stellt. Daher empfiehlt sich für diese Betrachtung eine eigene Studie, zumal mit steigender Anzahl an beteiligten Personen auch der Grad der Interaktivität zunimmt. Die Interaktivität wiederum ist ein eigener Forschungsbereich innerhalb der CvK (vgl. hierzu bspw. Kienle u. a. 2017, Dürscheid 2018).

2.1.1.2 Nähekommunikation

Der textbasierten Chatkommunikation werden in der Kommunikationsforschung Eigenschaften zugeschrieben, die im Sinne der Konzeption von Koch & Oesterreicher nur einem der beiden Pole zugeordnet werden können und sich somit auf den ersten Blick gegensätzlich gegenüberstehen. Den zentralen Punkt hebt Storrer (2001a: 446) hervor, die auf die konzeptionelle Mündlichkeit eines Textchats verweist. Dieses Merkmal ist im Modell von Koch & Oesterreicher im Bereich des Nähe-Pols angesiedelt, da der Chat mehrere entsprechende Versprachlichungsstrategien und Kommunikationsbedingungen aufweist. Es kann also von einer Ähnlichkeit mit der Kommunikation von Angesicht zu Angesicht ausgegangen werden. Dürscheid (2016: 35) folgt dieser Sicht nur bedingt, da sie die Chatkommunikation eher in die Nähe zu einem Telefongespräch als zu einer direkten, unmittelbaren Kommunikation von Angesicht zu Angesicht rückt. Bei einem Telefonat werde zwar räumliche Distanz offenkundig und die Kommunikation unterliege gewissen Einschränkungen, die Interaktion folge jedoch den Prinzipien der Mündlichkeit. So werde im Chat versucht, das Fehlen von para- und non-verbalen Kommunikationselementen (wie Gestik und Mimik) durch Emojis und andere Elemente auszugleichen (Storrer 2001b: 3 f.).

Orthographische Schwächen werden ebenfalls weitestgehend akzeptiert, was der Chatkommunikation einen vorläufigen Charakter verleiht (Storrer 2001a: 444 f.). Der geringe Planungsaufwand tritt dadurch zutage, dass eine Person eine Information in mehrere kleinere Nachrichten zerlegt, die sie dann versendet (Beißwenger 2010: 287). Allgemein weist die Chatkommunikation also eine unmittelbare Ausrichtung auf die jeweils gegenwärtige Situation auf. Daher kommt (Storrer 2001b: 5) als Zwischenfazit zu der Einschätzung, dass, wenn man zwischen den funktionalen Kategorien Text und Diskurs (Gespräch) differenziert, (...) die im Chat produzierten Kommunikate eindeutig dem Gespräch zuzurechnen“ sind.

2.1.1.3 Distanzkommunikation

Zugleich wird der Chat jedoch auch als Interaktionsform angesehen, die es ermöglicht, „über räumlich, zeitliche oder raumzeitliche Distanzen hinweg“ (Beck 2006: 12) zu kommunizieren, da er verschriftlicht vorliegt und dadurch „die Produktion und die Rezeption der Äußerungen (...) hier zeitlich entkoppelt“ (Dürscheid 2018: 35). Diese Gründe rücken den Chat in den Bereich der Distanzkommunikation. Auf diese Eigenschaft wird in der Forschung ebenfalls referiert, wenn es um die Beschreibung der Chatkommunikation aus technologischer Sicht

2 Linguistischer Aspekt

geht (s. in dieser Arbeit Abschnitt 2.2.1, S. 16). Beißwenger konstruiert eine Verbindung zwischen diesen beiden Polen durch die ergebnisorientierte Interaktion, die von allen Beteiligten erst dadurch als gemeinsames Produkt wahrgenommen wird, dass der Chat „zur Prozessierung synchronen und dialogischen Austauschs die Überlieferungsqualitäten von Textformen nutzt“ (Beißwenger 2010: 249). Wo im Gespräch von Angesicht zu Angesicht also zeitgleich kommuniziert werden kann, muss in der Chatkommunikation die Schrift genutzt werden, damit über die zeitliche und räumliche Distanz alle Beteiligten die jeweiligen Chatbeiträge als Dialog erfassen können (Beißwenger 2010: 249). Zusammenfassend stellt Beißwenger hierzu fest: „Während die Phonie irreversibel und flüchtig ist, ist die Graphie reversibel und persistent (...)“ (Beißwenger 2010: 249).

Zugleich merkt Storrer auch an, dass der Chatverlauf, also die Gesamtheit der getippten Beiträge, nicht dazu gedacht ist, mündlich vorgetragen zu werden.

Die mündliche Reproduktion von Chatprotokollen ist nicht intendiert. Mehr noch: Chat-Protokolle würden sich aufgrund ihrer sprachlichen Besonderheiten nur unter großen Schwierigkeiten mündlich vortragen lassen. (Storrer 2001b: 4)

Auch dieses Argument unterstreicht somit den Zwiespalt der Chatkommunikation zwischen Nähe und Distanz.

2.1.1.4 Ablauf der Chatkommunikation

Mehr noch erfordert die Chatkommunikation eine „konsekutive Abfolge von Aktivitäten“ (Kienle u. a. 2017: 145), die im Falle mündlicher Kommunikation zeitgleich erfolgen. Während die Prozesse des Verstehens, Verarbeitens und Produzierens von Gesprächsaußerungen gleichzeitig bei allen Beteiligten stattfinden, folgt eine gewöhnliche Chatkommunikation vielmehr dem hier dargestellten Schema: Zu Anfang steht bei einem oder einer der Chatbeteiligten die Entscheidung, einen Chatbeitrag verfassen zu wollen. Da die meisten Chatanwendungen nicht Zeichen für Zeichen der getippten Nachricht übermitteln, ist das explizite Senden (per Eingabetaste o. ä) erforderlich. Der Beitrag muss also weitestgehend vollständig schriftlich formuliert werden. Während dieser Zeit zeigen heutige Chatanwendungen der oder dem Gegenüber an, dass die andere Person gerade einen Beitrag verfasst (*Person XY schreibt gerade...*). Die daraufhin abgesendete Nachricht wird in den meisten Fällen zunächst an einen Server geleitet, der sie dann chronologisch sortiert nach Zeitstempel dem Empfangsrechner zuführt. So ist nahezu ausgeschlossen, dass es zu einer Überlappung von Beiträgen

kommt, wie es in der mündlichen Kommunikation der Fall sein kann (Storrer 2001b: 7). Die Übertragung dauert heutzutage nur wenige Millisekunden, selbst wenn zusätzliche Instanzen wie etwa der Skype Translator dazwischengeschaltet sind. Erst mit erfolgreicher Übermittlung kann dann die andere beteiligte Person die Nachricht erfassen und eine Reaktion hierauf bieten (Kienle u. a. 2017: 145 f.). Dürscheid (2016: 35 f.) erkennt hierin einen Vorteil gegenüber der Kommunikation von Angesicht zu Angesicht, da eine psychologische Hemmschwelle für die Chat-Teilnehmer·innen wegfällt. Die Personen werden nicht beobachtet, während sie ihre Beiträge formulieren und können sich dank der schriftlichen Kommunikation auch bei Bedarf noch bis zum Versand der Nachricht korrigieren. Sogar über den Versand hinaus ist eine Korrektur des Chat-Beitrags möglich.

Dahingegen erfordert die Chat-Interaktion als solche die tatsächliche synchrone Präsenz der beteiligten Personen vor bzw. an ihren jeweiligen Endgeräten. Chatbeiträge können zwar auch wie in einem (E-Mail-)Postfach für eine zeitverzögerte Erfassung zurückgehalten werden, verfehlten dann jedoch den Zweck der Chatkommunikation im Sinne von Beißwenger (2007: 30), nämlich den dialogischen, dynamischen Austausch. Daher kommen Storrer (2001b), Beißwenger (2007) und Hartmann & Krotz (2019) zu der Auffassung, es handele sich nicht um synchrone, sondern um quasi-synchrone Kommunikation: Chat-Beiträge können erst dann erfasst werden, sobald sie vollends produziert und an alle beteiligten Personen übermittelt worden sind. Zur Präzisierung ist noch festzuhalten, dass Beißwenger (2007: 23) *synchron* und *simultan* in eine hierarchische Beziehung zueinander setzt. *Synchron* bezieht sich auf die „zeitgleiche Verfügbarkeit von Kommunikanten für die Zwecke der Kommunikation“ (Beißwenger 2007: 23), wohingegen *simultan* als Hyponym dazu steht und zusätzlich noch das Erfassen der Chatbeiträge während der Interaktion umschließt.

Die Verbindung von Elementen der Nähe- und der Distanzkommunikation stellt deshalb besondere kognitive Anforderungen an die Beteiligten. Damit sich überhaupt erst eine zielgerichtete Interaktion entwickeln kann, müssen die am Chat teilnehmenden Personen „idealerweise permanent *gleichzeitig produzieren und rezipieren*“ (Beißwenger 2010: 260, Hervorhebung im Original), was jedoch unterschiedliche geistige Prozesse darstellt und somit „bestenfalls kurzzeitig möglich ist“.

2.1.2 Rollenverteilung in der Chatkommunikation

Aus den im vorausgehenden Abschnitt herausgearbeiteten Charakteristika der Chatkommunikation ergeben sich auch Konsequenzen für die Organisation des Sprecher·innenwechsels. Auch wenn eine Ähnlichkeit zwischen der Sprecher-

2 Linguistischer Aspekt

innenorganisation und dem Turn-Taking besteht, betont Storrer (2001a: 453), dass sich dessen Regeln nicht ohne Weiteres auf die Sprecher-innenorganisation anwenden lassen. Um sich eindeutig von der mündlichen Kommunikation abzgrenzen, bevorzugt Kienle u. a. (2017: 146) in diesem Zusammenhang die Bezeichnung „Wechsel der Produzentenrolle“. Im Gegensatz zum Sprecher-innenwechsel muss dieser aus o. g. Gründen seiner Auffassung nach „nicht notwendigerweise zwischen den Chattern ausgehandelt werden“ (Kienle u. a. 2017: 146). Beißwenger (2007: 117) verleiht dieser Auffassung Nachdruck, indem er gänzlich von der Nutzung von Konzepten wie Rederecht oder Turn-Taking abrät. Diese Konzepte implizierten mehrere „Möglichkeiten zur Verarbeitung von Verhaltensäußerungen“ (Beißwenger 2007: 117), was jedoch „nur unter radikalen Umdeutungen und Modifikationen möglich“ (Beißwenger 2007: 117) sei. Unter diesen Möglichkeiten sind etwa jene para- und non-verbalen Elemente zu verstehen, die dem Chat aufgrund der Charakterzüge der Distanzkommunikation fehlten.

In mündlichen Gesprächen redet „– von Überschneidungen an den Übergangspunkten abgesehen – nur ein Gesprächsteilnehmer (...), während die anderen schweigen und einen geeigneten Moment für die Ergreifung des Rederechts abwarten“ (Storrer 2001a: 453). Da diese Übergangspunkte im Chat nur selten eindeutig erkennbar sind und bestenfalls durch die Einblendung des o. g. „Person XY schreibt gerade...“ oder einer Art Lesebestätigung¹ strukturiert werden, besteht generell die gleiche und gleichzeitige Möglichkeit für alle, einen Chatbeitrag zu verfassen, was in Verbindung mit der chronologischen, linearen Anordnung aller eingehenden Chatbeiträge durch den Server auch als *Mühlen-Prinzip* bekannt ist: „Wer zuerst kommt, mahlt zuerst“ (Storrer 2001a: 452).

Storrer empfiehlt deshalb für die Chatkommunikation, deren Zweck über den eines sog. *Plauderchats* hinausgeht, die Regulierung der Interaktion. Dazu schlägt Storrer (2001b: 13) zwei Strategien vor: „Die moderierte Sequenzierung und die Regulierung über Konventionen“.

Die erstgenannte Strategie sieht den Einsatz einer Person im Chat vor, die Beiträge thematisch bündelt und geordnet veröffentlicht. Die automatische Sortierung durch den Server tritt hierhinter zurück. Diese Variante kommt vor allem in Chat-Konstellationen zum Einsatz, an denen mehrere Personen beteiligt

¹Diese Lesebestätigung existiert heutzutage in nahezu allen Chatanwendungen. In WhatsApp beispielsweise signalisiert ein kleiner Haken, dass die Nachricht erfolgreich versandt wurde. Zwei Haken bestätigen die erfolgreiche Übertragung an die andere Person und zwei blaue Haken gelten als Lesebestätigung. Dies heißt jedoch nur, dass die/der Chatpartner:in den Chatverlauf bis zur aktuellsten Nachricht gescrollt hat. Gleichermassen verhält es sich bei Skype, nur dass dort „gelesene“ Nachrichten durch die Markierung mit einer kleinen Version des Profilsbilds der Person gekennzeichnet werden.

sind (z. B. Fragerunden mit Politiker·innen). Storrer mahnt jedoch hierzu an, dass durch die Moderatorin oder den Moderator auch einzelne Beiträge (willentlich oder unabsichtlich) ignoriert bzw. censiert werden können. Weiterhin reicht diese Strategie häufig nicht über eine einzige Abfolge von Frage und Antwort hinaus (Storrer 2001b: 13 f.).

Die Regulierung über Konventionen hingegen bietet sich laut Storrer (2001b: 14) für „Gruppen mit überschaubarer Größe“ an. Diese Variante zielt auf institutionelle Settings ab, etwa online durchgeführte Lehrveranstaltungen an einer Universität², lässt sich jedoch problemlos auf kleinste Konstellationen wie einen Chat zwischen zwei Personen übertragen.

Die zwei zentralen Konventionen, die Storrer anführt, bedingen sich in gewisser Hinsicht gegenseitig. Die Aufteilung der Chatnachricht in Blöcke dient zunächst dazu, allen Beteiligten anzuzeigen, dass der eigene Beitrag noch nicht beendet ist. So soll vermieden werden, dass die anderen Personen sich langweilen, während sie auf die auf dem Bildschirm eingehenden Nachrichten warten. Damit jedoch auch deutlich wird, dass die Nachrichten als Blöcke mit Bezug auf einander verstanden werden sollen, bedarf es der sog. „Fortsetzungsmarkierung“ (Storrer 2001b: 15). Hierunter sind beispielsweise die drei Punkte am Ende einer Nachricht zu begreifen (...), aber auch das Aufteilen der gesamten Information an markanten semantischen oder syntaktischen Punkten.

Es ist vielmehr so, dass die einzelnen Aktivitäten im Chat durch besondere Aufmerksamkeit und eine Ökonomisierung der Interaktion erfolgen. Die Gesprächspartner·innen achten etwa durch die o. g. Anzeige von *schreibt gerade...* und durch eine zerstückelte Beitragssequenz darauf, dass die Person gegenüber mehr Gelegenheiten für einen eigenen Beitrag hat als in einem Gespräch von Angesicht zu Angesicht.

2.2 Medientheoretische Verortung

Nachdem sich der vorausgehende Abschnitt der Kommunikationsform des Textchats gewidmet hat, folgt nun eine Darstellung der Anforderungen aus technologischer und medientheoretischer Sicht. So ist einleitend hierzu festzuhalten, dass die Untersuchung von Chatkommunikation häufig mit dem Begriff des

²Obwohl das hier zugrundeliegende Werk bereits aus dem Jahre 2001 (Storrer 2001b) stammt, lassen sich Parallelen zu den sich im Rahmen der weltweiten Corona-Pandemie im Jahre 2020 herausgebildeten Konventionen der digitalen Kommunikation erkennen. Die Konvention, das Mikrofon nur dann einzuschalten, wenn man einen Redebeitrag leisten möchte, oder die Nutzung von Icons wie *Hand heben*, um einen Beitrag anzuzeigen, zählen sicherlich auch hierzu.

2 Linguistischer Aspekt

Individual- und des Massenmediums einhergeht. Der Medienbegriff ist dabei sehr weit gefasst und kann sich, wie Beck (2006: 12) anführt, neben der kommunikationsbezogenen Dimension (Zeitung, Radio, Fernsehen) auch auf Transportmittel (ergo: Medien) beziehen. Generell sei es Zweck von Medien, zur „Entgrenzung des Menschen [beizutragen], also seine zeitliche, räumliche und soziale Reichweite prothetisch [zu] erweitern“ (Beck 2006: 12). Die o.g. Aspekte der Distanzkommunikation (s. Abschnitt 2.1.1.3, S. 11) bestehen also auch bei der medientheoretischen Einordnung.

Weiterhin stellt Döring (2019: 170) kritisch fest, dass die Kontexte, in denen CvK genutzt werden kann, mittlerweile zu stark ausdifferenziert sind, als dass eine eindeutige Zuordnung zu einer einzigen Medientheorie möglich sei. Aus diesem Grund werden in Abschnitt 2.2.2 (S. 18) die gängigsten Ansätze vorgestellt.

2.2.1 Medientheoretische Verbindung von Internet und Kommunikation

Ein markantes Merkmal der CvK ist, dass man von der „Form des sprachlichen Zeigens“ (Weingarten 2001: 1142) hin zu sprachlichen „Formen, die auf ein definiertes Symbolfeld zurückgreifen“ (Weingarten 2001: 1142) gewechselt ist. „Technisch realisierte Kommunikation entsteht nicht quasi-natürlich und ist daher auch nicht für alle Menschen als selbstverständlicher Besitz verfügbar.“ (Weingarten 2001: 1142). Mit dieser Aussage bezieht sich Weingarten unter anderem auf die zwingend notwendige Verfügbarkeit von Soft- und Hardware zur Durchführung der Kommunikation, die als Zugangsvoraussetzungen zu dieser Kommunikationsform aufgefasst werden können. Konkret setzt Skype und der Skype Translators bestimmte Soft- und Hardware sowie ein Nutzerkonto voraus, um dieses Medium nutzen zu können. Andererseits kann der Funktionsumfang auch inklusive Folgen nachsichziehen, da etwa Analphabeten (durch Speech-to-Text) oder Sprecher, die in einem plurilingualen Land leben, aber nur einer der Sprachen mächtig sind, auch so Zugang zu Behörden o.ä. bekämen. Hierzu bedarf es jedoch der Entwicklung „weltweit gültiger Normen für Hardware und Software“ (Weingarten 2001: 1147).

Die technische Entwicklung baut zudem auf dem „vorhandenen sprachlichen Inventar und eingeübten Verwendungsformen bzw. Konventionen“ (Weingarten 2001: 1142) auf. Erst dann kann sich „ein eigenes Formeninventar“ (Weingarten 2001: 1142) entwickeln. Dies lässt sich plastisch an der Entwicklung von Emojis, also stilisierten Gemütsdarstellungen ausgehend vom Prototyp des Smileys,

nachvollziehen. Während Emojis vormals aus Platz- und Zeichenmangel auf kleinen mobilen Displays nur aus zwei bis drei Zeichen bestanden, haben heutige Emojis teilweise sogar die Sphäre der Normierung durchdrungen und sind etwa in der ASCII-Code-Tabelle enthalten. Damit einhergehend hat sich über die vergangenen Jahre auch ein reger Forschungszweig entwickelt, der sich der Kommunikation mit bzw. durch diese Emojis widmet (s. z. B. Pappert 2017).

Wenn bzw. da sie kontinuierlich weiterentwickelt und genutzt werden, könnten der Skype Translator und ähnliche Technologien also nur den Anfang dessen darstellen, was in Zukunft auf dem Gebiet der Kommunikationstechnologie noch zu erwarten ist. Die treibende Kraft hinter dieser Entwicklung ist bereits ausgiebig von der Kommunikationsforschung kartographiert, wie Beck gleich zu Anfang feststellt: Internet und Computer – und mittlerweile auch Smartphones sowie ähnliche Geräte und mit ihnen jegliche Form der Kommunikationstechnologie – bieten Lösungsmöglichkeiten für „kollektive Bedürfnisse und gesellschaftliche Problemlagen bzw. -wahrnehmungen“ (Beck 2006: 1).

Versteht man die Sprachbarriere daher ebenfalls als eine solche Problemlage, so stellt sich an dieser Stelle zwangsläufig die Frage, welchen Beitrag der Skype Translator zur Lösung leistet und welches (Kommunikations-)Bedürfnis er zu befriedigen im Stande ist.

Beißwenger betont in diesem Zusammenhang, dass man sich der Betrachtung der Rahmenbedingungen und der Gründe, die zur Wahl des jeweiligen Mittels führen, nicht entziehen kann, wenn man Kommunikation entlang ihrer Möglichkeiten – in seiner Wortwahl: entlang der Kommunikationstechnologien – beschreibt (Beißwenger 2007: 13). So kann bereits hier gemutmaßt werden, dass die Untersuchung von Chats mit Beteiligung des Skype Translators – im Sinne Beißwengers – notwendigerweise die Betrachtung der maschinellen Übersetzung und auch die Abgrenzung von anderen Kommunikationstechnologien erfordert, die ähnliche Funktionalität bieten. In diesem Rahmen sei beispielsweise – zumindest in der Theorie – Verb mobil (s. Abschnitt 3.1.4, S. 41) genannt. Unter dem Projektnamen *Verb mobil* (Wahlster u. a. 2000) wurde in den 1990er Jahren der Versuch unternommen, einen vollautomatischen Telefondolmetschdienst zu konzipieren, der es den Nutzer·innen ermöglichen sollte, Geschäftsgespräche zwischen verschiedenen Sprachen in Echtzeit maschinell übersetzen zu lassen.

Die Telekommunikationstechnologie setzt die „Schaffung sprachlicher Standards in dem jeweiligen Kommunikationsradius voraus bzw. befördert diesen Prozeß“ (Weingarten 2001: 1143). Die technischen Voraussetzungen für die Nutzung von Kommunikationstechnologien sind „Eingabemedien, Speicher- und Transportmedien sowie Ausgabemedien“ (Weingarten 2001: 1144). Diesem Gedanken folgen eindeutig Trepte & Reinecke bei der Definition von Kommuni-

2 Linguistischer Aspekt

kation in den eingangs angeführten Zitaten. „Durch die computerlinguistischen Methoden der automatischen Spracherkennung kann gesprochene Sprache denselben Code erhalten wie eine druck- bzw. heute: tastaturschriftliche Eingabe“ (Weingarten 2001: 1145). Die Kommunikationstechnologien verändern zudem auch den operativen Zugriff auf den Inhalt der Kommunikation: u.a. auch automatische Übersetzung. Durch die bereits o.g. Entgrenzung des Raumes kann die zeitliche Erfassung der Kommunikation irrelevant werden. Dieses Phänomen erkennt Beck (2006: 127) auch in der Verlagerung von Dienstleistung in jüngerer Vergangenheit. Sprechstunden und Beratungen können bei Bedarf ebenfalls per Text-, Audio- oder Videochat abgehalten werden.

2.2.2 Modelle und Theorien der computervermittelten Kommunikation

Die meisten Theorien und Modelle der CvK beruhen auf der Annahme, dass die Chatkommunikation durch einen – in unterschiedlichen Formen – ausgeprägten Mangel im Vergleich zur Kommunikation von Angesicht zu Angesicht gekennzeichnet ist (Trepte & Reinecke 2012: 159). Die Modelle weisen in mehreren Aspekten zum Teil enorme Schnittmengen auf und fokussieren sich meist lediglich auf einen Aspekt oder eine Dimension der CvK. Es ist daher umstritten, ob bei ihnen allgemein von *Theorie* gesprochen werden kann (Döring 2013: 429). Generell können die bestehenden Theorien und Modelle in drei übergeordnete Kategorien eingeteilt werden: Medien- und kanalbezogene Modelle rücken den Einfluss der gewählten Medien zur Kommunikation in den Mittelpunkt. Medienvahlmodelle betrachten die Abwägung für oder gegen ein bestimmtes analoges oder digitales Medium gemäß sozialer Normen, individueller Gewohnheiten oder der interpersonalen Abstimmung. Als dritter Typ rücken individuumsbezogene Modelle die beteiligten Personen in den Fokus und betrachten das Verhalten (Reaktion, Interaktion, Organisation) innerhalb der Kommunikation (Misoch 2006: 115 ff.).

2.2.2.1 Medien- und kanalbezogene Modelle

2.2.2.1.1 Kanalreduktionsmodell

Das ebenfalls unter dem Namen *Restriktionsmodell* (Misoch 2006: 68) bekannte, am häufigsten referenzierte Modell der CvK folgt dem Prinzip der *Kanalreduktion* (Döring 2019: 170). Hierbei wird aus technologischer Sicht auf die o. g. Charakteristika der Chatkommunikation verwiesen. Die Chatkommunikation bietet nicht die Möglichkeit zur Vermittlung von para- und non-verbalen Elementen (Döring

2019: 170). Die Bindung an die Schrift als Medium sowie die räumliche und zeitliche Entgrenzung bewirken, dass „dieser allgemeine Informations- und Aktionsverlust den zwischenmenschlichen Austausch verarmt“ (Döring 2013: 426 f.). Die Nutzung von Chattechnologien trotz dieser Mängel wird in diesem Modell durch äußere Zwänge, unreflektierte Gewohnheiten und unterschiedliche Verhaltensmuster in der Kommunikation erklärt (Döring 2013: 426 f.).

2.2.2.1.2 Filter-Modell

Das Filter-Modell baut ebenfalls auf der Annahme eines Mangels auf und kann als Verlängerung des Modells zur Kanalreduktion angesehen werden. Es ist deshalb auch unter dem englischen Namen *Reduced social cues approach* bekannt (Döring 2013: 427). Den am Chat beteiligten Personen liegen keine sozialen, biographischen oder demographischen Hinweise zum Gegenüber vor, wodurch es „zu *medialer Enthemmung* kommt“ (Döring 2019: 171, Kursivierung im Original). Im positiven Sinne bietet der Chat somit eine anonyme, unbefangene und vorurteilsfreie Umgebung für den Austausch. Negativ hingegen kennzeichnet sich diese Enthemmung häufig durch gesteigerte verbale Aggression (Döring 2013: 427).

2.2.2.1.3 Mediale Reichhaltigkeit

Unter der medialen Reichhaltigkeit können mehrere Modelle mit ähnlichem Ansatz zusammengefasst werden. Neben der Theorie der medialen Reichhaltigkeit (*media richness theory*, Trepte & Reinecke 2012: 161) ist dies besonders das *Digitalisierungs-Modell* (Döring 2013: 427). Beide setzen voraus, dass jedes Kommunikationsmedium sich für unterschiedliche Zwecke besonders eignet. Dabei steht die effiziente und zugleich (zeit-)ökonomische Gestaltung der Kommunikation als bestimmender Faktor im Vordergrund (Trepte & Reinecke 2012: 161). Döring (2019: 171) hält deshalb als Richtlinie fest, dass „komplexere Kommunikationsaufgaben reichhaltigere Medien erfordern“.

2.2.2.2 Medienwahlmodelle

2.2.2.2.1 Rationale Medienwahl

Das Modell der rationalen Medienwahl stellt die Annahme auf, dass Menschen ein bestimmtes Kommunikationsmedium nach sachlichen, inhaltlichen sowie sozialen Bedürfnissen auswählen. Die Gewichtung der Bedürfnisse findet dabei einerseits auf Grundlage des Modells zur medialen Reichhaltigkeit (vorausgehender Abschnitt 2.2.2.1) statt (Misoch 2006: 96). Ein reichhaltiges und soziales (i.S.v.

2 Linguistischer Aspekt

Lebhaftigkeit und Interaktivität sowie Vermeidung von Missverständnissen) Medium wird dabei über die Auswahlmöglichkeiten gestellt, die diese Aspekte weniger erfüllen. Hieraus lässt sich eine hierarchische Ordnung ableiten, an deren Spitze die Kommunikation von Angesicht zu Angesicht steht (Döring 2013: 425). Andererseits dient die Einschätzung der Kommunikationsteilnehmer·innen, wie effektiv und zugleich zufriedenstellend der Austausch wahrgenommen wird, als Entscheidungsgrundlage für die rationale Medienwahl. Hierbei werden vor allem die Schnelligkeit, Komplexität, Genauigkeit sowie Vertraulichkeit des jeweilig gewählten Mediums bewertet (Misoch 2006: 96).

2.2.2.2 Normative Medienwahl

Die normative Medienwahl geht über die rationale Entscheidungsfreiheit des Individuums im rationalen Modell hinaus (Döring 2013: 426) und stellt soziale Faktoren in den Fokus der Entscheidungsfindung. Gerade in einem institutionellen oder organisatorischen Umfeld bestimmen Konventionen und Normen die Wahl des Mediums, was zur Folge hat, dass nicht immer die optimale Gewichtung zwischen sachlichen, inhaltlichen und sozialen Bedürfnissen (s.o.) eingehalten werde, sondern vielmehr dann arbeitsprozess- oder organisationsbezogene Faktoren eine dominante Rolle spielen (Misoch 2006: 100 f.).

2.2.2.3 Interpersonale Medienwahl

Die interpersonale Medienwahl erfolgt in Abstimmung mit der jeweiligen Person, mit der die Kommunikation angestrebt wird. Dabei müssen die Medienpräferenzen aller Beteiligten in Einklang gebracht werden. Dies wird durch die Abwägung der sozialen Beziehung zueinander, der bevorzugten oder ablehnenden Nutzung eines bestimmten Mediums, der Persönlichkeit (z. B. schüchtern) sowie der demographischen Daten der beteiligten Personen erreicht. (Misoch 2006: 108 f.)

2.2.2.3 Individuumsbezogene Modelle

2.2.2.3.1 Theorie der sozialen Informationsverarbeitung

Die Theorie der sozialen Informationsverarbeitung (*social information processing theory*) stellt die Annahme auf, dass die unterschiedlichen Mängel der CvK, wie sie etwa im Rahmen der Kanalreduktion postuliert werden, durch das Nutzungsverhalten ausgeglichen werden. Hieraus hat sich beispielsweise über die vergangenen Jahre die Bandbreite an Emojis und anderen audiovisuellen Elementen

herausgebildet, die der üblicherweise als emotionsarm wahrgenommenen Chatkommunikation wiederum neue soziale und emotionale Dimensionen verleiht (Döring 2019: 171). Trepte & Reinecke (2012: 162) weisen jedoch zugleich darauf hin, dass dieses veränderte Kommunikationsverhalten trotzdem nicht bewirkt, dass sozio-emotionale Tiefe so schnell erreicht wird wie in einer Kommunikation von Angesicht zu Angesicht.

2.2.2.3.2 SIDE-Modell

Das SIDE-Modell (*Social Identity and De-Individuation*) kann als Negativ des Filter-Modells (Abschnitt 2.2.2.1.2, S. 19) angesehen werden (Döring 2013: 427). Das Modell entstammt der Sozialpsychologie und basiert auf zwei Säulen. Die eine Säule ist die Theorie der sozialen Identität, die zwischen personaler und sozialer Identität unterscheidet. In Wahrnehmung der personalen Identität „handeln Individuen aufgrund ihrer eigenen Eigenschaften und Erwägungen“ (Trepte & Reinecke 2012: 169). Im zweiten Fall ist das Handeln an der Gruppenzugehörigkeit und der entsprechenden Wahrnehmung ausgerichtet (Trepte & Reinecke 2012: 169). Durch die soziale Kategorisierung als eine bestimmte Gruppe (z. B. Nationalität, Geschlecht) kommt es zur Angleichung („Homogenisierung“, Trepte & Reinecke 2012: 169) der Mitglieder dieser *in-group*. In der Folge wird „[d]ie Ähnlichkeit der Mitglieder der eigenen Gruppe (...) überschätzt. Gleichzeitig werden die Mitglieder anderer Gruppen als *out-group* wahrgenommen und stereotypisiert“ (Trepte & Reinecke 2012: 169).

Die zweite Säule ist die De-Individualisierung. Das SIDE-Modell charakterisiert die CvK als anonyme Kommunikationsform, in der aufgrund des Mangels an individuellen demographischen und biographischen Hinweisen über die am Chat beteiligten Personen die soziale Identität dominiert. Erklärt wird dies mit der Annahme, dass „die Übergänge zwischen der personalen und der sozialen Identität fließend sind“ (Trepte & Reinecke 2012: 169), wodurch mitunter die individuelle Motivation im Chat hinter der Gruppenzugehörigkeit zurücktritt. Die Orientierung im Chat erfolgt dann „umso stärker an den wenigen erkennbaren Gruppenmerkmalen und Gruppennormen“ (Döring 2013: 427). Aus dieser anonymen Konstellation erwächst hieraus ein Umfeld, in dem die Identitätsmerkmale der Gruppe bestärkt werden (Trepte & Reinecke 2012: 170).

2.2.2.3.3 Hyperpersonale Kommunikation

Eine Kommunikationssituation im Sinne des Modells der hyperpersonalen Kommunikation ist im Vergleich mit der Kommunikation von Angesicht zu Angesicht

2 Linguistischer Aspekt

durch die Übersteigerung der erwartbaren sozialen Maßstäbe gekennzeichnet (Misoch 2006: 139). Es richtet die Aufmerksamkeit besonders auf die Beziehung, Interaktion und Auswirkung zwischen Sender-in, Empfänger-in, gewählten Kanal und der Rückmeldung (Trepte & Reinecke 2012: 164 f.). Da sich die CvK mehrheitlich durch Anonymität auszeichnet, erhalten eine Senderin oder ein Sender „damit die Möglichkeit der selektiven Selbstdarstellung, kann die selbstbezogenen Informationen frei auswählen, gestalten und sogar erfinden“ (Trepte & Reinecke 2012: 165). Auch die anderen beteiligten Personen können das Selbstbild auf diese Weise kontrollieren, da allen Beteiligten teilweise die Möglichkeiten zur Überprüfung der persönlichen Angaben fehlen. Das gilt ebenfalls für das Verfassen und Redigieren von Nachrichten im Chat. Die Tendenz dieses Verhaltens ist auf eine übermäßig positive Selbstdarstellung (Idealisierung) ausgerichtet, wobei unangenehme oder negative Eigenschaften stark zurückgehalten werden (Misoch 2006: 139 f.). Da die CvK nur quasi-synchron verläuft, bietet sie den am Austausch beteiligten Personen die Gelegenheit, intensiver über die eigenen Beiträge nachzudenken und diese ggf. vor dem Versenden noch einmal zu bearbeiten – was auch dem Prinzip der Idealisierung folgt (Trepte & Reinecke 2012: 165). Dadurch kommt es auf der anderen Seite dazu, dass die Personen, die die Nachrichten empfangen, ebenfalls die (begrenzten) ihnen zur Verfügung stehenden Informationen übermäßig idealisieren. Dies geschieht auf Grundlage von Projektionen, die von der Gruppenidentität oder Persönlichkeitsstereotypen ausgehen (Walther u. a. 2011: 4). Ist beispielsweise bekannt, dass eine der am Chat beteiligten Personen sportlich ist, so wird unter allen weiteren Chatteilnehmer·innen zunächst implizit davon ausgegangen, dass diese Person Attribute erfüllt, die mit dieser Eigenschaft einhergehen, wie etwa Ehrgeiz und Ausdauer (Trepte & Reinecke 2012: 167). Die Rückmeldung als letzte Komponente der hyperpersonalen Kommunikation verstärkt, intensiviert und erweitert die bestehenden Eindrücke der positiven Selbstdarstellung. Die am Chat teilnehmenden Personen glauben, dass die Eigenschaften, die sie den anderen Personen zuschreiben, tatsächlich zu 100 % so zutreffen und passen dementsprechend ihr gesamtes Kommunikationsverhalten an. Umgekehrt versuchen die Personen, die ihnen zugeschriebenen Eigenschaften im Chat zu erfüllen und abzubilden. Der Idealisierungsprozess wird folglich durch die Rückmeldung verstärkt (Walther u. a. 2011: 5).

2.3 Rückanbindung an konkrete Chat-Anwendungen

Die Auseinandersetzung mit dem Chat als Form der CvK ist deshalb interessant, weil hier Elemente aus den – zumindest auf den ersten Blick widersprüchlichen –

2.3 Rückanbindung an konkrete Chat-Anwendungen

Polen der Mündlichkeit und Schriftlichkeit aufeinandertreffen. Obwohl der Chat als Prozess aus schriftlicher Ein- und Wiedergabe besteht, weist er doch konzeptionelle Elemente der Mündlichkeit auf. Dazu zählt der Austausch *en bloc* zwischen beteiligten Chatter-innen, der nicht immer exakt den Satzgrenzen folgt, eine Orientierung am Wesen der Kommunikation von Angesicht zu Angesicht oder auch dem Versuch, so unmittelbar wie möglich zu wirken. Eine Äußerung, also ein Chatbeitrag, stellt jedoch im Vergleich zu einem medial mündlichen Gespräch eine permanente Nachricht dar, die noch nachträglich redigiert oder mit einem weiteren Chatbeitrag korrigiert werden kann.

Die in den vorausgehenden Abschnitten vorgestellten theoretischen Grundlagen zur CvK liegen schwerpunktmäßig auf Chatsituationen mit mehr als zwei beteiligten Personen. Deutlich wird dies einerseits dadurch, dass die zitierten Autor-innen die Chatkommunikation zumeist auf einer hierarchischen Ebene mit Massenmedien denken und dementsprechend auch Mailinglisten, Foren und ähnliche Kommunikationsformen mit vergleichbar großem Radius untersuchen. Andererseits sind einige der Charakteristika der Chatkommunikation ebenfalls im Umfeld der massenmedialen Untersuchung definiert worden, wie beispielsweise Chat-Rooms, Kanäle (*Channels*), moderierter Chat, das Chatverhalten an sich und auch die zeitliche sowie räumliche Nutzung von Chats.

Die im Rahmen dieser Arbeit zu untersuchende Chatkommunikation ist hingegen auf zwei beteiligte Personen begrenzt. Es besteht daher die Möglichkeit, dass die gesamte Kommunikation nicht alle Elemente aufweist, die in den vorausgehenden Abschnitten vorgestellt wurden. Ob dies auch generell auf die Komplexität der Kommunikation zutrifft, bleibt zu untersuchen. Aushandlungsstrategien des Rollenwechsels müssen beispielsweise lediglich zwischen zwei Personen ausgehandelt werden. Das Rederecht wechselt ebenfalls zwischen zwei und nicht noch mehreren Personen. Andererseits ist der Skype Translator hinzuzudenken, der in den folgenden Kapiteln noch genauer präsentiert wird. An dieser Stelle ist jedoch schon zu überlegen, in wie weit diese Technologie die Komplexität einer Kommunikationssituation mit zwei Personen wiederum erhöht und welchen Einfluss der Skype Translator auf den Rollenwechsel und alle weiteren o. g. Charakteristika der Chatkommunikation hat.

Es zeigt sich, dass der Chat zur Ausbildung neuer Strategien der Selbstdarstellung, Informationsvermittlung sowie Sprachverwendung führt. Dieser Prozess wurde über die letzten Jahrzehnte entlang nahezu aller aufkommenden Kommunikationsmittel untersucht. Mit Blick auf eine von der maschinellen Übersetzung vermittelten Kommunikationssituation wie dem Skype Translator ist somit sicherlich erneut eine Strategieentwicklung zu erwarten.

2.4 Der Textbegriff im Umfeld von Chatkommunikation

Koch & Oesterreicher (2011: 3) beginnen in ihrer Ausführung zum Wesen von Mündlichkeit und Schriftlichkeit mit den beiden Ebenen der Konzeption und des Mediums. Sprachliche Äußerungen können demnach auf Ebene der Konzeption *gesprochen* oder *geschrieben* realisiert und auf der Ebene des Mediums *graphisch* oder *phonisch* kodiert sein. Während zwischen beiden Medien eine klare Trennlinie besteht, ist die Konzeption als Kontinuum zwischen den beiden Polen aufzufassen.

Zwar erkennen Koch & Oesterreicher (2011: 4) gewisse Präferenzen bei der Kombination beider Merkmale, grundsätzlich erachten sie die auf den ersten Blick nur schwer miteinander zu vereinbaren Kombination von schriftlichen und mündlichen Merkmalen als durchaus möglich. Dies trifft auch auf den Chat zu, den die Autoren als konzeptuell gesprochen und graphisch kodiert auffassen: „Der *chat* ist sogar eines der schönsten Beispiele dafür, dass im graphischen Medium eine relative, allerdings auch in diesem Falle noch limitierte Annäherung an dialogische, spontane Nähersprachlichkeit möglich ist.“ (Koch & Oesterreicher 2011: 14, Kursivierung im Original)

Mit dieser Unterscheidung sei nun auf Coseriu (1988: 184) verwiesen, der Text als Redeakte, bzw. zusammenhängende Reihenfolge derselben auffasst. Auch Ehlich geht von derartigen Akten im Sinne einer sprachlichen Handlung aus: „Der Begriff „Text“ kann innerhalb einer Theorie des sprachlichen Handelns entfaltet werden. Ich entwickle Vorschläge für einen solchen Textbegriff. Texte beziehen sich auf zerdehnte Sprechhandlungssituationen.“ (Ehlich 2011: 532) Diese weite Definition greifen Koch & Oesterreicher (2011) auf und grenzen den Begriff *Text* in Umfeld der Kommunikation enger ein:

Vergegenwärtigen wir uns kurz die wichtigsten Instanzen und Faktoren der sprachlichen Kommunikation. Mindestens zwei Interaktionspartner treten miteinander in Kontakt, wobei sie – gegebenenfalls im Wechsel – die Gesprächsrollen des Produzenten und des Rezipienten einnehmen. Dabei entsteht eine Nachricht, ein Diskurs/Text, der sich auf Gegenstände und Sachverhalte der außersprachlichen Wirklichkeit bezieht. (Koch & Oesterreicher 2011: 6, Hervorhebungen ausgelassen)

Wie in dem Zitat ersichtlich unterscheiden die Autoren dabei zwischen dem medial phonisch kodierten Diskurs, der dem Gespräch in der Umgangssprache gleichgesetzt werden kann, und dem Text als medial graphisch gefasste Sprachhandlung. Beißwenger widmet sich in seiner Dissertation der Stellung dieser Instanzen und Faktoren zueinander. Er problematisiert dabei die zum Teil unscharf

vorgenommene Trennung zwischen Gespräch und Text im Rahmen einer Chatkommunikation:

Probleme bereitet aber nach wie vor eine genaue Verortung des Chat in seiner Stellung zum Gespräch. „Wenn Chat nicht Text ist“, so müsste man meinen, „dann kann Chat eigentlich nur Gespräch sein“. Ganz so einfach ist es aber nicht, da Chat andererseits auch Merkmale aufweist, die ihn in deutlicher Differenz zum Gespräch zeigen. (Beißwenger 2007: 2)

Brinker plädiert für einen integrativen Textbegriff, der einerseits den Satz als „wichtigste Struktureinheit“ (Brinker 2001: 17) ansieht, andererseits jedoch auch eine so weite Auslegung zulässt, dass „auch kleinere sprachliche Gebilde (...) unter ganz bestimmten situativen Bedingungen als Texte im kommunikativen Sinne fungieren können“ (Brinker 2001: 17).

Im Gegensatz zu Beißwenger liegt der Schwerpunkt von Brinkers Definition auf monologischen Texten, während „dialogische sprachliche Gebilde (Gespräche)“ (Brinker 2001: 17) im Rahmen der Dialog- und Gesprächsanalyse betrachtet werden. Diese Unterteilung ergibt sich aus dem „Merkmal der Kohärenz“ (Brinker 2001: 12), das im Falle von Gesprächen erst durch mehrere Kommunikationsteilnehmer·innen gegeben ist.

Der Textbegriff im Rahmen der Chatkommunikation sollte also nur als ein Baustein, als ein Akt innerhalb der ihm hierarchisch übergelagerten Kommunikation aufgefasst werden. Text kann daher als kohärente Aneinanderreihung von kohäsiven grammatischen Elementen verstanden werden, die einer kommunikativen Absicht folgen und die Erwartungshaltung im funktionellen Sinne eines Empfängers treffen (Bußmann u. a. 1998: 1187). Entgegen der eigentlichen Wahrnehmung als „anthropomorphisierend“ (Beck 2006: 7) verstandene Netzkomunikation beschränkt sich die Interaktion letztenendes auf die Auseinandersetzung „mit einer Maschine bzw. einem mediatisierten Datenuniversum, letztlich einem Text“ (Beck 2006: 7). Der Sinn als Eigenwert des Textes, von dem auch Coseriu (1988: 184 ff.) spricht, bildet sich erst in der globalen Betrachtung aller Sätze heraus.

Zugleich definieren im Rahmen von Eye-Tracking-Studien vergleichsweise wenige Beiträge konkret einen operationalisierten Textbegriff. In vielen Fällen wird stillschweigend von einzelnen Sätzen (Rayner u. a. 2003), einzelnen Passagen oder lediglich einzelnen Wörtern als Studienobjekt ausgegangen. Im Falle von Eye-Tracking-Studien im Bereich der Translationswissenschaften sind es meist Texte mit einer Länge von bis zu 300 Zeichen (Lykke Jakobsen & Jensen 2008) oder einzelne Segmente (Schaeffer u. a. 2017), die den Probanden präsentiert werden. So werden für Studien zum Postediting einzelne Segmente dargestellt, wie sie auch in gängigen CAT-Tools vorkommen (Schaeffer u. a. 2017,

2 Linguistischer Aspekt

Vardaro u. a. 2019). Ansätze, die die Wahrnehmung von Untertiteln untersuchen, greifen naturgemäß auf Einheiten der Länge von Untertiteln zurück (Bisson u. a. 2014, Fox 2018, Kruger u. a. 2018). Die fluide, dynamische Kommunikationssituation per Textchat über Skype erfordert demnach eine Anpassung des Textbegriffs. Es handelt sich einerseits um mehr als die üblicherweise verwendete Zeichenanzahl bei Eye-Tracking-Studien. Andererseits sind die Segmente ungleichmäßig umfangreich und stehen dennoch in einem kontextuellen Zusammenhang zueinander.

Es ist in diesem Zusammenhang wichtig, sich klarzumachen, dass (...) das Verhältnis von Materialität und Textualität im Vordergrund steht: Medien werden zugleich als Technologien und „materielle Kultur“ und damit auch als Bedeutungsträger, Symbole und Texte behandelt, die zudem als Medien noch Texte übermitteln – als Maschinen, als Programme und Programmierbares, als Konsumobjekte, Erlebnisräume etc. Es geht dementsprechend nicht allein um die Inhalte, sondern ebenso um die Objekte als Gegenstände sozialen Handelns und als Technologien (...). (Hartmann & Krotz 2019: 261)

In der Chatkommunikation kann folglich sowohl der einzelne Beitrag an sich als auch die gesamte Interaktion zwischen den beteiligten Personen als Text genommen werden.

2.5 Wissensrepräsentation in digitalen Räumen

Mit Blick auf den Skype Translator und die Studie dieser Arbeit ist eine Betrachtung der Wissensrepräsentation insofern wichtig, als durch den dialogischen Austausch zwischen Menschen neue Wissensstrukturen entstehen. Im Falle des Skype Translators und der Chatkommunikation im Allgemeinen geschieht dies in einem digitalen Raum. Viele der hier referenzierten Arbeiten legen nach wie vor die analoge Alltags- und Berufswelt als Betrachtungspunkt der Wissensrepräsentation zugrunde. Erst seit Mitte der 2000er widmet sich die Forschung auch der Abbildung von Wissensinstanzen im Internet. Umgekehrt jedoch hat der Begriff des *Cyberspace* das Verständnis von Digitalität nachhaltig geprägt. Beck erkennt hierin einen immateriellen Informationsraum,

der tendenziell alle Informationen in „Echtzeit“ bereit hält und die nahezu beliebige dynamische Verknüpfung von Daten erlaubt. Aus dieser Vorstellung, jederzeit und von jedem Ort aus alle Informationen abrufen zu können, speist sich der Mythos der Allwissenheit – wobei hier allerdings Information und Wissen vorschnell gleich gesetzt werden. (Beck 2006: 2)

Deshalb ist es zunächst notwendig, die Grundbegriffe zu definieren. Unter Wissen wird gemeinhin die kontextualisierte Akkumulierung von Informationen verstanden, die sich wiederum aus der Interpretation von zunächst losen Daten (gemeint ist nicht ausschließlich die digitale Konnotation) zusammensetzen, die keine Bedeutung über sich hinaus haben. Daten können als einfache fixierte Angaben verstanden werden oder im digitalen Umfeld auch als „jene Inhalte, welche mit traditionellen Technologien und Systemen (z. B. Datenbanken, Repositories, Expertensystemen) gespeichert werden können“ (Lehner 2014: 54). Ähnlich verstehen Schopflin & Walsh den Wissensbegriff, den sie entlang eines Kontinuums aus „data – information – knowledge – wisdom“ (Schopflin & Walsh 2018: 2) verorten, und weiterhin noch die Kompetenz als Element zwischen Wissen und Erfahrung (als mögliche deutsche Entsprechung des Begriffs *wisdom*, der nach Schopflin & Walsh die langfristige Aneignung von Wissen zur informationsbasierten Entscheidungsfindung, Erkenntnis und dem strategischen Handeln beschreibt).

Sowohl Lehner als auch Schopflin & Walsh weisen darauf hin, dass die eindeutige Definition von Wissen ein schwieriges Unterfangen sei, das bisher keine einheitliche Auffassung hervorgebracht hat. Dies mag auch der Tatsache geschuldet sein, dass die einzelnen Repräsentationsformen von Wissen nicht immer scharf voneinander getrennt werden können und es in den einzelnen Disziplinen unterschiedliche Definitionen gibt. Lehner nennt hier den Wissensbegriff in der KI-Forschung und in der Philosophie und den Informationsbegriff in der Informatik, der Physik oder der Chemie (Lehner 2014: 54).

In diesem Zusammenhang sollte der Umgang mit Wissen ausdifferenziert werden. Auch wenn diese Einteilung aus dem Bereich der Unternehmensführung stammt, so sei auf den Unterschied zwischen Wissens- und Informationsmanagement hingewiesen. Das Informationsmanagement betrifft die objektbezogene Organisierung und Strukturierung von Daten, die einen Bezug über die Daten hinaus aufbaut. Wissen hingegen wird in vielen Disziplinen als die Verbindung von Theorie und Praxis aufgefasst (Megill 2005: 2). Das Ziel des Wissensmanagements ist nach dieser Auffassung die Organisation und Strukturierung von Theorie und Praxis.³

Hier wird die Bedeutung eines operationalisierten Textbegriffs deutlich, da „Kommunikation (von Wissen) in jedem Fall an ein bestimmtes Medium gebunden ist“ (Streitz 1990: 15). Unter diesem Medium versteht Streitz sowohl gesprochene Sprache als auch Texte bzw. Dokumente, woraufhin er zu den Annahmen kommt, dass einerseits „Schreiben (...) nicht nur Textproduktion, sondern auch

³Um beispielsweise Arbeitsprozesse für einen Teil der Belegschaft eines Unternehmens gleichermaßen nachvollziehbar und transparent zu machen.

2 Linguistischer Aspekt

Wissensproduktion [ist]“ (Streitz 1990: 15) und andererseits „Lesen (...) nicht passive Textrezeption, sondern Integration und Konstruktion von Wissen“ (Streitz 1990: 15) darstellt. Eine ähnliche Auffassung vertreten auch Burchardt u. a., die auf den Wert der Schrift für die Wissensrepräsentation eingehen:

Gesprochene Sprache ist die älteste und im Hinblick auf die Evolution des Menschen die natürlicheste Form der sprachlichen Verständigung. Komplexe Informationen und der größte Teil des menschlichen Wissens werden jedoch in schriftlicher Form erfasst und vermittelt. (Burchardt u. a. 2012: 17)

Dieses Zitat verweist implizit auf Assmanns Theorie zum kulturellen Gedächtnis, die sich aus kulturtheoretischer Sicht der Wissensrepräsentation annähert. Ein wichtiger Aspekt dieser Theorie ist die Verfestigung von Wissen und Traditionen, die derart mit dem Übergang von oraler zur Schriftkultur möglich war. Die Schrift ermöglicht also erst die kontinuierliche, langfristige und umfassende Weitergabe von Informationen, die dadurch auch mit einem großen Personenkreis geteilt und somit zu Wissen verwoben werden können (Assmann 2002).

An dieser Stelle ist eine Rückanbindung an die Ausführungen zur CvK festzustellen. Da nicht nur unmittelbar gesprochene Sprache und Texte als Medium von Wissen gelten, sondern auch Druckerzeugnisse, Fernsehen, Radio und besonders das Internet, „beeinflussen die neuen Medien die Konstruktion von und den Umgang mit Wissen so intensiv und nachhaltig, dass soziale und kulturelle Umwälzungen offensichtlich sind, da auch Denkprozesse verändert werden.“ (Eckkrammer 2019: 345) Der Vorteil dabei liegt in der „Demokratisierung des Wissens“ (Eckkrammer 2019: 348), da der Zugang sowohl ökonomisch als auch strukturell erleichtert wird. Dies wird besonders durch freie und kollaborative Projekte wie der Online-Enzyklopädie Wikipedia, dem kostenfreien Office-Programm Libre Office, oder auch dem freien Betriebssystem Linux uvm. deutlich, die auch von einer interaktiven, freien Rechteverwaltung mittels z. B. Creative Commons profitieren. Problematisch und nicht zuletzt in den vergangenen Jahren offenkundig geworden ist jedoch die Schlagseite dieser Wissensverteilung. Die Validität, die Reliabilität, die Aktualität, die Autorenschaft und auch die Präsentation von Wissen und Informationen werden – unter dem Schlagwort *fake news* – für alle Beteiligten immer schwerer zu beurteilen (Eckkrammer 2019: 348). Unter diesen Gesichtspunkten muss also auch die CvK und somit der Skype Translator betrachtet werden, da er zu dieser Art Medium gezählt werden kann. Zugleich stammt die Datengrundlage aus der gleichen Hand wie die Softwareentwicklung – von Microsoft. Dementsprechend obliegt die Kuratierung als auch eine mögliche Einflussnahme auf den Leistungsumfang des Services dem Unternehmen.

Zwar ist der Service auf monetärer Ebene kostenfrei verfügbar, die Nutzerdaten werden jedoch zum Zwecke der Weiterentwicklung genutzt.

Für die CvK bedeutet das wiederum eine Beeinflussung des gemeinsamen Wissensbestandes, da der Grad der Vertrautheit zwischen den Gesprächspartnerinnen als ein weiterer Faktor anzusehen ist (Koch & Oesterreicher 2011: 7). Die vorausgehende, gemeinsame „Kommunikationserfahrung“ (Koch & Oesterreicher 2011: 7) ist für die Autoren eine wichtige Bedingung für die Verteilung von Wissen in Kommunikationssituationen. In Verbindung mit den vorausgegangenen Abschnitten zu den verschiedenen Kommunikationsmodellen kann die Vertrautheit der Gesprächspartnerinnen sowie die gemeinsame Kommunikationserfahrung sicherlich auch als die Freiheit aller am Chat beteiligten Personen interpretiert werden, willentlich und individuell persönliche Informationen entweder zu verschweigen oder besonders zu betonen.

Auf der Ebene der einzelnen Kommunikationsakte legen Koch & Oesterreicher weiterhin Faktoren fest, die die Verteilung von Wissen bedingen. Dies ist vor allem der „Wissenskontext“ (Koch & Oesterreicher 2011: 11), der sich einerseits individuell manifestiert als gemeinsamer Erfahrungswert der Gesprächspartner, und andererseits als allgemeine Dimension zutage tritt, in der „soziokulturelle und universale menschliche Wissensbestände“ (Koch & Oesterreicher 2011: 11) verortet sind. „Wissen ist in jedem Fall die Kenntnis von Beziehungen zwischen Ursache und Wirkung und basiert auf einer systematischen Vernetzung von Informationen. Man könnte Wissen kurz als den Zugang zu Möglichkeiten im Sinne von Wahlmöglichkeiten, Handlungsmöglichkeiten sowie Entwicklungs- und Entscheidungsmöglichkeiten charakterisieren.“ (Lehner 2014: 58) Auch bei der Definition durch Lehner tritt also wieder die Handlungskomponente zutage.

Wissen ist damit im Unterschied zu (passiven) Informationen proaktiv, unterstützt oder veranlasst somit Aktivitäten der Wissensinhaberinnen, kann dafür aber in vielen Fällen nicht einfach übertragen werden wie Informationen. Wissen entsteht allmählich und nicht ausschließlich aus Informationen. Die Fähigkeit, Wissen zu nutzen, zu adaptieren und weiterzuentwickeln, um es in interessante, innovative Produkte und Dienstleistungen umzusetzen, wird die Wettbewerbsposition von Technologieunternehmen auch in den kommenden Jahren prägen. Diese Punkte sind zu gegebener Zeit erneut aufzugreifen, wenn die Rolle der MÜ-Ausgabe innerhalb der CvK eingehender untersucht wird, da zu vermuten ist, dass die MÜ des Skype Translators einen Einfluss auf die einzelnen Kommunikationsakte sowie die Verteilung des Wissens innerhalb der Chatkommunikation hat.

In diesem Zusammenhang führen Döring & Pöschl (2007: 274) verschiedene Vorteile des Wissensaustauschs im Chat an, zu denen u. a. die Schnelligkeit

2 Linguistischer Aspekt

des Austauschs, die Vielfalt der Austauschmöglichkeiten sowie die Referenzierbarkeit zählen. Durch die bereits erwähnte synchrone Kommunikation werden Chatpartner·innen in die Lage versetzt, „sich ohne Medienbrüche von ihren jeweiligen Rechnern Protokolle, Programme oder Bilder zu[zu]spielen“ (Döring & Pöschl 2007: 274) und diese durch Referenzen und Verlinkungen in einen neuen Kontext zu setzen.

Als Nachteile halten Döring & Pöschl dem jedoch zugleich entgegen, dass digitale Kommunikation unpersönlich – teilweise gar anonym – sei und Chatpartner·innen somit nie die inhaltliche Autorität oder Legitimation ihrer Beiträge erreichen können, wie es von Angesicht zu Angesicht möglich wäre. Eine in einen Chat integrierte MÜ hat sicherlich Einfluss auf diesen Nachteil. Entweder vermag es die MÜ-Ausgabe, den Beiträgen mehr Autorität und Legitimation zu verleihen, da Sie in mehrsprachigen Chats neue Anhaltspunkte liefert. Oder Sie verstärkt das Misstrauen in die Kommunikationssituation durch offensichtliche Übertragungsfehler, die die Nutzer·innen zweifeln lassen. Weiterhin sei Chatkommunikation inkohärent und erfordere eine ständige Aufmerksamkeit, um keinen Beitrag zu verpassen. Dies könne gerade bei einem intensiven, stakkatoartigen Austausch von Nachrichten passieren, da die meisten Anwendungen die einzelnen Beiträge in linearer Reihenfolge auf dem Bildschirm präsentierten. Diese Drucksituation sei anstrengend und ermüdend (Döring & Pöschl 2007: 274 ff.).

Mit Blick auf alle diese Aspekte der Wissensrepräsentation zeichnen deshalb Bernardini u. a. (2020: 297) ein Bild von Übersetzer·innen und Dolmetscher·innen als diejenigen zentralen Figuren, die den Austausch an Ideen und Wissen sowohl in der realen Welt als auch in jüngerer Zeit auch intensiv im Internet anregen. Übersetzer·innen und Dolmetscher·innen vermögen es, sowohl das künstlerisch-musische menschliche Genie als auch die wissenschaftlich-fachliche Präzision zwischen Sprachen zu übertragen und reißen somit stetig Sprachbarrieren nieder. In Verbindung mit den Erkenntnissen von Döring & Pöschl (2007: 277 f.), die Wissenskommunikation sei derart effektiv möglich, dass etwa zwei Drittel aller sich stellenden Probleme im Chat gelöst werden könnten, ist der Stellenwert von sprachübergreifendem Austausch auch mit Blick auf eine Sprach-, und genauer, eine Übersetzungstechnologie im Gedächtnis zu behalten.

2.6 Der soziale Status einer Sprache im digitalen Raum

Das Konzept *Raum* ist in vielerlei Hinsicht ein interessantes. Der Schwerpunkt dieser Arbeit liegt eindeutig nicht auf der Auseinandersetzung, geschweige: der Definition, eines Raum-Begriffs. In den vorausgegangenen Abschnitten wurde

2.6 Der soziale Status einer Sprache im digitalen Raum

dieses Konzept bereits angeschnitten und zumindest mit Blick auf die CvK anfänglich skizziert.

Der Status einer Sprache hängt nicht nur von der Zahl der Sprecher ab und der Menge von Büchern, Filmen und Fernsehsendern, sondern immer mehr auch von der Präsenz der Sprache im digitalen Informationsraum und den verfügbaren Softwareprodukten. (Burchardt u. a. 2012: 2)

Mit diesem Zitat stellen Burchardt u. a. (2012) vom Forschungsnetzwerk *META NET* bereits 2012 fest, dass das soziale Ansehen einer Sprache nicht mehr allein durch kulturell-künstlerische Beiträge definiert wird. Es ist immer mehr die digitale und wirtschaftliche Präsenz, die den Stellenwert einer Sprache prägt. Dies ist in den vergangenen Jahren nicht mehr ausschließlich am Englischen zu beobachten, das im Umfeld digitaler Themen häufig als *lingua franca* betitelt wird. Mit den wachsenden wirtschaftlichen Ansprüchen Chinas wird auch zunehmend Mandarin zu einer Sprache, die Wissenschaft und Technologie – und somit auch das Internet – repräsentiert.

Die Beiträge des Forschungsnetzwerks *META NET* sprechen dabei nicht nur von Räumen, sondern auch von einer (europäischen) Informationsgesellschaft. In der ideellen Konzeption lassen sich dabei jedoch Parallelen entdecken. Denn einerseits wird der abhängige Teil *Informations-* des Kompositums schwerpunkt-mäßig in Bezug auf eine technologisierte und digitalisierte Gesellschaft gedacht, andererseits ergeben sich auch Schnittmengen zwischen Raum und Gesellschaft im (sozio-)linguistischen Sinne.

Eben dieses Verhältnis problematisiert Tesseur (2017: 630 f.) mit dem Hinweis, dass die Vernachlässigung von Übersetzung im globalisierungsempfindlichen Untersuchungsbereich der Soziolinguistik auch ein Problem für den Zugang zu Informationen, Möglichkeiten, der Wissensproduktion und letzten Endes der paritätischen Teilhabe darstellt. Weiterhin erkennt Tesseur in der Übersetzung einen Schauplatz für Konfrontationen darüber, wer entscheiden darf, was übersetzt wird, wie es übersetzt wird, welche Sprachen beteiligt sein dürfen und wer dadurch die Deutungshoheit über Wissen(-szugang), legitimierte Sprachen und rechtschaffendes Handeln erhält (Tesseur 2017: 631). Tesseur greift dabei ganz explizit ein räumliches Konzept auf, wenn er im Rahmen seiner Ausführung die Übersetzung als spezielle Form eines situierten Aufeinandertreffens, z. B. dem Ort, wo Übersetzer·in und Institution aufeinandertreffen, auffasst. Dort werden Entscheidungen getroffen, die auf den Werten und Glaubenssätzen, die dem (institutionellen) Setting innewohnen und in dem Übersetzer·innen arbeiten und interagieren (Tesseur 2017: 632).

2 Linguistischer Aspekt

Losgelöst von der Ausrichtung auf den institutionellen Bereich erwächst hieraus eine Forderung, die Burchardt u. a. wie folgt formulieren:

Dadurch ist der Einzelne angehalten, neue Sprachen zu lernen. Insbesondere sind aber Entwickler gefragt, neue Technologien zu schaffen, die das gegenseitige Verstehen und den Zugriff auf gemeinsam nutzbares Wissen sicherstellen. (Burchardt u. a. 2012: 5)

Bernardini u. a. (2020: 298) gehen in ihrem Beitrag soweit, die Existenz ganzer Institutionen, in diesem Fall: Europas, an eine gehegte und gepflegte Sprachdiversität auf dem Kontinent zu binden. Hierbei unterstreichen sie die Funktion, die die Verfügbarkeit von (digitalen) Inhalten in mehreren Sprachen und insbesondere in den Regional- bzw. Minderheitensprachen innerhalb Europas einnimmt: Sie stützt die Position der jeweiligen Sprache gegenüber einer anderen, übergeordneten Sprache. Umgekehrt ist die Bereitstellung von mehrsprachigen Inhalten besonders im Internet auch eine treibende Kraft der Diversifizierung von Sprachdienstleistungen. Da gerade häufig für „kleinere“ Sprachen oder Fachgebiete weniger professionelle Ressourcen bestehen, bilden sich informelle Laiendienstleistungsbranchen aus, um dennoch Inhalte für die Sprachgemeinschaft zu schaffen. Paradebeispiele hierfür sind etwa Fan-Subbings oder die Lokalisierung von Open-Source-Projekten. Aber auch die alltägliche Kommunikation muss derart gewährleistet sein: Sprecher jeder Gemeinschaft, auch von Minderheiten- bzw. Regionalsprachen, dürfen und müssen sich – zumindest innerhalb der EU – in ihrer bevorzugten Sprache mit der Verwaltung, Kultur, Bildung, Rechtssprechung und dem Gesundheitswesen auseinandersetzen können. Weiterhin heben Bernardini u. a. (2020: 298) hervor, dass gesamte Branchen von der mehrsprachigen Verfügbarkeit von Inhalten profitieren: Dies gilt insbesondere für die Marketing- und Unterhaltungsbranche sowie die Sozialen Netzwerke und Technologiedienstleister. In diesem Zusammenhang verweisen die Autoren auf eine Studie, nach der Konsumenten deutlich seltener auf Internetseiten einkaufen, die ausschließlich auf Englisch angeboten werden. Die eigene Sprache scheint also einen Anreiz für Konsum zu schaffen.

2.6.1 Das Katalanische im digitalen Raum

Katalanisch wird in der Autonomen Gemeinschaft Katalonien, in der Valenciaischen Gemeinschaft, in der französischen Catalunya Nord, auf Sardinien, auf den Balearen, in Andorra und in einem östlichen Gebiet der Region Aragón (*La Franja de Ponent*) gesprochen. In den genannten Autonomen Regionen Spaniens

2.6 Der soziale Status einer Sprache im digitalen Raum

ist das Katalanische im jeweiligen Statut als kooffizielle Sprache neben dem Kastilischen festgelegt, wobei die Spanische Verfassung von 1978 das Kastilische als Amtssprache des Staates festlegt und den einzelnen Regionalsprachen lediglich einen rechtlichen Status auf dem Gebiet der jeweiligen Region einräumt. Innerhalb dieses Sprachraumes sind fünf Varietäten auszumachen: *septentrional*, *central*, *nord-occidental*, *baleàric* und *valenciac* (Moreno u. a. 2012: 9). Daraus ergibt sich die Konstellation, dass in Katalonien die Mehrheit der Bewohner·innen mindestens zweisprachig sind. Erhebungen von 2018 zufolge⁴ sind von ca. 7,5 Mio. Einwohner·innen 7,0 Mio. (94,60 %) in der Lage, die Sprache zu verstehen und 6,0 Mio. (80,40 %) sie auch zu sprechen. Es finden sich allerdings zum Teil stark divergierende Sprecher·innenzahlen abhängig davon, wie die Erhebungskategorien gebildet wurden. So weisen Moreno u. a. (2012: 9) darauf hin, dass die Zahlen auf etwa 77,5 % sinke, wenn man nicht ausschließlich von Sprecher·innen ausgeinge, die in Katalonien geboren sind. Dennoch bescheinigen Casesnoves u. a. (2019: 527) dem Katalanischen eine positive Entwicklung, die am Übergang von einer Phase der Normalisierung nach Ende der Franco-Diktatur (*normalització*) zu einer der Nachhaltigkeit sichtbar wird.

Diese Tendenz wird auch an der Vielzahl an katalanischsprachigen öffentlichen Institutionen deutlich, wo es neben katalanischen Radio- und Fernsehsendern seit Mitte der 2000 das sog. *Institut Ramon Llull* gibt, eine Einrichtung deren Zielsetzung mit der des deutschen Goethe-Instituts ist, das die Förderung und Verbreitung der katalanischen Sprache zum Ziel hat. Hinzu kommen noch die Pflege der katalanischen Kultur durch *Omnium Cultural* und *TERMCAT*, ein Zentrum zur Verwaltung von Terminologie auf Katalanisch (Moreno u. a. 2012: 10).

Die beiden weiteren Autonomen Gemeinschaften, die Balearischen Inseln und die Valencianische Gemeinschaft, zeichnen sich durch eine zurückhaltendere, wenn nicht gar konservative Sprachpolitik aus. Das ist darauf zurückzuführen, dass das Katalanische im Statut der Valencianischen Gemeinschaft im Gegensatz zum Statut Kataloniens nicht als bevorzugte Sprache bezeichnet wird bzw. im Falle der Balearen erst viel später als in Katalonien mit einer nachhaltigen Sprachförderung und -entwicklung begonnen wurde (Casesnoves u. a. 2019: 527 f.).

Auch deshalb entfällt die Mehrheit der (populär-)wissenschaftlichen Auseinandersetzung mit diesem Thema auf die Situation des Katalanischen in der Autonomen Gemeinschaft Katalonien (Casesnoves u. a. 2019, Moreno u. a. 2012, Berché 2014).

⁴S. hierzu https://www.plataforma-llengua.cat/media/upload/pdf/informecat2018_1528713023.pdf, zuletzt abgerufen am 31. Januar 2022.

2 Linguistischer Aspekt

Die Förderung und Entwicklung von (Sprach-)Technologien, die das Katalanische betreffen, wird sowohl von der katalanischen Regionalregierung als auch vom spanischen Staat unterstützt. Nicht erst seit der Einführung der katalanischen Top-Level-Domain *.cat* im Jahre 2005, die explizit als Element der Förderung der katalanischen Sprache in der Welt gesehen wird, gibt es Ansätze, die Sprache in die Technologieforschung miteinzubeziehen. Die Einführung der Top-level-Domain war allerdings auch eine zutiefst politische Neuerung, da für gewöhnlich nur souveräne Staaten eine solche Adresse zugeteilt bekommen.

Die generelle Verteilung von Sprachtechnologien, die für das Katalanische verfügbar sind, ist allerdings unausgeglichen. So gibt es zwar einen lebhaften Forschungszweig, der sich mit der (Weiter-)Entwicklung der Textverarbeitung aus einandersetzt und auch die Spracherkennung für das Katalanische vorantreibt. Die Forschung zur maschinellen Übersetzung ist hingegen vergleichsweise bescheiden, wenn es über die Sprachkombination mit dem Spanischen hinausgeht (Moreno u. a. 2012: 22). Zugleich bieten jedoch die Internetgroßkonzerne wie Google (*Google Translate*) und Microsoft (*Bing Translate* bzw. mit Rückgriff darauf auch der *Skype Translator*) durch ihr Datenaufkommen auch maschinelle Übersetzungstools für das Katalanische in Kombination mit anderen Sprachen.

Die Generalitat de Catalunya führt auf ihrer Internetseite⁵ weitere im Netz frei verfügbare MÜ-Systeme auf, die sich in den meisten Fällen jedoch auf das Katalanische (oder eine seiner Varietäten) mit dem Spanischen und Englischen beschränken. Aus der Liste hervorzuheben ist das Projekt *Apertium*, das als erstes System über einen offenen Quellcode verfügt und bereits seit Anfang der 2000er entwickelt wird. Apertium gilt zudem als Wegbereiter für die MÜ im Sprachenpaar Spanisch-Katalanisch. (Moreno u. a. 2012: 24) Außerdem bietet es auch weiteren Minderheitensprachen eine Plattform: Apertium unterstützt weiterhin die Übersetzung in das und aus dem Aragonesischen, Asturianischen, Frankoprovenzalischen, Galicischen und Okzitanischen. Ferner kann in das Sardische übersetzt werden lassen⁶.

Ljubešić & Toral (2014) haben das – nach ihrem Wissen und zum Zeitpunkt der Veröffentlichung ihres Artikels (Kompilierung des Korpus 2013/Publikation des Artikels 2014) – größte katalanischsprachige Korpus (CaWaC) aus im Netz verfügbaren Texten kompiliert. Die Grundlage hierfür waren Webseiten mit der Top-Level-Domain *.cat*. Um ein möglichst robustes Korpus zu erstellen, wurden

⁵S. hierzu https://llengua.gencat.cat/ca/eines_i_serveis/traduccio-automatica-en-linia/, letzter Zugriff am 31. Januar 2022.

⁶Eine komplette Übersicht über die unterstützten Sprachen findet sich unter https://wiki.apertium.org/wiki/Main_Page, letzter Aufruf am 31. Januar 2022.

die Sprachdaten mehrfach dedupliziert und bereinigt. Weiterhin wurde das Korpus als Trainingsset für eine SMÜ im Sprachenpaar Spanisch-Katalanisch verwendet. Als Vergleichsgröße wurde einerseits ein in-domain Korpos für SMÜ und andererseits ein Korpus aus Einträgen der katalanischen Wikipedia verwendet. Die Evaluation des Korpus geschieht einerseits auf Grundlage des *language modeling* (LM) und andererseits auf Grundlage der SMÜ. Für die Anforderungen des LM war eine umfangreiche Deduplizierung der Daten notwendig, sodass nur noch einzigartige Sätze übrig bleiben. Als Testsubsets dienen Sammlungen an Nachrichten, offiziellen Publikationen der EU, Sätzen für Sprachenlerner und Literatur (Spanische Bestseller der Gegenwartsliteratur). Die Ausgangsversion der SMÜ wird auf Grundlage eines Parallelkorpus⁷ aus der Domäne Nachrichten in einem Zehnjahreszeitraum der Zeitung „El Periódico de Catalunya“ (ES-CAT) erstellt.

2.6.2 Das Katalanische im internationalen Kontext

2005 entschied der Rat der Europäischen Union, dass auch die weiteren kooffiziellen Sprachen Spaniens zur Kommunikation zwischen Bürger-innen und Institution genutzt werden dürfen (Moreno u. a. 2012: 11). Ebenso werden bestimmte Dokumente und Veröffentlichungen der EU ins Katalanische übersetzt. Diese Entscheidung wurde in mehreren bilateralen Verträgen zwischen dem spanischen Staat und den europäischen Institutionen festgelegt und gilt nicht uneingeschränkt, sondern hauptsächlich für die Kommunikation mit der Vertretung der EU in Barcelona. So wird das Katalanische nicht als offizielle Amtssprache der EU gezählt.⁷ Dieses Verhältnis wird auch mit Blick auf die europäischen Sprachtechnologien ganz offenkundig: Der im Rahmen der EU-Ratspräsidentschaft 2020 eingeführte und von Tilde, DeepL, eTranslation und dem DFKI entwickelte *EU Council Presidency Translator*⁸ umfasst selbstverständlich alle EU-Amtssprachen, jedoch weder Regional- noch Minderheitensprachen. Mit Rückbezug zum Deutschen als Übersetzungsrichtung finden sich ebenfalls kaum Ressourcen, die Regional- und Minderheitensprachen – und konkret: das Katalanische – unterstützen. Neben dem geringeren Umfang an vorhandenen authentischen Sprachdaten im Vergleich zu den „großen“ Sprachen wie Englisch, Spanisch und Französisch in Europa ist dies sicherlich der Tatsache geschuldet, dass Sprachdienstleistungen, die auf diese vermeintlich kleineren Sprachen abzielen, sensible Kundendaten beinhalten oder ohnehin aus dem Privatkundensegment stammen und somit nicht allgemein zugänglich sind.

⁷Vgl. hierzu https://ec.europa.eu/spain/barcelona/about-us/cat_UE2_es, letzter Zugriff am 31. Januar 2022.

⁸Siehe <https://presidencymt.eu/>, letzter Aufruf am 31. Januar 2022.

2.7 Zusammenfassung

In diesem Kapitel wurden die grundlegenden Merkmale der CvK dargestellt und mit einem Textbegriff im Umfeld der Chatkommunikation sowie der Wissenspräsentation in diesem Umfeld in Verbindung gesetzt. Die Stellung des Katalanischen im digitalen Raum als Studiensprache wurde ebenso betrachtet. Besonders hervorzuheben ist die Kurzlebigkeit der ausgetauschten Informationen, die von der quasi-synchronen Verfügbarkeit der am Chat beteiligten Personen abhängt. Zugleich führt die geringe Halbwertszeit von Informationen im Chat auch dazu, dass Nachrichten auch in mehrere Blöcke zerteilt werden, um dem Gegenüber zu signalisieren, dass ein Beitrag noch nicht abgeschlossen ist. Hieraus ergibt sich das Bild des linearen und zugleich vorläufigen, da fragmentarischen, Texts, das in diesem Kapitel abgezeichnet wird. Das wiederum beeinflusst die Art, wie Wissen und Informationen in eine solche Kommunikationssituation integriert werden.

3 Technologischer Aspekt

Zur theoretischen Ausarbeitung gehört die Darstellung der technologischen Aspekte. Im Folgenden soll daher ein Überblick über den Stand der Sprach- und Translationstechnologien gegeben werden. In einem ersten Teil werden die einzelnen Technologien vor dem historischen Hintergrund (folgender Abschnitt 3.1) kurz vorgestellt und ausdifferenziert. Konkret geht es um die Unterscheidung zwischen maschineller Übersetzung und Verdolmetschung (Abschnitte 3.4.3, S. 57 und 3.3, S. 52). Ein zweiter Teil (Abschnitt 3.5, S. 60) geht dann auf die Evaluationsmöglichkeiten ein, die zur Bewertung der Qualität angewendet werden. Damit verbunden ist auch das Postediting sowie die Usability (Abschnitte 3.6, S. 67 und 3.7, S. 68).

3.1 Geschichtlicher Überblick der Technologie

Die Entwicklung des Computers in den 1930er und 1940er Jahren wird gemeinhin als Beginn der Forschung zu Sprachtechnologien betrachtet. Was die Betrachtung vom Standpunkt der mechanischen bzw. maschinellen Verarbeitung von natürlicher Sprache angeht, so kann das Aufkommen des Computers als Wegpunkt entlang einer Geschichte, die bereits früher begann, verstanden werden. Sowohl Stein (2013: 5) als auch Hutchins (1995: 431, 434) weisen zu Beginn ihrer Beiträge darauf hin, dass die aktuelle Entwicklung auf diesem Gebiet ohne die früheren Überlegungen nicht zu verstehen ist. Da diese Arbeit jedoch einen hohen technologischen Anteil besitzt, soll die Entwicklung des Computers als definitorischer Ausgangspunkt genügen. Zeitlich vorausgehende Ansätze seien an dieser Stelle demnach nur kurz erwähnt.

3.1.1 Bis zum 2. Weltkrieg

Stein (2013: 5) führt den katalanischen Philosophen Ramon Llull (13. Jahrhundert) und den deutschen Philosophen Gottfried Wilhelm Leibniz (17. Jahrhundert) als Vordenker an, deren Beiträge zur maschinellen Übersetzung sich vor allem auf dem Gebiet der Universalssprache verorten lassen. Weiter ist der Alchemist Johann Joachim Becher (17. Jahrhundert) aufgeführt, dessen Überlegungen sich in

3 Technologischer Aspekt

erster Linie auf die Konzeption eines mechanischen Wörterbuches stützten, das in der Lage war, lateinische Wörter zu kodieren (Stein 2013: 5).

Erste praktische Umsetzungen im Bereich der maschinellen Sprachverarbeitung wurden durch den Franzosen George Artsrouni und den Russen Petr Petrovich Smirnov-Troyanskii allerdings erst in den 1930er Jahren unabhängig voneinander ausgearbeitet. Artsrouni konzipierte ein Art Lochkartensystem, das zur Eingabe die passende Übersetzung, sprich: die passende Perforation, auf einem mit Ausgangsbegriff und Übersetzung beschrifteten Papierstreifen suchte. Smirnov-Troyanskis Vorhaben umfasste eine dreiphasige Übersetzung. Zunächst sollte ein Mensch, der nur die Ausgangssprache beherrschte, den Ausgangstext in eine von der Maschine verarbeitbare Form bringen, die lediglich eine von Troyanskii entwickelte Symbolsprache und die Grundformen der Wörter im Text umfasste. In der zweiten Phase wandelte die Maschine die in der Symbolsprache kodierten Informationen in die Zielsprache um. Ein Posteditor, der nur die Zielsprache beherrschte, wandelte diese Kodierung im dritten Schritt wieder in eine korrekte sprachliche Form um (Hutchins 1995: 434).

Während des 2. Weltkriegs verlagerte sich das Forschungsinteresse auf diesem Gebiet weg von der Übertragung von Informationen von einer in die andere Sprache hin zur Kryptologie. Dabei hat das britische Vorhaben um den Mathematiker Alan Turing bis heute Symbolcharakter, der die Funktionsweise der deutschen Dechiffriermaschine *ENIGMA* zu verstehen versuchte (Stein 2013: 6).

3.1.2 Ab 1946

Das Jahr 1946 stellt ein Schlüsseldatum dar. Mit Ende des Weltkriegs richtete sich die Aufmerksamkeit dieses Forschungsbereiches wieder auf die maschinelle Übersetbarkeit von natürlicher Sprache. Als Ursprung gilt der Schriftverkehr zwischen Warren Weaver, Mitglied der Rockefeller Foundation, und A.D. Booth, einem Physiker. Weaver hatte einen Bericht Booths gelesen, in dem dieser über die Möglichkeit sprach, Computer für Übersetzungszwecken zu konzipieren (Locke 1956: 29). Booth selbst beschrieb dies wie folgt:

It arose because, in 1946, various new uses for automatic digital calculating machines were being considered and these ranged from the more obvious applications to problems of mathematics and physics, to philosophical problems such as the mechanization of human thought processes, the playing of games and the translation of language. (Booth 1958: 88)

Hieraus entwickelte sich ein reges Forschungsinteresse, das sich zunächst auf Wort-für-Wort-Übersetzungen auf Grundlage eines bilingualen Wörterbuches

konzentrierte und Syntax und Morphologie außer Acht ließ. Ziel der Bemühungen sollte zunächst sein, Wissenschaftlern (sprachlichen) Zugang zu Texten ihrer Kollegen aus anderen Ländern zu verschaffen (Delavenay 1960: 28). Um das Forschungsinteresse weiter zu schüren und gleichzeitig in Kontakt mit Kritikern zu treten, veröffentlichte Weaver 1947 seine Überlegungen in Form des bis heute berühmten *Weaver Memorandums*. Die darin enthaltenen Ausführungen veranlassten Forscher überall in den USA, sich intensiv mit der MÜ auseinanderzusetzen. So zählt Locke (1956: 30) folgende zentrale Namen aus jener Zeit auf:

At the University of Washington Erwin Reifler looked into the basic semantic equivalents of languages. At the University of California at Los Angeles Victor A. Oswald and Stuart L. Fletcher, Jr., analyzed German syntax and in 1951 published the first paper devoted to machine translation: "Proposals for the Mechanical Resolution of German Syntax Patterns." At M.I.T. Yehoshua Bar-Hillel began an attempt to identify the universal grammar elements in various languages and also gave some thought to translating idioms.

Besonders der Name Bar-Hillel ist einerseits eng mit der Erweiterung des Forschungsstands und andererseits auch mit der kritischen Auseinandersetzung der Möglichkeiten verbunden. Er war 1952 einer der Organisatoren der ersten nationalen Konferenz zu „mechanical translation“ (Delavenay 1960: 29) überhaupt. Finanziert wurde die Konferenz durch die Rockefeller Foundation. Im Jahre 1956 war er dann auch an der Ausrichtung der ersten internationalen Konferenz auf diesem Gebiet beteiligt. Etwa zeitgleich wurde das bis heute wegweisende *Georgetown-Experiment* (1954) durchgeführt, dessen Ausgang als Beleg genommen wurde, dass das Problem der maschinellen Übersetzung beinahe gelöst sei. Im Rahmen dieses Experiments wurden russische Sätze mit einem regelbasierten System ins Englische übertragen (Koehn 2009: 15). Überhaupt erfährt die Auseinandersetzung mit der MÜ in der Zeit nach dem 2. Weltkrieg bis in die 1960er Jahre einen enormen Aufschwung. Begünstigt durch das rege Forschungsinteresse entstehen mehrere beachtete Publikationen und regelmäßig erscheinende Fachmagazine, wie etwa die Zeitschrift *Mechanical Translation* von William Locke und Victor Yngve oder das Buch *Machine Translation of Languages* von Locke und Booth (Delavenay 1960: 30).

3.1.3 Der ALPAC-Bericht

Zu Beginn der 1960er Jahre musste man dann jedoch feststellen, dass die hohen Ziele und die Hoffnungen, die die MÜ-Forscher hatten, mit dem damaligen

3 Technologischer Aspekt

Stand der Technologie kaum zu erreichen seien. 1960 hatte Bar-Hillel in einem Bericht die technischen und linguistischen Probleme bei der MÜ angemerkt. Besonders den damaligen Anspruch, eine „fully automatic high quality translation (FAHQT)“ (Hutchins 1995: 438) erzielen zu wollen, wies Bar-Hillel als für den damaligen Stand utopisch zurück. 1961 veröffentlichte Mortimer Taube mit dem Buch *Computers and Common Sense* eine weitere kritische Auseinandersetzung mit der maschinellen Verarbeitung linguistischer Strukturen (Henisz-Dostert u. a. 1979: 9). Den härtesten Rückschlag erfuhr das Forschungsgebiet jedoch 1966 mit der Veröffentlichung des bis heute berühmten ALPAC-Berichts. Das *Automatic Language Processing Advisory Committee* war bereits 1964 von der *National Academy of Science* zur Evaluierung des gegenwärtigen Forschungsstandes gegründet worden. Kernaussagen des Berichtes waren zum einen die Feststellung, dass es mehr Übersetzerinnen als zu übersetzende Texte gäbe. Zum anderen gab das Expertenkommitee die Einschätzung ab, dass die als Maxime ausgegebene FAH-QT grundsätzlich möglich, mit den zur damaligen Zeit zur Verfügung stehenden technologischen Ressourcen jedoch nicht realisierbar sei. Weiterhin wog der ALPAC-Bericht das Kosten-Nutzen-Verhältnis der MÜ ab. Hierbei wurde festgestellt, dass die MÜ ohne menschliche Postedition zu kostspielig und langwierig sei. Zugleich enthielt der Bericht jedoch auch die Empfehlung, der Linguistik weitere Fördergelder zuzugestehen, vor allem mit der Zielsetzung, die Geschwindigkeit und Qualität zu verbessern und die Kosten zu senken (Automatic Language Processing Advisory Committee 1966: 1–34).

Rückblickend hält Hutchins dem entgegen, dass der ALPAC-Bericht viel zu unausgewogen und von den Absichten der beteiligten Institutionen gelenkt war.

The ALPAC report was widely condemned as narrow, biased and shortsighted. It is true that it failed to recognize, for example, that revision of manually produced translations is essential for high quality, and it was unfair to criticize MT for needing to post-edit output. (Hutchins 1995: 439)

Die Auswirkungen waren verheerend, sodass die Forschung in weiten Teilen der USA und Europas zum Erliegen kam oder zumindest auf die rein praktische Anwendung umgestellt wurde. Im Vordergrund stand nun die Entwicklung von digitalen oder maschinellen Hilfsmitteln, die Übersetzerinnen bei ihrer Tätigkeit unterstützen sollten. Für das folgende Jahrzehnt sollte die MÜ in den USA als kaum zukunftsträchtiges Feld stigmatisiert bleiben (Stein 2013: 6). In Kanada und Europa hingegen blieb das Forschungsinteresse – besonders in der freien Wirtschaft – bestehen (Hutchins 1995: 439). Hierbei wurden vor allem extrem spezialisierte Systeme entwickelt, die ein genau abgestecktes terminologisches

3.1 Geschichtlicher Überblick der Technologie

Feld bearbeiten konnten, was Krenz u. a. (2008: 33) als „Subsprachen-Systeme“ benennen. Als ein prominentes Beispiel wird häufig das MÉTÉO-Projekt der Universität von Montreal und des kanadischen Umweltministeriums genannt, das Wettervorhersagen aus dem Englischen ins Französische übersetzten sollte (Bowler & Ciro 2019: 38).

3.1.4 Technologische Weiterentwicklung

Erst Mitte der 1970er Jahre sollte die MÜ jedoch wieder intensivere Beachtung finden. Mit dem Aufkommen leistungsstärkerer Computer kam es auch zur Gründung neuer Forschungsgruppen auf dem Gebiet der Computerlinguistik und zur Entwicklung kommerzieller Systeme wie etwa *Systran*. Systran ist bis heute eine der bekanntesten Marken weltweit auf diesem Gebiet und wurde in den 1970er Jahren für das Sprachenpaar Englisch-Russisch von der US-Luftwaffe genutzt und wenig später auch von der Europäischen Kommission eingesetzt (Hutchins 1995: 139–142). Mit *Logos* und *METAL* kamen in den 1980er Jahren zudem zwei weitere kommerzielle Systeme auf den Markt (Koehn 2020: 35). Dies geschah zunächst in Form der regelbasierten maschinellen Übersetzung (eng. *rule-based machine translation*, RBMT, dt. *RBMÜ*, s. S. 43), die insofern aufwändig zu betreiben war, als ein möglichst umfassender und präziser Fundus an grammatischen und morphosyntaktischen Regeln ausgearbeitet und ein umfangreiches zweisprachiges Wörterbuch bereitgestellt werden musste (Stein 2013: 7). Weiterhin erfuhr das Sprachangebot eine Ausdehnung. Während anfangs Französisch, Deutsch, Englisch und Russisch im Fokus standen, kamen im Laufe der Zeit auch Projekte zur Übersetzung des Chinesischen ins Englische (*CULT*) oder gar zu mehrsprachigen Übersetzungen (*TITUS*) hinzu (Krenz u. a. 2008: 34). Als gemeinsames europäisches Projekt wurde zwischen 1970 und 1994 *EUROTRA* betrieben, das zwar das eigentliche Ziel „des Aufbaus eines modernen Übersetzungssystems“ (Burchardt u. a. 2012: 30) verfehlte, dennoch nachhaltig zur Entwicklung der Sprachtechnologieforschung in Europa beitrug.

Mit der weltweit steigenden Menge an zu übersetzen Texten wuchs zu Beginn der 1980er Jahre auch der Bedarf an maschinellen Übersetzungssystemen. Eine Forschungsgruppe von IBM unter der Leitung von Peter F. Brown rückte die statistische maschinelle Übersetzung in den Fokus des Interesses (Stein 2009: 7).

Binnen kürzester Zeit konzentrierte sich die Mehrheit der Forschungen auf die statistischen Ansätze, mit denen man Erfolge erzielen konnte, die mit denen der etablierten, regelbasierten Systeme vergleichbar waren – nur dass

3 Technologischer Aspekt

man zu deren Erstellung keine 10 Jahre Entwicklungszeit und kein Fachwissen von Linguisten benötigte. Ein paar Tage Zeit und große bilinguale Korpora (Bitexte) genügten für einen Prototypen. (Stein 2009: 7)

Neben den großen Computer-Unternehmen investierten einige japanische Unternehmen intensiv in den Sektor, so etwa Mitsubishi mit dem System *MELTRAN*, Sanyo und Toshiba mit *AS-TRANSAC* oder Fujitsu mit *ATLAS*. Alle Systeme beschränkten sich jedoch auf die Analyse und Übertragung von morphologischen und syntaktischen Strukturen und bedurften der Nachbearbeitung durch den Menschen (Hutchins 1995: 442).

Hutchins (1995: 443) zählt weiterhin die Untersuchung und Theoriebildung zur Interlingua-Übersetzung (s. S. 44), die Weiterentwicklung der bestehenden Systeme und die Arbeit mit künstlicher Intelligenz als wichtigste Forschungsbereiche der Zeit auf dem Gebiet der MÜ. Dies geschah meist auf Grundlage linguistischer Regeln, weshalb RBMÜ zu jenem Zeitpunkt der verbreitetste Ansatz der MÜ war. Zu Beginn der 1990er Jahre kam zudem das Forschungsinteresse am maschinengestützten Dolmetschen auf. Das populärste Projekt dieser Zeit stammt mit *Verbmobil* aus Deutschland. Das Ziel des Vorhabens war die Schaffung eines automatischen Echtzeit-Telefondolmetschsystems, das den jeweiligen Kommunikationsparteien zwischengeschaltet war (Wahlster u. a. 2000). Damit einher ging die Auseinandersetzung mit digitaler Spracherkennung und -reproduktion (Krenz u. a. 2008: 35).

3.1.5 Ausweitung des Produktportfolios

In den frühen 1990er Jahren wandte sich die Forschung und die Softwareentwicklung den *Computer-Aided Translation Tools* (CAT-Tools) zu. Immer kostengünstigere Heimrechner und die Ausbreitung des Internets rechtfertigten diese Tendenz nicht als Ersatz des menschlichen Übersetzers, sondern vielmehr als Ergänzung und Unterstützung der Tätigkeit, da sie neben der eigentlichen Übersetzungsoberfläche zusätzliche Hilfsmittel wie Terminologiedatenbank, Translation Memory und elektronische Wörterbücher boten (und bieten, Bowker & Ciro 2019: 38). Ebenfalls in dieser Zeit stellten Unternehmen erste MÜ-Systeme bereit, die online erreichbar waren, so etwa *Minitel* von Systran (1988), *CompuServe* (1992) und als erstes für alle Internetnutzer·innen kostenfrei verfügbares Angebot *BabelFish* wiederum ebenfalls von Systran und AltaVista (1997).

3.1.6 Das neue Jahrtausend

Die Zeit seit der Jahrtausendwende ist vor allem durch eine rasante Weiterentwicklung der Effizienz und einer Ausweitung der möglichen Sprachenkombinationen geprägt. Durch die Kombination von regelbasierter und statistischer maschineller Übersetzung verbesserte sich seit Anfang der 2000er die Ausgabequalität merklich und wurde zunehmend für professionelle Sprachdienstleister attraktiv.

Das bis heute am weitesten verbreitetste, (kosten-)frei zugängliche MÜ-System ist *Google Translate*. 2006 mit lediglich zwei Sprachen ins Leben gerufen, bietet der Dienst mittlerweile die maschinelle Übersetzung in 103 verschiedene Sprachen. Seit 2016 werden einzelne Sprachen sogar auf Basis der neuronalen maschinellen Übersetzung verarbeitet. Weiterhin können Internetseiten oder – im Rahmen von Google Lens – auch Straßenschilder u. ä. übersetzt werden lassen.

3.2 Maschinelle Übersetzung

Maschinelle Übersetzungssysteme können grundlegend in zwei große Kategorien eingeteilt werden: *regelbasierte*, die auf Grundlage grammatischer Regelsätze und mehrsprachiger Wörterbücher arbeiten, und *datenbasierte* Systeme, die statistisch-algorithmisch vorgehen. Die Kombination beider zu sog. *hybriden* Systemen ist ein Ansatz, der seit Mitte der 2000er verfolgt wird, um die jeweiligen Stärken der beiden Kategorien zu verbinden. Seit Mitte der 2010er kommen vermehrt *neuronale* Systeme zur Anwendung, die in großen Mengen datenbasiert arbeiten und – grob umrissen – die Funktionsweise des menschlichen Gehirns imitieren. Als theoretisches Konzept wird zudem häufig auf sog. *Interlingua-Systeme* verwiesen. Der Ansatz gründet auf dem syntaktischen Formalismus und dem Versuch, die Bedeutung einer Aussage, eines Satzes oder eines ganzen Textes in einer Tiefenstruktur zu formalisieren, die dann in eine andere Sprache übertragen werden kann (Koehn 2009: 16 f.). Mit Hutchins & Somers (1997), Wilks (2009), Koehn (2009, 2020) sind über die letzten Jahrzehnte umfassende Werke entstanden, die die Entwicklung aus internationaler Perspektive betrachten. Für eine deutschsprachige Sicht auf die MÜ bestehen mit Laisiepen u. a. (1996), Ramlow (2009), Stein (2013), Porsiel (2017) Werke bereit, die sich einer generellen Übersicht verschrieben haben.

3.2.1 Regelbasierte Maschinelle Übersetzung

Der simpelste Ansatz zur maschinellen Übersetzung basiert auf vom Menschen zusammengetragenen Grammatik- und Morphologieregeln. Auch heute noch

3 Technologischer Aspekt

kommt eine hybride Form aus regelbasiertem und statistischem Vorgehen in Übersetzungssystemen vor. Der regelbasierten maschinellen Übersetzung liegt dabei neben dem Repertoire an Regeln auch ein bilinguales Wörterbuch zugrunde, aus dessen Fundus aus der Ausgangs- in die Zielsprache übersetzt wird. Ein großes Problem bei diesem Ansatz ist die geringe Flexibilität der Übersetzung und die je nach Sprachenpaar stark variierende Qualität. Die Ergebnisse sind häufig unbrauchbar, was auch Carstensen u. a. (2010: 647) kritisieren: „Die Abhängigkeiten der linguistischen Regeln und die umfassende Behandlung von sprachlichen Ausnahmen führen jedoch immer wieder zu Problemen bei der Robustheit und Wartbarkeit.“

Stein (2009: 8) charakterisiert RBMÜ zunächst als ein dreistufiges System aus „Analyse, Transfer und Synthese (bzw. Generierung)“. Diese drei Stufen werden weiterhin nach ihrem Grad an Komplexität in *direkte Übersetzung*, *Transferübersetzung* und *Interlingua-Übersetzung* unterschieden (Stein 2009: 8).

- Die direkte Übersetzung ist als Wort-für-Wort-Übersetzung zu verstehen. Im Analyseschritt werden die einzelnen Wörter im Ausgangstext identifiziert und mit den Einträgen im bilingualen Wörterbuch abgeglichen. Die Transfer- und Syntheseprozesse gehen gemeinsam einher. Zunächst erfolgt die Ersetzung der einzelnen Wörter in der Zielsprache und dann ggf. die Anpassung der Wortreihenfolge. Als Vorteile werden die Zeit- und Kostensteneffizient angeführt (Carstensen u. a. 2010: 645 f.).
- Die Transfer-Übersetzung kann als eine Erweiterung der direkten Übersetzung angesehen werden. Der Analyse- und Transferprozess umfasst in diesem Fall mehr als den Abgleich mit dem Wörterbuch: Hinzu kommt eine semantische und morpho-syntaktische Analyse (*Parsing*). Die so erhaltenen Informationen ermöglichen die Definition von genaueren Regeln. (Stein 2009: 8, Carstensen u. a. 2010: 646)
- Der Grundgedanke hinter der Interlingua-Übersetzung ist die Annahme, jede Sprache habe die gleiche tiefenstrukturelle Kodierung und nur das oberflächliche Erscheinungsbild, sprich: Lexik und Grammatik, seien individuell. „Diese abstrakte universalsprachliche Repräsentation würde dann das Ziel und die Quelle sämtlicher Übersetzungssysteme sein.“ (Stein 2009: 8) Der Ausgangstext würde demnach zunächst auf seine semantischen und syntaktischen Funktionen hin analysiert und dann eine gleichwertige Entsprechung im Zieltext erzeugt.

3.2.2 Statistische Maschinelle Übersetzung

Im Gegensatz zu dem manuell verfassten Repertoire an Grammatikregeln der RBMÜ ist die Grundvoraussetzung für die statistische maschinelle Übersetzung (eng. *SMT*, dt. *SMÜ*) ein umfangreiches, gepflegtes Parallelkorpus für das zu übersetzende Sprachenpaar, auf dessen Grundlage das System trainiert werden kann. Ist die Größe dieses Korpus ausreichend, lässt sich das SMÜ-System sogar automatisch erstellen (Carstensen u. a. 2010: 647). Das Parallelkorpus erfordert eine umfassende Vorarbeit seitens der menschlichen Übersetzer-innen, da die Texte in beiden Sprachen zunächst auf Satzebene aligniert, d. h.: ihren jeweiligen Entsprechungen in Ausgangs- und Zielsprache zugeordnet werden müssen. Die Alignierung wird in einem weiteren Arbeitsschritt auch auf Wortebene durchgeführt, um die Möglichkeiten zur Übersetzung eines Wortes flexibler zu gestalten. Durch diesen Prozess entsteht ein sog. Trainingskorpus und ein zweisprachiges Wörterbuch, das jedoch bislang nur die flektierten Formen und nicht die Lemmata enthält. In diesem Trainingskorpus sind zudem die Wahrscheinlichkeiten aufgeführt, mit denen ein Wort in der Ausgangssprache einem Wort in der Zielsprache entspricht. Diesen sog. probabilistischen Ansatz konkretisiert Koehn (2009: 6) mit den Worten: „Probabilities are used when we have to deal with events with uncertain outcomes, such as a foreign word that may translate into one of many possible English words.“ So kann in einem folgenden Schritt ein Wortfolgenlexikon erstellt werden, das auf Grundlage der o. g. Wahrscheinlichkeiten das Auftreten einzelner Wortfolgen dokumentiert.

Ausgangspunkt der SMÜ ist also ein großes Korpus humanübersetzter Texte, wenn möglich aus dem intendierten Anwendungsbereich. Wenn das SMÜ-System Dokumente einer Maschinenbau-Firma übersetzen soll, sollte auch das Eingabekorpus aus diesem Bereich stammen. (Carstensen u. a. 2010: 648)

Die Qualität der Ausgabe bei SMÜ ist also in hohem Maße einerseits von der Menge der Daten im bilingualen Trainingskorpus und andererseits der Häufigkeit der Wortfolge im monolingualen Sprachenmodell abhängig.

Stein (2009: 11) weist hierbei noch auf eine gängige Differenzierung innerhalb der SMT hin: „Man unterscheidet allgemein zwischen wortbasierter und phrasenbasierter SMÜ.“

- Wie die Bezeichnung der wortbasierten maschinellen Übersetzung vermuten lässt, wird hierbei zunächst eine rein lexikalische Übersetzung vorgenommen. Im Gegensatz zu wortbasierten Ansätzen bei der RBMÜ wird hier

3 Technologischer Aspekt

jedoch nicht immer nur der erstbeste Begriff aus dem bilingualen Wörterbuch verwendet. Auf Grundlage statistischer Daten kann für einen Begriff im Ausgangstext abgewogen werden, welcher Begriff in der Zielsprache die wahrscheinlichste Entsprechung ist (Koehn 2009: 82). Stein (2009: 11) hält dem jedoch entgegen, dass so zwar die Möglichkeit gegeben sei, einen Begriff mit einer Mehrworteinheit zu übersetzen, umgekehrt jedoch eine Mehrworteinheit Wort für Wort übersetzt und ggf. nicht nach einem einzelnen Begriff gesucht werde. „Jedem Wort in der Quellsprache muss also mindestens ein Wort in der Zielsprache entsprechen.“ (Stein 2009: 11)

- Koehn (2009: 127) charakterisiert den phrasenbasierten Ansatz als die gegenwärtig beste Möglichkeit zur maschinellen Übersetzung mit den Worten: „The currently best performing statistical machine translation systems are based on phrase-based models: models that translate small word sequences at a time.“ Ein Ausgangstext wird also nicht Wort für Wort verarbeitet, sondern in größere Einheiten aufgeteilt, die jedoch kleiner sind als ein Satz. Dabei handelt es sich um die *Phrasen*. Als Begründung führt Koehn (2009: 127) an, dass es zwischen Ausgangs- und Zielsprache häufig Wörter gibt, die sich nicht mit einem Wort allein in die Zielsprache übersetzen lassen. Solche Mehrworteinheiten, zuweilen auch als *n-Gramme* bezeichnet, kann die phrasenbasierte maschinelle Übersetzung erfassen. Einen weiteren Vorteil sieht Stein (2009: 12) in der Möglichkeit, „bestimmte Disambiguierungsentscheidungen zu treffen.“

3.2.3 Neuronale Maschinelle Übersetzung

Die Anwendung von neuronalen Netzwerken zur maschinellen Übersetzungen ist der jüngste und gegenwärtig am intensivsten untersuchte Bereich in der Sprachtechnologieforschung. So beschrieb ein Beitrag in der Ausgabe 01/17 des MDÜ diesen Ansatz als nächsten „Hype bei der maschinellen Übersetzung“ (Krüger 2017: 38), nicht zuletzt unter Verweis auf die Ankündigung von Google, die 2016 die neuronale maschinelle Übersetzung (eng. *NMT*, dt. *NMÜ*) in den Übersetzungsdienst des Konzerns eingebunden zu haben.

Die Ankündigung wurde begleitet von einem wissenschaftlichen Paper, das den ambitioniert klingenden Titel „Bridging the Gap between Human and Machine Translation“ trug. Nicht weniger ambitioniert waren die Behauptungen, die Google in diesem Paper aufstellte. So könnten Evaluatoren stellenweise kaum noch Unterschiede zwischen menschlichen Übersetzern und der Google-NMÜ feststellen, [...]. (Krüger 2017: 38)

Bevor jedoch auf die Funktionsweise der NMÜ eingegangen werden kann, sollte eine Erklärung zu künstlichen neuronalen Netzwerken im Allgemeinen erfolgen. Um ihre Funktion (KNN, bzw. auf englisch *artificial neural network, ANN*) zu verstehen, wird häufig der Vergleich zu dem neuronalen Netzwerk im Nervensystem eines Lebewesens gezogen. Dort gibt es Zellen, die Informationen in Form von elektrischen und chemischen Impulsen untereinander weiterleiten können, die sog. *Neuronen*. Sie sind über Verbindungen, den *Synapsen*, miteinander verbunden. Je häufiger nun eine bestimmte Neuronenverbindung im Gehirn durch einen Impuls aktiviert wird, um so stärker und leichter wird die Impulsverarbeitung im Folgenden (Kriesel 2005: 19 f.). Das neuronale Netzwerk bildet somit „relativ stabile Aktivierungsmuster heraus“ (Krüger 2017: 38), die als Lernprozess des Lebewesens verstanden werden. Der Fluchtreflex eines Tieres, das Sprachen(er)lernen und viele weitere musterhafte Tätigkeiten aktivieren so die immer gleichen Regionen im Nervensystem. So ist es Lebewesen wiederum möglich, unbekannte Sachverhalte aufzunehmen und zu verarbeiten, die den bereits bekannten Mustern ähnlich sind.

3.2.3.1 Künstliche neuronale Netze

Das Konzept der zerebralen Vernetzung kann künstlich nachgebildet werden. Abbildung 3.1 zeigt ein solches künstliches neuronales Netzwerk (eng. *ANN*, dt. *KNN*). Auch hier gibt es Neuronen, die Kriesel als „simple Recheneinheiten“ (Kriesel 2005: 35) beschreibt, die über gerichtete, gewichtete Verbindungen verknüpft sind (Kriesel 2005: 35). Diese Verbindungen entsprechen den *Synapsen* und sind in Abbildung 3.1 als dünne Pfeile dargestellt. Die Neuronen eines KNN sind in Schichten (eng. *layers*) angeordnet. Erfährt dieses Netzwerk nun also eine Eingabe (z. B. externe Informationen), nimmt die Eingabeschicht (*input layer*) diese auf und verarbeitet sie zu Signalen, die über die unterschiedlich stark oder schwach gewichteten Verbindungen an die verdeckte Schicht (*hidden layer*) weitergeleitet werden. „Je nach Gewicht [...] der Verbindungen, über die bestimmte Signale übertragen werden, empfängt ein „aufnehmendes“ Neuron diese Signale in abgeschwächter oder verstärkter Form, wodurch sich der Aktivierungsstatus dieses Neurons verändert.“ (Krüger 2017: 39) Verändert sich der Aktivierungsstatus ausreichend, sodass ein gewisser Schwellenwert überschritten wird, löst dies die Aktivierung des Neurons in der verdeckten Schicht aus. Dieses Neuron „feuert“ (Kriesel 2005: 38) also ein Signal ebenfalls über die gewichteten Verbindungen in abgeschwächter oder verstärkter Form zur Ausgangsschicht (*output layer*). Wird dieser Prozess häufig genug wiederholt, bilden sich die eingangs erwähnten Aktivierungsmuster zwischen den künstlichen Neuronen und Verbindungen

3 Technologischer Aspekt

heraus. Das Netzwerk lernt also allmählich, „in großen Mengen unstrukturierter Daten komplexe Muster zu erkennen, diese zu klassifizieren und mit anderen Mustern in Verbindung zu setzen.“ (Krüger 2017: 39)

In ihrem Webinar zur neuronalen maschinellen Übersetzung vom 21. Juni 2017 wiesen Cattelan & Federico in diesem Zusammenhang jedoch explizit darauf hin, dass die Vorgänge, die sich innerhalb der verdeckten Schicht abspielen, bislang unerschlossen sind und dass aufgrund der Komplexität dieser Vorgänge auch davon ausgegangen werden kann, dass es in jüngerer Zukunft keine ausführliche Erklärung hierzu geben wird (Cattelan & Federico 2017). Außerdem sei darauf hingewiesen, dass Abbildung 3.1 lediglich ein vereinfachtes Netzwerk darstellt. Für eine realitätsnahe Abbildung müssen noch weitere (verdeckte) Schichten und Neuronen hinzugedacht werden, was für Nutzer*innen die Notwendigkeit einer enorm hohen Rechenleistung nach sich zieht (Krüger 2017: 39).

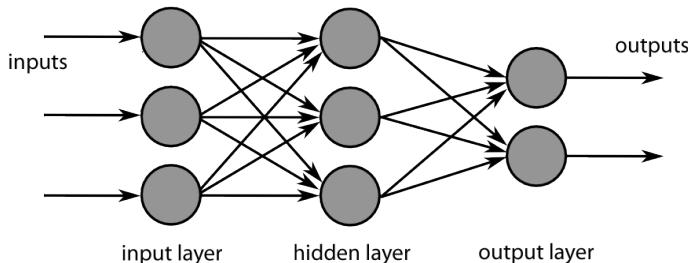


Abbildung 3.1: Einfaches künstliches neuronales Netzwerk (CC-BY-SA 3.0 H3llkn0wz)

3.2.3.2 Ein einfaches Beispiel

In besagtem Webinar haben Cattelan & Federico (2017) weiterhin die Verarbeitungsprozesse anschaulich am Beispiel einer automatischen Bilderkennung, wie sie auch Google anbietet, erklärt. Aufgabe des Beispielnetzwerkes sollte es sein, eingegebene Bilder einer von vier verschiedenen, stilisierten Tierabbildungen zuzuordnen. Zeigte das eingegebene Bild (*input*) etwa eine Katze, wurden Bildinformationen (Farbwerte, Farbverteilung, Auflösung, usw.) in Zahlenwerte umgewandelt und in das ANN eingespeist (*Signale an der Eingabeschicht*). Anhand dieser Werte konnte das neuronale Netzwerk Aussagen über die am wahrscheinlichsten passende stilisierte Tierabbildung der Eingabe treffen. Durch Eingabe von bislang nicht in dem Bilderkorpus des Beispielnetzwerks enthaltenen Tierfotos kann dann die Präzision trainiert werden (Cattelan & Federico 2017). Im Grunde basieren ANN also auch auf einem probabilistischen Ansatz.

Auf diesem einfachen Beispiel aufbauend lässt sich nun auch die NMÜ erklären. Wie auch bei SMÜ-Systemen ist zunächst ein qualitativ hochwertiges, umfangreiches, bilinguales Textkorpus notwendig. In der Trainingsphase dieses NMÜ-Systems werden nun – ausgehend von zu Beginn zufälligen Gewichtungen an den Verbindungen der Neuronen – die Ausgangstexte mit ihren jeweiligen, bereits bekannten Übersetzungen eingegeben. Das ANN ordnet dann anhand von Mustern, die es in den Ausgangstexten identifiziert, passende Zieltexte zu. Nun wird die tatsächliche Ist-Ausgabe mit der gewünschten Soll-Ausgabe abgeglichen, wobei der Unterschied zwischen Ist und Soll als Fehler angesehen wird (Krüger 2017: 40).

3.2.3.3 Backpropagation

An diesem Punkt gewinnt der sog. *Backpropagation*-Algorithmus an Bedeutung, da er eines der Elemente ist, die dem ANN die Lernfähigkeit überhaupt erst ermöglichen.

When neural networks are used to model a set of existing data so that predictions can be made on new data, the main challenge is to find the set of weight and bias values that generate the outputs that best match the existing data. The most common technique for estimating optimal neural network weights and biases is called back-propagation. (McCaffrey 2012)

Im Bemühen darum, die Differenz zwischen Ist und Soll zu minimieren (und im Optimalfall komplett aufzulösen), überprüft der Algorithmus also stetig die Ausgabe und passt – gemessen am Goldstandard des Trainingskorpus – dementsprechend die Gewichtung der einzelnen Verbindungen innerhalb des ANN an. Dieser Vorgang wird so oft wiederholt, bis das ANN die gewünschte Ausgabe liefert (Habra 2017).

Für die Nutzung im Umfeld von kleineren Sprachen¹ weist Krüger auf die vielversprechenden Forschungsansätze, eine sog. *Zero Shot Translation* zu schaffen. Im konkreten Fall wurde ein NMÜ-System von Google auf die Übersetzung von Texten aus dem Portugiesischen ins Englische und aus dem Englischen ins Spanische trainiert. Danach war dieses System ebenfalls in der Lage, aus dem Portugiesischen ins Spanische zu übersetzen, ohne dafür trainiert worden zu sein (Krüger 2017: 40).

¹Wobei in diesem Buch jedoch nicht weiter darauf eingegangen wird, was genau *kleine Sprache* bedeutet. Die Definition hängt auch immer vom Kontinent ab, auf dem man sich befindet.

3 Technologischer Aspekt

3.2.3.4 Vektoren

Diese Beschreibung geht allerdings noch nicht auf einen weiteren wichtigen Prozess ein. Um die Arbeitsschritte eines NMÜ-Systems zu verstehen, fehlt noch eine Erklärung, wie die Texteingabe in Informationen umgewandelt wird, die das Netzwerk verarbeiten kann, um daraus dann erneut einen Zieltext zu erzeugen. Ausgangspunkt hierfür ist zunächst die Aussage von Bahdanau u. a. (2017: 2):

This neural machine translation approach typically consists of two components, the first of which encodes a source sentence x and the second decodes to a target sentence y . For instance, two recurrent neural networks (RNN) were used [...] to encode a variable-length source sentence into a fixed-length vector and to decode the vector into a variable-length target sentence.

Der Vorgang des Dekodierens und Enkodierens bezieht sich auf die Prozesse, die an der Eingabe- und an der Ausgabeschicht stattfinden. Die Begriffe, auf denen hier der Fokus liegt, sind viel mehr *recurrent* und *vector*. Vektoren fasst Seidel (2005: 2) als eine Matrix auf, „die nur aus einer Zeile $\vec{z}_i = (a_{i1}, a_{i2}, \dots, a_{in})$ (Zeilenvektor) oder nur aus einer Spalte $\vec{s}_k = (a_{1k}, a_{2k}, \dots, a_{mk})$ (Spaltenvektor) besteht.“

Matrix definiert Seidel (2005: 1) als „eine Tabelle, die in Zeilen und Spalten unterteilt ist. Die Einträge in der Matrix lassen sich über die Koordinaten aus Zeilen und Spalten mit den Indizes m und n direkt ansprechen.“

Damit wird verständlich, was Kriesel (2005: 51) meint, wenn er den Eingabevektor (respektiv auch Ausgabevektor) definiert mit den Worten:

Ein Netz mit n vielen Eingabeneuronen benötigt n Eingaben x_1, x_2, \dots, x_n . Wir fassen diese als Eingabevektor $x = (x_1, x_2, \dots, x_n)$ auf. Die Eingabedimension bezeichnen wir also mit n . Daten werden in ein Neuronales Netz eingegeben, indem die Komponenten des Eingabevektors einfach bei den Eingabeneuronen als Netzeingabe verwendet werden.

Dies eröffnet dem NMÜ-System die Möglichkeit, „semantische und grammatische Relationen zwischen diesen Wörtern“ (Krüger 2017: 40) im Vektorraum anhand ihrer Nähe oder Distanz zu erfassen. Je näher die Wörter zueinander stehen, um so ähnlicher sind sie. Diese Relationen werden auch *word embeddings* genannt und werden in einer Tabelle abgelegt, auf die dann wiederum das NMÜ-System zurückgreift (Krüger 2017: 40).

3.2.3.5 Rekurrenz

Neben Vektoren war in dem Zitat von Bahdanau u. a. (S. 50) die Rede von rekurrenten neuronalen Netzwerken, (eng. *recurrent neural networks, RNN*). Durch die Rekurrenz erhält das Netzwerk die Fähigkeit, „die Outputs der Neuronen einer Schicht wieder als neue Inputs in das Netz“ (Krüger 2017: 41) einzuspeisen. So können Sprachdaten verarbeitet werden. Cattelan & Federico (2017) skizzieren diesen Prozess folgendermaßen: Signale werden einerseits von einer zur nächsten Schicht weitergeleitet, zugleich jedoch als erneuter Input an einen Zwischenspeicher der vorherigen Schicht zurückgeführt, usw. So ist es in Bezug auf Sprachdaten dem RNN möglich, den Kontext eines eingegebenen Textes zu erfassen und ausgehend von den einzelnen Einheiten eines Satzes darauf aufbauende Übersetzungen zu generieren.

3.2.4 Leistungsmerkmale der Systeme

Die hier aufgeführten maschinellen Übersetzungssysteme zeichnen sich durch charakteristische Stärken und Schwächen aus. Ganz allgemein weist Koehn (2009: 20) zunächst darauf hin, dass die maschinelle Übersetzung nicht perfekt sein muss, um einen Nutzen für den Anwender darzustellen. Mit dieser Bemerkung betritt er das Feld des *information retrieval* bzw. der Informationsextraktion. Ausgerichtet am Leitgedanken, generell Verständnis zu schaffen, kann der Nutzen der MÜ in drei Bereiche unterteilt werden: *Assimilation*, *Dissemination* und *Kommunikation*. Erstes Konzept versteht die MÜ als Möglichkeit, fremdsprachliche Inhalte in einer anderen Sprache verfügbar zu machen. Zweiter Begriff verweist auf die Verbreitung von übersetzten Inhalten zwecks Publikation und letztgenannter Begriff dient dem zwischenmenschlichen Austausch in Form von Mails, Chats und ähnlichen eher informell ausgerichteten Bereichen (Koehn 2009: 20). Gerade letzter Aspekt ist auch mit Blick auf den Skype Translator ein wertvoller Hinweis.

Die Stärke der RBMÜ liegt dank der notwendigen Wörterbücher als Datengrundlage auf der hohen Terminologiekonsistenz. Allerdings stoßen die Grammatik- und Syntaxregeln auch bei noch so minutiöser Ausarbeitung bei Sprachen mit offenen Satzstrukturen rasch an ihre Grenzen, sodass die Ausgabe je nach Domäne und Sprachkombination nur schwer leserlich und fehlerbehaftet ist. Russisch, Baskisch oder Türkisch stellen RBMÜ-Systeme beispielsweise regelmäßig vor größte Herausforderungen.

Die SMÜ produziert zwar demgegenüber eine besser lesbare Ausgabe, jedoch bedarf es eines umfassenden Trainings auf Basis eines sauberen zweisprachigen

3 Technologischer Aspekt

Goldstandard-Korpus' (Bowker & Ciro 2019: 44). Je domänenspezifischer die Anpassung des Systems ist, desto genauere Ergebnisse produziert die SMÜ.

Die neuronale MÜ machte in jüngster Zeit besonders von sich reden. Mehrere Publikationen suggerierten auf Grundlage einzelner neuronal übersetzter Beispieldokumente einen Abgesang auf die Übersetzung durch den Menschen. Mildere Prognosen sehen zumindest eine erhebliche Konkurrenzsituation für die Übersetzungsbranche, da die getesteten Systeme schnell, flüssig und kosteneffizient Ergebnisse liefern, die für einen Laien durchaus brauchbar erscheinen.² Die NMÜ zeichnet sich vor allem durch eine gute Verarbeitung offener Satzstrukturen aus und erweitert sich durch jede Eingabe selbstständig (Schlagwort: *machine learning*, Bowker & Ciro 2019: 45). Allerdings ist die NMÜ auch die ressourcenaufwändigste Variante, da die neuronalen Netze auf Grundlage großer mehrsprachiger Korpora gewichtet werden müssen und allein dieser Prozess rechenintensiv ist. Weiterhin führen unbekannte Lexeme und Realia zu Auslassungen von z. T. ganzen Phrasen oder gar Sätzen in der Ausgabe. Ein ebenso schwerwiegendes Defizit ist die Invertierung des Sinns einer Aussage. Diese Bereiche problematisiert auch Koehn mehrfach: Bei der Aufbereitung der notwendigen Textquellen ist auf das jeweilige Urheberrecht zu achten (Koehn 2020: 7). Sprachenpaarabhängige Probleme in Bezug auf Lexik, Semantik und Syntax sind individuell zu betrachten (Koehn 2020: 9). Die maschinelle Übersetzung kann manche realweltliche Probleme schlichtweg nicht erkennen, lösen und übersetzen. Nicht zuletzt bedarf es für einen sinnvollen Einsatz von MÜ im professionellen Kontext eines domänenspezifischen Trainings, was erneut viel Zeit und Ressourcen erfordert. Gerade an diesem Punkt werde es sogar für Sprachdienstleister in puncto Ausstattung und Fachkenntnissen schwierig, allen Erfordernissen nachzukommen (Koehn 2020: 22).

3.3 CAT und CAI

Wie eingangs (s. S. 37) bereits erwähnt, soll sich an dieser Stelle nur eine kurze Darstellung der Systeme zur computerbasierten Unterstützung von Übersetzerinnen und Dolmetschern finden. Computerunterstützte Übersetzungs- und Dolmetschsysteme (jeweils *CAT*- und *CAI*-Tools) stehen dem Anwender unterstützend bei seiner Tätigkeit zur Hilfe, verarbeiten die Eingabe jedoch – im Gegensatz zu den vorausgehend beschriebenen MÜ-Systemen – nicht vollautomatisch. Die Grenzziehung an dieser Stelle scheint sinnvoll, da sich *CAT*- und *CAI*-Systeme offensichtlich am Randbereich der maschinellen Übersetzung befinden.

²Vgl. hierzu beispielsweise die Beiträge von Holzki (2020), Himmlein (2019).

So zählt sie Zimmermann (2012: 244) beispielsweise nur „[i]m weiteren Sinne“ dazu und Carstensen u. a. (2010: 642, 654) widmen den Systemen zwar ein Unterkapitel, unterscheiden übergeordnet jedoch explizit zwischen maschineller und computergestützter Übersetzung. Ganz klar unterscheidet Hutchins (1995: 431) zwischen MÜ und CAT- bzw. CAI-Tools mit den Worten: „It excludes computer-based translation tools which support translators by providing access to on-line dictionaries, remote terminology databanks, transmission and reception of texts, etc.“ Auch wenn er im Folgenden ebenfalls darauf hinweist, dass sich entlang dieser Grenze viele Gemeinsamkeiten finden lassen (Hutchins 1995: 431).

Bowker (2002: 4) nimmt an dieser Stelle ebenfalls eine klare Unterscheidung vor, die sie mit Verweis auf die bestehende Terminologie der *human-assisted machine translation* (HAMT) und der *machine-assisted human translation* (MAHT) verdeutlicht. Diese zwei Begriffe sieht sie als Standard, die weithin als MÜ und CAT bekannt sind. Als Unterscheidungsmerkmal dient ihr vor allem die Verantwortung (wobei auf die urheberrechtliche Dimension hierbei nicht weiter eingegangen wird), wer für die Übersetzung zuständig ist. Während bei der MÜ hauptsächlich der Computer die Verarbeitung des Textes vornimmt und lediglich durch Pre- und Postediting vom Menschen begleitet wird, liegt die Verantwortung bei CAT-Systemen durchweg bei menschlichen Übersetzer·innen, deren Arbeitsabläufe durch die Technologie beschleunigt und vereinfacht werden.

Ihre Definition umfasst allerdings auch Hilfsmittel, die möglicherweise zu selbstverständlich im Alltag verankert sind, so etwa die Rechtschreibprüfung und das Internet:

In its broadest definition, CAT technology can be understood to include any type of computerized tool that translators use to help them do their job. This could encompass tools such as word processors, grammar checkers, e-mail, and the World Wide Web (WWW). (Bowker 2002: 6)

Carstensen u. a. (2010: 655) betrachten hingegen in erster Linie die an die beruflichen Bedürfnisse angepassten Arbeitsumgebungen, Alignierungsprogramme und die Möglichkeit zur Verwaltung von Terminologiedatenbanken als Kernelemente.

In diese Richtung tendiert auch Seewald-Heeg (2005: 1), die das Wesen von CAT-Tools, neben den bereits o. g. Hilfsmitteln zur Kommunikation, Textverarbeitung und auch Angebotserstellung, über die *Translation Memory*-Systeme (TM) definiert, die vorgenommene Übersetzungen speichern und sie mit dem aktuellen Auftrag abgleichen, sodass Benutzer·innen teilweise (*fuzzy matches*)

3 Technologischer Aspekt

oder vollständige übereinstimmungen einfach übernehmen können. Zimmermann (2012: 246) erwähnt außerdem noch die Glossarfunktion, mit der die Nachbearbeitung (*Post Editing*) verbessert werden soll. Auch Koehn (2020: 21 f.) verweist in diese Richtung, indem er noch einmal besonders den Mehrwert der NMÜ als jüngstes System in diesem Bereich hervorhebt: Zwar reiche die Qualität der Übersetzung nicht an die einer professionellen, menschlichen Leistung heran, jedoch biete gerade die NMÜ kreative Übersetzungsvorschläge, die ein statistisches oder regelbasiertes System so nicht erwägen würde.

In Bezug auf das Dolmetschen sind andere Grundvoraussetzungen gegeben, die wiederum andere Ansprüche an die Fähigkeiten der CAI-Tools stellen. So spricht Fantinuoli (2009: 411) von Dolmetschern als Fachgebietslaien, die sich „auf einen bevorstehenden technischen Einsatz gezielt vorbereiten und sowohl fachliches als auch terminologisches Wissen aneignen“ müssen. Allerdings bemängelt er die bislang fehlenden technologischen Möglichkeiten, die den Dolmetschern zur Vorbereitung und während der Aufträge zur Verfügung stehen. Während sich allmählich zwar ein Forschungs- und Entwicklungsinteresse in dieser Richtung herausbilden würde, bestünden die gegenwärtigen Werkzeuge eines Übersetzers noch aus Stift, Papier und allenfalls Word und Excel (Fantinuoli 2009: 2).

Eine Kommunikationsanwendung wie der Skype Translator, der die Nachrichten maschinell übersetzt, ragt also in gewissem Maße in den Bereich der übersetzerischen Hilfsmittel hinein. Welche Merkmale in der Software wiederzufinden sind, zeigt der folgende Abschnitt auf.

3.4 Dienste für Voice- und Video-Chat

Skype mag lange Zeit die bekannteste Software für Voice- und Video-Chats gewesen sein. Eine Monopolstellung hat die Anwendung dadurch jedoch nicht mehr. Die Konkurrenten Facebook und Google betreiben jeweils mit ihrem *Facebook Messenger* und *Google Hangouts*³ Anwendungen, die Voice- und Video-Chats unterstützen. Auch Apple bietet mit *Facetime* einen eigenen Dienst in diesem Segment.

³Die Studie wurde knapp vor der Bekanntgabe durch Google konzipiert, Google Hangouts in seiner damaligen Form abzuschalten und sowohl für Privat- als auch Firmenkunden als „Hangouts Chat“ anzubieten. Zum Zeitpunkt des Verfassens dieser Arbeit war der Migrationsprozess der Kunden jedoch noch im Gange. Daher wurde der Name „Google Hangouts“ weiterhin im Fragebogen belassen. Für die genaueren Hintergründe s. z. B. <https://arstechnica.com/gadgets/2019/01/the-great-google-hangouts-shutdown-begins-october-2019/>, abgerufen am 31. Januar 2022.

Ferner existieren kleinere Dienste wie etwa die Smartphone-App *Viber* oder die browserbasierte Anwendung *appear.in*⁴. In den 2000er Jahren waren zudem die Dienste *ICQ* und der *Microsoft Windows Live Messenger* beliebte Plattformen für Video- und Voice-, aber vor allem für Textchats. Die meisten Dienste erfordern eine Registrierung und die Erstellung eines Nutzeraccounts. Hierbei werden mitunter personenspezifische Daten wie bevorzugte Sprache, Geschlecht oder das Alter erhoben. Die Angabe ist in den meisten Fällen jedoch freiwillig. Da viele der Programme auch als mobile Version zur Verfügung stehen, wird bei der Anmeldung auch die Handynummer entweder zwingend oder optional erfasst. Nutzungsgebühren fallen bei Verwendung der Standardversionen keine an, jedoch bieten die meisten Dienste erweiterte Funktionen wie Gruppenchats, Add-Ons oder Festnetztelefonie, welche dann wiederum kostenpflichtig sind.

Die Recherche nach verlässlichen Nutzungszahlen gestaltet sich als schwierig, wenn nicht nur auf die firmeneigenen Angaben vertraut werden soll. Oftmals fehlen Angaben zur Nutzungsdauer und -frequenz. Ein weiteres Problem sind die schwankenden Nutzungszahlen nach Region. Einer Datenerhebung von Statista zufolge nutzten im Jahr 2017 etwa 300 Mio. Menschen monatlich den Facebook Messenger für Voice- und Videochats (Statista 2017).

In enger Verwandtschaft zu diesen Arten der Kommunikationssoftware stehen zweierlei Konzepte, die immer stärker in die Lebenswelt des Menschen drängen. Dabei handelt es sich einerseits um Chat-Bots, also textbasierte Dialogsysteme, die den Nutzer·innen eine menschliche Kommunikation simulieren. Sie werden dort eingesetzt, wo es zu häufigen Anfragen kommt, die wiederum durch einfache Verschlagwortung und das sog. *Information Retrieval* gelöst werden können. Internationale Unternehmen etwa haben auf ihren Webseiten eben diese Chat-Bots implementiert, die ohne die Kosten von menschlichen Sachbearbeitern die Anfragen der Kund·innen kanalisieren und geläufige Probleme beantworten können.

Andererseits wird das Prinzip des Chat-Bots von sog. *Personal Assistants* aufgegriffen, die über die einfache Frage-Antwort-Verarbeitung hinaus auch (Sprach-) Befehle der Anwender·innen aufgreifen und ausführen können. Die Paradebeispiele hierfür stammen von den global bekannten Technologiekonzernen Amazon, Google und Apple und sind unter den Namen *Alexa*, *Google Assistant* und *Siri* bekannt. Durch die Installation des physischen Lautsprechers im Haushalt ist es möglich, gekoppelte Geräte im selben Netzwerk zu steuern, Informationen und Nachrichten abzurufen oder gar Bestellungen im Internet zu tätigen.

⁴<https://www.appear.in>: Offenbar aus rechtlichen Gründen wurde der Dienst umbenannt in *Whereby*, <https://whereby.com/>. Während der Ausarbeitung dieses Buches agierte der Dienst jedoch noch unter altem Namen. Beide Links abgerufen am 31. Januar 2022.

3 Technologischer Aspekt

3.4.1 Skype

Zu Beginn des neuen Jahrtausends war Skype eine Zeit lang die wohl bekannteste Instant-Messaging-Software für Video- und Voice-Chats. 2003 wurde der Dienst von Skype Technologies entwickelt, bevor er 2011 von Microsoft aufgekauft wurde. Die Verwendung des Dienstes bedarf lediglich der Registrierung und Erstellung eines kostenfreien Nutzerkontos. Für die Kommunikation über Skype sind jeweils ein Gerät zur Toneingabe und zur -ausgabe von Nöten. Dabei kann es sich sowohl um ein Headset mit Mikrofon und Kopfhörern als auch um ein separates Mikrofon und Lautsprecher handeln. Der Videochat erfordert eine Webcam. Derzeit steht die Software für die Betriebssysteme MacOS, Windows und Linux zur Verfügung. Darüber hinaus ist sie als App für Android und iOS erhältlich, oder kann plattformunabhängig in ausgewählten Browsern gestartet werden.

Neben den bereits angesprochenen Voice- und Videochats bietet der Dienst eine Konferenzschaltung mit bis zu 25 Personen, Dateienübertragung, Anrufe und SMS (beides kostenpflichtig) auf Festnetz und Mobiltelefon, Bildschirmübertragung und eine Business-Version mit erweiterten Konferenzeinstellungen. Für die vorliegende Monographie ist die Funktion des Skype Translators von zentralem Interesse, die gegenwärtig (Stand: 31. Januar 2022) Arabisch, Mandarin, Deutsch, Französisch, Englisch, Italienisch, Japanisch, südamerikanisches Portugiesisch, Russisch und Spanisch für Sprach- und Videoanrufe⁵ sowie 60 weitere Sprachen für Textchats unterstützt. Der Skype Translator läuft auf allen gängigen Plattformen ab Windows 7 sowie auf MacOS X, iOS, Android bzw. Linux. Alternativ lässt sich die Funktion auch in der Browsersversion – dort allerdings nur mit Google Chrome, Safari oder Microsoft Edge – zuschalten.

Eine Veröffentlichung des Fraunhofer-Instituts für Eingebettete Systeme und Kommunikationstechnik (ESK) aus dem Jahre 2013 beziffert die Nutzer·innen auf insgesamt 650 Mio. weltweit (Messerer & Eickhoff 2013: 6). Weiterhin beschreibt selbige Publikation die Skype zugrunde liegende Technologie. So handelt es sich bei dem Dienst um ein sog. Peer-to-Peer-System, das die dynamische Erweiterung und Strukturierung des Netzwerkes ermöglicht. Im Gegensatz zu Festnetz- und Mobiltelefonen ist bei Skype keine Notruffunktion vorhanden. Außerdem hängt die Kommunikationsqualität stark von der zur Verfügung stehenden Bandbreite und der Netzwerkauslastung ab (Messerer & Eickhoff 2013: 4).

Zum Kreis der Nutzer·innen gehören vor allem Privatpersonen. Durch die in jüngerer Zeit implementierten Office-Anwendungen und die Ausweitung der Business-Version bemüht sich Skype jedoch auch verstärkt um Unternehmen.

⁵Auf der firmeneigenen Website geht Microsoft nicht auf die einzelnen Sprachvarietäten ein. Daher ist die Aufzählung hier ebenso oberflächlich gehalten.

3.4.2 Skype Translator

Die Funktion des Skype Translators konnte in seinem alten Konzept (bis zum 13. Januar 2020) als sog. Chat-Bot während der Kommunikation mit einem Gesprächspartner aufgerufen werden. Mit dem Befehl *Skypetranslator languages* ließen sich die zur Verfügung stehenden Sprachen auswählen (Abb. 3.2 auf S. 58 zeigt dies z. B. für die Browser-Version). Seit dem 13. Januar 2020 ist der Skype Translator tiefer in die Chat-Umgebung integriert: Eine Aktivierung ist nicht mehr nötig, der automatische Übersetzungsdiensst orientiert sich an der Sprach-einstellung des jeweiligen Endgerätes, sodass er aktiv wird, sobald die Sprach-einstellungen der beteiligten Parteien voneinander abweichen. Der eingegebene Text wird dann automatisch als Übersetzung ausgegeben, wobei mit einem Klick auf den Schriftzug *Original anzeigen*, der sich unmittelbar unterhalb der Nach-richt befindet, auch der ursprüngliche Text betrachtet werden kann.

Für die weitere Betrachtung des Skype Translators ist der Hinweis wichtig, dass das Chatfenster keinen automatischen Zeilenumbruch besitzt. Mit Chat-fenster ist im Folgenden der Bereich gemeint, in dem die Chatbeiträge auf dem Bildschirm angezeigt werden. Zwar verfügt die Desktopanwendung ebenso wie die Browser-Version über *responsive design*, sprich: die Größe und Anordnung der Textnachrichten sowie das generelle Layout werden auf Grundlage der Bild-schirmgröße proportional skaliert, jedoch kann es je nach Nachrichtenlänge den-noch zu Überlappungen der beiden Bereiche für ein- und ausgehende Nachrichten kommen (s. Abbildung 3.3, S. 59). Eine ausführliche Diskussion dieses Auf-baus findet in den Abschnitten 5.2.2.1 (S. 82), 6.3.1 (S. 122) sowie 6.5.1 (S. 170) statt.

Laut Unternehmensdarstellung basiert der Skype Translator auf einem maschi-nellen Lernprozess und einem proprietären neuronalen MÜ-System. So soll die Qualität der Übersetzung durch häufige Nutzung gesteigert werden, da sich die Präzision der Ausgabe an die wiederholte Eingabe und das Verhalten der Nut-zer-innen anpasst.⁶

3.4.3 Differenzierung der Technologie

Die Betrachtung des Skype Translators als Technologie macht es erforderlich, ei-ne grundlegende Unterscheidung zu treffen. Der Dienst wird zwar als *translator* be-titelt, sprich: Texteingaben im Chatfenster werden in Echtzeit dem Kommu-nikationspartner übersetzt, jedoch bietet er auch die Übersetzung der Gesprächs-situation bzw. des Video-Anrufs. Die Wortwahl im vorausgehenden Satz ist da-bei bewusst gewählt. Skype selbst begeht diese terminologische Ungenauigkeit.⁷

⁶Vgl. <https://support.skype.com/en/faq/FA34583/skype-translator-privacy-faq>, abgerufen am 31. Januar 2022.

⁷Vgl. <https://www.skype.com/de/features/skype-translator/>, abgerufen am 31. Januar 2022.

3 Technologischer Aspekt

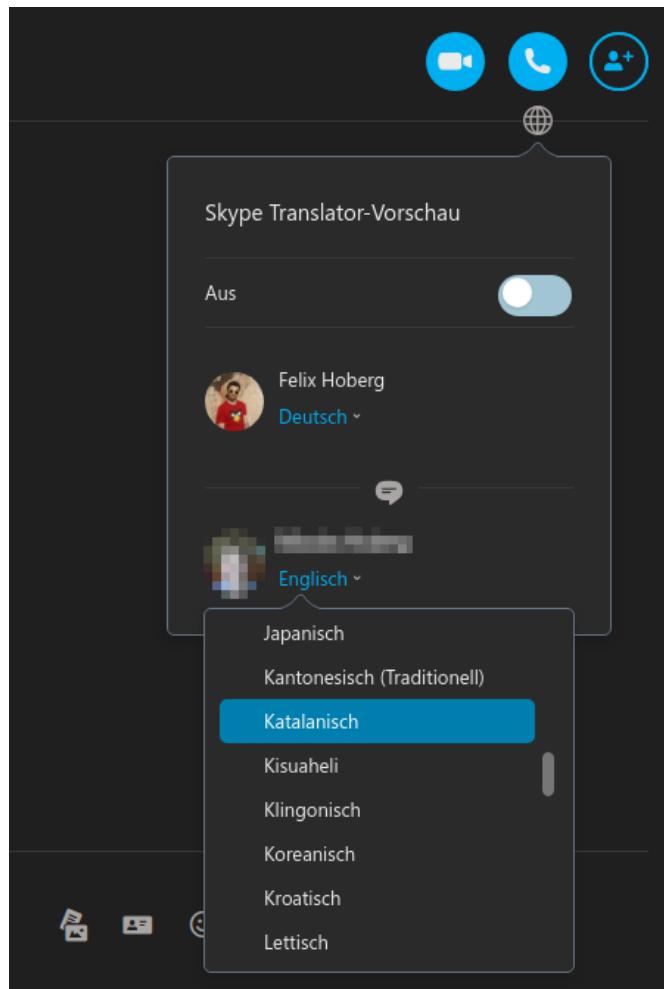


Abbildung 3.2: Sprachauswahl in der Browerversion des ST

3.4 Dienste für Voice- und Video-Chat



Abbildung 3.3: Ausgabe des Skype Translators

Diese Eigenvorstellung macht einen grundsätzlichen Unterschied zwischen der intendierten Nutzer.innenwahrnehmung und der faktischen, wissenschaftlich-technologischen Konzeption des Skype Translators deutlich: Die möglichen Anwender.innen sollen den Skype Translator wohl eher wie einen vollautomatischen Dolmetscher wahrnehmen, der dezent im Hintergrund agiert und verlässliche Ausgaben produziert. So bewirbt Microsoft den Skype Translator auf der Projektwebseite mit dem Anspruch, „Sprachbarrieren in Echtzeit“⁸ überwinden zu können, um sich „nahtlos mit Freunden, Familienmitgliedern, Kunden und Kollegen“ (s. Fußnote 8) auszutauschen.

Das wird auch mit Blick auf die Darstellung von Microsofts artverwandtem Projekt *Conversations* deutlich.⁹ *Conversations* beruht auf der gleichen Systemarchitektur wie der Skype Translator und wird (vermutlich) aus den gleichen Resourcen gespeist. Auf der Webseite des Projektes wird *Conversations* als dynamisches, interaktives System präsentiert, das dafür sorgt, „dass sich alle willkommen fühlen“. Tatsächlich handelt es sich allerdings um eine maschinelle Übersetzung auf Grundlage von statistischen und neuronalen Systemen, die mit einer Spracherkennung und Sprachsynthese kombiniert sind.

Ein Unterscheidungsmerkmal zwischen Dolmetschen und Übersetzen ist die Lebensdauer des zu verarbeitenden Textes. Kade weist darauf hin, dass der Ausgangstext beim Dolmetschen flüchtiger Natur sei und meist ausschließlich mündlich vorläge, wohingegen bei der Übersetzung stets schriftlich fixierte Texte aus einer in eine andere Sprache übertragen würden (Kade 1968: 34). Dieses Konzept mag heutzutage beinahe als selbstverständlich gelten und den Ausgangspunkt

⁸Vgl. <https://www.skype.com/de/features/skype-translator/>, abgerufen am 31. Januar 2022.

⁹Vgl. <https://translator.microsoft.com/> bzw. translate.it, abgerufen am 31. Januar 2022.

3 Technologischer Aspekt

der Dolmetscher- und Übersetzer.innenausbildung darstellen. Für die Funktionsweise des Skype Translators ist der Verweis auf diese Unterscheidung jedoch notwendig, da sich die vorliegende Arbeit ausschließlich dem Textchat widmet. Die Chatnachrichten sind also auch im Nachhinein einsehbar und können bei Bedarf von den am Chat beteiligten Personen erneut gelesen werden – ganz im Gegenteil zu der flüchtigen Sprachausgabe während des Video- oder Voicechats.

3.5 Zum Qualitätsbegriff

Die vorausgehenden Abschnitte haben den Stand der Technologie grob umrissen. Nun gilt es, die Möglichkeiten zur Evaluation eben dieser aufzuzeigen. Deshalb sei hier zunächst einführend auf den Begriff der *Qualität* geschaut. Der Duden¹⁰ schreibt dem Wort vier semantische Felder zu: 1) Beschaffenheit, 2) Eigenschaft, 3) Güte und 4) Schach – wobei letzteres für die vorliegende Arbeit keine Rolle spielt. Die Punkte 1), 2) und 3) haben gemein, dass sie sowohl in Bezug auf Lebewesen als auch auf Materialien verwendet werden. Erwähnenswert ist zudem, dass der Begriff *Qualität* eine Wertsteigerung von neutral zu positiv erfahren hat. Darauf verweist Zech (2015: 23), da der lateinische Ursprung ursprünglich nur ein Merkmal oder einen Zustand beschrieb. Erst heutzutage, gerade durch in Verwendung in Kombination mit dem Attributpaar *gut* und *schlecht*, hat sich der Wert des Wortes *Qualität* erhöht (Zech 2015: 23). Dies ist womöglich auf zweierlei Gründe zurückzuführen: Erstens ist es gegenwärtigen Strategien der Werbebranche geschuldet, Produkte in Qualitätsklassen wie *Standard* und *Premium* zu unterteilen. Zweitens zementiert die Forderung nach Qualitätsmanagement und die Normierung von Prozessen eine gehobene Erwartungshaltung an eben jenen Begriff.

Die Bewertung der Qualität von Übersetzungen, wie sie im Folgenden betrachtet wird, ist zunächst einmal zweierlei: ein sehr breit gefächertes Feld und ein Bereich mit langer Tradition innerhalb der Disziplin. Die Betrachtung des Verhältnisses von Ausgangs- und Zieltext steht nicht erst seit kurzem im Mittelpunkt des Forschungsinteresses. In der Routledge Encyclopedia of Translation wird, ausgehend vom Konzept des *translation quality assessments (TQA)*, zunächst zwischen verschiedenen Ansätzen und Betrachtungsebenen unterschieden, wobei der Fokus in erster Linie auf der rein menschlichen Tätigkeit liegt. Bei der Qualitätsbewertung wird dort zwischen vier Typen von Ansätzen unterschieden: anekdotische und subjektive, zu denen auch neu-hermeneutische Ansätze zählen, sowie weiterhin antwortorientierte sowie textbasierte Ansätze (Baker & Saldanha

¹⁰Vgl. hierzu <https://www.duden.de/rechtschreibung/Qualitaet>, abgerufen am 31. Januar 2022.

2011: 222). Ein weiteres kritisches Element bei der Evaluation ist darüber hinaus die Einheit, die betrachtet wird: Werden die einzelnen Wörter des Zieltextes als Bewertungsgrundlage verwendet oder doch die Sätze oder ausschließlich der gesamte Text.

Auch Castilho u. a. (2018: 11) heben den Stellenwert der Auseinandersetzung mit dem TQA hervor. Neben der vielfältigen Definition des Qualitätsbegriffs gebe es auch in Hinblick auf die Einsatzmöglichkeiten und den Umfang der Evaluation noch zu viele Unterschiede.

Martín Martín (2010: 230) weist deshalb gleich zu Beginn seines Artikels auf eine der zentralen Herausforderungen hin: Die Tätigkeit des Übersetzens hängt in ihrer Komplexität von vielen Faktoren ab. Diese beeinflussen auch die Bewertung. So gebe es nicht die eine perfekte Übersetzung; es existieren unterschiedliche Schulen; es herrsche kein Konsens über die Bewertungsmethoden (Martín Martín 2010: 230). Ähnlich fassen auch Moorkens u. a. das Qualitätsmanagement im Bereich der Sprachdienstleistung auf. Der Übersetzungsprozess in Verbindung mit der Qualitätsbewertung stellt für sie einen hochkomplexen Vorgang dar, der in den Translationswissenschaften, den Translationstechnologien, unter Sprachdienstleistern und in der Lokalisierungsbranche bislang vielseitig diskutiert wird. Fest steht für die Autoren, dass eine präzise Definition des Qualitätsbegriffs schwierig zu operationalisieren und zu messen ist (Moorkens u. a. 2018: 2).

Heutzutage wird sich sowohl manueller als auch automatisierter Verfahren bedient, die in Teilen in international anerkannten Normen fixiert sind. Die folgenden Abschnitte haben somit nicht das Ziel, alle bestehenden Ansätze – derer es unbestritten viele gibt – darzustellen. Vielmehr soll es darum gehen, das Gros der einzelnen Theorien herauszuarbeiten und zu präsentieren.

3.5.1 Qualitätsbewertung

Der folgende Abschnitt ist daher der Qualitätsbewertung gewidmet. Die Qualitätsbewertung stellt bislang den Oberbegriff dar, den es zu differenzieren und in Bezug zum Skype Translator zu setzen gilt. Allein die Erwähnung des Wortes *Qualität* in der Translationswissenschaft impliziert häufig neben der Evaluation auch das Qualitäts- und Risikomanagement.

Das wiederum betrifft die Bewertungsmöglichkeiten sowohl für Dolmetscher als auch Übersetzungstechnologien, in deren Rahmen der Skype Translator sowohl mit einem Angebot für gesprochene als auch geschriebene Kommunikation beworben wird. Erst nach der Ausarbeitung all dieser Faktoren wird es möglich

3 Technologischer Aspekt

sein, die Wahrnehmung des Skype Translators zu beschreiben, auf der – in Kombination mit Kapitel 5 und dem empirischen Teil in Kapitel 6 – eine Analyse zum Umgang mit Informationen in einem mehrsprachigen Chat stattfindet.

Das Qualitätsmanagement ist kein Konzept, das exklusiv der Translationswissenschaft vorbehalten ist. Beinahe jeder Wirtschaftszweig hat im Laufe seiner Entwicklung Mechanismen ausgebildet, die ein reibungsloses und effizientes Zusammenspiel aller Beteiligten (z. B. Auftraggeber-innen und -nehmer-innen), Prozesse und Objekte (z. B. Produkte) gewährleistet. Außerdem ist das Qualitätsmanagement der Evaluation als hierarchisch übergeordnet zu verstehen. Die Evaluation stellt mitunter das zentrale Element für Fortschritt in der Entwicklung – in diesem Fall in der Entwicklung des Skype Translators – dar. Durch sie erhalten Forschungsansätze die zwingend notwendige Rückmeldung sowohl über Bereiche, die bereits hohe Qualität aufweisen als auch solchen, die noch einmal überdacht und rekonzipiert werden sollten.

Das Risikomanagement soll an dieser Stelle weitestgehend ausgespart werden. Der Schwerpunkt des Risikomanagements in der Translationswissenschaft liegt vor allem auf der technischen Dokumentation und der Vermeidung von gravierenden, physischen Unfällen (vgl. z. B. Canfora & Ottmann 2015b, Altena & Moritz 2014). Inwiefern oder ob der Skype Translator bei unsachgemäßer Verwendung ein Risiko birgt, bleibt selbstverständlich zu untersuchen. Was nun im Hinblick auf das Übersetzen und Dolmetschen unter Qualitätsmanagement verstanden wird, zeigen die folgenden Abschnitte jeweils für die menschliche Tätigkeit (folgender Abschnitt 3.5.3) und maschinelle Verarbeitung (3.5.4, S. 63) auf.

3.5.2 Qualitätsnormen

Eine prominente Rolle nehmen dabei die unterschiedlichen Normen ein, die das Qualitätsmanagement sowie den Begriff Qualität definieren. In der Übersetzungsbranche ist dies die internationale ISO 17100, die generelle Qualitätsanforderungen an die professionelle Übersetzung und den Prozess stellt. Schmitt & Jüngst stellen dem voran die Normenreihe ISO 9000ff. sowie die Definitionen nach ISO 8402 bzw. ISO 9004-2 aus dem Qualitätswesen, nach denen „Qualität (...) also weder etwas Absolutes noch das maximal Machbare, sondern die Erfüllung definierter Erwartungen [ist]“ (Schmitt & Jüngst 2007: 394 f.). In Bezug auf die Arbeitsprozesse in der Übersetzungsbranche versucht DIN 2345, „die Abwicklung von Übersetzungsaufträgen zu vereinfachen und Empfehlungen auszusprechen, zu welchen Punkten Auftraggeber und Übersetzer (durchaus freiwillige) vertragliche Vereinbarungen treffen (...)“ (Schmitt & Jüngst 2007: 396).

3.5.3 Evaluation von menschlichen Übersetzungen

Die Evaluierung von Übersetzungen beginnt bereits während der Ausbildung, wenn Lehrende die Leistungen der angehenden Sprachmittler·innen beurteilen. Ein rudimentärer Ansatz ist dabei die manuelle Evaluierung durch professionelle Übersetzer·innen bzw. durch die Kursleitung, indem Ausgangs- und Zieltext miteinander verglichen und anhand einer Skala bewertet werden. Ein zentrales Element ist die Kategorisierung und Gewichtung der Fehler. Die einfache Fehlerannotation mittels Kürzeln dürfte dabei noch aus der Schulzeit bekannt sein:

- /O für Orthographie
- /Z für Zeichensetzung
- /Präp für Präposition
- *Inhalts- und Sinnverschiebungen*
- usw.

Mitunter kommen auch Zeichen aus dem Umfeld professioneller KorrekturleseR·innen und Verleger·innen zum Einsatz, die wie o. g. Korrekturmarker sowohl am Seitenrand als auch am betreffenden Zeilenabschnitt angebracht werden. Im Gegensatz zu den Zeichen aus der Schule handelt es sich nicht um Buchstaben, sondern um Symbole wie etwa Kreise oder Pfeile.

Martín Martín (2010: 236) stellt jedoch fest, dass bei der Bewertung – zumindest im akademischen Umfeld – viel zu stark auf den Fehler fokussiert wird. Positive Aspekte einer Übersetzung finden hingegen kaum Beachtung. Dies ist in Verbindung mit dem Riskomanagement gerade bei Fachübersetzungen nachvollziehbar: Eine gut gestaltete Übersetzung erfüllt schlichtweg ihren Zweck, beispielsweise als Dokumentation oder Bedienungsanleitung eines Produktes. Eine schlechte Übersetzung kommt dem Anspruch nicht nahe und führt im schlimmsten Fall zu einer gesundheitsgefährdenden Verwendung.

In der Praxis kommt bei der Qualitätsbewertung neben dem Bewertungsmaßstab der Textfunktion also noch das potenzielle Risiko für Endnutzer·innen der jeweiligen Übersetzung hinzu. In Kombination können diese beiden Faktoren mit einem Quotienten abgebildet werden, um so die Qualität der Übersetzung auszudrücken. Dalla-Zuanna (2010: 24) beschreibt dieses Vorgehen exemplarisch an den Abläufen in der Übersetzungsabteilung des VW-Konzerns. Die Fehlerkategorien im Beispiel des Autoherstellers sind:

3 Technologischer Aspekt

- Falsche Benennung (*WT, Wrong Term*)
- Falsche Bedeutung (*WM, Wrong Meaning*)
- Auslassung (*OM, Omission*)
- Strukturfehler (*SE, Structural Error*)
- Rechtschreibfehler (*SP, Misspelling*)
- Interpunktionsfehler (*PE, Punctuation Error*)
- Diverses (*ME, Miscellaneous Error*)

Jede dieser Kategorien wird zudem graduell abgebildet durch die Einteilung in mindere und schwere (*minor* und *serious*) Fehler (Dalla-Zuanna 2010: 24). Diese Unterteilung ist keineswegs neu, wie Martín Martín (2010: 237) mit Verweis auf das kanadische *Translation Bureau* feststellt, das derart schon in den 1980er Jahren vorging. Carstensen u. a. (2010: 653) merkt an, dass die Bewertung häufig durch mehrere Übersetzer·innen durchgeführt wird, damit sie nicht auf einer einzigen subjektiven Wahrnehmung basiert. In diesem Zusammenhang stellt Williams (2009: 3) fest, dass es zwar einen Bedarf an guten, zufriedenstellenden oder akzeptablen Übersetzungen gibt, die Maßstäbe, mit denen diese Bewertungen festgelegt werden, jedoch national und international undurchsichtig und Gegenstand fortwährender Diskussion sind. Auch Koehn (2009: 217) und King (1997: 251) weisen auf das Fehlen einer allgemeingültigen Antwort auf die Frage nach der korrekten Bewertungsmethode hin. Ferner ist es Drugan (2013: 36), die die unübersichtliche Lage innerhalb des akademischen und wirtschaftlichen Bereichs aufgreift. So könnten selbst Theoretiker·innen nicht exakt bestimmen, wie viele Ansätze zur Qualitätsbewertung es überhaupt gäbe Drugan (2013: 36). Als Hauptgrund führt Drugan die stark variierenden Bedürfnisse in der Praxis an. Während die Qualitätsbewertung in der Theorie an jeden Einzelfall exemplarisch angepasst werden kann, findet sich in der Praxis eine „huge diversity in real-world needs and requirements“ (Drugan 2013: 37), was sie an Übersetzungen im medizinischen Bereich exemplifiziert. Dort werden Qualitätsanforderungen nicht nur durch die eigentliche sprachmittlerische Tätigkeit, sondern auch durch die Regularien des Fachgebietes selbst gestellt (Drugan 2013: 37). Aber auch die theoretische Anpassung und Verfeinerung der Modelle stören ein effizientes Qualitätsmanagement.

Williams schlägt daher vor, sich stärker auf das *translation quality assessment* zu konzentrieren. Der Translationsforschung sei zwar viel daran gelegen, Quellen, Protagonist·innen, Ausgangs- und Zieltexte – kurzum: das Übersetzen in seiner gesamten Breite – zu evaluieren, eine einheitliche Festlegung auf Gütegrade habe es bislang jedoch nicht gegeben (Williams 2009: 4).

Die Gründe hierfür sieht Pym (1992: 279) mitunter in der Subjektivität des Forschungsfeldes und in der mangelnden Representativität der durchgeföhrten Studien. Es sei schlichtweg schwierig, professionelle Sprachmittler·innen als Proband·innen zu rekrutieren, weshalb Studien zur Qualitätsbewertung sich häufig auf Studierende stützten.

Das TQA liefert somit ein Rahmenwerk, innerhalb dessen die Qualität von Übersetzungen eindeutig und objektiv bewertet werden soll. Dabei kann der Ansatz allerdings ebenso variieren wie die Bedeutung des Begriffs *Evaluation*: entweder qualitativ oder quantitativ, diagnostisch, formativ oder summativ. Dieser Ansatz soll es dennoch ermöglichen, allgemein anerkannte Gütekriterien zu benennen (Williams 2009: 4).

Williams (2009: 5) identifiziert acht Faktoren, die es bei der Erstellung eines TQA-Models zu beachten gilt: den/die Übersetzer·in, Stilanforderungen in der Zielsprache, Schwere der Transferfehler, Umfang der Textanalyse, Art der Qualitätsbewertung, Einteilung der Schweregrade von Fehlern, Anzahl an Bewertungsebenen, Eigenanspruch bzw. Zweck der TQA.

Diesen Aussagen ist jedoch entgegenzusetzen, dass es sehr wohl Normen und Standards gibt, die zumindest europaweit Anwendung finden. Diese enthalten hauptsächlich Empfehlungen und Hinweise zum Qualitäts- und Risikomanagement im Umfeld der menschlichen übersetzerischen Tätigkeit. In ihrem Artikel zu derartigen, übersetzungsbezogenen Normen führen Canfora & Ottmann (2015a) drei Normen auf, die innerhalb der Branche weit verbreitet sind. Dabei handelt es sich um die sog. *Übersetzungsnorm ISO 17100*, die internationale *Risikomanagementnorm ISO 31000* und die unternehmensbezogene Norm für Qualitätsstandards *DIN EN ISO 9001* (Canfora & Ottmann 2015a).

3.5.4 Evaluation von maschineller Übersetzung

Die *Knowledge Base* der *Translation Automation User Society (TAUS)* listet mehrere Methoden zur Evaluation von maschineller Übersetzung auf. Die zwei größten Felder dabei sind die Evaluation durch den Menschen und automatisierte Verfahren. Beide Felder unterteilen sich weiterhin in die direkte und indirekte Bewertung. Ferner zählt die TAUS noch die Rückmeldung von Kund·innen und die

3 Technologischer Aspekt

Evaluation durch die Endnutzer·innen (*community evaluation*) zu den gängigen Methoden (Translation Automation User Society (TAUS) 2017a).

Die Evaluation der MÜ durch den Menschen beruht in erster Linie auf Vergleichen. In dem Fall wird der von der Maschine produzierte Zieltext mit der Referenzübersetzung von menschlichen Übersetzer·innen verglichen. Bei dieser Methode müssen „verschiedene Dimensionen der Übersetzungsqualität getrennt betrachtet (z. B. grammatische Korrektheit, Flüssigkeit, Übersetzungstreue“ (Carstensen u. a. 2010: 653) werden. Koehn (2009: 218) empfiehlt für die Bewertung mittels Referenzübersetzung den Einsatz von zweisprachigen Gutachter·innen, auch wenn dies häufig aufgrund mangelnder Verfügbarkeit nicht möglich ist.

Wie auch schon bei der Bewertung von menschlichen Übersetzungen kann auch in diesem Fall das Verhältnis von Lesbarkeit (*fluency*) und Angemessenheit (*adequacy*) des Zieltextes betrachtet werden (Koehn 2009: 219). Die Beurteilung erfolgt dabei auf Grundlage einer Skala, die beispielsweise von 0 bis 5 reicht und im Falle der Lesbarkeit von einer einsprachigen Person und im Falle der Angemessenheit von einer bilingualen Person durchgeführt wird (Läubli u. a. 2018: 2).

Automatisierte Verfahren bedienen sich hingegen Algorithmen, die die Nähe von Ausgangs- und Zieltext in Relation zu einer Referenzübersetzung abbilden sollen. Auch in diesem Fall stammen die Referenzübersetzungen von einem Menschen. Diese dienen zugleich häufig der Verbesserung der MÜ-Ausgabe (Lommel, Burchardt u. a. 2014, Läubli u. a. 2018, Ott u. a. 2018, Khayrallah u. a. 2020).

Die bekanntesten Metriken sind der sog. *Bi-Lingual Evaluation Understudy-Score* (BLEU-Score) und seine Weiterentwicklung *METEOR*, der *General Text Matcher* (GTM) sowie die *Translation Error Rate* (TER). Weiterhin hat das *US National Institute of Standard and Technology* (NIST) mit dem gleichnamigen Bewertungssystem eine weitverbreitete Methode eingeführt (Translation Automation User Society (TAUS) 2017b).

Der Einsatz automatisierter Verfahren birgt sowohl Vor- als auch Nachteile. Einerseits sind sie für die jeweiligen Unternehmen sowohl in wirtschaftlicher als auch zeitlicher Hinsicht effizienter und aufgrund der algorithmischen Natur auch weitaus konsistenter als eine menschliche Evaluation. Andererseits liefern diese Metriken ausschließlich numerische Werte, die entlang einer entsprechenden Skala interpretiert werden können. Eine individuelle Analyse der Übersetzung bzw. des Übersetzungssystems in puncto Stärken und Schwächen ist hingegen nicht möglich. Daher ist die Forschung in diesem Bereich zu klassifizierenden Ansätzen übergegangen, die zu den bewertenden Systemen Profile und Fehlerklassen erstellt (Popović 2018: 130 f.).

Eng mit den Ausführungen der TAUS verbunden ist die Multidimensionale Qualitätsmetrik (MQM), welche vom Deutschen Forschungszentrum für Künst-

liche Intelligenz entwickelt wird.¹¹ Die Metrik steht unter einer Creative-Commons-Lizenz, welche die Verbreitung unter Namensnennung und ohne Derivate erlaubt.¹² Damit ermöglicht es diese flexibel konzipierte Metrik, die Qualität von Übersetzungsprojekten zu bewerten und die einzelnen Bereiche der Metrik an die jeweiligen Anforderungen anzupassen (Lommel, Uszkoreit u. a. 2014). Die Konzeption einer derart flexiblen Metrik ist zugleich auch der größte Kritikpunkt: Im Gegensatz zu vielen wertbasierten Ansätzen wirkt die MQM zu umfangreich, um kurzfristige Bewertungen vorzunehmen.

3.6 Postediting

Das Postediting stellt die nachträgliche Korrektur der rohen MÜ-Ausgabe durch menschliche Übersetzer·innen gemäß festgelegter Richtlinien und Qualitätskriterien dar (O'Brien u. a. 2014: 197 f.). Durch das kontinuierlich wachsende Übersetzungsvolumen nimmt auch der Einsatz von MÜ zu, wodurch sich primär Zeitersparnis erhofft wird. Das wiederum führt professionelle Übersetzer·innen in die Situation, mit Rohübersetzungen arbeiten zu müssen, was bis zu Beginn der 2000er Jahre noch ein gänzlich neues Feld darstellte. Mit der ISO 18587 (2017) besteht mittlerweile jedoch sogar eine Norm, die standardisierte Anforderungen an den Prozess des Posteditierens sowie an die Fähigkeiten der Posteditor·innen stellt.

Übersetzer·innen finden bei der Postedition ein bereits bestehendes textuelles Grüst vor, das sie nachträglich weiterverarbeiten und einen funktionierenden, fehlerfreien Text formen. Verschiedene Studien und Erfahrungsberichte aus der Praxis untersuchen den möglichen Produktivitätsgewinn durch den Einsatz von MÜ und Postediting (Castilho & Guerberof Arenas 2018, Macken u. a. 2020). Je nach Güte des zur Verfügung stehenden MÜ-Systems kann das Postediting jedoch auch ins Gegenteil schlagen und zeitlichen Mehraufwand sowie höhere Kosten bedeuten. Daher hat sich analog zum Postediting als Nachbearbeitung das sog. Preediting herausgebildet, dessen Ziel die regularisierte, kontrollierte Vorbereitung des Ausgangstexts für die MÜ ist (Baker & Saldanha 2011: 164).

Das Postediting und die menschliche Übersetzung sind als verschiedene Aufgaben anzusehen. Das ist nicht zuletzt darin begründet, dass die MÜ-Ausgabe andere Merkmale aufweist als die menschliche Übersetzung. Postediting wird in

¹¹S. hierzu <http://www.qt21.eu/mqm-definition/definition-2015-06-16.html>, abgerufen am 31. Januar 2022.

¹²S. hierzu https://creativecommons.org/licenses/by-nd/4.0/deed.en_US, abgerufen am 31. Januar 2022.

3 Technologischer Aspekt

vielen Forschungsbeiträgen deshalb nicht als Art des Korrekturlesens angesehen, sondern strikt davon abgegrenzt.

Massardo u. a. (2016: 8 f.) stellen die geläufigste Kategorisierung dieser Tätigkeit vor: das vollständige (*full*) und das leichte (*light*) Postediting. Das vollständige Postediting ist auf ein druck- und veröffentlichtungsreifes Endprodukt ausgerichtet. Daher findet die Überprüfung auf nahezu allen Ebenen statt, von der Anpassung von Punktuation, Maßeinheiten und Formaten, über den Umgang mit Eigennamen bis hin zur morpho-syntaktischen Korrektur. Das leichte Postediting hingegen ist auf die generelle Lesbarkeit des Textes ungeachtet des Stils ausgelegt, wobei auch hier bereits auf terminologische Konsistenz und die Vollständigkeit der MÜ-Ausgabe geachtet wird (Massardo u. a. 2016: 16 f.).

Die Geschichte des Posteditings ist zudem eng verbunden mit der Entwicklung der MÜ sowie der Evaluation von Übersetzungen. Sie lässt sich an den o. g. Wegmarken der MÜ (s. Abschnitt 3.1, S. 37) nachverfolgen. Eine detaillierte Übersicht präsentiert Garcia (2012: 293), der das Georgetown-Experiment und den ALPAC-Bericht (S. 3.1.3, S. 39) als Ausgangspunkt für die wissenschaftliche und praktische Auseinandersetzung mit nachträglich durch den Menschen bearbeitete MÜ-Ausgaben sieht. Der ALPAC-Bericht führte, wie o. g., zu einer drastischen Reduzierung der Forschungsmittel für die MÜ, wodurch auch das Postediting zunächst vernachlässigt wurde. MÜ und Postediting galten zu der Zeit als zu teuer und zeitaufwändig im Vergleich zu menschlichen Übersetzungen (Garcia 2012: 295). Erst in den 1980er Jahren wurde die (Nach-)Bearbeitung der MÜ-Ausgabe aus wirtschaftlicher Sicht für Unternehmen attraktiv (Garcia 2012: 297). Ungefähr zur selben Zeit bildeten sich auch die ersten Richtlinien zum Einsatz von Postediting im Berufsalltag aus (z. B. TAUS, Massardo u. a. 2016), während besonders über die vergangenen zwei Jahrzehnte der Anteil des Posteditings an der akademischen Übersetzer-innenausbildung und am übersetzungswissenschaftlichen Forschungsbetrieb deutlich gesteigert wurde. Auch die überwiegende Mehrheit der professionellen CAT-Tools hat in jüngster Vergangenheit Schnittstellen zur Integration von MÜ erhalten, wodurch dem Postediting im Berufsalltag ein größerer Stellenwert zukommt.

3.7 Usability

Unter Usability versteht man die Nutzbarkeit oder auch die Nutzer.innenfreundlichkeit eines Produktes oder einer Software.¹³ Bei der Untersuchung von Usability geht es in erster Linie häufig um die Qualität der Interaktion zwischen

¹³S. hierzu <https://www.duden.de/rechtschreibung/Usability>, abgerufen am 31. Januar 2022.

Anwender und Produkt. An diesem Punkt wird die Untersuchung der Nutzbarkeit und Nutzer-innenfreundlichkeit bereits in zwei grundlegende Felder unterteilt: Einerseits werden Produkte in der analogen Welt in puncto Ergonomie und Design untersucht. Andererseits bezieht sich das Konzept auf die qualitative Be trachtung von Software, um die es sich auch schwerpunktmäßig im Folgenden handelt.

Die Untersuchung von Nutzer-innenfreundlichkeit ist weiterhin zu differenzieren und auf vielen verschiedenen Ebenen zu verstehen. Stapelkamp (2007) etwa stellt im Umfeld des Screen- bzw. Interfacedesigns die einzelnen Punkte der Projektentwicklung, des Gestaltungsprozesses und der Umsetzung heraus. Ein (Software-)Produkt wird demnach an seiner Form, Funktion und Zielgruppe gemessen.

Da es wünschenswert ist, dass jedes Produkt die Ansprüche und Bedürfnisse seiner Anwender befriedigt und auch die Inbetriebnahme und Nutzung dieser Produkte keine Herausforderung, sondern im Idealfall eine Erleichterung und Bereicherung darstellen sollte, spielt Usability für die Entwicklung von allen Produkten eine entscheidende Rolle. (Stapelkamp 2007: 514)

Die Forschung zur Nutzbarkeit bzw. der Nutzer-innenfreundlichkeit von maschineller Übersetzung ist nur spärlich ausgeprägt. Der Schwerpunkt des Interesses liegt vor allem auf dem Umgang mit Postediting. In diesem Bereich stützen sich Castilho & O'Brien auf eine Adaption der ISO/TR 16982. Drei wesentliche Aspekte hieraus bilden den Kern der Definition: Effektivität, Effizienz und Zufriedenheit (der Nutzer-innen). Castilho & O'Brien (2016: 310 f.) verweisen zwar auf verschiedene Studien in den vergangenen 25 Jahren zum Nutzen von maschinell übersetzten Inhalten. Die meisten von ihnen weisen jedoch normativ-evaluierenden Charakter auf. Castilho & Guerberof Arenas (2018: 79) merken in diesem Zusammenhang an, dass trotz eines starken Fokus auf der Evaluation von MÜ-Qualität Untersuchungen zum Einfluss von maschineller Übersetzung auf Endnutzer-innen bislang unterrepräsentiert sind.

Neben dem Einsatz von Eye-Tracking-Methoden zur Untersuchung des Nutzungsverhaltens und der Interaktivität werden bei Usability-Studien häufig Fragebögen oder Leseverständnistests eingesetzt, die eine subjektiv-individuelle Rückmeldung von den Versuchspersonen ermöglichen sollen. So arbeiten Daems & Macken (2019: 128) mit einem Fragebogen, um die bevorzugte Nutzung von interaktiven SMÜ- oder NMÜ-Systemen beim Postediting zu erheben.

Der Nutzen einer rohen (eng. *raw*) MÜ-Ausgabe wurde mittels Eye-Tracking-Methoden erstmals von Doherty u. a. untersucht. Hierzu wurden französischen

3 Technologischer Aspekt

Muttersprachler:innen einzelne von der MÜ übersetzten Sätze präsentiert. Die Versuchspersonen sollten die Sätze nach ihrer Verständlichkeit bewerten und wurden während der Leseaufgabe von einem Eye-Tracker erfasst. Die gesammelten Daten wurden mit den Ergebnissen der HTER-Metrik verglichen. Doherty u. a. (2010) kommen zu dem Schluss, dass der Eye-Tracking-Einsatz eine wertvolle Ergänzung bei der Evaluation von MÜ sei. Besonders die Möglichkeit, die Interaktion mit der MÜ-Ausgabe zu beobachten, stellt für die Autoren einen besonderen Mehrwert dar.

Unter Rückgriff auf Stapelkamp ist es daher notwendig, den Skype Translator besonders unter den folgenden Aspekten zu betrachten: Aus dem Bereich *Form* sind Orientierung, Gestaltungslayout sowie das Screen- und Informationsdesign zu betrachten. Aus dem Bereich *Funktion* fällt dem Interaktions- und Interface-design eine wichtige Rolle zu.

3.8 Zusammenfassung

Das vorausgehende Kapitel hat zunächst die gängigen MÜ-Systeme sowohl historisch als auch funktionell umrissen. Der Schwerpunkt lag auf dem jüngsten Ansatz, der NMÜ, wie sie in Kombination mit der SMÜ mutmaßlich nach allen verfügbaren Quellen für den Skype Translator angewendet wird. Zur genaueren Ein- bzw. Abgrenzung wurde einerseits die Funktion von im professionellen Übersetzungsbereich verwendeten CAT- und CAI-Tools dargestellt. Andererseits wurde genauer auf Dienste für Voice- und Videochats eingegangen, um hierüber eine Brücke zu geläufigen Evaluationsmöglichkeiten von Translationstechnologie zu schlagen. Die Usability spielt in diesem Zusammenhang ebenfalls eine wichtige Rolle, weshalb sie in Verbindung mit dem Postediting den Abschluss dieses Kapitels darstellt.

4 Arbeitshypthesen

Das folgende Kapitel stellt die Arbeitshypothesen ausgehend von der Forschungsfrage dar und erklärt sie eingehend.

Während die Untersuchung der Nutzbarkeit von maschinell übersetzten Texten besonders im Umfeld des Posteditings über eine lange Forschungstradition verfügt, ist die Betrachtung im Spannungsfeld von Mensch und Maschine noch weitestgehend Neuland. Zwar gibt es immer wieder Versuche, maschinelle übersetzte Inhalte in den Wahrnehmungsbereich der Endnutzer·innen von Kommunikationsmedien zu schieben, jedoch sind diese spärlich vertreten und lassen bislang eine spürbare Aufmerksamkeit vermissen. So bieten alle großen Social-Media-Plattformen seit mehreren Jahren bereits maschinell übersetzte Inhalte in anderen als der Systemsprache des verwendeten Endgeräts. Eine intensive Auseinandersetzung mit dem Mehrwert, den eine solche Erweiterung der Interaktions- und Kommunikationsmöglichkeiten bietet, hat bislang aus übersetzungswissenschaftlicher Sicht nur begrenzt, und dort eher im professionellen Rahmen, stattgefunden, sodass sich für die nachfolgende Analyse die nachstehenden Arbeitshypothesen definieren lassen.

Da die computervermittelte Kommunikation zwischen Menschen eine zutiefst subjektive Komponente beinhaltet, werden über die kognitiven Prozesse hinaus auch Hypothesen auf Grundlage der zur Verfügung stehenden Fragebögen getroffen. Deshalb stellt sich hier die Frage, wie und ob sich die Nutzer·innen des Skype Translators überhaupt bewusst sind, dass ein MÜ-System der Kommunikation zwischengeschaltet ist. Da die verwendete Kohorte aus gegenwärtig eingeschriebenen Studierenden besteht, die darüber hinaus allesamt aus einer Generation stammen, die ohnehin permanent vom medialen und technologischen Wandel begleitet ist, kann angenommen werden, dass die Kohorte generell affin mit derartiger Software umgeht. Konkreter bedeutet dies, dass die Proband·innen eine Gesprächssituation über die gestellte Aufgabendauer aufrechterhalten können. Dies geschieht trotz bzw. wegen der Implikationen, die sich aus den im Theorieteil präsentierten Kommunikationsmodellen (Abschnitt 2.2.2, S. 18) sowie der Charakterisierung als Nähe- und Distanzkommunikation in Bezug auf die Chat-situation ergeben. Sollten Verständigungsprobleme eintreten, werden diese von ihnen vermutlich wahrgenommen, aber auch ausgeglichen.

4 Arbeitshypothesen

Die Verbindung von Fragebögen und Eye-Tracking-Studie bietet Raum für weitere Annahmen. Die Kommunikationssituation wird auf dem Niveau der gestellten Aufgabe erfolgreich bewältigt. Eine fehlerfreie, publikationsreife Qualität der Kommunikation ist für die Teilnehmer·innen nebensächlich. Der Erfolg der Kommunikation liegt vielmehr in dem flüssigen, reibungslosen Austausch zwischen den beteiligten Personen.

Der monolinguale Versuchsaufbau dient als Referenzpunkt für die Betrachtung der Eye-Tracking-Daten. Hier interagieren zwei Personen über Skype, allerdings in derselben Sprache und ohne sonstige technologische Mittelung. Es ist daher davon auszugehen, dass sich Elemente der Chatkommunikation, wie sie im Theorieteil dargestellt wurden, auch hier aufzeigen lassen. Dazu gehört im einsprachigen Chat sicherlich besonders die Fokussierung auf die Beiträge des Gegenübers, da diese neue bzw. bislang unbekannte Informationen darstellen, die die Versuchspersonen noch verarbeiten müssen. Der Skype Translator hingegen erweitert diese Kommunikation um die entsprechende Ausgabe der MÜ in beide Sprachen. Die Versuchsteilnehmer·innen im Setting Katalanisch-Deutsch werden deshalb vermutlich ihre Aufmerksamkeit auf alle vier verschiedenen Beitragsarten richten und das Kommunikations- sowie Leseverhalten dementsprechend anpassen. Mehr noch werden die Studienteilnehmer·innen durch eben diese informative Erweiterung ein hohes Maß an Struktur innerhalb des Chats erkennen lassen, dass sich insbesondere an den aktuellsten Nachrichten orientiert.

Weiterhin ergeben sich mehrere Annahmen aus dem Aufeinandertreffen von Mensch und Maschine. Die erste ist, dass sich die Elemente der Chatkommunikation, wie sie im theoretischen Teil aufbereitet wurden, auch in dieser um die MÜ-Ausgabe erweiterten Konstellation wiederfinden lassen. Deshalb kann zweitens erwartet werden, dass die unterschiedlichen Beitragsarten einen Einfluss auf das Kommunikationsverhalten haben. Dies lässt sich sowohl ausgehend von den absoluten Werten der Eye-Tracking-Daten als auch auf Grundlage der inferenzstatistischen Untersuchung erkennen.

Das Leseverhalten ist sowohl markant für die jeweiligen Originalbeiträge als auch deren MÜ-Ausgabe und für die jeweilige Sprache, in der sie verfasst sind. Es stellt sich also die Frage, wie diese vier verschiedenen Beitragsarten konkret wahrgenommen werden. Von besonderem Interesse ist dabei die Wahrnehmung der MÜ-Ausgabe, da zu vermuten ist, dass diese trotz aller Fortschritte in den vergangenen Jahren noch immer in irgendeiner Weise fehlerbehaftet ist und sich auch im Leseverhalten der Studienteilnehmer·innen widerspiegelt.

Da bewusst nur Personen ohne Vorkenntnisse im Katalanischen ausgewählt wurden, kann angenommen werden, dass die Proband·innen ein Leseverhalten an den Tag legen, das sich zwischen Original und MÜ-Ausgabe unterscheidet

und zugleich jedoch aufeinander bezieht. Mit Rückbezug auf den Umgang mit Wissen (Abschnitt 2.5, S. 26), besonders die Tatsache, dass Lesen und Schreiben auch immer Produktion, Integration und Konstruktion von Wissen seien, werden sich dementsprechend eindeutige Muster in der Betrachtung der einzelnen Indikatoren finden lassen. In diesem Zusammenhang steht die Vermutung, dass katalanischsprachige Originalbeiträge ebenfalls von den Proband·innen in den Strukturierungsprozess der Kommunikation integriert werden.

5 Empirischer Teil

Der empirische Teil der vorliegenden Arbeit umfasst eine allgemeine Umfrage zur Nutzung von Skype (s. Kap. 6.1 und Anhang A, S. 237) sowie eine Fallstudie, mit der die Wahrnehmung der Chat-Kommunikation unter Einsatz des Skype Translators untersucht werden soll. Auch wenn die Online-Umfrage nicht den Anspruch erhebt, umfassend und repräsentativ zu sein, sollen die Ergebnisse dazu dienen, die Wahrnehmung und die Nutzung von Skype – und im Speziellen des Skype Translators – besser einordnen zu können. An dieser Stelle sei bereits darauf hingewiesen, dass der gesamte empirische Teil pilotiert wurde. Um inhaltliche, organisatorische, logische und orthographische Fehler zu erkennen und zu beheben, wurden die Fragebögen vorweg an nicht beteiligte Personen weitergeleitet. Auch die Kommunikationssituation über Skype wurde – um nicht zuletzt die reibungslose Integration des Eye-Trackers in den Versuchsaufbau zu gewährleisten – vorab mit unbeteiligten Personen überprüft. Auf Grundlage der während der Pilotierung gewonnenen Erkenntnisse wurde für die Teilnehmer·innen ein Einweisungstext samt rechtlicher Hinweise ausgearbeitet, der im Anhang ersichtlich ist (s. Anhang B, S. 251). Alle weiteren Details den Aufbau und die Durchführung betreffend werden in den entsprechenden Unterkapiteln dargelegt. Zitate aus den Angaben der Proband·innen werden im Folgenden als Original mit Orthographie- und Interpunktionsfehlern dargestellt.

5.1 Begrifflicher Hinweis

Um Missverständnisse bei der Darstellung der nachfolgenden Studie zu vermeiden, sei hier noch kurz die verwendete Terminologie aufgeführt:

Chatbeitrag: Jegliche textuelle Äußerung von einer der beiden am Gespräch beteiligten Personen, die über die Eingabemaske von Skype in den Chat eingebracht und versendet wurde, sodass sie für die beiden Gesprächspartner·innen sichtbar wurde. Üblicherweise löst ein Chatbeitrag unter Verwendung des Skype Translators automatisch die maschinelle Übersetzung in die jeweils andere Sprache aus.

5 Empirischer Teil

eingehende Nachrichten: Unter eingehenden Nachrichten sind alle Nachrichten zu verstehen, die als neue Information an die Proband·innen im Rahmen der Kommunikationssituation herangetragen werden. Die MÜ-Ausgabe der von den Proband·innen selbst verfassten Nachrichten zählt nicht zu eingehenden Nachrichten.

ausgehende Nachrichten: Ausgehende Nachrichten wiederum sind die Chatbeiträge, die von den Proband·innen selbst verfasst wurden. Hierzu zählt auch die MÜ-Ausgabe der von den Proband·innen selbst verfassten Nachrichten.

5.2 Methodik

5.2.1 Online-Befragung

Das Erhebungsverfahren der wissenschaftlichen Umfrage ist ein vielseitiges Feld. Fragebögen werden heutzutage nicht nur in gedruckter Form verteilt, sondern können auch als Online-Version den zu Befragenden zugänglich gemacht werden. Eine Telefonumfrage sowie der Versand von Fragebögen per Post kann dabei ebenfalls zu dieser Kategorie gezählt werden.

Für die Wahl des Erhebungsverfahrens sind mehrere Faktoren entscheidend. Zunächst einmal steht die Aufwand-Nutzen-Abwägung im Vordergrund. Der Fragebogen bzw. die Umfrage gelten als zeit- und ressourcensparend sowie einfach durchzuführen (Albert & Marx 2014: 59). Weiterhin ist die Grundgesamtheit der Zielgruppe ebenso zu beachten wie die generelle Erreichbarkeit (sprich: der Zugang zu) dieser Gruppe. Auch die Reaktivität sollte vorab in die Konzeption miteinbezogen werden, da nicht jede befragte Person gleichermaßen gerne Online-Fragebögen, analoge oder Telefonumfragen beantwortet. Bei der persönlichen Befragung bzw. der Papierumfrage kann es zum sog. „Interviewer-Effekt“ (ebd., auch *observer's paradox*) kommen, bei der die befragte Person von der Gesamtsituation beeinflusst wird und dies zu einer Verzerrung der Antworten führt. Dieser Effekt tritt bei Online-Umfragen selten bis überhaupt nicht auf. Zugleich sollte bei der Konzeption aber bedacht werden, dass Fragebögen lediglich „die subjektiven Einschätzungen der befragten Person zu den Themen der Befragung (...) erfassen“ (Aeppli u. a. 2016: 164) und somit einer möglichen Verzerrung der Antworten durch die gegenwärtige Gefüllslage der befragten Personen unterliegen.

Ein Nachteil der Online-Befragung ist die Kontrollierbarkeit. Das Ausfüllen findet ohne die Möglichkeit für die befragte Person statt, bei Unklarheiten Rückfragen an den Wissenschaftler zu stellen. Auch kann es zu Missverständnissen

bei der Formulierung der einzelnen Fragen kommen, sodass die Ergebnisse durch eine unzureichende Konzeption von vornherein verzerrt sind. Dieses Problem kann durch eine tiefgreifende, ausführliche Pilotierung, also einem Testlauf bzw. Pretest, umfassend eingegrenzt werden (Aeppli u. a. 2016: 61). Nicht zuletzt liegt die Rücklaufquote von Online-Befragungen häufig unter der anderen Umfragearten (Aeppli u. a. 2016: 165). Ferner kann bei der Bereitstellung von Online-Umfragen nie vollständig gewährleistet werden, dass auch nur die intendierte Zielgruppe angesprochen wird. Auf Seiten des Wissenschaftlers drängt sich in diesem Zusammenhang nicht erst seit Inkrafttreten der Datenschutzgrundverordnung im Jahre 2018 die Frage nach der vollständigen Anonymisierung der erhobenen Daten auf (Kirchhoff u. a. 2010: 16). Die Anonymisierung bietet allerdings die Möglichkeit, Aussagen über heikle Themen zu erhalten, die in einer persönlichen Kontaktsituation so nicht einholbar wären (Aeppli u. a. 2016: 165).

Bei der inhaltlichen Ausarbeitung des Fragebogens ist die womöglich größte Herausforderung die Wahl der passenden Fragetypen mitsamt präziser Formulierung. Die zwei übergeordneten Typen sind die offene und die geschlossene Frage. Erstere erlaubt freie, umfangreichere Antworten, muss jedoch im Nachgang qualitativ analysiert werden. Zweitere hingegen gibt die Antwortmöglichkeiten in Form einer Skala (z. B. Likert-Skala), einer Matrix oder einer Multiple-Choice-Tabelle vor. Die so erhobenen Daten können einfach und schnell mit einem Kalkulationsprogramm statistisch aufbereitet werden.

Atteslander u. a. (2010: 139 ff.) differenziert neben o. g. Fragetypen weiterhin zwischen:

- direkten und indirekten Fragen
- Meinungsfragen
- Einstellungsfragen
- Sonntagsfragen
- Bilanzfragen

In diesem Zusammenhang haben sich Gütekriterien herausgebildet, die bei der Konzeption der einzelnen Fragen beachtet werden sollten. So wird gemeinhin empfohlen, demographische Angaben – sofern sie überhaupt notwendig sind – an das Ende der Umfrage zu setzen, um mögliche Teilnehmer·innen nicht zu Beginn mit der Erfassung von zutiefst persönlichen Daten zu verschrecken. Des Weiteren sollten Fragen so kurz, eindeutig und neutral wie möglich formuliert

5 Empirischer Teil

werden, um Missverständnisse und Unklarheiten oder auch subjektive Tendenzen zu vermeiden. Hiermit einher geht bei geschlossenen Fragen die Forderung nach einer vollständigen Angabe aller Antwortmöglichkeiten. Das Merkmal *Kürze* gilt auch als Kriterium für den gesamten Fragebogen, da die Aufmerksamkeitsspanne und das Interesse der Mehrheit der Teilnehmer·innen bereits nach 10–15 Minuten schwindet. Eine fortlaufende Nummerierung bzw. ein Fortschrittsanzeiger kann diesem Effekt begrenzt entgegenwirken (Aeppli u. a. 2016: 170).

Der zeitliche Rahmen wird weiterhin verengt durch organisatorische Elemente, die zum transparenten, ethischen wissenschaftlichen Arbeiten dazugehören. Damit sind neben dem Datenschutz- und Anonymisierungshinweis besonders die Einleitung und der Abschluss (bzw. bei einer mündlichen Umfrage die jeweilige Aufklärung) gemeint. Der verantwortliche Wissenschaftler muss den Teilnehmer·innen gegenüber offenlegen, zu welchem Zweck die Erhebung stattfindet, wer die Verantwortlichen sind und welche Verpflichtungen sich ggf. aus einer Teilnahme ergeben (häufig darf die Teilnahme ohne Konsequenzen und ohne Nennung von Gründen jederzeit unter- oder ganz abgebrochen werden. Außerdem darf die Person auf eine vollständige Löschung ihrer Daten bestehen, Atteslander u. a. 2010: 147).

5.2.2 Eye-Tracking

Unter Eye-Tracking wird die Erfassung der Augenbewegung und Blickposition verstanden. Auch wenn es je nach Untersuchungsgegenstand und Disziplin noch weitere im Einsatz befindliche Systeme gibt, erfolgt die Erfassung heutzutage im Bereich der Translationswissenschaft in den meisten Fällen mittels einer hochauflösenden Kamera, die die Lichtreflexionen der Augen, genau gesagt: der Cornea, einfängt. Diese Kamera wird vor Studienbeginn sowohl auf die Versuchsperson als auch auf das Untersuchungsobjekt (häufig an einem Bildschirm) kalibriert. In diesem Zusammenhang ist die verwendete Hardware zu beachten, da zwischen freistehenden Eye-Tracking-Systemen und fest in einen Bildschirm montierten Geräten zu unterscheiden ist. Hinzu kommen außerdem noch Eye-Tracking-Brillen und Ausführungen, bei denen der Kopf der Versuchsperson auf einem Stativ fixiert wird.

Eine Vielzahl an Forschungsfeldern bedient sich des Eye-Trackings, um Erkenntnisse zu visuellen Reizen (Stimulus) des jeweiligen Gebietes zu gewinnen (Bergstrom & Schall 2014: 3). So kann mittels Eye-Tracking beispielsweise die Nutzerfreundlichkeit von Software getestet, Marketingelemente auf ihre Wirksamkeit bei den Kund·innen untersucht, oder auch das Leseverhalten beim Übersetzen erforscht werden. Dabei ist allerdings zu beachten, dass die Durchfüh-

rung von Eye-Tracking-Studien sowohl in Bezug auf die notwendige Hardware als auch die konzeptionelle Vor- und Nachbereitung weitaus komplexer, kostspieliger und ressourcenaufwändiger ist als Versuchsdesigns mit vergleichbarem Erhebungszweck.

Man unterscheidet beim Eye-Tracking zwischen verschiedenen Typen von Augenbewegungen. Die in diesem Zusammenhang wichtigsten sind die sog. Fixationen und die Sakkaden. Fixationen repräsentieren die Fokussierung des Auges auf einen bestimmten Punkt über einen bestimmten Zeitraum hinweg – wenn auch das Auge an diesem sog. Fixationspunkt nie absolut still steht und es immer wieder zu kaum wahrnehmbaren, zitternden Bewegungen (Tremor), kleinsten Abweichungen vom Fixationspunkt (Drift) und deren Korrektur (Microsakkaden) kommt (Holmqvist 2011: 22).

Die Dauer einer Fixation kann zwischen wenigen Millisekunden und Sekunden liegen (Holmqvist 2011: 377 f.). Bereits in den 1980er Jahren stellten Just & Carpenter die auch als *Eye-Mind-Hypothese* bekannte Annahme auf, dass der Mensch immer dem Bereich die Aufmerksamkeit widmet, der mit den Augen gerade fixiert wird. Dagegen spricht, dass eine vom Eye-Tracker registrierte Fixation allein nicht bedeuten muss, dass eine Versuchsperson an der Stelle der Fixation wirklich einen Reiz wahrgenommen und kognitiv verarbeitet hat (Bergstrom & Schall 2014: 7). Fixationen können sich technisch aus mehreren räumlich eng beisammenliegenden Augenpositionen zusammensetzen, die vom Eye-Tracking-System aggregiert werden. Die einzelnen registrierten Augenbewegungen heißen *gaze points* und werden typischerweise als Koordinaten auf der x- und y-Achse des Bildschirms angegeben (Raschke u. a. 2014: 7)

Die Fixationsdauer ist ein Indikator für die kognitive Arbeitsbelastung Rayner (1998: 376 f.). Je länger eine Fixation dauert, desto anspruchsvoller und intensiver ist die kognitive Auslastung. Im Falle von Lesestudien bedeutet dies, dass z. B. ein Text intensiver gelesen wird. Gerade die Dauer der ersten Fixation (eng. *first fixation duration*) in einem Bereich wird deshalb häufig verwendet, um die unmittelbare Informationsverarbeitung zu untersuchen (Holmqvist 2011: 384). Die Anzahl an Fixationen hingegen ist ein rein quantitativer Indikator und erlaubt Rückschlüsse auf die Lesetiefe (eng. *reading depth*, Holmqvist 2011: 390).

Weiterhin benötigen die Proband·innen bei Lesestudien mehr Zeit für die Verarbeitung schwieriger Wörter, weil diese entweder angesichts des gegebenen Kontexts unvorhergesehen auftauchen oder eine niedrige Frequenz innerhalb der Sprache haben. Damit vergleichbar deutet eine höhere Fixationsdauer in Lesestudien zu Untertiteln in der Fremdsprache auf erhöhte kognitive Aktivität hin (Bisson u. a. 2014: 412). Deshalb sollte die Betrachtung der Regressionen (s. u.) im-

5 Empirischer Teil

mer in Kombination mit der Dauer der Augenbewegungen erfolgen, um kognitive und linguistische Prozesse zu analysieren (Eskenazi & Folk 2017: 2).

Bei der durchschnittlichen Gesamtzeit des ersten Durchlaufs pro AOI (eng. *first run dwell time*) zeichnet sich ein anders Bild ab. Auf Wortebene wird diese Durchlaufdauer als Indikator für die auf der Dauer der ersten Fixation aufbauende Verarbeitung angesehen. Je länger die Durchlaufdauer ist, desto länger, seltener und somit schwieriger zu verarbeiten ist das betrachtete Wort. Auf einen Chatbeitrag übertragen, kann die Durchlaufzeit als Indikator für die Verarbeitung des gesamten Beitrags angesehen werden.

Regressionen stellen je nach rezipierter Literatur eine Unterkategorie der Fixationen oder Sakkaden dar (Duchowski 2017, Holmqvist 2011, Iqbal u. a. 2005), die entgegen der gewöhnlichen Leserichtung von einem AOI zu einem vorherigen zurückspringen (bei Deutsch also nach links). Derartige Rücksprünge des Auges zu bereits wahrgenommenen Punkten verweisen auf eine unzureichende oder fehlerhafte Interpretation des Stimulus (Eskenazi & Folk 2017: 1). Dabei wird kategorial zwischen eingehenden (*regressions in*) und ausgehenden Regressionen (*regressions out*) unterschieden. Erstere, also eingehende Regressionen, beschreiben die Augenbewegung zu einem Bereich zurück von einer bereits fixierten Stelle, während letztere, also ausgehende Regressionen, die Augenbewegung aus einem Bereich zurück zu einem anderen darstellt (Schaeffer u. a. 2017: 24). Weiterhin ist zwischen zurückkehrenden Regressionen *return regressions (RR)* und nicht-zurückkehrenden Regressionen *nonreturn regressions (NRR)* zu unterscheiden. RR stellen Regressionen auf einen Bereich zurück dar, der bereits zuvor fixiert worden ist. NRR hingegen bezeichnen Regressionen, die zu einem zuvor übersprungenen Bereich zurückkehren (Eskenazi & Folk 2017: 3).

Die regressive Durchlaufdauer (*regressive path duration*) beschreibt die Zeitspanne, die die Proband:innen erstmalig in einem AOI verbringen, bevor sie ein nachfolgendes AOI höherer Ordnungszahl besuchen. Je kürzer diese Spanne also ist, desto früher wurde das AOI verlassen, um zu einem nachfolgenden AOI zu springen. Die nachfolgend vorgestellten AOI-Kategorien im Rahmen der hier durchgeföhrten Studie sind im Wechsel die Ausgabe der MÜ oder die Originale. Somit kann die regressive Durchlaufdauer als Hinweis auf einen sprunghaften Leseabbruch bzw. im Umkehrschluss auf die komplette Wahrnehmung in einem AOI genommen werden.

Die Pupillengröße erlaubt Rückschlüsse auf die kognitive Arbeitsbelastung der Versuchsperson während der Verarbeitung der Stimuli und wird laut Schaeffer u. a. (2017: 25) bei translationsbezogenen Eye-Tracking-Studien eher selten allein verwendet. Doherty u. a. (2010: 2) begründen dies vor allem mit der Störanfälligkeit bei der Vermessung und Erfassung der Pupillengröße (meist in Millimetern)

als Indikator durch die Lichtverhältnisse während der Studie, Geräuschen, Koffeinkonsum der Studienteilnehmer·innen sowie der Augenfarbe. Dem entgegenzusetzen ist jedoch, dass die Pupillengröße bereits über die vergangenen Jahrzehnte mehrfach als Indikator in Lesestudien verwendet wurde, um Aufschluss über die kognitive Auslastung zu erhalten (vgl. z. B. Hess & Polt 1964, Beatty & Kahneman 1966, Kahneman & Beatty 1966, Beatty 1982, Hyönä u. a. 1995).

Mit Hess & Polt (1964), Beatty & Kahneman (1966), Kahneman & Beatty (1966) und Beatty (1982) wiesen dieselben Autoren bereits in den 1960er bis 1980er Jahren einen Zusammenhang zwischen der gegenwärtigen kognitiven Arbeitsbelastung und der Pupillengröße nach. Die Pupillenerweiterung weist daher eine kurze Latenz nach Beginn der kognitiven Verarbeitung auf und ebbt ebenso schnell wieder ab, nachdem die Verarbeitung beendet ist (Beatty 1982: 276).

Weiterhin beschreiben Krejtz u. a. in ihrem Beitrag den Zusammenhang von Pupillengröße und Microsakkaden zwecks Untersuchung der Benutzerfreundlichkeit und Anwenderorientiertheit (s. hierzu in dieser Arbeit Kap. 3.7, S. 68). Demnach korreliert die Schwierigkeit einer gestellten Aufgabe (sprich: die höhere Auslastung des Arbeitsgedächtnisses) mit der Aktivierungsschwelle zur Ausführung von Fixationen, was sich zugleich wiederum in weniger Microsakkaden widerspiegelt (Krejtz u. a. 2018: 17 ff.).

Sakkaden hingegen sind schnelle Augenbewegungen zwischen zwei Fixationen. Es besteht die starke Vermutung, dass in dieser Zeit keine visuellen Informationen erfasst werden können. Sie sind zugleich die schnellsten Bewegungen überhaupt, zu denen der menschliche Körper imstande ist (Holmqvist 2011: 23). Am Ende einer Sakkade werden häufig sog. *Glissaden* beobachtet, kleine Rejustierungen der Augenbewegungen, da die Sakkade den ursprünglich anvisierten Punkt nicht genau getroffen hat und sich daher nun anpasst. Gerade die Dauer von ersten Fixationen auf einen Bereich kann in Teilen stark durch Glissaden beeinflusst werden (Holmqvist 2011: 165). Dies hängt mit der generellen visuellen Aufmerksamkeit zusammen, bei der zwischen offener Orientierung (*overt orienting*), während der Augenbewegung, und geschlossener Orientierung (*covert orienting*), während des Stillstandes der Augen, unterschieden wird (Le Meur & Liu 2015: 152). Bei Lesestudien findet die Mehrzahl der Sakkaden in Leserichtung statt. Die räumliche Abfolge von Fixationen und Sakkaden über den zeitlichen Verlauf der Erhebung (pro Teilnehmer·in) ergibt den sog. *scan path*, der eine Visualisierung der Augenbewegungen ermöglicht.

5.2.2.1 Dynamische Areas of Interest

Die sog. *Areas of Interest* (AOI) sind ein wesentlicher Bestandteil der Analyse von Eye-Tracking-Daten. Durch sie werden bei der Aufbereitung der Rohdaten jene Bereiche markiert, in denen der präsentierte Stimulus auffällige Reaktionen erwarten lässt. In den meisten Fällen handelt es sich um rechteckige Formen, die ein bestimmtes Areal (auf dem Computerbildschirm) markieren. Im Falle von Lesestudien wird häufig jedem Wort ein eigenes AOI zugewiesen, das zugleich die Wortgrenzen markiert. Daher wird mit Blick auf derartige Studiendesigns empfohlen, eine möglichst große Schrift und nicht mehr als 300 Wörter gleichzeitig auf dem Bildschirm zu präsentieren. Für Studien, bei denen die Teilnehmer·innen sowohl lesen als auch schreiben, wird weiterhin nahegelegt, entweder die Rekrutierung auf Personen zu beschränken, die ohne Blick auf die Tastatur schreiben können, oder bei der Sichtung der Rohdaten auf das Blickverhalten zu achten, um zu gewährleisten, dass ausreichend Datenpunkte innerhalb der AOI gesammelt wurden. Die meisten Eye-Tracking-Systeme registrieren dabei die sog. Blinzler (eng.: *blinks*), also die Zeitpunkte, an denen keine Augenbewegung registriert werden konnte, sei es durch Kontaktverlust der Kamera oder durch Blickbewegungen außerhalb des kalibrierten Bereichs (O'Brien 2009: 253 ff.).

Abhängig von der Gestalt des präsentierten Stimulus sind auch andere Polygone möglich. Für die Feldstudie im Rahmen dieser Arbeit (s. Abschnitt 5.2.5, S. 88) war es zur Aufbereitung der gewonnenen Daten notwendig, die Bildschirmaufnahmen mit sog. *dynamischen AOI* zu versehen. Im Gegensatz zu statischen AOI können diese Bereiche mit dem Stimulus „mitwandern“. Die Größe, Form, Position sowie der Verlauf am Bildschirm müssen manuell festgelegt und vorab geplant werden (Holmqvist 2011: 209). Dies ist mit Blick auf Textchats notwendig, da sich die einzelnen Chatbeiträge zweispaltig und auf der vertikalen Achse zum oberen Bildschirmrand hin verschieben (s. exemplarisch dazu vorausgehende Abbildung 5.1). Dieses Vorgehen ist für die Analyse im Folgenden insofern erschwerlich, als durch die händische Annotation eine kongruente Abdeckung von AOI-Fläche und tatsächlicher Chatnachricht nahezu unmöglich ist. Die einzelnen AOI der Nachrichten sind somit nahezu immer etwas größer als die am Bildschirm angezeigte Nachricht.

Zugleich wurde die Eingabemaske mit einem statischen AOI annotiert. Im Gegensatz zu den einzelnen Chatbeiträgen behält sie ihre Position und Größe auf dem Bildschirm bei, weshalb daher keine dynamischen AOI benötigt werden.

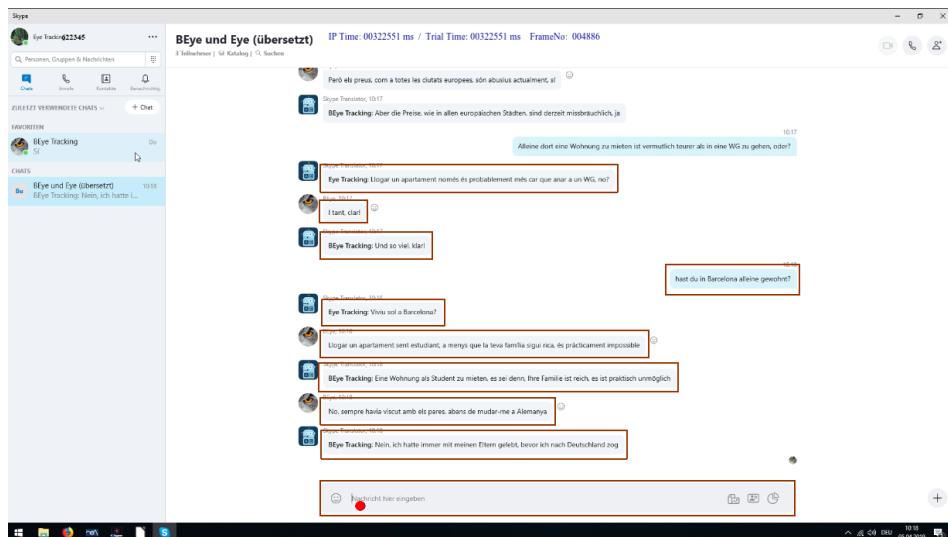


Abbildung 5.1: Areas of Interest der Feldstudie

5.2.2.2 Einsatz von Eye-Tracking zur MÜ-Evaluation

Der Einsatz von Eye-Tracking-Methoden zur Evaluation der MÜ-Qualität ist besonders in Verbindung mit dem Postediting ein kleines, jedoch reges Forschungsfeld. Die MÜ-Ausgabe wird dabei aus unterschiedlichen Blickwinkeln untersucht. Die Wahrnehmung der einzelnen MÜ-Systeme ist ebenso Gegenstand der Untersuchungen wie das Leseverhalten entlang der MÜ-Ausgabe und die Wahrnehmung von Fehlern im rohen Output durch Posteditor·innen. Daher ergeben sich besonders aus diesem Bereich wichtige Anhaltspunkte für die Ausarbeitung der Eye-Tracking-Studie.

Die zweiphasige Studie von Doherty & O'Brien (2012) wird weithin als erste angesehen, die Eye-Tracking zur Untersuchung der Nutzbarkeit von übersetzten Texten durch den/die Endnutzer·in verwendet. Hierzu wurde die Nutzbarkeit der rohen MÜ-Ausgabe auf Deutsch, Französisch, Japanisch und Spanisch mit der des englischsprachigen Originaltextes verglichen. Die Ergebnisse deuten darauf hin, dass wesentliche Unterschiede in puncto Zufriedenheit, Effizienz und Zielführung bestehen, abhängig davon, ob Nutzer·innen das Original oder eine MÜ verwenden. In der zweiten Phase der Studie griffen Doherty & O'Brien (2014) die Ergebnisse auf und untersuchten diese auf den Grad der kognitiven Auslastung bei der Verwendung von Original und MÜ. Die Erkenntnisse dieses Studienteils deuten darauf hin, dass die rohe MÜ kognitiv fordernder ist und als weniger nützlich wahrgenommen wird als der Originaltext.

5 Empirischer Teil

Doherty & O'Brien (2014: 14) befürworten in diesem Zusammenhang explizit den Einsatz von Eye-Tracking-Methoden zur Evaluierung von MÜ-Systemen. Eye-Tracking-Ansätze vermögen es, unbewusste und individuelle kognitive Prozesse der Probanden weitaus deutlicher abzubilden als ähnliche, ebenfalls häufig verwendete Methoden wie Think-Aloud-Protokolle oder stimulerte Post-Studien-Interviews.

Sajjad u. a. befassen sich mit dem Leseverhalten bei der Bewertung von maschinellem Übersetzung. Die Hypothese ist es, dass eine qualitativ gute MÜ-Ausgabe weitaus weniger kognitiven Aufwand erfordert als schlechte. Dabei folgen die Autoren der Annahme, das Leseverhalten von der Komplexität der syntaktischen Strukturen und der Schwierigkeit des Satzes abhängt. Die Studie greift dabei auf Datenmaterial des WMT'16-Evaluation-Task innerhalb des Sprachenpaars Spanisch-Englisch zurück. 60 mittellange Sätze wurden ausgewählt und von mindestens zwei Annotatoren bewertet. Die jeweils beste und schlechteste Übersetzung wurde mit einem Score (*expected wins*) versehen. So entstanden 120 Evaluationsaufgaben, jeweils 60 gute und 60 schlechte Übersetzungen zu überprüfen. Diese wurden von jeweils sechs Revisor·innen bewertet (Sajjad u. a. 2016: 1084). Zur Untersuchung wählten Sajjad u. a. die Indikatoren *Verweildauer*, *Sakkadenanzahl*, jeweils als *Progressionen* sowie *Regressionen*, und die Sakkadenrichtung (Sajjad u. a. 2016: 1083). Sajjad u. a. kommen zu dem Ergebnis, dass sich eindeutige Blickmuster in der Art erkennen lassen, wie die Revisor·innen bei der Bewertung der MÜ-Ausgabe vorgehen.

Die Studie von Klerke u. a. (2015) untersucht unter Einsatz von Eye-Tracking-Methoden das Problemlösungsverhalten von Proband·innen, die Logiktests in der Ausgangssprache, als MÜ-Version, in einfacher Sprache oder als MÜ-Version in vereinfachter Sprache bearbeiten sollen. Dabei wird einerseits eine höhere Anzahl an Fixationen sowie eine längere Verweildauer auf der maschinell übersetzten Version beobachtet als auf dem Original. Andererseits lösen die Proband·innen die MÜ-Aufgabe weniger effizient als das Original. Die vereinfachte und maschinell übersetzte Version jedoch scheint diesen Effekt aufzuheben.

Castilho & Guerberof Arenas (2018) untersuchen in einer Lesestudie unter Einsatz des Eye-Trackers die qualitative Wahrnehmung der Ausgabe verschiedener MÜ-Systeme. Hierzu dient weiterhin die Bewertung der Zufriedenheit mit dem gelesenen Text und ein Fragebogen zum Leseverständnis. Die Pilotstudie arbeitet mit insgesamt sechs Proband·innen mit Englisch, Spanisch oder Chinesisch als Muttersprache. Die Versuchspersonen sollen die englischsprachigen Ausgangstexte oder die Ausgabe auf Spanisch oder Chinesisch jeweils eines SMÜ- bzw. eines NMÜ-Systems in ihrer Muttersprache lesen. Castilho & Guerberof Arenas

kommen zu dem Schluss, dass die NMÜ-Ausgabe zwar zufriedenstellendere Ergebnisse als die der SMÜ für die Versuchspersonen liefert. Die kognitive Auslastung bei der Erfassung ist jedoch für beide Systeme ähnlich.

Vardaro u. a. (2019) wiederum untersuchen sowohl mit Eye-Tracking-Methoden als auch auf Grundlage einer Korpusanalyse, wie professionelle Übersetzungsexpert·innen¹ der Generaldirektion Übersetzung des Europäischen Parlaments Fehler in der NMÜ sowie der entsprechenden posteditierten Version identifizieren und korrigieren. Die 30 Proband·innen weisen dabei eine hohe fremdsprachliche Kompetenz mit bis zu sechs verschiedenen Ausgangssprachen auf (Vardaro u. a. 2019: 10). Zur Identifikation und Korrektur der MÜ-Fehler werden diese in zwei grundlegende Klassen eingeteilt. Unter den Lesefluss betreffende Aspekte (*fluency*) werden Ortographie-, Grammatik-, Konsistenz- sowie Kohärenzfehler gezählt. Die Angemessenheit (*accuracy*) wird durch Sinn- und Auslassungsfehler gekennzeichnet. Die Fehler in der maschinellen Übersetzung wurden auf Grundlage des automatischen Fehlerannotationswerkzeugs *Hjerson* sowie der manuellen *MQM* annotiert. Die Eye-Tracking-Studie erhebt die Dauer der ersten Fixation, die Dauer des ersten Durchlaufs, die regressive Durchlaufdauer sowie die Verweildauer auf dem jeweiligen Ausgangstoken und auf dem Zieltoken. Die zentrale Erkenntnis der Studie ist, dass die sprachlich umfassend versierten Versuchspersonen jegliche MÜ-Fehler gleichermaßen schnell identifizieren und keiner Fehlerkategorie Priorität einräumen. Weiterhin scheinen die Versuchspersonen ihre Lese- bzw. Korrekturstrategien den Erfordernissen der jeweiligen Situation anzupassen.

5.2.3 Konzept der Online-Umfrage

Mit der Umfrage sollen zunächst aktuelle Daten zum Nutzerkreis, zum Nutzungsverhalten und – mit Fokus auf dem Skype Translator – zum Kenntnisstand im Umgang mit dieser Technologie erhoben werden. Diese aktuellen Daten sollen eine Zuarbeit für die anschließende Feldstudie mit dem Skype Translator liefern, die in 5.2.5 (S. 88) genau beschrieben wird. Eine Interviewsituation, sowohl strukturiert als auch semi-strukturiert, wurde als Möglichkeit der Erhebung verworfen, da der zeitliche Aufwand in keiner Relation zu der benötigten Teilnehmerzahl und den erwarteten Antworten steht und somit schlichtweg kaum durchführbar ist.

Mit Blick auf den intendierten Proband·innenkreis der Feldstudie um den Skype Translator wurde als Zielgruppe für die Umfrage das gleiche Umfeld ausge-

¹Eng.: *translation expert* – In der Studie verwendete Sammelbezeichnung für eine Tätigkeit, die sowohl Übersetzung, Postediting als auch Revision umfasst.

5 Empirischer Teil

wählt: Student-innen jeglicher Fachrichtung an deutschen Universitäten. Um eine möglichst weitläufige Stichprobe zu ziehen, wurde der Fragebogen als Online-Version über Verteilerlisten an Universitäten in Deutschland versendet und als Link in entsprechenden Hochschulgruppen² innerhalb von Facebook verbreitet. Aeppli u. a. sprechen in diesem Fall von einer Klumpenstichprobe, die auf „natürlich vorkommende Gruppen“ (Aeppli u. a. 2016: 174) abzielt. Das einzige einschränkende Kriterium ist die Verwendung von Skype, wobei auch die Verwendung des Internets zum Zwecke der Diffusion des Fragebogens als Zugangsbarriere angesehen werden kann (Aeppli u. a. 2016: 174). Allerdings kann dieses Manko insofern vernachlässigt werden, als für die Nutzung von Skype ohnehin ein Internetzugang bestehen muss.

Ein generelles Problem bei Online-Umfragen stellt jedoch die Zugangsbegrenzung bzw. die zielgruppengenaue Veröffentlichung dar. Dies wird sowohl mit Blick auf die Altersangaben der Teilnehmer-innen, den höchsten Abschluss als auch die Auslandserfahrung deutlich. Die genauen Angaben zu diesen Fragen sind im Abschnitt 6.1.1 (S. 97) detailliert offengelegt. Hierbei ist allerdings zu erwähnen, dass auch Institutionen, die keine Rückmeldung auf die Anfrage zur Verbreitung des Fragebogens gegeben haben, dennoch den Aufruf über entsprechende Verteilerlisten veröffentlicht haben. Diese – zugegebenermaßen vage – Annahme basiert auf dem Teilnahmezähler des verwendeten Online-Tools.

Der Fragebogen für diese Arbeit konzipierte Fragebogen umfasst insgesamt 38 Items. Die Erhebung findet dabei sowohl in Form von Ja-Nein-Fragen (z. B. S. 239), offenen Fragen mit Freitextfeldern als auch geschlossenen Fragen mit Mehrfachantworten (z. B. S. 242) oder Skalen (z. B. S. 241) statt (Atteslander u. a. 2010: 138 f.). Die Skalen umfassen eine Abstufung von 1–4, was sprachlich mit Oppositionspaaren unterstützt wird. Es wurde bewusst eine geringe Spanne ohne Mittelwert gewählt, um so wenig aussagekräftige, unentschlossene Antworten zu vermeiden und die Teilnehmer-innen zu einer positionierten Aussage zu bewegen. Weiterhin sind die Items des Fragebogens in sechs Unterthemen aufgeteilt: *Nutzung von Skype* (S. 240), *Nutzungsbewertung* (S. 241), *Nutzungsbewertung Skype Translator* (S. 241), *Alternativen* (S. 242), *Auslandserfahrung – gegenwärtig & vergangen* (S. 245) und *individuelle Angaben* (S. 247). Alle Fragetypen sind so eingestellt, dass dem Teilnehmer-innen auch die Möglichkeit geboten wurde, sich entweder komplett der Beantwortung zu verweigern oder zumindest keine Angaben zu machen. Ebenfalls war es möglich, die Eingangsfrage *Nutzen Sie Skype?*

²Gerade zu Beginn des jeweiligen akademischen Jahres, also dann, wenn neue Erstsemester an die Universitäten drängen, sind die umgangssprachlich als *Ersti-Gruppen* bezeichneten Gemeinschaften auf Facebook ein guter Anlaufpunkt für die Verbreitung von derartigen Umfragen. Dies hängt jedoch auch in großen Teilen vom Wohlwollen der Administratoren ab.

mit *Nein* zu beantworten. In diesem Falle wurde die Online-Umfrage so konditioniert, alle weiteren Fragen, die in direktem Zusammenhang mit Skype stehen, zu überspringen und sogleich zu den weiteren Aspekten zu leiten.

Im Umfrageteil zur generellen Nutzung von Skype werden konkret Nutzungs-dauer, -länge, -modus, sowie die Skypeversion und die Art der Gesprächspartner-innen erhoben. Neben einem Freitextfeld zur Angabe weiterer Anwendungen führt der Umfrageteil zur Nutzung von Alternativen folgende Dienste auf: *Appear.in, Facebook Messenger, Google Hangouts, Apple Facetime, WhatsApp, Telegram, Viber, Twitter, Snapchat, Instagram-Chat* und *ICQ*. Die Auswahl dieser Anwendungen erfolgte dabei aufgrund des Bekanntheitsgrades und wurde großzügig weit gefasst, um Umfrageteilnehmer-innen nicht durch eine zu begrenzte Auswahl auf *Skype* zu fokussieren. Zusätzlich zur generellen Nutzung der angebotenen Alternativen wurde ebenfalls nach der am häufigsten verwendeten Software – unterschieden nach Modus Video-, Voice- und Textchat – gefragt.

Im Bereich der Auslandserfahrung wurde zwischen gegenwärtigen und vergangenen Reisen unterschieden. Als Auslandserfahrung wurden dabei Aufenthalte definiert, die einerseits keinen Urlaub darstellten und andererseits länger als vier Wochen dauerten. Somit sollte sichergestellt werden, dass die Umfrage teilnehmer-innen eindeutig einen Landes- und Sprachkontakt nachweisen konnten. Die Unterscheidung zwischen gegenwärtig und vergangen wurde vorgenommen, da aufgrund der intendierten Zielgruppe nicht ausgeschlossen werden konnte, ob Studierende den Aufruf zur Teilnahme im Laufe ihres Auslandsaufenthaltes erhielten. Neben der Frage nach dem Grund der Reise – *Spracherwerb, Sprachvertiefung, Kulturkontakt, Pflichtaufenthalt, Fortbildung, beruflich und sonstiges* – wurde das Land, der (schwerpunktmaßige) Hauptaufenthaltsort sowie die Dauer in Monaten erhoben.

Im Rahmen der demographischen Angaben wurden das Geschlecht, das Alter und der höchste Abschluss (*kein Abschluss, bis mittlere Reife, Hochschulzugangsberechtigung, abgeschlossene Ausbildung, höherer Berufsabschluss, Hochschulabschluss sowie sonstiges*) erhoben. Auch die Sprachkenntnisse der Teilnehmer-innen wurde abgefragt. Dies geschah aufgeteilt nach Sprachen, mit denen man primär aufgewachsen war, die man hauptsächlich im Alltag und hauptsächlich im Beruf verwendete. Eine Selbsteinschätzung erhob abschließend die Kompetenz gemäß des Europäischen Referenzrahmens für die Sprachen Spanisch, Katalanisch, Galicisch, Portugiesisch, Italienisch, Französisch und Rumänisch, wobei die Teilnehmer-innen nur Angaben machen sollten, sofern sie eine nachweisliche Qualifikation (Zeugnis, Zertifikat o. ä.) hierrüber besaßen.

5.2.4 Datenschutz und Anonymität

Die Verbreitung und Bearbeitung dieses Fragebogens erfolgen ausschließlich online. Daher enthält die hier abgedruckte Fassung zwar Fragen zur Person, jedoch keine Elemente wie sie bei einer manuellen Aufbereitung notwendig sind, so z. B. die anonymisierten Teilnahmecodes oder fortlaufende Ziffern zur Systematisierung. Die Anonymisierung erfolgt durch das verwendete Online-Umfragetool der Universität Leipzig auf Basis von *LimeSurvey*³. Unabhängig von der eigens ausgearbeiteten Eingangsbelehrung weist auch das Onlinewerkzeug auf geltende Datenschutzrichtlinien hin (s. Anhang A, S. 237).

Zusätzlich wurde ein individueller Hinweis auf die Teilnahmebedingungen der Umfrage ausgearbeitet. Dieser wird, ebenso wie die Angaben zum Untersuchungsvorhaben und eine Anleitung, vor Beginn der Online-Umfrage eingebettet (s. Anhang A, S. 237).

5.2.5 Konzept der Feldstudie

Die Feldstudie fand in den Räumlichkeiten des IALT statt. O'Brien (2009: 263) empfiehlt für optimale Versuchsbedingungen die Nutzung eines Eye-Tracking-Labors, in dem die Lichtverhältnisse kontrolliert werden können. Da es am IALT zum Zeitpunkt dieser Arbeit noch kein solches Labor gab, wurde ein gewöhnlicher Büroraum genutzt. Um dennoch die Lichtverhältnisse kontrollieren zu können, wurde jede Sitzung mit heruntergelassenen Sonnenblenden und bei eingeschalteter Deckenlampe durchgeführt.

Der Eye-Tracker, der dabei verwendet wurde, war der *Eye Link Portable Duo* von SR Research. Der gesamte Aufbau besteht aus einem *Display-PC*, einem *Host-PC* und der Eye-Tracking-Kamera. Erster enthält eine gewöhnliche Windows-Distribution, Skype und die Software zur Bildschirmaufnahme. Über ein LAN-Kabel ist der Display-PC mit dem Host-PC verbunden, von dem aus der Eye-Tracker gesteuert werden kann. Neben Tastatur und Maus wurden keine weiteren peripheren Eingabegeräte verwendet. Für die Durchführung wurde der Eye-Tracker im freien Modus (sog. *head-free-to-move mode*) verwendet, bei dem die Kamera zwischen Bildschirm und Tastatur auf einem Stativ platziert wird. Bei dieser Einstellung ist es nicht nötig, den Kopf der Proband·innen zu fixieren. Ein Sticker in Form einer kleinen Zielscheibe auf der Stirn der Person reicht dem System aus, um die Augen zu triangulieren. Eine Kalibrierung des Systems auf die jeweilige Person ist jedoch weiterhin erforderlich. Als Bildwiederholungsrate wurde mit 1.000 Hz der höchstmögliche Wert im freien Modus gewählt. Dabei zeich-

³Verwendet in der Version 3.22.27+200720.

nete das System die Bewegungen beider Augen auf (*binocular recording*), nachdem eine 9-Punkt-Kalibrierung durchgeführt wurde. Die Abbildung 5.2 (S. 89) skizziert den Versuchsaufbau aus der Vogelperspektive. Der Versuchsleiter und die jeweilige Versuchsperson saßen sich gegenüber. Das gesamte Eye-Tracking-Equipment befand sich ebenfalls auf dem Bürotisch zwischen beiden Personen. Der Versuchsleiter konnte nur die Ausgabe des Host-PCs sehen. Die Versuchsperson hatte wiederum nur Sicht auf den Display-PC.

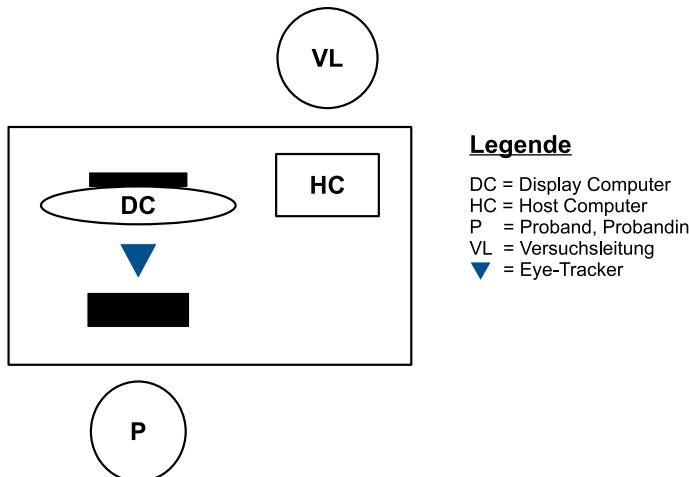


Abbildung 5.2: Versuchsaufbau aus der Vogelperspektive

Balling u. a. (2014: 240) weisen darauf hin, dass die gegebene Bewegungsfreiheit der Proband·innen im gewählten freien Aufnahmemodus zu Präzisionsverlusten bei der Erfassung der Augenbewegung führen kann. Eine mögliche Lösung wäre die Nutzung einer Kinnstütze, die jedoch den Bewegungsradius der Versuchsperson einschränkt. Im Rahmen der naturalistisch orientierten Translationsprozessforschung ist von diesem Vorgehen daher abzuraten.

Für die Studie wurde mit der jeweils aktuellsten Skype-Version für Windows 10 gearbeitet. Die Bildschirmaufnahmen und die Datenaufbereitung wurde mit proprietärer Software von SR Research durchgeführt. Die Mitschnitte der einzelnen Sitzungen liegen im MP4-Format vor. Außerdem registrierte das System neben den Augenbewegungen auch alle Eingabesignale der Peripheriegeräte. Diese *Key-Logs* können für weitere Untersuchungen an den Daten verwendet werden.

Die proprietäre Software *SR Research Data Viewer*⁴ ermöglichte die Aufbereitung der gesammelten Rohdaten. Ohne vorherige Bearbeitung liefert die Soft-

⁴Verwendet in der Version 4.1.1.

5 Empirischer Teil

ware bereits sog. *Heatmaps* und Screenshots zu den Fixationen, Sakkaden und Blinzlern. Eine kombinierte Ansicht aus den Events und allen Eingabesignalen ist ebenfalls möglich. Darüber hinaus liefert die Betrachtung der Rohdaten Angaben zu den grundlegenden Einstellungen der Aufnahme, wie etwa Bildschirmgröße in Pixeln, Sitzungsdauer, verwendeter Modus des Eye-Trackers usw.

Es ist weiterhin möglich, statische Stimuli automatisch mit AOI versehen zu lassen. Dies ist bei der Aufbereitung von Texten oder Bildern hilfreich. Bei Bildschirmmitschnitten mit häufigen Positionswechseln der AOI, wie etwa bei der Feldstudie dieser Arbeit (s. hierzu die Abbildungen 3.3, S. 59 und 5.1, S. 83), mussten die Rohdaten zunächst manuell annotiert werden. Dabei wurde mit fünf Kategorien an statischen und dynamischen AOI (s. hierzu S. 82) gearbeitet: Die Eingabemaske des Skype Translators wurde als statisches AOI mit dem Label *Eingabe* versehen. In den Rohdaten wurden originale, deutschsprachige Chatbeiträge der/des Proband·in zunächst mit dem fortlaufend nummerierten Label *AEyeO-#* erfasst, die entsprechende katalanische maschinelle Übersetzung mit *AEyeMT-#*. Die Beiträge des katalanischsprachigen Gegenübers erhielten respektive hierzu die Marken *BEyeO-#* und *BEyeMT-#*. Um für die Analyse eine eindeutige Grundlage mit Blick auf Übersetzungsrichtung und Sprache zu schaffen, wurden die Labels im Zuge der Auswertung gemäß der Länderkürzel der ISO-639-1⁵ abgeändert. Dementsprechend finden sich in dieser Arbeit nun die Labels

GERO für den originalen, deutschsprachigen Beitrag der Versuchsperson,

CATMT für die maschinelle Übersetzung ins Katalanische,

CATO für den originalen, katalanischsprachigen Beitrag des Gegenübers und

GERMT für die maschinelle Übersetzung ins Deutsche

für die dynamischen AOI im Setting Katalanisch-Deutsch und

A für die originalen, deutschsprachigen Beiträge der Versuchsperson,

B für die originalen, deutschsprachigen Beiträge des Gegenübers

für die dynamischen AOI im Setting Deutsch-Deutsch.

Bei der Annotation wurden die Chatbeiträge dabei bis zu einer Höhe von etwa zwei Dritteln der Bildschirmanzeige mit dynamischen AOI versehen, da die

⁵S. hierzu <https://www.bib-bvb.de/web/kkb-online/rda-sprachencode-nach-iso-639>, Abruf am 31. Januar 2022.

Mehrheit der Eye-Tracking-Events sich hauptsächlich in diesem Bereich verteilte. Dies wird in der Darstellung als Heatmap deutlich, wie Abbildung 5.3 (S. 91) zeigt. Als weiteren Beleg für die Annotation bis zu einer Höhe von zwei Dritteln des Bildschirms dient die graphische Darstellung der Fixationen, Sakkaden und Blinzler jeweils für beide Studienversionen in den Abschnitten 6.3.1 (S. 122) und 6.5.1 (S. 170). Gemessen an den Hinweisen von O'Brien (s. in dieser Arbeit Kap. 5.2.2.1, S. 82), kann die vorliegende Studie keine Analyse auf Wortebene liefern, da die Auflösung des Bildschirms und die Standard-Schriftgröße des Chats bei Skype zu klein für eine sinnvolle Annotation von AOI entlang der Wortgrenzen ist.

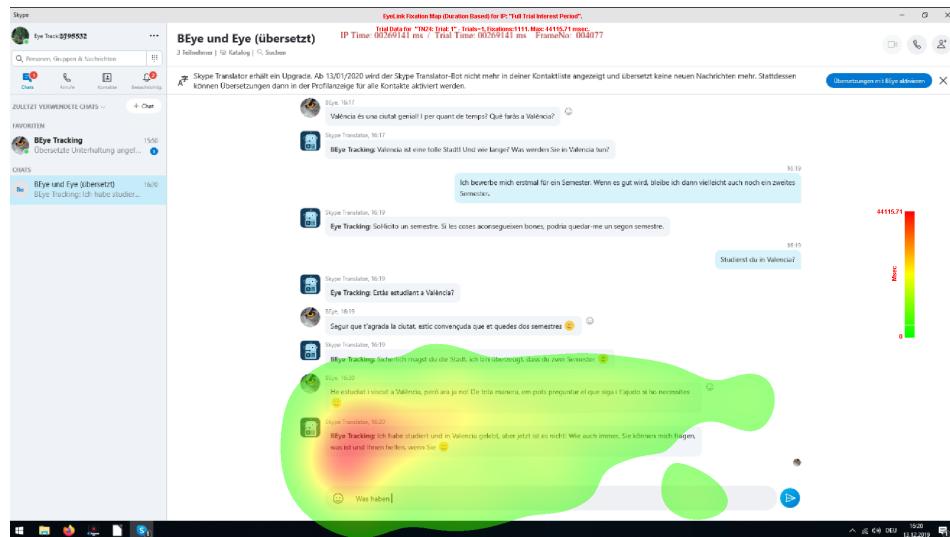


Abbildung 5.3: Heatmap einer Bildschirmaufnahme

Da mit zwei Versuchsgruppen und sich leicht voneinander unterscheidenden Abläufen gearbeitet wurde, werden die grundlegenden Informationen zu den Kohorten im jeweiligen Abschnitt – für das Setting Katalanisch-Deutsch ab S. 108 und für das Setting Deutsch-Deutsch ab S. 169 – dargelegt.

5.3 Die Feldstudie

Die Feldstudie wurde in zwei Settings mit sich nicht überschneidenden Teilnehmer.innengruppen und einem mehrteiligen Aufbau durchgeführt. Eine Variante sah den Textchat von deutschen Muttersprachlern mit katalanischen Mutter-

5 Empirischer Teil

sprachlern über Skype und bei aktiviertem Skype Translator vor, wobei die deutsche Seite mit einem Eye-Tracking-System erfasst und ein Videomitschnitt der Bildschirmaktivität (eng.: *screen recording*) erzeugt wurde.

Diese Gesprächssituation war flankiert von einem digitalen Eingangs- und einem Ausgangsfragebogen, die die Versuchspersonen zu beantworten hatten, und deren Ergebnisse im weiteren Verlauf dieser Arbeit vorgestellt werden. Der Ablauf dieser Studienvariante kann daher wie folgt skizziert werden: Die Proband·innen wurden zunächst gebeten, den Eingangsfragebogen (s. Abschnitt 6.2.1, S. 108, und Anhang B, S. 251) zum Kenntnisstand im Umgang mit Skype und zur eigenen Auslandserfahrung auszufüllen. Dann folgte der aufgezeichnete Textchat über Skype bei aktiviertem Skype Translator. Danach erhielten die Proband·innen den Ausgangsfragebogen (s. Anhang C, S. 263), der Angaben zur Nutzungserfahrung erheben sollte. Die Ergebnisse dieses Ausgangsfragebogens werden ab Kapitel 6.2.3, Seite 111 betrachtet.

Die zweite Studienvariante sah Textchats von deutschen Muttersprachlern untereinander über Skype vor. Auch hier kamen das Eye-Tracking-System sowie die Bildschirmaufzeichnung zum Einsatz. Die Proband·innen erhielten bei dieser Variante jedoch nur den Eingangsfragebogen (s. Anhang B, S. 251). Dieser Aufbau diente dem Vergleich. Ein monolinguale Durchführung ohne den zentralen Stimulus, den die maschinelle Übersetzung des Skype Translators liefert, sollte daher Aufschluss über das Nutzer·innenverhalten in einem reinen Chat-Gespräch geben.

5.3.1 Abhängige Variablen der statistischen Analyse

Schaeffer u. a. (2017: 25) weisen darauf hin, dass man die Variablen einer kognitionswissenschaftlichen Studie auf zwei Arten kontrollieren kann: Auf natürlichem Wege kann zu Beginn der Studie festgelegt werden, welche Faktoren eine Auswirkung auf die Untersuchungssituation haben. Hierbei muss jedoch von vornherein sicher sein, dass diese Faktoren die Situation zweifelsfrei beeinflussen und welche Auswirkungen zu erwarten sind. Da dies bei Studien mit mehreren Variablen häufig nur kaum bis überhaupt nicht möglich ist, bauen viele Analysen auf dem sog. *linear mixed-effects regression model* auf, einer statistischen Betrachtung der unterschiedlichen Faktoren und ihrer Auswirkung auf eine abhängige Variable. Gegenüber anderen statistischen Methoden wie der Varianzanalyse (ANOVA) besitzen derartige Regressionsmodelle die Eigenschaft, flexibel auf unausgewogene Daten zu reagieren. So können beispielsweise die individuellen Unterschiede zwischen Proband·innen als Zufallsvariablen kontrolliert werden, die im

Rahmen der natürlichen Konzeption nur schwer zu beherrschen sind. Gleichzeitig können fixierte Variablen wie das Geschlecht in das Modell integriert werden. Gerade im Rahmen einer naturalistisch orientierten Translationsprozessforschung wird diese statistische Methode häufig verwendet (vgl. z. B. Schaeffer u. a. 2017, Hvelplund 2017). Eine Grundvoraussetzung dafür ist jedoch die Annahme der Normalverteilung der Daten. Sollte diese Annahme nicht möglich sein, können sog. nicht-parametrische Tests verwendet werden. Diese stellen ebenfalls gewisse Anforderungen an die Beschaffenheit der Datengrundlage, können aber auch dann eingesetzt werden, wenn die Daten nicht normal verteilt sind. Bei der Analyse in Kapitel 6 wird deshalb jede verwendete Variable auf ihre Verteilung hin überprüft und das statistische Testverfahren entsprechend gewählt. Um die Bedeutsamkeit des Testergebnisses zu beurteilen, wird bei signifikanten Unterschieden zudem die Effektstärke angegeben. Diese verdeutlicht die praktische Relevanz des Testergebnisses.

Die Unterscheidung zwischen unabhängigen und abhängigen Variablen stammt ursprünglich aus der Psychologie, wird jedoch auch bei der Operationalisierung von Eye-Tracking-Variablen verwendet (Raschke u. a. 2014: 6). Unabhängige Variablen stellen dabei den veränderbaren Teil eines Experiments bzw. einer Hypothese dar, da nur sie vom Forscher beeinflusst und verändert werden können. Abhängige Variablen hingegen zeigen den Effekt, den die Manipulation der unabhängigen Variablen haben. Die Darstellung des Vorgehens zur Messung dieser Effekte inkl. Begriffsdefinition und Angabe der Messverfahren ist die Operationalisierung (Albert & Marx 2014: 81, Aeppli u. a. 2016: 125 ff.). Soweit möglich und sinnvoll, werden im Folgenden die Variablen mit einer deutschen Bezeichnung aufgeführt. Eine Übernahme der geläufigen englischen Bezeichnungen hätte den Lesefluss gestört.

Duchowski (2017: 218) kategorisiert folgende Indikatoren als traditionelle Eye-Tracking-Maße:

- Fixation inkl. Regressionen
- Dauer der Fixation (Fixation Duration)
- Anzahl an Fixationen (global)
- Fixationsrate (global)
- durchschnittliche Fixationsdauer (global)
- Blickverlauf (Fixationssequenz)

5 Empirischer Teil

- Fixierte Anzahl an Area of Interest (AOI)
- Prozentualer Anteil der Fixation pro AOI
- Anzahl an Fixationen pro AOI (Fixation Count)
- durchschnittliche Verweildauer inkl. Sakkaden und Fixationen pro AOI (Dwell Time)

Neben diesen Indikatoren werden in Lesestudien ferner Sakkaden und die Pupillengröße zur Untersuchung herangezogen. Allerdings ist sich die Forschung bei diesen beiden Variablen nicht einig, wie präzise sie die kognitive Auslastung abbilden und wie zuverlässig mögliche Aussagen sind. Aus der Gesamtheit an Maßen werden für die vorliegende Arbeit einzelne Indikatoren ausgewählt und im Kapitel 6.2 ab Seite 105 eingehend untersucht. Die Auswahl wurde ausgehend von den in Abschnitt 5.2.2.2 (S. 83) kurz umrissenen Studien getroffen, auch wenn darüberhinaus noch weitere Maße (s. o., s. u.) miteinbezogen werden. Die Definition und Abkürzung ist dabei der Dokumentation der proprietären Analysesoftware von SR Research Ltd. (2019) entlehnt:

Fixationsanzahl (FixC): Anzahl aller Fixationen in einem AOI in absoluten Zahlen

Dauer der ersten Fixation (IAFFD) in einem AOI

Regressionen (Reg): Rückkehr in ein AOI, wenn zuvor bereits ein AOI höherer Ordnungszahl betreten wurde. Die Variable ist nominalskaliert und besitzt lediglich zwei Zustände: ja (1) und nein (0).

Eingangsregression (RegIn): Rückfall in ein AOI

Ausgangsregression (RegOut): Rückfall aus einem AOI

Verweildauer im ersten Durchlauf (IAFRD): Dauer des ersten Durchlaufs in Millisekunden, in Form der Summe aller Fixationen ohne Sakkaden, ab Betreten des AOI bis zum Verlassen hin zu einem AOI höherer Ordnungszahl

Verweildauer (Dwell): Aufsummierte Dauer aller Fixationen ungeachtet der Sakkaden in einem AOI in Millisekunden

Regressive Durchlaufdauer (IARegPD): Aufsummierte Dauer in Millisekunden aller Fixationen innerhalb eines AOI ab dem erstmaligen Betreten bis zum erstmaligen Verlassen hin zu einem AOI höherer Ordnungszahl. Wurden in dieser Zwischenzeit AOI mit niedrigerer Ordnungszahl fixiert, wird auch diese Dauer hinzugezählt.

Selektive regressive Durchlaufdauer (IASelRegPD): Aufsummierte Dauer in Millisekunden aller Fixationen innerhalb eines AOI ab dem erstmaligen Betreten bis zum erstmaligen Verlassen hin zu einem AOI höherer Ordnungszahl, ohne die Dauer der Fixationen, die in dieser Zwischenzeit in AOI mit niedrigerer Ordnungszahl verbracht wurde

Sakkadenamplitude (SacAmp): Größe der gegenwärtigen Sakkade, angegeben in Grad des sichtbaren Winkels (*degrees of visual angle*)

Sakkadendauer (SacDur): Dauer der gegenwärtigen Sakkade, angegeben in Millisekunden

Pupillengröße (PSize): Pupillengröße in willkürlichen Einheiten (*arbitrary units*)

5.3.2 Unabhängige Variablen der statistischen Analyse

Folgende unabhängige Variablen werden in die anschließende Analyse miteinbezogen:

AOI-Kategorie: Alle Chatbeiträge sowie die Eingabemaske wurden mit AOI versehen. Die Gesamtheit aller Beiträge mit dem selben Label wird als AOI-Kategorie gesehen und stellt eine nominalskalierte Variable für die statistischen Tests dar.

Progressive erste Fixation (IAFFixPro): Gibt in Form von 0 (nein) oder 1 (ja) an, ob es sich um eine tatsächliche erstmalige Fixation (*first-pass fixation*) handelt. Eine solche Fixation ist als erstmalige Fixation in einem AOI zu verstehen, ohne dass vorher AOI mit höherer Ordnungszahl betreten wurden. Diese Variable ist ebenfalls nominalskaliert.

Größe des AOI (IASize): Gibt die Größe des AOI in Pixeln an. Diese Variable ist intervallskaliert. Werte kleiner als Null sind nicht möglich, im positiven Raum sind – zumindest theoretisch – unendlich große Werte denkbar. Die meisten Chat-Systeme besitzen jedoch über eine Obergrenze an Zeichen pro einzelner Nachricht, was somit auch die Größe des AOI begrenzt.

6 Analyse

Dieses Kapitel ist der Analyse der einzelnen Studienelemente gewidmet. Zu Anfang werden die Ergebnisse der Online-Umfrage zum Nutzungsverhalten von Skype dargestellt und situiert (Abschnitt 6.1). Danach werden die Angaben im Ein- und Ausgangsfragebogen der Proband·innen im katalanisch-deutschen Versuchsaufbau betrachtet (Abschnitt 6.2, S. 105), bevor sich die Analyse den Eye-Tracking-Daten dieses Versuchsteils zuwendet (Abschnitt 6.3, S. 121). Zuletzt werden in gleicher Abfolge die Angaben und Daten aus dem Setting Deutsch-Deutsch betrachtet (Abschnitte 6.4, S. 166 und 6.5, S. 170).

6.1 Online-Umfrage zur allgemeinen Nutzung von Skype

Die Umfrage zum Nutzungsverhalten von Skype wurde im Zeitraum von Mai bis Juli 2018 durchgeführt. In diesem Zeitraum nahmen 329 Personen teil, von denen 290 den Fragebogen komplett ausfüllten. Verglichen mit den Angaben des Statistischen Bundesamtes, nach denen im Zeitraum des Wintersemesters des betreffenden Zeitraums (2017/18) 2.847.000 Student·innen an deutschen Hochschulen eingeschrieben waren, stellt diese Erhebung keine repräsentative Stichprobe dar¹. Die folgenden Angaben stützen sich außerdem – sofern nicht anders angegeben – ausschließlich auf diese 290 vollständigen Datensätze.

6.1.1 Ergebnisse der Umfrage

Von den Teilnehmer·innen gaben 222 an, weiblich und 64 männlich zu sein. 4 enthielten sich der Angabe. Das entspricht einer Geschlechterverteilung von 77 % zu 22 % zu 1%.²

¹Vgl. https://www.destatis.de/DE/Presse/Pressemitteilungen/2017/11/PD17_427_213.html, letzter Abruf am 31. Januar 2022.

²Die Abfrage des Geschlechts erfolgte zu einem Zeitpunkt, als es in Deutschland noch nicht absehbar war, dass ein sog. *drittes Geschlecht* mit der Bezeichnung *divers* eingeführt würde. Daher konnten die teilnehmenden Personen nur zwischen *männlich*, *weiblich* und *sonstiges* wählen, oder sich der Antwort enthalten (s. a. Anhang A, Frage 36).

6 Analyse

Teilt man die Datensätze nach Alter in Dekaden ein, so erhält man eine Verteilung wie in der Alterspyramide in Abbildung 6.1 dargestellt. Der Altersdurchschnitt liegt bei 25,72, der Median bei 24,00 Jahren. Damit liegt der Altersschnitt über dem der im Wintersemester 2019/20 regulär in Deutschland eingeschriebenen Studierenden. Dieser beträgt 23,4 Jahre³. Das erste und dritte Quartil liegt bei jeweils 22,00 und 27,00 Jahren. Eine Person gab ein Alter von 2 Jahren an. Hier lässt sich eine Falscheingabe oder ein Anonymisierungsversuch vermuten. Am oberen Ende der Alterspyramide liegt eine Person mit 71 Jahren. Ein Erklärungsansatz könnte hier zum emeritierten Lehrpersonal oder zu den Angeboten der sog. Senioren-Universität führen. Auch wenn die überwiegende Mehrheit der Teilnehmer·innen im für Student·innen typischen Alter liegt, zeigt die demographische Angabe ebenfalls, dass eine ausschließliche Adressierung der Online-Umfrage an Student·innen kaum bis unmöglich ist.

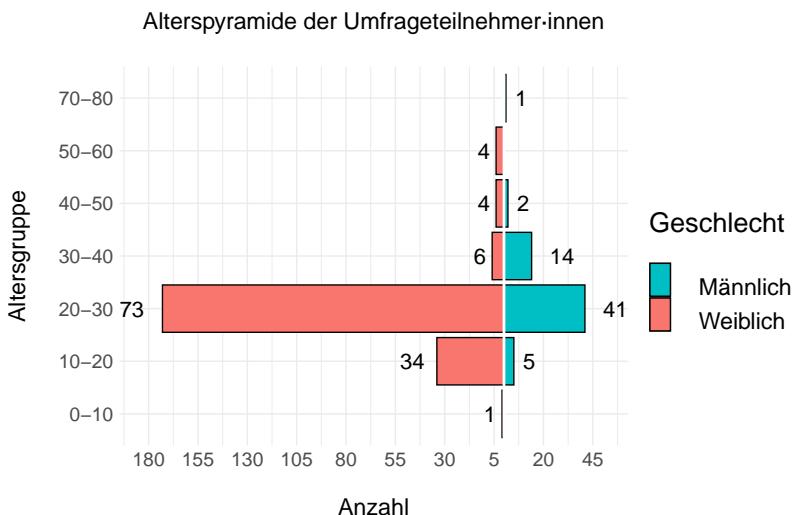


Abbildung 6.1: Alterspyramide der Teilnehmer

Dies lässt sich auch anhand der Angaben zum höchsten Abschluss erkennen. Die Abkürzungen entlang der X-Achse stehen der Reihenfolge nach für eine abgeschlossene Ausbildung (*Abg.A.*), die mittlere Reife (*mit.R.*), einen höheren Berufsabschluss (*höh.B.*), die Hochschulzugangsberechtigung (*HZB*) und einen Hochschulabschluss (*HSA*). In Bezug auf die Frage gaben 270 Personen an, entweder die Hochschulzugangsberechtigung erlangt oder bereits ein Studium (i. S. v.

³S. <https://www.forschung-und-lehre.de/lehre/studierende-im-schnitt-23-4-jahre-alt-3163/>, letzter Abruf am 31. Januar 2022.

6.1 Online-Umfrage zur allgemeinen Nutzung von Skype

Bachelor, Master, Diplom o. ä.) abgeschlossen zu haben. Weitere 12 der verbleibenden 20 Teilnehmer:innen gaben an, eine Ausbildung abgeschlossen zu haben. Somit ist auch hier erkennbar, dass zwar einerseits überwiegend die Zielgruppe angesprochen, zugleich jedoch auch vereinzelt Personen teilgenommen haben, die nachweislich nicht zur intendierten Gruppe gehören (s. Abbildung 6.2).

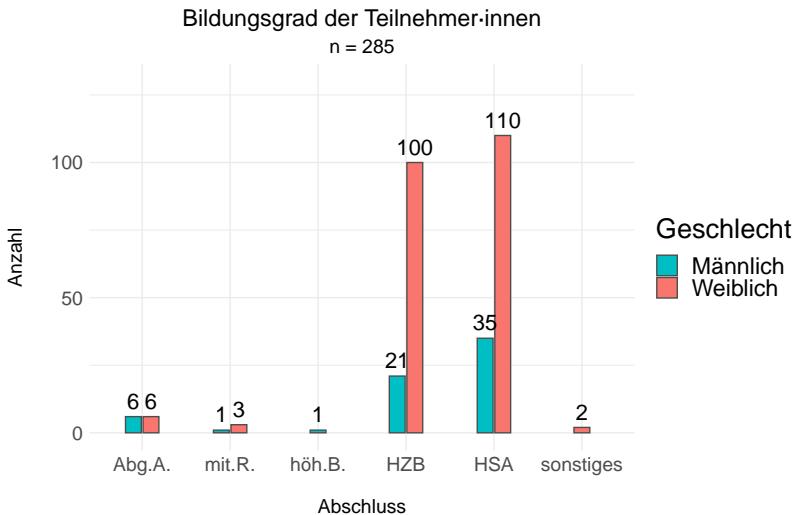


Abbildung 6.2: Höchster Abschluss nach Geschlecht

Die Mehrheit der Teilnehmer:innen (254/290) gab Deutsch als die Sprache an, mit der sie primär aufgewachsen seien. Diese Werte sind nicht deckungsgleich mit der im Alltag hauptsächlich verwendeten Sprache. Hier gaben 238 Personen an, hauptsächlich Deutsch zu verwenden, weitere zehn setzten Deutsch und Englisch gleich. Die verbleibenden Umfrageteilnehmer:innen gaben ausschließlich Englisch oder eine gänzlich andere Sprache an. Ebenso verhielt es sich bei der im Beruf hauptsächlich verwendeten Sprache, als welche 208 Personen Deutsch angaben, sechs sowohl Deutsch und Englisch und 19 ausschließlich Englisch. Die verbleibenden Personen gaben auch hier eine andere Sprache an. Eine dritte Frage konzentrierte sich auf weitere Fremdsprachen, sofern die Umfrageteilnehmer:innen einen offiziellen Nachweis über das Niveau besaßen. Hierbei wurden 36 verschiedene Sprachen angegeben, die in Abbildung 6.3 aufgeschlüsselt sind. Bereits während der Auswertung wurden die Angaben bereinigt: Anmerkungen zur eigenen Einschätzung des Kompetenzniveaus der Proband:innen und Angaben zum offiziellen Nachweis wurden ebenso wie Orthographiefehler entfernt.

167 von 290 Umfrageteilnehmer:innen gaben weiterhin an, bereits längere Zeit (mehr als vier Wochen Minimum, kein Urlaub) im Ausland verbracht zu haben.

6 Analyse

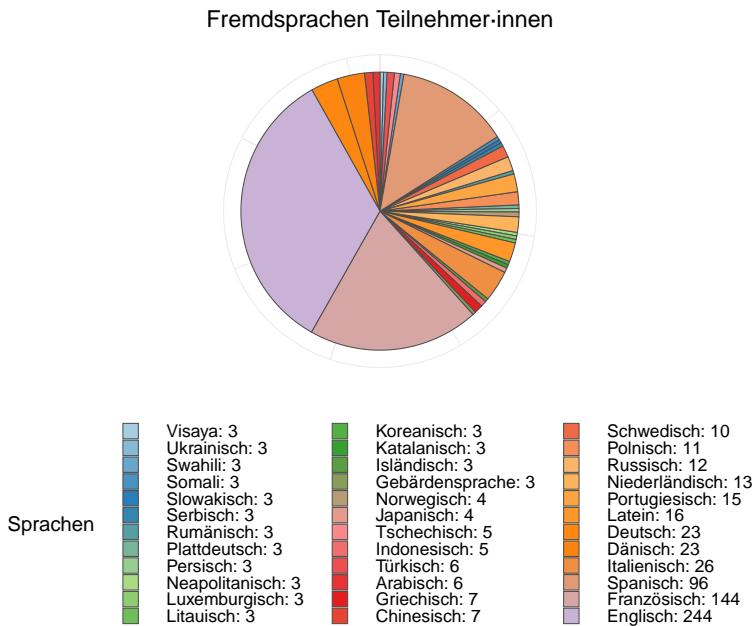


Abbildung 6.3: Weitere Fremdsprachen der Teilnehmer·innen

Insgesamt wurden 120 verschiedene Länder genannt. Die in den jeweiligen Ländern verbrachte Dauer reicht von einem bis 576 Monate (entspricht 48 Jahren, $\bar{x} = 34,5$ Monate, Median = 13 Monate, SD = 77,17). Als Gründe wurden Spracherwerb (52/290), Sprachvertiefung (91/290), Kulturkontakt (84/290), ein Pflichtaufenthalt (20/290), eine Fortbildung (20/290), berufliche Motivation (39/290) sowie die (zeitweise) Verlagerung des Lebensmittelpunktes (35/290) genannt. Die Aufenthalte lagen dabei zwischen weniger als einem und 360 Monaten zurück ($\bar{x} = 32,5$ Monate, Median = 22,5 Monate, SD = 42,5).

6.1.2 Generelle Nutzungsangaben

Von 290 Teilnehmer·innen an der Umfrage gaben 227 an, Skype überhaupt zu nutzen (s. Abbildung 6.4). Da die Online-Umfrage so konditioniert war, dass nur die Teilnehmer·innen an der Umfrage den Fragenblock zu den Nutzungsgewohnheiten im Umgang mit Skype eingeblendet bekamen, beziehen sich zunächst alle folgenden Angaben auf diese 227 Personen. Mit ca. 73,6 % nutzte deutlich mehr als die Hälfte der Umfrageteilnehmer·innen Skype entweder seltener als monatlich (143/227) oder nahezu monatlich (43/227). 40 Personen (18,5 %) gaben an, Skype wöchentlich nutzen und weitere 18 (7,9 %) nutzen es laut eigenen Angaben

6.1 Online-Umfrage zur allgemeinen Nutzung von Skype

täglich. Auffälligerweise wählten zwei Personen (0,9 %) die Antwortmöglichkeit *Nie*, auch wenn sie zuvor auf die Frage, ob sie Skype nutzten, mit *Ja* geantwortet haben müssen (s. Abbildung 6.4).

In Bezug auf die Frage nach der Nutzungsduer pro Sitzung gaben 165 Teilnehmer·innen (72,7 %) an, die Anwendung entweder länger als eine Stunde (85/227) oder mindestens eine Stunde (80/227) zu verwenden. Eine kürzere Sitzungsduer in den Kategorien *bis zu 30 Minuten*, *bis zu 15 Minuten* und *bis zu 5 Minuten* wurde jeweils von 34 (15 %), 16 (7 %) bzw. 8 (3,5 %) Personen angegeben. Vier Personen enthielten sich der Antwort.

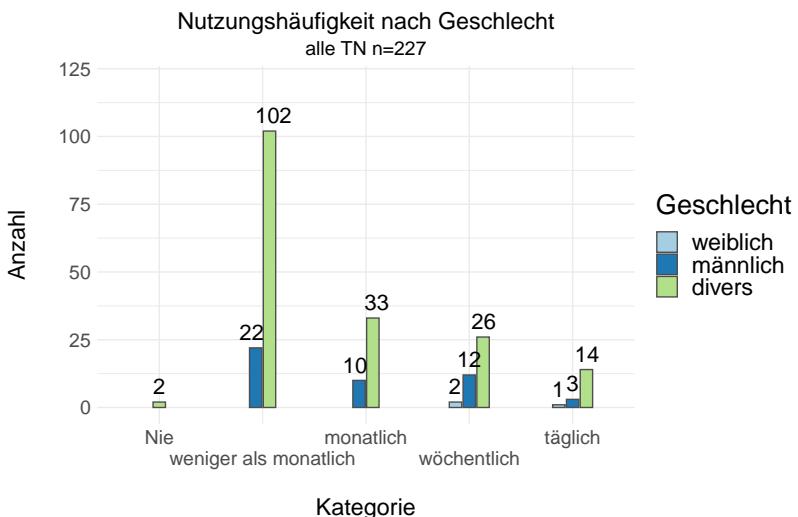


Abbildung 6.4: Nutzungshäufigkeit nach Geschlecht

Wie in Abbildung 6.5 dargestellt, wurde Skype am häufigsten auf Windows verwendet (190/227, 83,7 %). Da bei der Frage nach der Plattform Mehrfachantworten möglich waren, gaben zugleich 72 Teilnehmer·innen (31,7 %) an, Skype ebenfalls als mobile Anwendung auf dem Smartphone oder Tablet zu nutzen. 34 Personen (15 %) gaben an, den Dienst auf MacOS zu verwenden.⁴ Dabei nutzten die Teilnehmer·innen vor allem den Video-Chat (199/227, 87,7 %) bzw. den Textchat (136/227, 60 %) noch vor dem Voice-Chat (104/227, 45,8 %).

Die weiteren Funktionen *Bildschirmübertragung*, *Konferenzmodus*, *Dateiübertragung* und *Festnetzanruf* wurden deutlich weniger verwendet, wie in Abbildung 6.6 ersichtlich wird. Auf die Frage, ob der *Skype Translator* Verwendung

⁴Diese Frage hat im Verlauf der Ausarbeitung etwas an Relevanz verloren, da der Skype Translator mittlerweile auf allen Plattformen nutzbar ist. Dies war zu Beginn des Forschungsvorhabens 2017 entweder nur begrenzt (MacOS, Browser) oder gar nicht (Linux, Mobil) der Fall.

6 Analyse

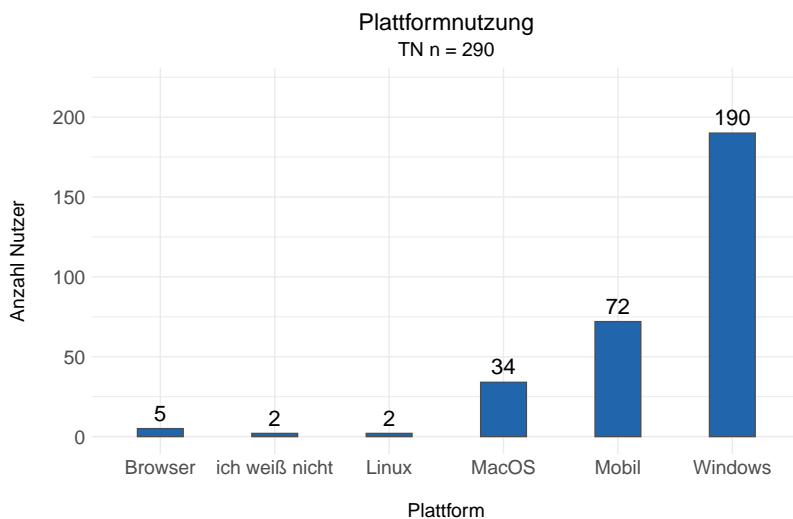


Abbildung 6.5: Anzahl Nutzer·innen nach Plattform

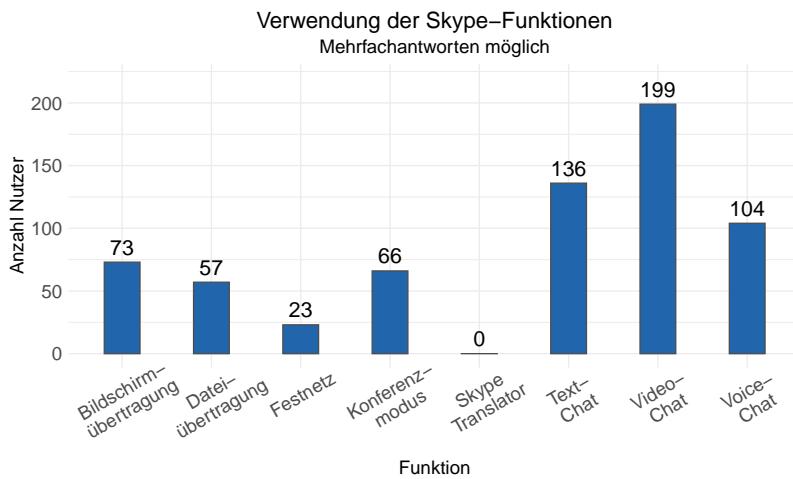


Abbildung 6.6: Anzahl an Nutzer·innen nach Skype-Funktion

6.1 Online-Umfrage zur allgemeinen Nutzung von Skype

findet, antwortete keine Person mit *Ja*. Nur drei Personen gaben allgemein an, jemals Erfahrungen mit dem Skype Translator gemacht zu haben.

In Bezug auf die Gesprächspartner·innen konnten die Umfrageteilnehmer·innen aus acht Kategorien mehrfach auswählen: *Familie*, *Freunde*, *berufliche Zwecke* und *Institutionen*, wobei jede dieser Kategorien noch einmal mit *Inland* und *Ausland* spezifiziert wurde. Besonders häufig sind demnach Freunde im In- und Ausland die Gesprächspartner·innen bei einer Kommunikation über Skype. Jeweils 126 (55,5 %) bzw. 143 (63 %) Personen gaben dies an. Mehr als ein Drittel der Skypenutzer·innen der Umfrage verwenden den Dienst zudem für Gespräche mit ihrer Familie im In- und Ausland (jeweils 74/227, 32,6 %, bzw. 93/227, 41 %).

Deutlich weniger wird Skype für den Kontakt mit Institutionen und für berufliche Zwecke verwendet. Jeweils 43 (18,9 %) und 34 (15 %) Personen gaben an, den Dienst für berufliche Kontakte im In- und Ausland zu nutzen. Nur 9 (4 %) bzw. 5 (2,2 %) Personen verwendeten Skype im Umgang mit Institutionen im In- und Ausland.

6.1.3 Nutzung von Alternativen zu Skype

Die nachfolgenden Angaben beziehen sich nun wieder auf die eingangs erwähnten 290 vollständigen Datensätze der Umfrageteilnehmer·innen. Mit der Erhebung der verwendeten Alternativen zu Skype begann in der Online-Umfrage ein neuer Block, zu dem die Teilnehmer·innen direkt weitergeleitet wurden, die die Frage nach der generellen Skype-Nutzung verneint hatten.

217 Personen gaben in diesem Teil an, generell Alternativen zu Skype zu verwenden, die ebenfalls einen Voice-, Video- oder Textchat bieten. Abbildung 6.7 zeigt dabei in Diagramm A (linke Seite), wie viele der Teilnehmer·innen die zur Auswahl stehende Alternative nutzen. Diagramm B (rechte Seite) zeigt dazu im Vergleich die am häufigsten verwendete Anwendung für den jeweiligen Kommunikationsmodus.

6.1.4 Situierung der Ergebnisse

Um die intendierte Gruppe der möglichen Versuchsteilnehmer·innen besser einordnen zu können und einen groben Überblick über das Kommunikationsverhalten in der anvisierten Altersgruppe zu erhalten, wurde die in den voranstehenden Abschnitten vorgestellte Online-Umfrage konzipiert. Lässt man die in Abschnitt 6.1.1 (S. 97) angesprochenen auffälligen Altersangaben außer Acht, so ergibt sich ein überwiegend weibliches und junges, sich in den Zwanzigern befindliches Feld an Teilnehmer·innen. Die meisten der Umfrageteilnehmer·innen besitzen mindestens die Hochschulzugangsberechtigung oder sogar bereits einen

6 Analyse

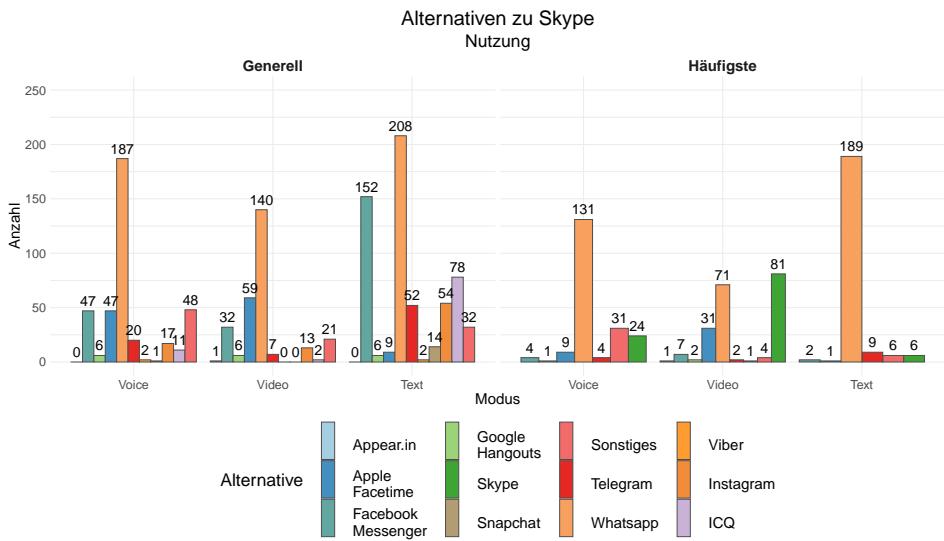


Abbildung 6.7: Vergleich: Generelle vs. häufigste Nutzung der Alternativen zu Skype nach Modus

ersten akademischen Abschluss. Die Kerngruppe kann daher als formal gut gebildet definiert werden. Die Mehrheit der Umfrageteilnehmer·innen gibt Deutsch als Muttersprache an, auch wenn häufig noch mindestens eine weitere, wenn nicht sogar zwei oder mehr Sprachen genannt werden. Insgesamt werden 36 verschiedene Sprachen aufgeführt.

Skype wird generell selten (mehrheitlich monatlich und seltener als monatlich), aber dafür lange (länger als eine Stunde oder mindestens eine Stunde) genutzt. Das häufigste Betriebssystem, auf dem Skype genutzt wird, ist Windows, auch wenn ein Drittel der Teilnehmer·innen darüber hinaus auch noch die App für das Smartphone verwendet. Die meistgenutzte Funktion von Skype ist der Videochat. Die Gesprächspartner·innen sind überwiegend Freunde im In- und Ausland. Allerdings geben die Befragten mehrheitlich an, parallel dazu auch Alternativen zu Skype zu verwenden. Für den Voice-, Video- und Textchat ist dies WhatsApp. Ebenfalls häufig wird für Textnachrichten der Facebook Messenger genutzt. Die jeweils am meisten genutzte Software für die jeweilige Funktion ist für Voice- und Textchats WhatsApp, für Videochats Skype.

Die Nutzungsangaben in der Umfrage decken sich mit repräsentativen Erhebungen zum Kommunikationsverhalten und zur Mediennutzung der vergangenen Jahre. So kam eine von Deloitte durchgeführte Studie (Deloitte 2017) zur Nutzung von Messaging-Diensten nach Altersgruppen aus dem Jahre 2017 zu dem

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Ergebnis, dass jeweils über 80 % der 18–24-Jährigen und 25–34-Jährigen *Mobile Instant Messaging* nutzt. Andere Antwortmöglichkeiten wie E-Mail, *Social Network Messaging* oder SMS fallen hinter diesem Ergebnis deutlich zurück. In diesem Zusammenhang ergibt die Studie zudem, dass über den Zeitraum von 2013 bis 2018 die Nutzung von Mobile Instant Messaging, E-Mail und Social Network zugenommen hat. Die Nutzung von SMS hingegen ging zurück (Deloitte 2019).

Laut einer von Bitkom durchgeführten Erhebung zur primär verwendeten Messenger-App von August 2018 nutzen über 80 % der Befragten vor allem WhatsApp. An zweiter Stelle mit 42 % steht der Facebook Messenger noch vor Skype mit 22 %. Dieselbe Studie erhebt weiterhin die meistgenutzten Funktionen bei Messenger-Apps. Die Teilnehmer·innen geben hierbei an, vor allem Nachrichten zu schreiben (82 %) und Bilder, Videos u.ä. Inhalte zu versenden (70 %). Telefonie und Videochats werden erst an dritter (50 %) bzw. sechster Stelle (27 %) genannt. Eine Auswahl der App nach der internationalen Verfügbarkeit ist den Umfrageteilnehmer·innen dabei weniger wichtig (Joachim Thommes 2018).

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Als Zielgruppe für die Studie wurden eingeschriebene Student·innen der Leipziger Hochschulen ungeachtet des Studiengangs erachtet. Als eine für dieses Experiment hinreichend große Gruppe wurde eine Anzahl von 15–20 Proband·innen erachtet. Zwei Ausschlusskriterien mussten mögliche Teilnehmer·innen zwingend erfüllen: Sie sollten über keine, allenfalls wenige (max. Niveau A gem. Europ. Referenzrahmen), Katalanischkenntnisse verfügen und Deutsch als Muttersprache ansehen. Ein Studienaufruf wurde an die Studierendenvertretung der Universität Leipzig und der HTWK gesendet. Die Vergütung bestand aus einer Aufwandsentschädigung von 10 Euro.

Als katalanischsprachige Gesprächspartner·innen fungierten drei Frauen und ein Mann aus unterschiedlichen Gebieten der katalanischsprachigen Länder (*paisos catalans*) – aus Valencia, Girona und Barcelona. Die Personen waren jeweils 26, 24, 25 und 26 Jahre alt. Weiterhin wiesen alle vier Personen Deutschkenntnisse durch einen studienbedingten Aufenthalt in Deutschland auf. Die katalanischen Muttersprachler·innen erhielten jedoch weder den Fragebogen vor oder nach dem Experiment, noch wurden sie selbst okulometrisch erfasst.

Auch wenn der Einsatz einer einzigen katalanischsprachigen Person als Gesprächspartner·in die Vergleichbarkeit der Daten untereinander sicherlich erhöht hätte, wurde von dieser Möglichkeit bewusst abgesehen. Bereits in der Pilotierungsphase hatte sich herausgestellt, dass die Koordination einer Studie, die

6 Analyse

von der gleichzeitigen Verfügbarkeit zweier Personen (jeweils Proband·in und Gesprächspartner·in) sowie vom Zugang zum Eye-Tracker abhängt, ein kommunikationsintensives Vorhaben wird. Um den Versuchsteilnehmer·innen möglichst viel Spielraum bei der Terminfindung zu bieten und so den unnötigen Wegfall von potenziellen Proband·innen aus Termingründen zu vermeiden, wurden mehrere katalanische Gesprächspartner·innen rekrutiert.

Insgesamt wurde das Experiment, beginnend mit der Einweisung und dem Fragebogen, über die Kommunikationssituation bis hin zum abschließenden Fragebogen und Debriefing, auf eine Länge von 60 Minuten pro Proband·in ausgelegt, wobei die Gesprächssituation sich auf etwa 15 Minuten beschränken sollte. Um den Studienteilnehmer·innen nicht gänzlich unvorbereitet in eine derartige Situation hineinzuwerfen und um mit Blick auf die Weiterverwendung der gewonnenen Sprachdaten eine einheitliche Ausgangssituation zu schaffen, wurde ein generelles Gesprächsthema angekündigt: Die Proband·innen sollten so tun, als bereiteten sie sich auf ein Auslandssemester in der Heimatstadt des katalanischen Gegenübers vor. Um Informationen zur Wohnungssuche (und zur Universität usw.) zu erhalten, kontaktierten sie deshalb die jeweilige Person.

Im Zeitraum zwischen Februar und Dezember 2019 wurde die Studie im Setting Katalanisch-Deutsch mit 25 Personen durchgeführt. Vier der dabei erhaltenen Datensätze wurden jedoch im Zuge der Datensichtung aufgrund von entweder schlechter Aufnahmequalität oder technischen Problemen bei der Aufnahme ausgeschlossen. So ergibt sich eine Kohorte von 21 Personen: 13 Frauen und acht Männer im Alter von 20 bis 32 Jahren ($\varnothing 23,7$ Jahre, Median = 23 Jahre, SD = 4,0).

Zwei der teilnehmenden Student·innen gaben an, die Hochschulzugangsbe rechtigung und eine abgeschlossene Ausbildung zu besitzen, eine weitere Person konnte sowohl die allgemeine Hochschulreife als auch eine abgeschlossene Ausbildung vorweisen. Bis auf eine Person von der HTWK kamen alle Teilnehmer·innen von der Universität Leipzig und waren zum Zeitpunkt der Erhebung regulär eingeschrieben.

Die Studienfächer der Proband·innen sind in Tabelle 6.1 und die Fachsemesteranzahl in Tabelle 6.2 aufgeschlüsselt. Studierende des Instituts für Angewandte Linguistik und Translatologie durften ebenfalls an der Studie teilnehmen, sofern Sie explizit keinerlei Vorkenntnisse im Katalanischen besaßen.

Auf die Frage nach der Sprache, mit der die Proband·innen primär aufgewachsen seien, gaben alle Personen Deutsch an. Jeweils eine Person gab darüber hin aus noch Russisch oder Englisch an, eine weitere sowohl Russisch als auch Englisch.

Die hauptsächlich im Alltag und im Beruf verwendete Sprache war für alle Teilnehmer·innen Deutsch, wobei zwei Proband·innen sich der Frage nach der

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Tabelle 6.1: Studienfächer der Teilnehmer·innen

	Bachelor	Master	Staatsexamen
Konferenzdolmetschen	-	1	-
Übersetzen	1	0	-
Lehramt	-	-	5
Interkult. Komm.	1	0	-
BWL	1	0	-
Medizin	-	-	5
DaF/DaZ	3	1	-
Linguistik	3	0	-
gesamt	9	2	10
		21	

Tabelle 6.2: Fachsemester der Teilnehmer·innen

Fachsemester	1	2	3	4	5	6	7	8	9	10	11
Konferenzdolmetschen	-	-	-	-	-	-	1	-	-	-	-
Übersetzen	-	-	-	-	-	-	1	-	-	-	-
Lehramt	2	-	-	1	-	1	-	1	-	-	-
Interkult. Komm.	-	-	1	-	-	-	-	-	-	-	-
BWL	-	-	-	-	-	1	-	-	-	-	-
Medizin	-	-	-	1	-	2	-	1	-	-	1
DaF/DaZ	1	-	2	-	-	1	-	-	-	-	-
Linguistik	-	-	-	-	3	-	-	-	-	-	-

6 Analyse

Berufssprache enthielten. Als weitere Fremdsprachen gaben die Proband·innen überwiegend Englisch an (17/21). Weitere Sprachen können Abbildung 6.8 entnommen werden, wobei auch hier wieder eine Mehrfachnennung möglich war.

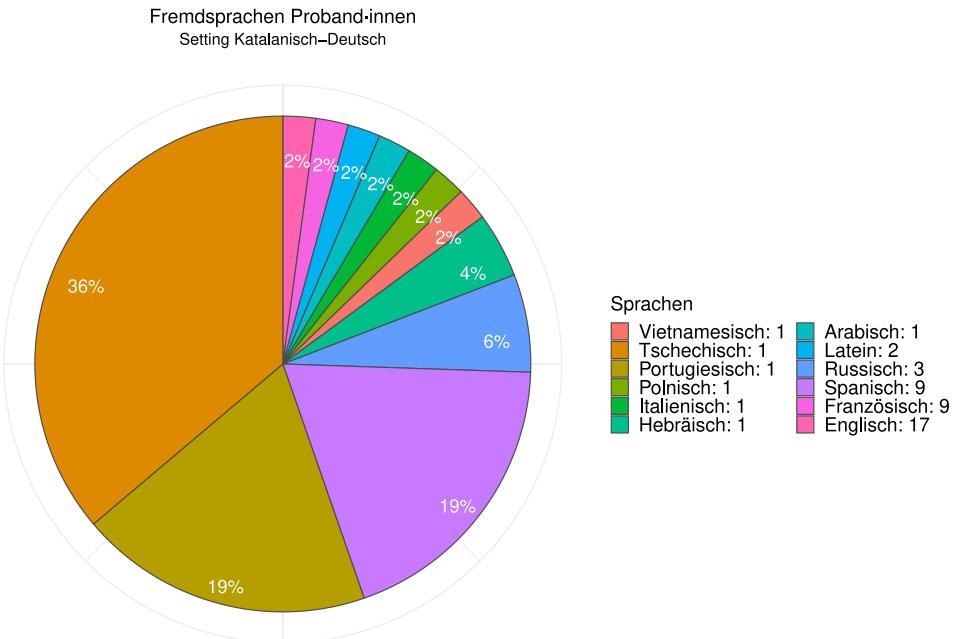


Abbildung 6.8: Fremdsprachen der Proband·innen

6.2.1 Angaben der Proband·innen im Eingangsfragebogen

Zu Beginn der Erhebung, vor der Textchatkommunikation über Skype, wurden die Teilnehmer·innen gebeten, einen Eingangsfragebogen auszufüllen. Die Angaben in diesem Fragebogen dienen der Verortung der Kohorte in Bezug auf Nutzungsverhalten derartiger Kommunikationssoftware.

Von den 21 Studienteilnehmer·innen gaben 17 an, Skype zu nutzen. Die weiteren vier Teilnehmer·innen erhielten daher nicht den Fragenteil zum Nutzungsverhalten, sondern nur zur Auslandserfahrung und Fremdsprachenbeherrschung sowie die demographischen Angaben. Deshalb wird im Folgenden zunächst nur von 17 Proband·innen ausgegangen. Von diesen 17 verwendeten 15 Skype seltener als monatlich, eine Person monatlich und eine wöchentlich. In Bezug auf die Nutzungsdauer pro Sitzung gaben fünf an, die Anwendung bis zu 15 Minuten zu nutzen, weitere fünf bis zu 30 Minuten, drei bis zu einer Stunde und vier länger als eine Stunde. 14 Personen gaben an, Skype für Windows zu verwenden,

6.2 Die Proband·innen im Setting Katalanisch–Deutsch

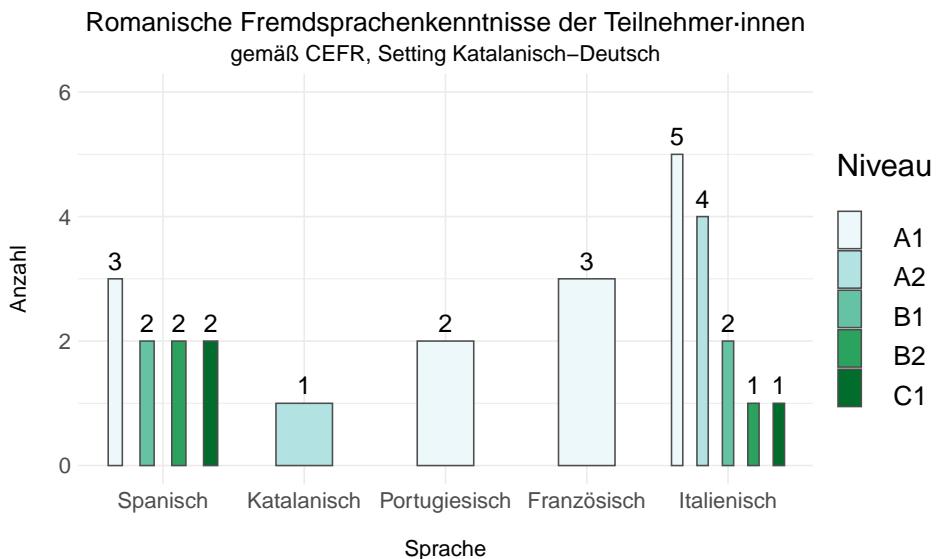


Abbildung 6.9: Kompetenz in ausgewählten romanischen Sprachen

jeweils eine für MacOS und im Browser und sechs als mobile App. Bei dieser Frage waren Mehrfachantworten möglich.

Die Mehrheit der Proband·innen nutzte Skype für Video- (15/17) oder Textchats (11/17), wohingegen der Voice-Chat von nur vier Proband·innen genutzt wurde. Auch die weiteren Funktionen wurden nur von wenigen Studienteilnehmer·innen verwendet: Festnetzanruf (3/17), Bildschirmübertragung (4/17), Konferenzmodus (5/17), Dateiübertragung (4/17). Der Skype Translator wurde von niemandem genutzt. Auch bei dieser Frage waren Mehrfachantworten möglich.

Die Proband·innen geben ihre Erfahrungen in Bezug auf den Textchat durchweg mit gut (11/17) bzw. überwiegend gut (3/17) an, wobei sich die verbleibenden Teilnehmer·innen der Antwort enthielten. Sowohl die Videochat- als auch die Voicechat-Funktion werden hingegen etwas schlechter wahrgenommen: Zwei Proband·innen haben gute Erfahrungen mit der Videoübertragung gemacht, zehn waren der Ansicht, die Übertragung sei überwiegend gut und fünf empfanden sie als eher schlecht. Der Voicechat wurde nur von einer Person als gut empfunden, zehn bewerteten ihn überwiegend gut und zwei eher schlecht.

In puncto Gesprächspartner·innen gaben fünf der Proband·innen an, mit Familienmitgliedern im Inland und zehn mit Familienmitgliedern im Ausland zu kommunizieren. Acht verwendeten Skype für die Kommunikation mit Freunden im Inland, 12 für den Kontakt mit Freunden im Ausland. Lediglich zwei bzw. eine

6 Analyse

Person gab an, Skype für berufliche Zwecke im In- bzw. Ausland zu nutzen. Eine Person gab an, Skype sowohl für die Kommunikation mit Institutionen im In- und Ausland zu verwenden. Die häufigste Sprache, die neben Deutsch (17/17) in diesen Situationen verwendet wurde, ist Englisch (8/17). Weiterhin wurden Russisch (3/17), Spanisch (2/17), Französisch (1/17) und Ungarisch (1/17) angegeben. Auch bei diesem Fragenteil waren Mehrfachantworten möglich.

13 der 21 Proband·innen gaben weiterhin an, bereits längere Zeit (mehr als vier Wochen Minimum, kein Urlaub) im Ausland verbracht zu haben. Die bereisten Länder der Teilnehmer·innen verteilen sich über die gesamte Erde, wobei ein Schwerpunkt auf Europa auszumachen ist. Sieben Versuchsteilnehmer·innen geben zwei oder mehr Länder an, in denen sie mehr als vier Wochen verbracht haben. Die in den jeweiligen Ländern verbrachte Dauer reicht von einem bis 110 Monate (9,17 Jahre, \bar{x} 30,5 Monate, Median = 16 Monate, SD = 37,85). Als Gründe wurden Spracherwerb (7/21), Sprachvertiefung (7/21), Kulturkontakt (11/21), ein Pflichtaufenthalt (1/21), eine Fortbildung (1/21), berufliche Motive (2/21) sowie die (zeitweise) Verlagerung des Lebensmittelpunktes (3/21) genannt. Dabei lagen die Aufenthalte zwischen drei und 166 Monaten zurück (entspricht 13,83 Jahren, \bar{x} 34,2 Monate, Median = 16 Monate, SD = 49,9).

6.2.2 Nutzung von Alternativen zu Skype

Im Folgenden beziehen sich die Angaben zur Nutzung von Alternativen zu Skype nun wieder auf alle 21 Studienteilnehmer·innen. Analog zu den Ergebnissen der Online-Umfrage gaben auch die Proband·innen der Studie an, bislang keine Erfahrungen mit dem Skype Translator gemacht zu haben. Wohlwollend interpretiert, könnte dies als favorable Ausgangslage angesehen werden, da so keine Beeinträchtigungen durch Vorerfahrung in die Studie einfließen. Andererseits jedoch zeichnet sich ein recht unpopuläres Gesamtbild des Features ab.

17 Teilnehmer·innen gaben an, generell Software zu verwenden, die ebenfalls eine Voice- oder Videochat-Funktion bietet. Sowohl in Hinblick auf Voice- als auch auf den Videochat gab die Mehrheit an, WhatsApp zu verwenden (jeweils 17 bzw. 14/21). Die anderen angebotenen Alternativen in dieser Frage (s. Frage im Anhang B, S. 256) wurden allesamt stets von weniger als sieben Proband·innen überhaupt verwendet. Beim Textchat wurden sowohl der Facebook Messenger als auch WhatsApp als generelle Alternative zu Skype angegeben (jeweils 14 bzw. 16/21). Auch hier erhielten die anderen angebotenen Alternativen nie mehr als sechs Angaben. Im Vergleich dazu gaben die Proband·innen antworteten auf die Frage nach dem Dienst, der für Voice-, Video- und Textchat jeweils am Häufigsten verwendet wird, deutlich diverser.

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Die für Videochats am häufigsten verwendeten Dienste sind demnach WhatsApp (7/21), Skype (3/21), Facebook Messenger sowie Apple Facetime (jeweils 2/21) bzw. Discord (1/21). Auch bei Voice-Chats wurde WhatsApp am häufigsten verwendet (11/21). Ebenfalls wurden Facebook Messenger (2/21), Discord (2/21), Viber (1/21) sowie das Haustelefon (1/21) genannt. Gleichermaßen ist WhatsApp ebenso bei Textchats die am häufigsten verwendete Alternative (15/21), bei jeweils einer Nennung von Facebook Messenger und Google Hangouts.

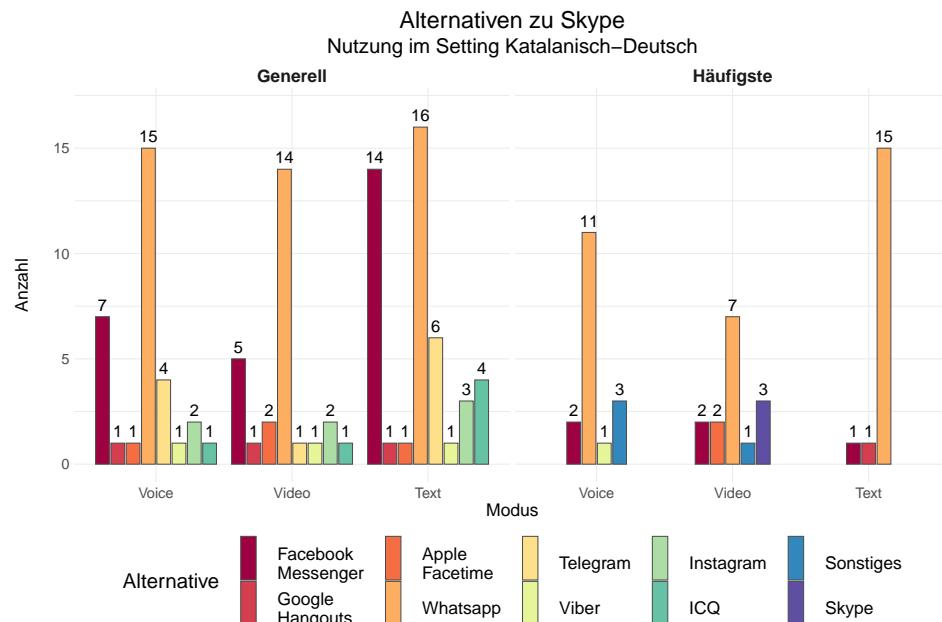


Abbildung 6.10: Vergleich Generelle und häufigste Nutzung der Alternativen zu Skype nach Modus

6.2.3 Angaben der Proband·innen im Ausgangsfragebogen

Unmittelbar nach Beendigung der Eye-Tracking-Sitzung wurden die Proband·innen gebeten, den Ausgangsfragebogen auszufüllen (s. Anhang C, S. 263 ff.). Zweck dieses Fragebogens war es, die unmittelbare Einschätzung der Studienteilnehmer·innen zur gerade erlebten Kommunikationssituation unter Beteiligung des Skype Translators zu erfassen. Auch an diesem Teil der Erhebung nahmen alle 21 Personen teil. Die Teilnahmebedingungen waren die gleichen wie auch beim Eingangsfragebogen (zum datenschutzrechtlichen Hinweis s. Anhang B, S. 251).

6 Analyse

Zunächst wurde um eine generelle Einschätzung der Aspekte *Gesprächsqualität*, *Übersetzungsqualität* und *Versuchsaufbau* gebeten. Die Aspekte *Tonqualität* sowie *Videoqualität* wurden vernachlässigt, da die Kommunikationssituation im Modus Textchat stattfand. Dies geschah durch eine vorangestellte Ja-Nein-Frage, zu welchen Aspekten die Teilnehmer·innen sich in der Lage sahen, eine Einschätzung abzugeben. Daher wird im Folgenden stets von der Personenanzahl ausgegangen, die sich zur Beurteilung des jeweiligen Aspektes in der Lage sah. Die Beantwortung der Einschätzungsfragen ließ Angaben in Form einer Likert-Skala von 1 – schlecht bis 4 – gut sowie die Möglichkeit der Enthaltung zu. Es wurde bewusst darauf verzichtet, die Aspekte genauer zu spezifizieren, um einen größtmöglichen Antwortspielraum zu bieten.

So gingen elf der Studienteilnehmer·innen auf die Gesprächsqualität ein. Vier von Ihnen bewerteten die Gesprächsqualität als gut, fünf als eher gut und zwei als eher schlecht. Die Übersetzungsqualität wurde von 20 Personen bewertet. Auch hier wurde die Qualität vier Mal als gut bewertet, zehn Mal als eher gut, fünf Mal als eher schlecht und einmal als schlecht. Der Versuchsaufbau wurde von 13 Teilnehmer·innen bewertet. Neun von ihnen erachteten das Setting als gut und jeweils zwei von ihnen als eher gut bzw. eher schlecht.

Die Erhebung der Einschätzung wurde von zwei Freitextfragen begleitet. Die Proband·innen wurden um die Einschätzung gebeten, wie sie sich in der Gesprächssituation gefühlt und wie sie den zweispaltigen Aufbau des Skype Translators (s. hierzu Abbildung 3.3, S. 59) wahrgenommen haben.

In Bezug auf die Wahrnehmung der Gesprächssituation lassen sich die Teilnehmer·innen in zwei Kategorien einteilen: Ein Teil der Proband·innen gab an, sich merkwürdig bzw. komisch gefühlt zu haben, ein anderer Teil sprach durchweg von einer angenehmen bzw. guten Situation. Viele der Teilnehmer·innen beschreiben zugleich einen Wandel ihrer Wahrnehmung über das Gespräch hinweg, sodass die binäre kategoriale Einteilung nicht trennscharf durchzuführen ist. Ausgehend von der Dichotomie *merkwürdiges – angenehmes* Gefühl während des Gesprächs lassen sich dennoch neun bzw. 15 entsprechende Aussagen identifizieren.

Auffällig ist, dass nur zwei Personen ihr Gefühl mit der maschinellen Übersetzung in Verbindung setzen (s. Beispiele 1 und 2).

- (1) insgesamt gut auch wenn man sich fühlt als spräche man mit Robotern, wirkt es gleichzeitig sehr unverfänglich
- (2) Ich habe mich zunächst etwas komisch gefühlt, weil ich primär ja mit einer Maschine geschrieben habe und nicht mit einer realen Person. Aber irgendwann habe ich es als sehr praktisch empfunden, weil ich meinen Gegenüber sonst nicht verstehen könnten.

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Die meisten Klartextantworten verweisen auf den/die unbekannte Gesprächspartner·in und die hypothetische Aufgabenstellung, die die Situation für sie eher unangenehm machte.

- (3) Erstaunlich wohl dafür, dass ich mit einer mir völlig unbekannten Person über eine fiktive Situation chatten sollte und sogar inhaltlich ganz gut mitgekommen bin.
- (4) Ich habe mich sehr seltsam und merkwürdig gefühlt, weil ich ein sinnvolles, höfliches Gespräch mit einem wildfremden Menschen aufrechterhalten musste, wobei ich allerdings keinen blassen Schimmer hatte, wie dieser Mensch denkt und was seine geistige Haltung und seine Weltanschauung sind.

Die Antworten auf die Frage zur Einschätzung der Kommunikationssituation bieten außerdem Anknüpfungspunkte an die in den Abschnitten 2.1 (S. 7) und 2.2 (S. 15) dargestellten theoretischen Hintergründe zur CvK. So weisen mehrere Versuchspersonen darauf hin, dass ihnen die Anonymität in der Situation zunächst befremdlich vorkam. Besonders deutlich wird dies anhand des folgenden Beispiels:

- (5) Ich habe mich zunächst etwas komisch gefühlt, weil ich primär ja mit einer Maschine geschrieben habe und nicht mit einer realen Person. Aber irgendwann habe ich es als sehr praktisch empfunden, weil ich meinen Gegenüber sonst nicht verstehen können.

Diese Person ist offenbar stark von der CvK-Situation beeinflusst und fühlt sich dadurch von dem eigentlichen, menschlichen Gegenüber entfernt. Ähnlich beschreibt diese unpersönliche Distanzierung eine andere Person mit der Aussage:

- (6) Es hat sich etwas gezogen. An sich aber eine gute Idee. Mir persönlich hat es aber nicht gefallen. Ich fühlte mich sehr distanziert von der Person und es war nicht wirklich persönlich.. eher so wie auf dem Amt.

Zusammenfassend sind an mehreren Stellen in den Antworten Indizien zu finden, die darauf hinweisen, dass die Versuchspersonen zwischen der im Theorieteil angesprochenen Nähe und Distanz schwanken, nicht zuletzt bedingt durch die Vermittlung des Skype Translators.

Die direkte Folgefrage knüpfte an die Beschreibung der Wahrnehmung an. Die Proband·innen wurden gebeten, Angaben zur Wahrnehmung des zweisprachigen

6 Analyse

Aufbaus des Skype Translators zu machen. Da der Skype Translator, wie in Abbildung 3.3 (S. 59) dargestellt, den Originalbeitrag der Versuchsperson zwar linksbündig, alle weiteren Beiträge – also sowohl Original des Gegenübers als auch alle jeweiligen maschinellen Übersetzungen aller Beteiligten – rechtsbündig ausgibt, zielte diese Frage nicht nur auf den empfundenen Nutzen ab. Die Antworten liefern auch Aufschluss über die Art, wie die Versuchsteilnehmer-innen sich während der Situation verhalten haben, um Informationen im Chat zu verarbeiten.

Zehn Teilnehmer-innen sahen den Aufbau als gut an, da sie so während des Gesprächs direkte Vergleiche ziehen konnten zwischen Original und Übersetzung bzw. bereits Geschriebenes noch einmal Revue passieren lassen konnten.

(7) gut, dann kann man katalanische und deutsche Version vergleichen

Zugleich berichten 12 Personen von einer (anfänglichen) Verwirrung, die durch die Anordnung der Übersetzung und des Originals entsteht. Auch wurde mehrfach angemerkt, dass die permanente Einblendung sowohl der Originale als auch deren Übersetzungen diese Verwirrung auslösten bzw. die Aufmerksamkeit verringerten. Hier hätten sich zwei Personen die Möglichkeit gewünscht, entweder innerhalb des gleichen Textfeldes zwischen Original und Übersetzung wechseln zu können oder eines der beiden auszublenden. Dabei ist zugleich der Hinweis wichtig, dass die Person eine Neugierde für die katalanischsprachigen Beiträge hegt und diese offenbar mit dem deutschen Beitrag vergleicht.

- (8) Okay. Viel Text. Es verwirrt, dass meine deutsche Übersetzung links bei der Chat Seite vom gegenüber eingebendet wird. Obwohl, ich es auch interessant finde, zu lesen was mein Text auf Catalan bedeutet...ist aber tendenziell etwas ablenkend.
- (9) Es [der zweispaltige Aufbau] ist sehr gut. Die Simultanübersetzung erleichtert die Kommunikation. Was verwirrend war, dass die Übersetzung meiner Aussagen auf der linken Seite erschienen.. ich fände es besser, wenn sie direkt unter meinem Text auf der rechten Seite erscheinen.

Zum Verhältnis von MÜ-Ausgabe zu Originalbeitrag lassen sich aus diesem Bereich Antworten erkennen, die auf Neugierde (s. Beispiel 8) und ein vergleichendes Verhalten hindeuten. So gibt eine Person implizit an, dass sie die Vermittlung des Skype Translators durchaus in ihr eigenes Chatverhalten integriert:

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

- (10) Gut, man kann die Fremdsprache auch lesen und eventuell aus beiden Sprachen die Verständlichkeit für einen selbst erhöhen

Zur genaueren Kontextualisierung dieser freien Antworten folgten vier Einstellungsfragen zur Qualität der Kommunikation, zur Informationsdichte, zur Leistung des Skype Translators und zur Dauer der Kommunikationssituation in Form einer Likert-Skala. 20 Personen gaben zum erstgenannten Aspekt ihre Einschätzung ab. Zehn von ihnen bewerteten die Qualität der Kommunikation als gut, sechs als eher gut, drei als eher schlecht und eine Person als schlecht.

Die Informationsdichte wurde ebenfalls von 20 Personen bewertet. Auch hier waren zehn Personen der Ansicht, die Informationsdichte über die Kommunikation hinweg sei als gut zu bewerten. Vier Teilnehmer·innen nahmen diese als eher gut wahr, drei als eher schlecht und ebenfalls drei als schlecht.

Die Leistung des Skype Translators wurde von allen 21 Teilnehmer·innen bewertet. Sechs sahen diese als gut an, elf als eher gut, drei als eher schlecht und eine Person als schlecht.

Die Dauer der Kommunikationssituation wurde ebenfalls von allen 21 Teilnehmer·innen beurteilt. Im Gegensatz zu den vorausgehenden Fragen besaß die Likert-Skala dieser Frage fünf Ausprägungen, die von *1 – zu kurz* über *3 – angemessen* bis *5 – zu lang* reichten. Eine Person empfand die Dauer als eher zu lang, 16 Proband·innen nahmen die Dauer als angemessen wahr, zwei als eher zu kurz und eine Person als zu kurz.

Die Proband·innen wurden danach gebeten, Angaben zu fünf Aspekten der Ausgabequalität der maschinellen Übersetzung zu machen: semantische, syntaktische, lexikalische, stilistische und das Register betreffende Auffälligkeiten. Diese Einteilung stellt einen Kompromiss dar, da sie einerseits Übersichtlich und verständlich ist, andererseits jedoch auch gängige Kategorien bei der Bewertung von MÜ-Systemen darstellt. Auch diese offenen Fragen wurden im weiteren Verlauf unmittelbar durch die präzisierende Frage kontextualisiert, welchen Einfluss die jeweiligen Aspekte auf die Kommunikation hatten. Hierbei wurden den Proband·innen Antwortmöglichkeiten in Form einer Likert-Skala mit der Ausprägung von *1 – kein Einfluss* bis *4 – großer Einfluss* angeboten.

14 Teilnehmer·innen machten Angaben zu semantischen Auffälligkeiten, zehn zur Syntax der Ausgabe, 12 jeweils zu Lexik und zum Stil, sieben zum Register und fünf nutzen das Feld für „weitere Angaben“.

Die überwiegende Mehrheit der zur Semantik abgegebenen Kommentare bezieht sich auf die fehlerhafte Übersetzung einzelner Begriffe während der jeweiligen Chat-Konversation. So bemängeln die Proband·innen vor allem die Wortwahl bzw. die Präzision der gewählten Wörter.

6 Analyse

- (11) Die Übersetzung eines Wortes war nicht bekannt (No conduisc=kein duisc)
- (12) Die Wortwahl war meistens etwas vereinfacht. An einigen Stellen wurde auch über die Satzbedeutung nachgefragt, bzw. ich habe es mir aus dem Kontext erschlossen

Für 20 Teilnehmer·innen hatten die semantischen Auffälligkeiten der MÜ-Ausgabe Einfluss auf die Kommunikation. Jeweils vier von ihnen sahen dies als großen bzw. eher großen Einfluss an, sieben als eher geringen Einfluss und drei vermerkten keinen Einfluss.

Die Anmerkungen zur Syntax zeugen von einer generellen Skepsis gegenüber der MÜ-Ausgabe. An diesem Punkt scheint den Studienteilnehmer·innen offenbar besonders deutlich geworden zu sein, dass eine Maschine in die Kommunikation eingebunden war. Zwar werden Syntaxfehler einerseits als akzeptabel bzw. erwartbar hingenommen, wie Beispiel 13 verdeutlicht, andererseits bestehen Differenzen zwischen der subjektiv erwarteten Satzstellung im Rahmen einer solchen, informellen Kommunikationssituation seitens der Proband·innen und der faktischen Ausgabe (s. Beispiel 14).

- (13) Oftmalige Fehlstellung der Sätze und Phrasen, Richtigkeit des deutschen Satzgefüges konnte nicht eingehalten werden, wobei das bei Maschinellen Übersetzern fast schon mit einzuberechnen ist
- (14) Einige Sätze waren einfach etwas falsch gebildet, inhaltlich aber verständlich.

In diesem Zusammenhang merkten zwei Proband·innen an, dass die Möglichkeit zum Rückgriff auf das deutsche bzw. katalanische Original hilfreich war (s. Beispiele 15 und 16).

- (15) gelegentliche Fehler, dann habe ich noch einmal in die katalanische Version geschaut
- (16) Der Satzbau wurde einfach von Catalan ins Deutsche übernommen, obwohl die Satzbauteile in unterschiedlicher Reihenfolge sind. Klingt demnach eher so übersetzt wie bei Google Übersetzer bzw. schlechter.

15 Personen gaben hieran anschließend ihre Einschätzung zum Einfluss der syntaktischen Auffälligkeiten auf die Kommunikationssituation ab. Für eine Person hatte dieser Faktor großen Einfluss, für drei eher großen, für fünf eher geringen und für ebenfalls fünf keine Einfluss.

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Die Angaben zu lexikalischen Auffälligkeiten bei der MÜ-Ausgabe deuten auf eine generelle Inkonsistenz und den Mangel an kulturellen Hintergrundinformationen hin. Wie auch zur Frage nach semantischen Auffälligkeiten findet hier der häufig verwendete katalanische Interjektionspartikel *ostres* Erwähnung, der als Ausdruck von Überraschung dient. Die maschinelle Übersetzung hat diesen Ausruf jedoch stets im eigentlichen Sinne von *Astern* (s. Beispiele 17 und 18) übertragen.

- (17) Das Wort „Austern“, welches im Katalanischen einen Ausruf der Überraschung darstellt, wurde eins zu eins ins Deutsche übertragen, anstatt ein gleichwertiges deutschen Gegenstück zu finden.
- (18) Ja, z. B. das Wort Ostrés wurde mit Austern übersetzt oder Pero Sants mit Heilige oder Studentenwohnungen wurde ins Englisch übersetzt mit Student Apartments. Das alles ist natürlich nicht richtig.

16 Studienteilnehmer·innen bewerteten darüber hinaus den Einfluss, den die lexikalischen Auffälligkeiten auf die Kommunikationssituation hatten. Drei Personen nahmen hierbei einen großen, drei einen eher großen, fünf einen eher geringen und fünf keine Einfluss wahr.

Der Stil der Ausgabe wird sechs Studienteilnehmer·innen als umgangssprachlich und einfach beschrieben. Allen Personen, die zu dieser Frage Angaben gemacht haben, schien eine Differenz zwischen den beteiligten Sprachen aufgefallen zu sein. So wird die generelle Umgangsform zwischen Gesprächspartner·innen als Unsicherheitsfaktor erwähnt (s. Beispiele 19 und 20). Zugleich wird darauf hingewiesen, dass es sich bei der Ausgabe um einen einfachen Stil handele.

- (19) Ich habe auf Deutsch die formelle Anrede „Sie“ benutzt. Ich bin mir nicht sicher, ob das so ins Katalanische übersetzt wurde.
- (20) Der Stil war im Katalanischen etwas gesenkter als im Deutschen. Zudem hat meine Gesprächspartnerin eher informellen Wortschatz und Redewendungen gebraucht.

Bei der Betrachtung der Angaben zum Register der MÜ-Ausgabe werden Überschneidungen mit dem Bereich Stil ersichtlich. Mehrere Proband·innen geben auch hier den Wechsel innerhalb des Gesprächs zwischen duzen und siezen als Auffälligkeit an.

- (21) Ich war mir unsicher, ob das Siezen und Duzen so übersetzt wurde, wie es geschrieben wurde, weil ich mit „du“ gestartet habe und die Person mit „Sie“ geantwortet hat. Außerdem klangen die Antworten eher formell.

6 Analyse

- (22) Ich wurde das ganze Gespräch hindurch gesiezt, was mir in der Chat-Situation komisch vorkam. Ich habe am Anfang geduzt, zwischendurch gesiezt und es dann wieder vergessen und weiter geduzt. ich weiß überhaupt nicht ob es diese Unterscheidung im Katalanischen gibt

Aufbauend auf diesen Eindrücken wurden die Proband·innen im folgenden Abschnitt der Erhebung nach fehlenden bzw. unzureichenden Informationen befragt. Sechs der Teilnehmer·innen gab hier an, dass ihnen im Nachhinein noch immer Informationen fehlten. Zwei dieser Personen führten das Fehlen auf individuelle, persönliche Gründe zurück. Einer Person schien die Dauer der Studie zu knapp bemessen, eine andere gab an, nicht genau gewusst zu haben, für welche Stadt sie sich fiktiv hätte interessieren sollen. Selbige Person gab in der Folgefrage zu individuellen Gründen, die die Kommunikationspartner·in betreffen, zudem an, ebenfalls unsicher im Umgang mit dem Gegenüber gewesen zu sein, da dieser „sicherlich auch kein insiderwissen“ hatte. Eine zweite Person bemängelte die lange Antwortlatenz des katalanischen Gegenübers.

Drei Proband·innen führten das Fehlen an Informationen auf technische Gründe zurück. Hier wurde angemerkt, dass die MÜ-Ausgabe entweder eine zu hohe Latenz aufwies oder als unvollständig wahrgenommen wurde.

Um daraufhin den Informationsgewinn durch die Konversation über Skype mit anderen Quellen zu vergleichen, schloss hieran die Frage an, wo die Proband·innen womöglich fehlende Informationen nachträglich einholen würden. In Form einer Ja-Nein-Auswahl standen, neben einem erneuten Gespräch mittels Skype Translator, die Möglichkeiten *Freunde, Heimatuniversität, Internet, Gastuniversität, bis zum Beginn des Auslandsaufenthaltes warten und sonstiges* zur Auswahl.

Drei Personen gaben an, fehlende Informationen über ein erneutes Gespräch mittels Skype Translator einzuholen. Die Möglichkeit, weitere Angaben bei Freunden einzuholen, zogen acht Personen in Erwägung. Drei Teilnehmer·innen würden diese Informationen an der Heimat- und 12 an der Gastuniversität eинholen. 16 Personen, und damit die Mehrheit, würde ferner das Internet konsultieren. Drei Studienteilnehmer·innen würden bis zum Beginn des Auslandsaufenthaltes warten. Zwei Personen gaben weiterhin sonstige Quellen an.

Dieser Themenbereich wurde durch die Frage in Form einer Likert-Skala von 1 - *unangemessen* bis 4 – *angemessen* abgeschlossen, wie angemessen die Teilnehmer·innen den Einsatz des Skype Translators in dieser Situation fanden. Keine Person gab an, die Situation gänzlich unangemessen wahrgenommen zu haben.

6.2 Die Proband·innen im Setting Katalanisch-Deutsch

Zwei Teilnehmer·innen fanden die Situation teilweise unangemessen, sechs eher angemessen und 13 durchweg angemessen.

16 Teilnehmer·innen geben an, sich vorstellen zu können, den Skype Translator auch in anderen Situationen zu nutzen. Die Mehrheit der Angaben beschränkt sich dabei jedoch auf ungezwungene, informelle Situationen, die selten präzisen, eindeutigen Wortschatz erfordern. So findet die Vorbereitung auf einen Auslandsaufenthalt (Urlaub, Studium u. a.) mehrfach Erwähnung. Auch eine anfängliche Kontaktaufnahme mit Personen, deren Muttersprache man nicht beherrscht, wird mehrfach vorgeschlagen. Nur zwei Personen gehen auf die Möglichkeit ein, die Technologie auch für berufliche bzw. offizielle Anlässe zu nutzen (s. Beispiele 23 und 24), während eine andere Person dies gänzlich ablehnt (s. Beispiel 25).

- (23) Weitere Recherchen zum Auslandsaufenthalt, Berufliche Kontakte
- (24) Bereich der Arbeit
- (25) Ein Chat mit Freunden/Bekannten – auf keinen Fall ein „Gespräch“ mit Geschäftspartnern

Der abschließende Fragenkomplex ging auf die Zukunftsaussichten einer solchen Technologie ein. Zunächst wurden die Studienteilnehmer·innen gebeten, ihre Einschätzung zum Verhältnis von Skype Translator zu menschlichen Übersetzer·innen und Dolmetscher·innen in Form einer Likert-Skala von *1 – verschiedene Welten bis 4 – Konkurrenz* anzugeben. Keine Person nimmt das Verhältnis demnach als starke Konkurrenzsituation wahr, sieben Proband·innen jedoch waren der Ansicht, es handele sich tendenziell eher um Konkurrenz. Weitere sieben sahen das Verhältnis eher als zwei verschiedene Welten an, sechs als eindeutig verschiedene Bereiche.

Auch in diesem Fall schloss eine Frage mit Freitextantwort an, um die Einschätzung zu präzisieren. Innerhalb der Antworten, die das Verhältnis von Skype Translator zu menschlichen Übersetzer·innen und Dolmetscher·innen eher als Konkurrenzsituation wahrnehmen, dominieren Gründe technischer und funktionaler Natur. So benötige man keine weiteren Personen für die Übertragung des Gesprächs und die Technologie sei „billiger“ als der Einsatz von qualifizierten Übersetzer·innen und Dolmetscher·innen (s. Beispiel 26). Auch wird hervorgehoben, dass der Skype Translator einfach zu handhaben ist (s. Beispiel 27).

- (26) Skype Translator ist billiger, einfacher zu installieren (in dem Sinne sich zu organisieren).

6 Analyse

- (27) Skype ist sehr einfach in der Handhabung und kostenlos nutzbar, schnell, akzeptables Sprachniveau. Einen Dolmetscher würde ich bevorzugen, wenn es um ernste Sachverhalte wie z. B. das Erstellen von Dokumenten geht.

Zugleich muss an dieser Stelle angemerkt sein, dass die Einschätzung von Personen kommt, die wenig Berührungspunkte mit der Übersetzungs- und Dolmetschbranche haben. Dies wird in der konzeptuellen Vermischung der Tätigkeit von Übersetzer·innen und Dolmetscher·innen (s. Beispiel 27) widergespiegelt.

Die Gründe für ein eher distanziertes Verhältnis von Skype Translator zu menschlichen Übersetzer·innen und Dolmetscher·innen werden von den Proband·innen vor allem im unterschiedlichen Einsatzzweck gesucht. Hier differenzieren die Studienteilnehmer·innen zwischen informellen und komplexen Situationen (s. Beispiel 28). Während der Skype Translator für informelle Zwecke und einfachen Themen in der Wahrnehmung der Teilnehmer·innen eine gute Leistung erbringt, befürworten sie den Einsatz von menschlichen Übersetzer·innen und Dolmetscher·innen in Situationen, in denen Empathie benötigt wird, in denen es um präzise, wichtige Informationen geht oder in denen fach(-sprach-)liche Kenntnisse erforderlich sind (s. Beispiel 29).

- (28) Die Zielsetzung ist anders, einen Dolmetscher/Übersetzer würde ich für wichtige Anliegen in Anspruch nehmen, bei denen auch formale Korrektheit wichtig ist. Der Translator erfüllt seinen Zweck, wenn es um die reine Verständigung auf informeller Ebene geht.
- (29) Eigentlich sollte man sich schon als Übersetzer und/oder Dolmetscher der zukünftigen Konkurrenz und Gefahr bewusst sein. Da aber Maschinen nie bestimmte Vokabeln im richtigen Kontext finden werden, oder es ihnen schwer fallen wird umgangssprachlich zu kommunizieren, denke ich, dass es weiterhin Übersetzer und Dolmetscher geben wird. Institutionen wie die UNO oder EU werden hoffentlich nie auf Übersetzer verzichten. Wenn dann eher Unternehmen, die sparen wollen und die Arbeit nicht als professionell genug ansehen.

Die Gründe für die Bewertung des Verhältnisses von Skype Translator und menschlichen Übersetzer·innen und Dolmetscher·innen als verschiedene Welten beruhen vor allem auf den wahrgenommenen sprachlichen Unzulänglichkeiten der MÜ-Ausgabe. Alle Proband·innen, die sich dieser Einschätzung angeschlossen haben, liefern Argumente mit lexikalischen Bezug. So wird mangelnder Kontextbezug der MÜ-Ausgabe angemerkt. Auch die Übertragung von Redewendungen wird als Schwachstelle identifiziert.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

- (30) Übersetzer können sich dem Gesprächskontext bewusst sein, was dem Translator nicht möglich ist. Manche Wörter haben in verschiedenen Kontexten ebenso verschiedene Bedeutungen, was von dem Translator möglicherweise nicht erkannt werden kann.
- (31) Da ich selbst schon einige Male dolmetschen musste (deutsch-englisch) meine ich, dass ein viel gewichtigerer Teil auf der Überbringung einer Botschaft und eines sinnvollen Inhaltes beruht, als nur auf der wort-für-wort Übersetzung eines Inhaltes.

Als letzten Punkt des Ausgangsfragebogens sollten die Studienteilnehmerinnen angeben, ob sie bereit wären, Geld für derartige Technologie auszugeben. Hier antworteten nur vier Personen mit Ja, während alle weiteren sich enthielten bzw. diese Bereitschaft verneinten. Zwei Personen gaben an, ein monatliches Bezahlungsmodell unterstützen zu würdigen. Hierbei wären sie bereit, sieben bzw. zehn Euro zu bezahlen. Dieselben Personen gaben an, für ein Jahresabonnement 50 bzw. 60 Euro bezahlen zu wollen. In Form einer Einmalzahlung gaben insgesamt vier Personen eine Spanne von 20 bis 100 Euro an.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

In Tabelle 6.3 sind grundlegende Informationen zu den einzelnen Bildschirmmitschnitten im Setting Katalanisch-Deutsch aufgeführt. Die Mehrheit der Proband·innen hat länger als die ursprünglich veranschlagten 15 Minuten mit der Textchat-Kommunikation verbracht, lediglich die Person mit dem Pseudonym TN4 hat das Gespräch bereits nach ca. 12 Minuten beendet. Jeder einzelne Chatbeitrag wurde dabei – wie in Abschnitt 5.2.5 (S. 88) dargelegt – mit einem dynamischen AOI versehen und entsprechend mit einem Label versehen. Die divergierende Anzahl an originalen Beiträgen und maschineller Übersetzung ergibt sich aus einer bislang nicht erklärbaren Eigenheit des Skype Translators, zuweilen die MÜ-Ausgabe der Beiträge zu verschmelzen. Generell schwankt die Anzahl an Beiträgen auf Seiten der Proband·innen zwischen 6 und 48 ($\bar{x} 21$, Median = 20, SD = 9,60) und auf Seiten der katalanischen Gesprächspartner·innen zwischen 13 und 46 ($\bar{x} 27$, Median = 25, SD = 10,85). Die starken Unterschiede und die hohe Standardabweichung zeugen von einer unterschiedlichen Produktivität und Aktivität während der Kommunikationssituation. Eine Durchsicht der Bildschirmmitschnitte zeigt, dass keine·r der Studienteilnehmer·innen willentlich während der Bildschirmaufzeichnung das Chatfenster minimiert oder die Scroll-Funktion verwendet hat. Lediglich bei einer Aufnahme ist ersichtlich, dass die betreffende

6 Analyse

Person – vermutlich unabsichtlich – einmalig im Chat nach oben gescrollt ist, sodass kurzzeitig die jüngsten Nachrichten nicht von der Bildschirmaufnahme erfasst wurden.

In Kapitel 5.2.5 (S. 91) wurde bereits die Möglichkeit offenbart, sog. *fixation heatmaps* auf Grundlage der Fixationen zu generieren. Neben der Möglichkeit der Eingrenzung des notwendigen Annotationsbereichs für dynamische AOI können derartige Visualisierungen auch für eine grundlegende Inspektion des Leseverhaltens genutzt werden. Die Proband·innen schauten überwiegend in das untere linke drittel des Chatfensters, knapp oberhalb der Eingabemaske. Dort wird, wenn der Gegenüber tippt, der Schriftzug „schreibt gerade...“ eingeblenkt. Knapp darüber wiederum erscheinen die Chatbeiträge. Die visuelle Inspektion der Eye-Tracking-Daten deutet darauf hin, dass das rechte untere Drittel des Bildschirms, dort, wo die eigenen abgesendeten Chatbeiträge erscheinen, wenig beachtet wird. Ein vergleichbares Leseverhalten haben Schlösser u. a. (2018: 4) in Bezug auf kooperative, geteilte Chat-Systeme festgestellt. Bisson u. a. (2014: 413) kommen ferner zu dem Schluss, dass das Leseverhalten auch als Ergebnis eines Gewöhnungsprozesses angesehen werden kann: Die Teilnehmer·innen der Lesestudie von Untertiteln zeigten je nach Sprache der Untertitel ein unterschiedliches Leseverhalten. Auf den Skype Translator bezogen, könnte dementsprechend ebenfalls angenommen werden, dass die Proband·innen ebenfalls durch ihre Kommunikationsgewohnheiten im Vorhinein geprägt sind. Dies könnte sich beispielsweise am Umgang mit der relativ kleinen Schrift im Messenger wider spiegeln oder dem Aufbau der Beiträge als Blöcke.

Weiterhin wägen Bisson u. a. (2014: 414 f.) die Salienz des untersuchten Elements ab. Während es in der Studie um die sich stetig ändernden Untertitel in Kombination mit dem jeweiligen Bild ging, ist der saliente Bereich des Skype Translators vor allem – erneut – der untere, linke Bereich des Bildschirms, knapp oberhalb der Eingabemaske.

6.3.1 Visuelle Inspektion der Bildschirmmitschnitte

Die visuelle Inspektion der Bildschirmmitschnitte dient der einführenden Verortung der gewonnenen Eye-Tracking-Daten. Nachfolgend werden deshalb die Fixationsmuster (Abbildung 6.11), die Sakkaden (Abbildung 6.12) und die Blinzler (Abbildung 6.13, S. 126) einer exemplarisch ausgewählten Versuchsperson analysiert. Die bereits in Abschnitt 5.2.5 dargestellte Abbildung 5.3 (S. 91) gehört ebenfalls dazu.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.3: Übersicht über die Bildschirmaufzeichnungen im Setting Kat-De ohne AOI der Eingabe

Pseudonym	Dauer in min	Anzahl AOI	Anzahl Chatbeiträge			
			GerO	CatMT	CatO	GerMT
TN3	22,20	44	10	10	12	12
TN4	11,79	52	16	16	10	10
TN5	21,41	36	7	7	11	11
TN6	18,56	86	13	13	30	30
TN7	16,98	76	18	18	20	20
TN8	16,27	150	26	26	49	49
TN9	19,30	112	23	23	33	33
TN11	16,43	176	42	42	46	46
TN12	17,69	118	24	24	35	35
TN13	15,90	94	22	22	25	25
TN15	17,77	94	15	15	32	32
TN16	20,18	106	25	25	28	28
TN17	16,71	76	20	20	18	18
TN18	17,73	58	8	8	21	21
TN19	20,18	126	25	25	38	38
TN20	17,71	109	20	19	35	34
TN21	17,28	84	20	20	22	22
TN22	18,41	78	18	18	21	21
TN23	16,63	88	24	24	20	20
TN24	20,23	60	14	14	16	16
TN25	15,77	78	21	21	17	17
Gesamt	375,13	1.901	411	411	540	539
Ø Anzahl Beiträge (gerundet)						
20 20 26 26						
Median Anzahl Beiträge						
20 20 22 22						
Standardabweichung (SD)						
7,61 7,61 10,91 10,87						

6 Analyse

Charakteristisch für alle drei Darstellungen ist die Dichte der erfassten Daten im unteren Drittel des Bildschirmfotos, jedoch knapp oberhalb der Eingabemaske. Am dichtesten sind die Markierungen mittig links verteilt, dort wo neue Nachrichten des Gegenübers und der MÜ erscheinen. Selbstverständlich werden über den Verlauf der Studie auch vereinzelt Fixationen, Sakkaden oder Blinzler außerhalb dieses eng umrissenen Bereichs erfasst, bilden in Relation dazu jedoch eher die Ausnahme.

Die einzelnen Zahlen, die an den kreisrunden Markern in Abbildung 6.11 erkennbar sind, stellen die chronologische Abfolge der Fixation dar. Je größer die Markierung ist, desto länger wurde von der Versuchsperson auf den Bereich geblickt. Geht man rein von dem Muster aus, das die Fixationen annehmen, so fällt neben der bereits oben angesprochenen Fokussierung auf den unteren Bildschirmbereich oberhalb der Eingabemaske weiterhin auf, dass die Mehrzahl der Nachrichten relativ kurz gewesen sein muss. Dafür spricht der beinahe komplett undurchsichtige blaue Bereich.

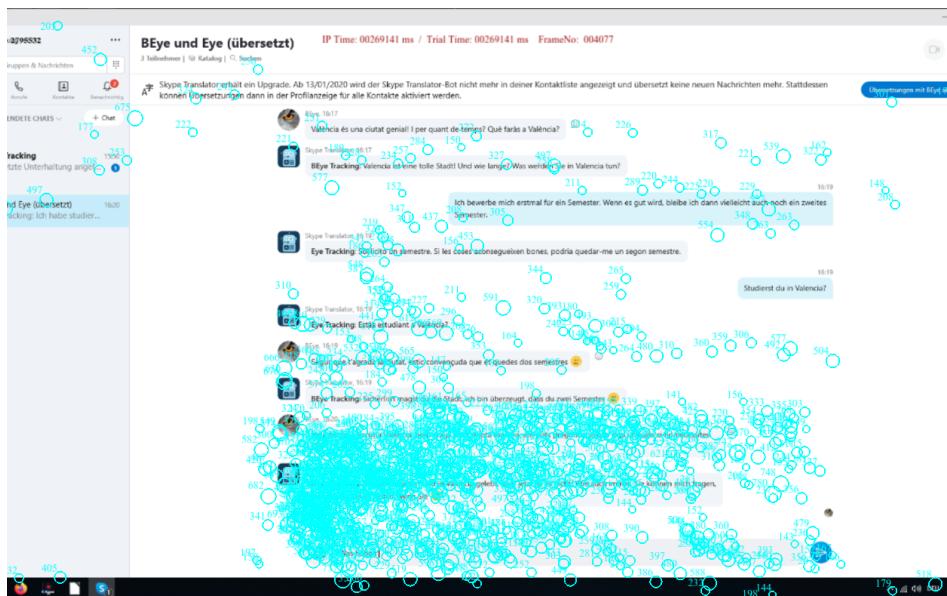


Abbildung 6.11: Fixationen während einer Bildschirmaufnahme

6.3 Der Datensatz im Setting Katalanisch-Deutsch

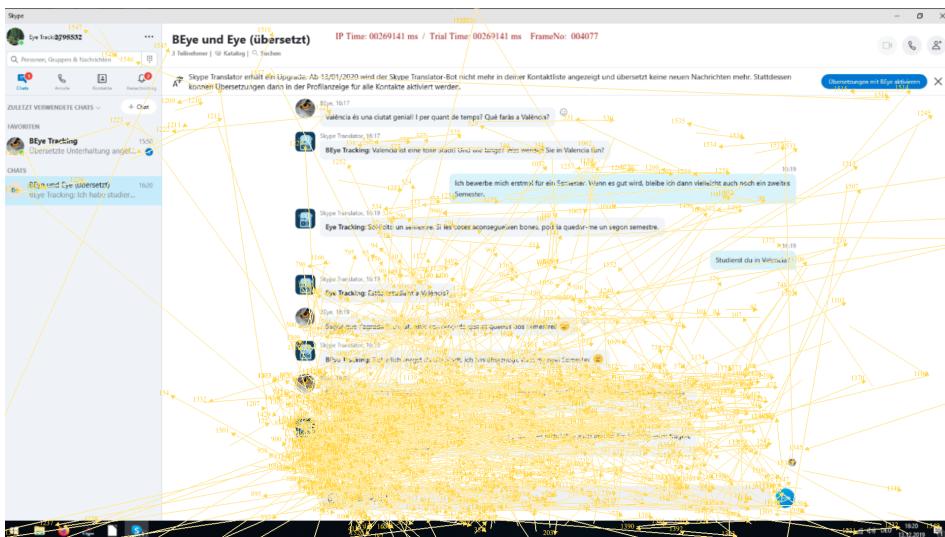


Abbildung 6.12: Sakkaden während einer Bildschirmaufnahme

Die Sakkaden in der Abbildung 6.12 bilden ebenfalls ein dichtes Netz analog zu dem Muster, das die Fixationen abgeben. Vereinzelt führen Sakkaden über den Bereich hinaus. Das kann verschiedenen Gründe haben. Die Proband·innen mussten womöglich auf eingehende Nachrichten warten und ließen den Blick über die bestehenden Nachrichten wandern. Ebenso können diese Sakkaden ein vergleichendes Leseverhalten beschreiben, bei dem zu bereits erfassten Informationen zurückgesprungen wurde. Die Nummern an den Spitzen der Feile geben den chronologischen Verlauf der Sakkaden an (den sog. *scan path*). Der wichtigste Bereich ist hier ebenfalls mittig links im unteren Bildschirmbereich oberhalb der Eingabemaske.

Die visuelle Aufbereitung der Blinzler während einer Eye-Tracking-Session kann ebenfalls als Indikator für kognitive Auslastung genommen werden. Die roten Linien bilden die Distanz ab, die der Eye-Tracker nicht erfassen konnte, sprich: Der Beginn der Linie ist der Punkt, an dem letztmalig eine Reflektion der Pupille aufgezeichnet wurde und das Ende der Linie der Punkt, an dem darauf folgend wieder eine Reflektion eingefangen wurde. Somit erklären sich auch die roten Linien, die über den Bildschirmrand hinausgehen. Hier ist zu vermuten, dass die Versuchsperson durch einen Reiz im Versuchslabor abgelenkt wurde. Zunächst ist jedoch auch hier festzuhalten, dass sich die erfassten Blinzler oberhalb der Eingabemaske mittig links verdichten. Wie bereits weiter oben erwähnt, werden an dieser Stelle die MÜ-Ausgabe sowie die eingehenden Nachrichten des Gegenübers angezeigt.

6 Analyse

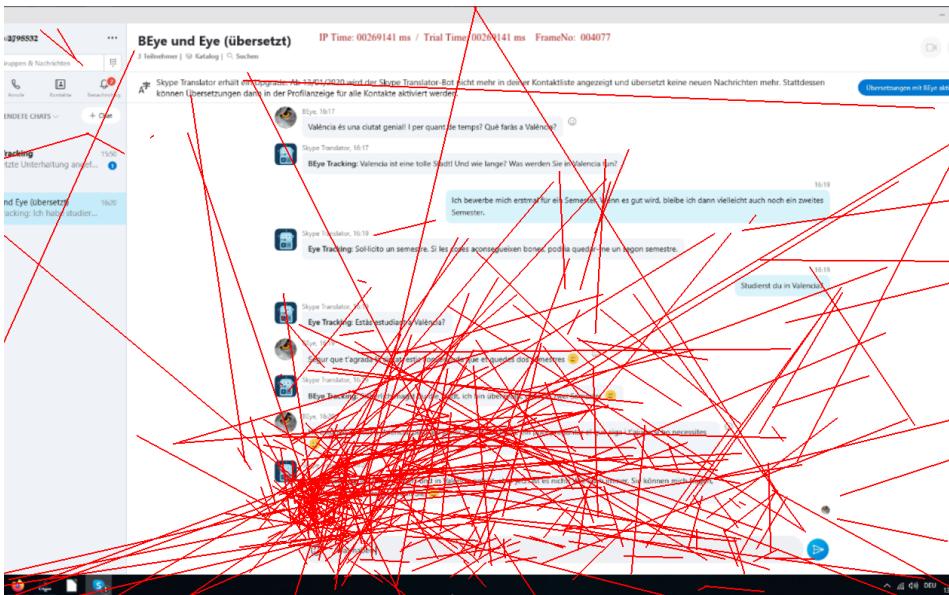


Abbildung 6.13: Blinzler während einer Bildschirmaufnahme

6.3.2 Fixationen der Probanden

6.3.2.1 Fixationsanzahl

Tabelle 6.4 zeigt die aufsummierte Anzahl an Fixationen pro AOI-Kategorie. Insgesamt wurden 21.196 Fixationen in den mit AOI versehenen Bereichen erfasst. Der Gesamtschnitt beträgt 12,96 Fixationen pro AOI. Die maschinelle Übersetzung ins Deutsche (*GerMT*) erhielt mit 7.037 Fixationen etwa zweieinhalb Mal so viele wie die deutschsprachigen Originale (2.469). Auch die maschinelle Übersetzung ins Katalanische erhielt mehr Fixationen (4.553) als die katalanischsprachigen Originale (3.046). Das Eingabefeld wurde 4.097 Mal fixiert. Der Durchschnittswert für die katalanischen Originalbeiträge ist mit 6,99 am kleinsten. Auch die durchschnittliche Fixationsanzahl auf die deutschen Originalbeiträge liegt mit 8,20 unter dem Gesamtschnitt. Nur die mittlere Anzahl an Fixationen, die auf die MÜ ins Deutsche und auf die Eingabemaske verfallen, liegen über dem Schnitt (jeweils 13,8 und 195,1).

Sowohl die visuelle Inspektion als auch die statistische Untersuchung deuten darauf hin, dass die Fixationsanzahl nicht normalverteilt ist. Ein Shapiro-Wilk-Test ergibt für den gesamten Datensatz ($W = 0,36, p < 0,001$). Eine logarithmische Transformation ändert an dieser Beobachtung nichts (Shapiro-Wilk

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.4: Summe, Mittelwert, Median und SD der Fixationsanzahl pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
GerO	2.469	8,20	4	15,05
CatMT	4.552	12,34	7	17,85
CatO	3.041	6,99	3	10,97
GerMT	7.037	13,8	9	16,45
Eingabe	4.097	195,1	188	102,29
Global	21.196	12,96	6	28,24

Test, \log_{10} : $W = 0,97$, $p < 0,001$). Deshalb wird die Fixationsanzahl mit nicht-parametrischen Tests analysiert.

Ein Kruskal-Wallis-Test zeigt, dass die Fixationsanzahl pro AOI nach den einzelnen Teilnehmer:innen variiert ($\chi^2(20) = 252,11$, $p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 132 von 210 TN-Gruppen (62,86 %) signifikant unterscheiden. Ein zweiter Kruskal-Wallis-Test zeigt, dass die Fixationsanzahl von der betrachteten AOI-Kategorie beeinflusst wird ($\chi^2(4) = 245,9$, $p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.5) zeigen, dass sich alle bis auf eine Gruppierung (CatO-GerO) voneinander unterscheiden, sodass gefolgert werden kann, dass sich die zentrale Tendenz aller AOI-Kategorien mit Ausnahme der Gruppierung CatO-GerO signifikant unterscheiden. Es handelt sich um schwache bis mittlere Effekte nach Cohen (1992) mit $0,1 < r < 0,5$.

Weiterhin ergibt ein durchgeführter Mann-Whitney-U-Test, dass sich die zentrale Tendenz der Fixationsanzahl pro AOI-Kategorie unter Beteiligung der progressiven ersten Fixation signifikant unterscheidet. Im Falle des gesamten Datensatzes ($U = 360.541,00$, $p < 0,001$) wurden mehr Fixationen getätigkt, wenn die AOI in konsekutiver Reihenfolge betreten wurden (Median: 7) als wenn zunächst ein AOI mit höherer Ordnungszahl betrachtet wurde (Median: 4). Nach AOI-Kategorie unterteilt betrachtet ergeben sich folgende Werte: Der Mann-Whitney-U-Test für das deutsche Original ($U = 21.050,00$, $p < 0,01$) deutet ebenfalls darauf hin, dass signifikante Unterschiede in der zentralen Tendenz der Fixationsanzahl unter Beachtung der progressiven ersten Fixation vorliegen. In konsekutiver Reihenfolge beträgt der Median der Fixationsanzahl 5, nach Betrachtung eines AOI mit höherer Ordnungszahl 3. Auch bei der Untersuchung der MÜ ins Katalanische weist der Mann-Whitney-U-Test signifikante Werte ($U = 11.844,00$,

6 Analyse

Tabelle 6.5: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Fixationsanzahl nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	8,105397	0,0000 ***	0,29
CatMT-Eingabe	-6,479862	0,0000 ***	0,33
CatO-Eingabe	-9,074030	0,0000 ***	0,42
CatMT-GerMT	-3,327846	0,0005 **	0,11
CatO-GerMT	-12,27414	0,0000 ***	0,4
Eingabe-GerMT	5,507214	0,0000 ***	0,24
CatMT-GerO	6,185499	0,0000 ***	0,21
CatO-GerO	-1,243531	0,1068	
Eingabe-GerO	8,569340	0,0000 ***	0,37
GerMT-GerO	9,738661	0,0000 ***	0,34

$p < 0,05$) auf. Hier beträgt der Median für Fixationen in konsekutiver Abfolge 7, nach Betrachtung von AOI höheren Ranges 5. Im Falle des katalanischen Originals ergibt der Test ebenso signifikante Unterschiede ($U = 44.397,50, p < 0,001$), wobei der Median für die Fixationsanzahl in chronologischer 4 und in ungeordneter Reihenfolge 3 beträgt. Die abschließende Betrachtung der MÜ ins Deutsche weist auch signifikante Werte auf ($U = 21.282,00, p < 0,01$). Der Median der Fixationsanzahl bei konsekutiver Abfolge beträgt 10, bei ungeordneter Reihenfolge 7.

Zur genaueren Einordnung sind in Tabelle 6.6 (S. 129) Summe, Mittelwert und Standardabweichung der AOI-Größe in Pixeln pro AOI-Kategorie aufgeführt. Die Analyse der AOI-Größe ist jedoch aufgrund der in Abschnitt 5.2.2.1 (S. 82) beschriebenen technischen Ungenauigkeit der Annotation nur eine Annäherung. Dabei zeigt ein Korrelationstest nach Spearman, dass die Anzahl an Fixationen pro AOI-Kategorie positiv und signifikant mit der Größe des AOI zusammenhängt ($r_s = 0,52, p < 0,01, n = 349.055.310$). Das Bestimmtheitsmaß beträgt 27,22 %. Der Test geht dabei von einer nicht normalen Verteilung der Daten aus. Spearman's Rho nimmt Werte zwischen -1 und 1 ein: -1 deutet auf eine stark negative Korrelation hin, 0 zeigt keinerlei Beziehung zwischen den Variablen an und 1 verweist auf eine stark positive Korrelation⁵.

⁵S. hierzu <http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r>, letzter Aufruf am 31. Januar 2022.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.6: Summe, Mittelwert und SD der AOI-Größe in px pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	SD
GerO	8.437.392	28.031,20	19.022,38
CatMT	10.410.362	28.212,36	16.537,87
CatO	10.702.559	24.603,58	18.337,66
GerMT	15.975.212	31.323,95	17.908,85
Eingabe	1.462.981	69.665,76	5.164,34
Gesamt	46.988.507	28.721,58	18.592,74

6.3.2.2 Dauer der ersten Fixation

Wie in Tabelle 6.7 und Abbildung 6.14 ersichtlich wird, liegt die durchschnittliche Dauer der ersten Fixation auf den eingehenden Nachrichten 15 ms unter dem Gesamtschnitt, wohingegen die Fixationsdauer auf den eigenen Nachrichten 7 ms überhalb des Schnittes liegt. Der Wert bei Betrachtung der Eingabemaske liegt lediglich 2 ms vom Durchschnitt entfernt. Mit durchschnittlich 304 ms ist die Dauer der ersten Fixation auf den katalanischen Originalen auffällig lange.

Tabelle 6.7: Mittelwert, Median und SD der Dauer der ersten Fixation pro AOI-Kategorie

AOI-Kategorie	Mittelwert	Median	SD
GerO	298,98	245,0	157,89
CatMT	288,61	234,5	162,41
CatO	304,13	257,0	152,42
GerMT	277,94	234,0	141,96
Eingabe	290,26	228,0	174,23
Global	292,01	242,0	153,23

Sowohl die graphische als auch statistische Inspektion der Daten deutet darauf hin, dass keine Normalverteilung vorliegt. Eine logarithmische Transformation nähert die Verteilung zwar der Gausschen Glockenkurve an, ist jedoch noch immer leicht verschoben (s. Abbildung 6.15). Auch eine Berechnung der Schiefe (normal: 1,67, log-transformiert: 0,46) und der Kurtosis (normal: 3,30, log-transformiert: -0,25) sowie der Shapiro-Wilk-Test (normal: $W = 0,86, p < 0,001$,

6 Analyse

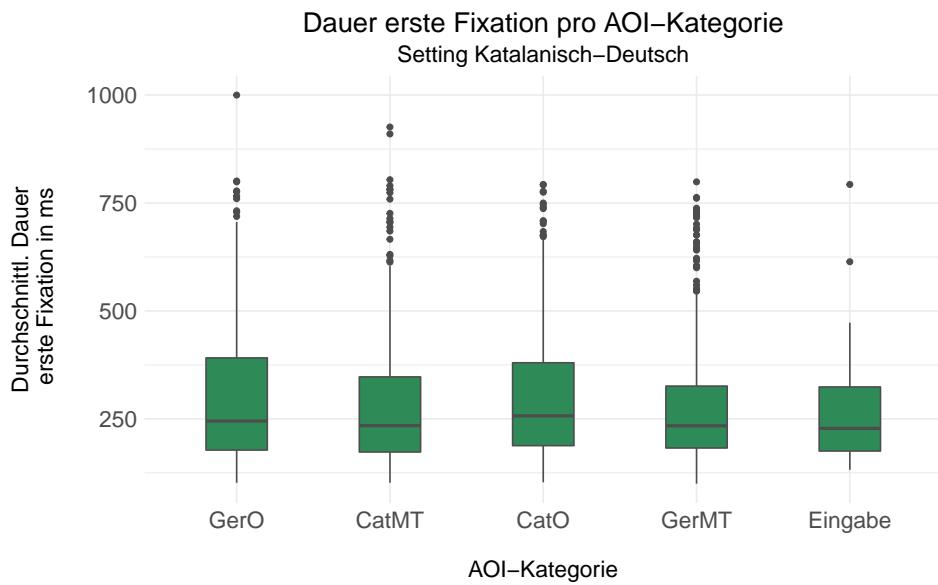


Abbildung 6.14: Durchschnittliche Fixationsdauer pro AOI

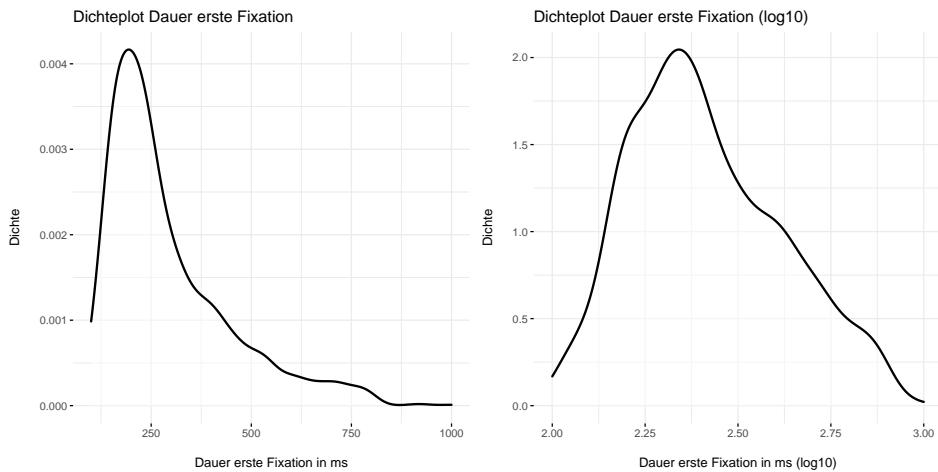


Abbildung 6.15: Verteilung der Dauer der ersten Fixation: normal (links), logarithmisch transformiert (rechts)

6.3 Der Datensatz im Setting Katalanisch-Deutsch

log-transformiert: $W = 0,96, p < 0,001$) deuten darauf hin, dass der Datensatz nicht normalverteilt ist. Deshalb werden für die statistische Analyse dieser Daten nicht-parametrische Tests verwendet.

Ein Kruskal-Wallis-Test zeigt, dass Unterschiede in der zentralen Tendenz der Fixationsdauer zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(6) = 76,06, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 67 von 210 TN-Gruppen (31,9 %) signifikant unterscheiden. Es kann gefolgert werden, dass die Fixationsdauer in gewissem Maße durch die Teilnehmer·innen beeinflusst wird.

Ein weiterer Kruskal-Wallis-Test in Kombination mit anschließend durchgeführten Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.8) zeigt, dass keine signifikanten Unterschiede in der zentralen Tendenz der Dauer der ersten Fixation zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 8,73, p = 0,07$). Auch bei testweisem Ausschluss der Kategorie „Eingabe“ verändern sich die Werte kaum. Hierbei ist jedoch zu beachten, dass – wie bereits im empirischen Teil problematisiert – die Schrift- bzw. Fenstergröße der einzelnen AOI zu gering ist, als dass der Eye-Tracker die Fixationen pro Wort hätte präzise festhalten können. Dem entgegenzuhalten ist jedoch, dass die dynamischen AOI ohnehin nicht auf Wortebene konzipiert sind, sodass die Dauer der ersten Fixation pro AOI weniger für die anfängliche Verarbeitung eines Wortes steht, sondern vielmehr auf den ganzen Textbaustein zu beziehen ist.

Tabelle 6.8: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Dauer der ersten Fixation nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	-2,327723	0,0498	-
CatMT-Eingabe	0,080594	0,5849	-
CatO-Eingabe	0,812292	0,3472	-
CatMT-GerMT	0,049745	0,4802	-
CatO-GerMT	2,561243	0,0521	-
Eingabe-GerMT	-0,065733	0,5264	-
CatMT-GerO	-1,274422	0,2531	-
CatO-GerO	0,880663	0,3785	-
Eingabe-GerO	-0,516368	0,4326	-
GerMT-GerO	-1,401717	0,2683	-

6 Analyse

Weiterhin sollen die Unterschiede zwischen der Fixationsdauer und der progressiven ersten Fixation untersucht werden, die allerdings nur in zwei Ausprägungen vorliegt (0 und 1). Es wird daher nicht der Kruskal-Wallis-Test, sondern der Mann-Whitney-U-Test verwendet. Es besteht ein signifikanter Unterschied in der zentralen Tendenz der Dauer zwischen den AOI, bei denen zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, und denen, die chronologisch betreten wurden. Die eigenen Chatbeiträge der Versuchspersonen werden kürzer fixiert, wenn zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, als wenn es sich um eine tatsächliche chronologische erstmalige Fixation handelt (Rangmediane 808,64 zu 706,78). Ein exakter Mann-Whitney-U-Test ($U = 426.963,00, p < 0,01, r = 0,11$) in Verbindung mit einer Untersuchung der Effektstärke nach Cohen (1992) zeigt dabei allerdings nur einen schwachen Effekt.

Betrachtet man die Art der ersten Fixation nach AOI-Kategorie getrennt, so weist lediglich die zentrale Tendenz der progressiven ersten Fixation bei Betrachtung der maschinellen Übersetzung ins Deutsche (*GerMT*) einen signifikanten Unterschied auf ($U = 22.014,00, r < 0,01, r = 0,16$, s. Tabelle 6.9). Auch hier ist die Effektstärke nach Cohen (1992) nur schwach. Erfolgt die Fixation, nachdem ein AOI mit höherer Ordnungszahl betreten wurde, so dauert sie länger als wenn die Fixation chronologisch erfolgte (Rangmediane 212,03 zu 265,23).

Tabelle 6.9: Ergebnisse des Mann-Whitney-U-Tests: Vergleiche der Dauer der ersten Fixation nach AOI-Kategorie und progressiver ersten Fixation

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	Effekt
GerO	21.981,0	0,26	
CatMT	12.603,0	0,25	
CatO	47.480,5	0,07	
GerMT	22.014,0	0,001 ***	0,16
Eingabe	n/a	n/a	
Global	426.963,0	0,001 ***	0,11

Ein durchgeführter Korrelationstests (Spearman's Rho) belegt, dass signifikante Zusammenhänge zwischen Fixationsdauer und der Größe der AOI bestehen ($r_s = -0,14, p < 0,01, n = 6,31 \times 10^8$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 38,02 %.

6.3.2.3 Regressionen

6.3.2.3.1 Eingehende Regressionen

Abbildung 6.16 und Tabelle 6.10 (S. 134) zeigen die aufsummierte Anzahl aller Regressionen in ein AOI nach Kategorie sowie deren Mittelwert und Standardabweichung. Auf das katalanische Original und die Eingabemaske verfielen dabei die meisten Regressionen (*CatO*: 504, \bar{x} 1,16, SD: 1,52, *Eingabe*: 548, \bar{x} 26,1, SD: 14,6), die wenigsten Regressionen wurden in AOI der MÜ ins Katalanische gemacht (*CatMT*: 249, \bar{x} 0,67, SD: 1,09). Das deutsche Original und die maschinelle Übersetzung ins Deutsche erhielten jeweils 312 (\bar{x} 1,04, SD: 2,39) bzw. 332 (\bar{x} 0,65, SD: 1,07) Regressionen. Im Schnitt wird jeweils auf das deutsche und das katalanische Original also somit mindestens ein Rücksprung ausgeführt. Die Eingabemaske weist einerseits viele eingehende Regressionen in absoluten Zahlen auf, andererseits sind auch der Mittelwert mit 26,1 und die Standardabweichung mit 14,6 auffällig hoch.

Zur Untersuchung der Regressionen in ein AOI wurde mit einem binomialen logistischen Regressionsmodell mit gemischten Effekten gearbeitet, das die Teilnehmer:innen (*SessionLabel*) als Zufallseffekte beinhaltete (s. Tabelle 6.11). Die Variablen wurden in einer bedingten Vorwärtsauswahl in das Modell integriert. Das simpelste und zugleich genauste Modell erweist sich dabei ebenso wie einzelne Koeffizienten der integrierten Variablen als signifikant und liefert bessere Ergebnisse als ein einfaches lineares Regressionsmodell ($\chi^2(9) = 50,91, p < 0,001$). Zugleich zeigt sich, dass das Modell noch nicht optimal angepasst (*C*: 0,73, Somers' *D_{XY}*: 0,46) ist. Daher wurden die Ergebnisse noch einmal mittels nicht-parametrischer Tests auf Plausibilität überprüft.

Während die Kategorien *GerO* bzw. *GerMT* keinen signifikanten Einfluss auf die Wahrscheinlichkeit haben, dass eine Regression in ein AOI gemacht wird, steigt die relative Wahrscheinlichkeit bei *Eingabe* und *Intercept* (*CatMT* in diesem Fall) um jeweils 2.474 % bzw. 59 %. Für die Kategorie *CatO* steigt die relative Wahrscheinlichkeit, dass eine Regression in dieses AOI gemacht wird, um 70 %. Weiterhin steigt die Wahrscheinlichkeit um 42 % mit der Größe der AOI. Hingegen nimmt die relative Wahrscheinlichkeit mit jeder Einheit der *IARegPD* um 16 % und mit jeder progressiven ersten Fixation um 63 % ab.

Die zur Überprüfung durchgeführten Chi-Quadrat-Tests deuten ebenfalls in die von dem Modell aufgezeigte Richtung: Es besteht ein signifikanter Unterschied zwischen der jeweils betrachteten AOI-Kategorie und der eingehenden Regression ($\chi^2(4) = 72,56, p < 0,001$). Die progressive erste Fixation steht zudem ebenfalls in einem signifikanten Verhältnis zu den eingehenden Regressionen.

6 Analyse

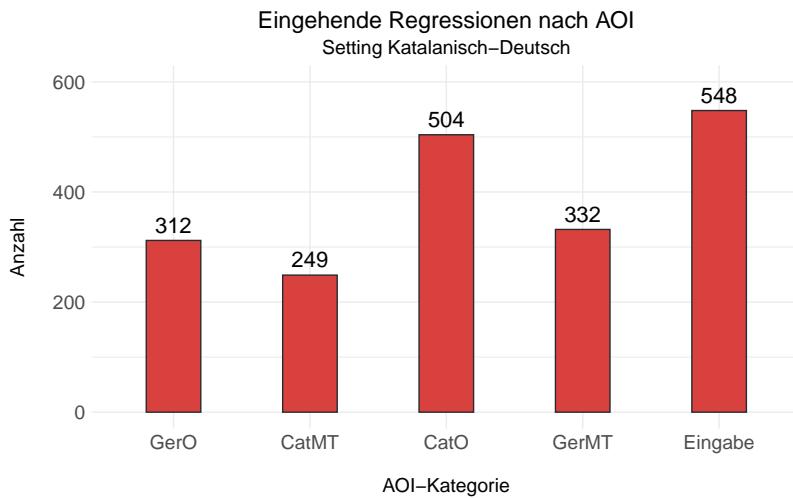


Abbildung 6.16: Anzahl Regressionen in ein AOI

Tabelle 6.10: Summe, Mittelwert und SD der eingehenden Regressionen pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	SD
GerO	312	1,04	2,39
CatMT	249	0,67	1,09
CatO	504	1,16	1,53
GerMT	332	0,65	1,07
Eingabe	548	26,1	14,6
Global	1.945	1,19	3,61

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.11: Werte des Regressionsmodells für eingehende Regressio-
nen im Setting Katalanisch-Deutsch

	Gruppen	Varianz	SD	L.R. X2	DF	Pr
Zufallseffekt(e)	SessionLabel	0,3167	0,5628	50,907	21	<0,001***
Fixe Effekt(e)	Schätzung	VIF	OR	SE	z	Pr(> z)
(Intercept)	0,49460		1,59	0,52049	1,976	<0,05*
IATagCatO	0,55883	1,6827326	1,70	0,16221	3,445	<0,001***
IATagEingabe	4,85341	1,8523138	25,74	1,93224	2,512	<0,05*
IATagGerMT	-0,16912	1,6189873	0,81	0,14887	-1,136	0,225943
IATagGerO	-0,09574	1,5730919	0,90	0,17642	-0,543	0,587328
IAFFixPro1	-0,95428	1,1235513	0,37	0,14752	-6,469	<0,001***
IARegPD_z	-0,29912	1,9849059	0,84	0,10146	-2,948	<0,01**
IA_AREA_z	0,29316	1,1041541	1,42	0,06123	4,788	<0,001***

Model statistics	Wert
No. Groups	21
Number of cases in model	1,636
Observed misses	
Observed deviance	2021,0
Residual deviance	
R2 (Nagelkerke)	0,1212020
R2 (McFadden)	0,0702412
R2 (Cox & Snell)	0,891032
C	0,7282527
Somers' D _{XY}	0,4565054
AIC	2039,0
BIC	2087,6
Prediction accuracy	
Model Likelihood Ratio Test	
x2:	152,68
df:	7
	p < 0,001***

6 Analyse

Sowohl die statistische Untersuchung des globale Datensatzes ($\chi^2(1) = 99,38, p < 0,001$) als auch nach den Kategorien *GerO* ($\chi^2(1) = 15,97, p < 0,001$), *CatMT* ($\chi^2(1) = 26,16, p < 0,001$) und *GerMT* ($\chi^2(1) = 27,84, p < 0,001$) weist signifikante Unterschiede auf. Die weiteren im Modell verwendeten Variablen führen hingegen zu keinerlei signifikantem Ergebnis.

6.3.2.3.2 Ausgehende Regressionen

Abbildung 6.17 und Tabelle 6.12 zeigen die aufsummierte Anzahl aller Regressionen aus einem AOI nach Kategorie sowie deren Mittelwert und Standardabweichung. Aus der maschinellen Übersetzung ins Deutsche und ins Katalanische wurden dabei die meisten Regressionen (*GerMT*: 627, $\bar{x} 1,23$, $SD: 1,68$ und *CatMT*: 397, $\bar{x} 1,08$, $SD: 2,31$) getätigt, die wenigsten Regressionen kamen aus dem AOI der Kategorie deutsches Original (*GerO*: 45, $\bar{x} 0,15$, $SD: 0,46$). Aus den Kategorien *Eingabe* und katalanisches Original wurden jeweils 108 ($\bar{x} 5,14$, $SD: 11,71$) bzw. 77 ($\bar{x} 0,18$, $SD: 0,64$) Regressionen vorgenommen. Im Durchschnitt wird jeweils mindestens eine Regression aus den maschinell übersetzten Nachrichten ins Katalanische und ins Deutsche getätigt. Aus der Eingabemaske werden im Schnitt sogar fünf Regressionen ausgeführt, der absolute Wert liegt hier jedoch mit 108 im unteren Bereich.

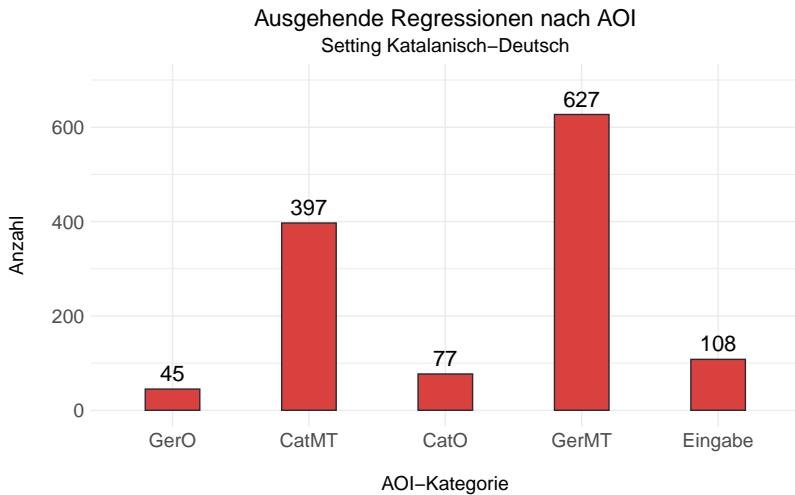


Abbildung 6.17: Anzahl Regressionen aus einem AOI

Zur Untersuchung der Regressionen aus einem AOI wurde mit einem binomialen logistischen Regressionsmodell mit gemischten Effekten gearbeitet, das

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.12: Summe, Mittelwert und SD der ausgehenden Regressio-nen pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	SD
GerO	45	0,15	0,46
CatMT	397	1,08	2,31
CatO	77	0,18	0,64
GerMT	627	1,23	1,68
Eingabe	108	5,14	11,71
Global	1.254	0,77	2,1

die Teilnehmer-innen (*SessionLabel*) als Zufallseffekte beachtete (s. Tabelle 6.13). Die Variablen AOI-Kategorie, AOI-Größe, progressive erste Fixation, Dauer des ersten Durchlaufs und regressive Durchlaufdauer wurden in einer bedingten Vor-wärtsauswahl in das Modell integriert. Das simpelste und zugleich genauste Mo-dell bietet dabei bessere Resultate als ein einfaches lineares Regressionsmodell ($\chi^2(9) = 191,82, p < 0,001$), erweist sich allerdings auch als nicht optimal ange-passt (C: 0,90, Somers' D_{XY}: 0,80). Das endgültige, simpelste Modell zeigt signifi-kante Werte bei einzelnen Kategorien der unabhängigen Variable *IATag*.

Während Regressionen aus der Eingabemaske und aus dem *Intercept* (*CatMT* in diesem Fall) keinen bzw. nur einen marginalen Effekt auf die Wahrscheinlichkeit haben, dass eine Regression aus dem AOI gemacht wird, weisen die Katego-rien *GerMT*, *GerO* und *CatO* eine jeweils signifikante Wahrscheinlichkeit auf. Aus der Kategorie *GerMT* steigt die Wahrscheinlichkeit um 93 %, aus den Katego-rien *GerO* und *CatO* hingegen sinkt die relative Wahrscheinlichkeit um jeweils 65 % bzw. 71 %. Die Größe der AOI ist signifikant ($p < 0,05$), die Wahrscheinlich-keit einer Regression aus dem AOI sinkt mit einer Einheit um 1 %.

Wie auch bei den eingehenden Regressionen wurde hier zur Überprüfung der Plausibilität mit Chi-Quadrat-Tests gearbeitet. So bestehen signifikante Unter-schiede zwischen den ausgehenden Regressionen pro AOI-Kategorie ($\chi^2(4) = 331,16, p < 0,001$). Die progressive erste Fixation ergibt ebenfalls auffällige Ergebnisse. Der gesamte Datensatz ($\chi^2(1) = 454,2, p < 0,001$) sowie die geson-derte Betrachtung nach AOI-Kategorie *GerO* ($\chi^2(1) = 43,25, p < 0,001$), *CatMT* ($\chi^2(1) = 76,77, p < 0,001$), *CatO* ($\chi^2(1) = 65,44, p < 0,001$) und *GerMT* ($\chi^2(1) = 177,42, p < 0,001$) liefern hier signifikante Resultate. Alle weiteren im Modell verwen-deteten Variablen erweisen sich darüber hinaus als statistisch unauffällig.

6 Analyse

Tabelle 6.13: Werte des Regressionsmodells für ausgehende Regressionen im Setting Katalanisch-Deutsch

	Gruppen	Varianz	SD	L.R. X2	DF	Pr
Zufallseffekte	SessionLabel	5,854	2,419	< 0,001 ***		
Fixe Effekte	Schätzung	VIF	OR	SE	<i>z</i>	Pr(> <i>z</i>)
(Intercept)	-0,93610		0,794	0,55865	-1,676	0,0938.
IATagCatO	-1,21271	1,3833037	0,292	0,21022	-5,769	<0,001 ***
IATagEingabe	-12,80930	1,0005803	0,000	57,24434	-0,224	0,8229
IATagGerMT	0,91356	1,454387	1,933	0,17529	5,212	<0,001 ***
IATagGerO	-1,09895	1,3365114	0,349	0,23442	-4,688	<0,001 ***
IA_AREA_z	0,17517	1,1163253	0,988	0,08277	2,116	0,0343*
IARegPD_z	4,36986	1,3180099	32,037	0,43993	9,933	<0,001 ***
IARunDwell_z	-0,31179	1,2305214	0,808	0,07934	-0,3930	<0,001 ***

Model statistics	Wert
No. Groups	21
Number of cases in model	1,636
Observed misses	
Observed deviance	1366,8
Residual deviance	
R2 (Nagelkerke)	0,452491
R2 (McFadden)	0,314851
R2 (Cox & Snell)	0,318824
C	0,9008282
Somers' D _{XY}	0,8016564
AIC	1384,8
BIC	1433,4
Prediction accuracy	
Model Likelihood Ratio Test	
x2:	628,12
df:	7
p:	< 0,001 ***

6.3.2.4 Dauer des ersten Durchlaufs

Tabelle 6.14 zeigt die aufsummierte Dauer des ersten Durchlaufs pro AOI-Kategorie und die entsprechenden Mittelwerte sowie Standardabweichungen. Während die Durchlaufdauer innerhalb der Eingabemaske und den deutschen Originalnachrichten mit jeweils 10.728 ms bzw. 179.586 ms am kürzesten ist, erreicht der Durchlauf der MÜ ins Deutsche den Höchstwert von 391.664 ms. Dies gilt ebenfalls für die durchschnittliche Zeit, die der erste Durchlauf pro AOI benötigt. Auch hier liegt die MÜ ins Deutsche mit 815,97 ms deutlich über dem Schnitt von 641,98 ms und auch dem deutschen Original mit 600,62 ms. Die beiden katalanischsprachigen Beitragsarten liegen mit durchschnittlich 611,18 ms (*CatMT*) und 508,86 ms (*CatO*) hingegen unter dem Gesamtschnitt. Die jeweilige Gesamt-durchlaufdauer liegt zwischen der des deutschen Originals und der der MÜ ins Deutsche.

Tabelle 6.14: Summe, Mittelwert, Median und SD der Dauer des ersten Durchlaufs in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
GerO	179.586	600,62	427,0	562,68
CatMT	221.246	611,18	426,5	558,06
CatO	218.808	508,86	380,0	451,91
GerMT	391.664	815,97	513,0	738,49
Eingabe	10.728	510,86	335,0	643,96
Global	1.022.032	641,98	427,0	607,56

Wie die graphische Inspektion des Datensatzes anhand von Abbildung 6.19 darstellt, kann nicht von einer Normalverteilung ausgegangen werden. Eine logarithmische Transformation (rechte Seite der Abbildung) ändert nichts an dieser Beobachtung. Auch die statistische Untersuchung des Datensatzes mittels Shapiro-Wilk-Test lehnt die Annahme einer Normalverteilung ab (roh: $W = 0,79$, $p < 0,001$ und log10: $W = 0,98$, $p < 0,001$).

Ein Kruskal-Wallis-Test bestätigt, dass Unterschiede in der zentralen Tendenz der Betrachtungsdauer im ersten Durchlauf pro AOI zwischen den einzelnen Teilnehmer*innen bestehen ($\chi^2(20) = 85,85$, $p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 70 von 210 TN-Gruppen (33,33 %) signifikant unterscheiden, sodass gefolgert werden kann, dass die Dauer im ersten Durchlauf in gewissem Maße von der Individualität der TN abhängt.

6 Analyse

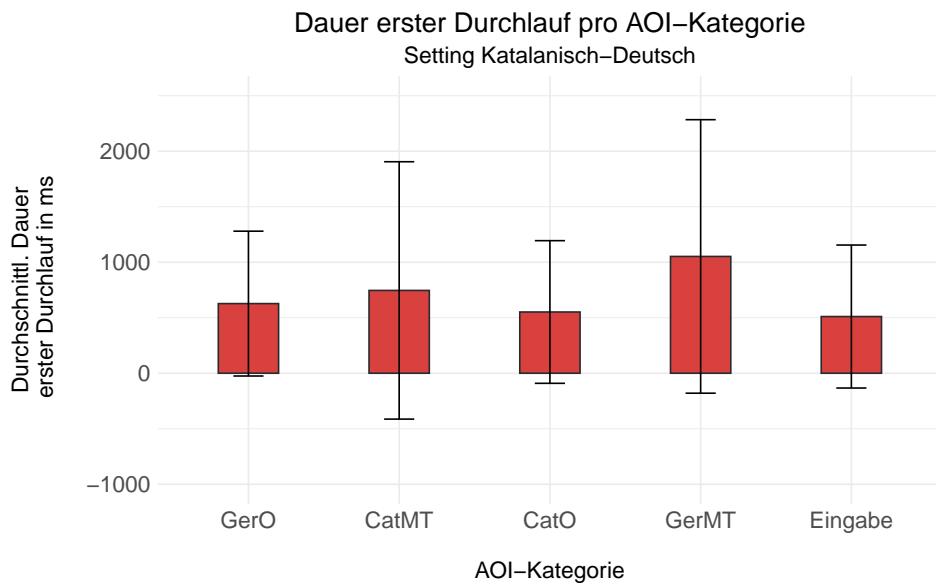


Abbildung 6.18: Durchschnittliche Betrachtungsdauer im ersten Durchlauf pro AOI-Kategorie im Setting Katalanisch–Deutsch

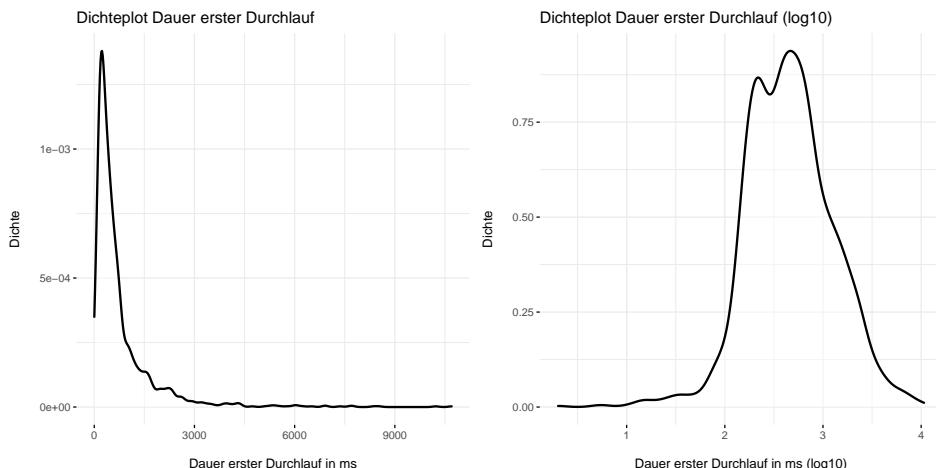


Abbildung 6.19: Verteilung der Betrachtungsdauer im ersten Durchlauf in ms: normal (links), logarithmisch transformiert (rechts)

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Ein weiterer Kruskal-Wallis-Test ergibt, dass Unterschiede in der zentralen Tendenz der Dauer des ersten Durchlaufs zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 34,05, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich vor allem die Gruppierungen mit Beteiligung der maschinellen Übersetzung ins Deutsche (*CatMT-GerMT*, *CatO-GerMT*, *Eingabe-GerMT* sowie *GerMT-GerO*) unterscheiden, sodass gefolgert werden kann, dass die Durchlaufdauer von der AOI-Kategorie, und speziell von der Richtung (ins Deutsche), abhängt (s. Tabelle 6.15). Die Effektstärke nach Cohen (1992) liegt bei allen Gruppierungen nur im schwachen Bereich.

Tabelle 6.15: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Dauer des ersten Durchlaufs nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	1,851890	0,640	
CatMT-Eingabe	1,467788	0,1015	
CatO-Eingabe	0,883115	0,2095	
CatMT-GerMT	-3,294221	0,0016 **	0,11
CatO-GerMT	-5,442975	0,0000 ***	0,20
Eingabe-GerMT	-2,506379	0,0152 *	0,11
CatMT-GerO	0,223453	0,4116	
CatO-GerO	-1,522364	0,1066	
Eingabe-GerO	-1,382031	0,1044	
GerMT-GerO	3,349606	0,0020 **	0,12

Weiterhin wurden die Unterschiede in der Dauer des ersten Durchlaufs zwischen den beiden Ausprägungen der progressiven ersten Fixation untersucht. Da die progressive erste Fixation nur in zwei Ausprägungen vorliegt (0 und 1), wird nicht der Kruskal-Wallis-Test, sondern der Mann-Whitney-U-Test verwendet. Es besteht allerdings kein signifikanter Unterschied in der zentralen Tendenz der Dauer des ersten Durchlaufs zwischen den AOI, bei denen zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, und denen, die chronologisch betreten wurden. Die Testergebnisse des gesamten Datensatzes und der einzelnen AOI-Kategorien sind in der Tabelle 6.16 aufgeführt.

Die Dauer der Fixationen im ersten Durchlauf korreliert signifikant mit der Größe der AOI ($r_s = 0,10, p < 0,01, n = 602.248.371$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 1,09 %.

6 Analyse

Tabelle 6.16: Ergebnisse des Mann-Whitney-U-Tests zur Dauer des ersten Durchlaufs nach progressiver erster Fixation und AOI-Kategorie

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	<i>r</i>
GerO	24.004,0	0,23	0,07
CatMT	13.022,5	0,61	0,03
CatO	50.083,5	0,81	0,01
GerMT	24.098,0	0,30	0,05
Gesamt	437.979,5	0,58	0,01

6.3.2.5 Gesamtverweildauer

In Tabelle 6.17 (S. 142) sind die Summe sowie die Mittelwert und Standardabweichung der Gesamtverweildauer in Millisekunden pro AOI-Kategorie dargestellt. Neben der Eingabemaske (12.000 ms) verzeichnen die deutschsprachigen Originalbeiträge die niedrigste Summe (667.000 ms).

Im Vergleich dazu ist die Dauer im Falle der MÜ ins Katalanische beinahe doppelt so lang (1.305.000 ms) bzw. im Falle der MÜ ins Deutsche sogar drei Mal so lang (1.944.000 ms). Die aufsummierte Gesamtdauer beträgt 4.848.000 ms. Die Werte der Eingabemaske lassen allerdings vermuten, dass es sich um einen Messfehler handelt oder die statische Annotation des AOI die Werte verzerrt. Auch mit Rückgriff auf die visuelle Inspektion, bei der sich eindeutig die Aufmerksamkeit auf den Bereich der Eingabemaske sowie leicht darüber konzentriert, lässt sich nur so der Mittelwert von etwa 5.800 ms bei einer aufsummierten Dauer von etwa 12.000 ms und einer Standardabweichung von 8.000 erklären.

Tabelle 6.17: Summe, Mittelwert, Median und SD der Gesamtverweildauer in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
GerO	666.787	2.230,06	1.137,0	3.311,21
CatMT	1.305.425	3.596,21	2.005,0	4.219,83
CatO	919.508	2.123,58	1.038,0	3.081,15
GerMT	1.944.324	3.865,46	2.766,0	3.815,79
Eingabe	11.954	5.977,00	5.977,0	8.066,67
Global	4.847.998	3.030,0	1.728,5	3.727,83

6.3 Der Datensatz im Setting Katalanisch-Deutsch

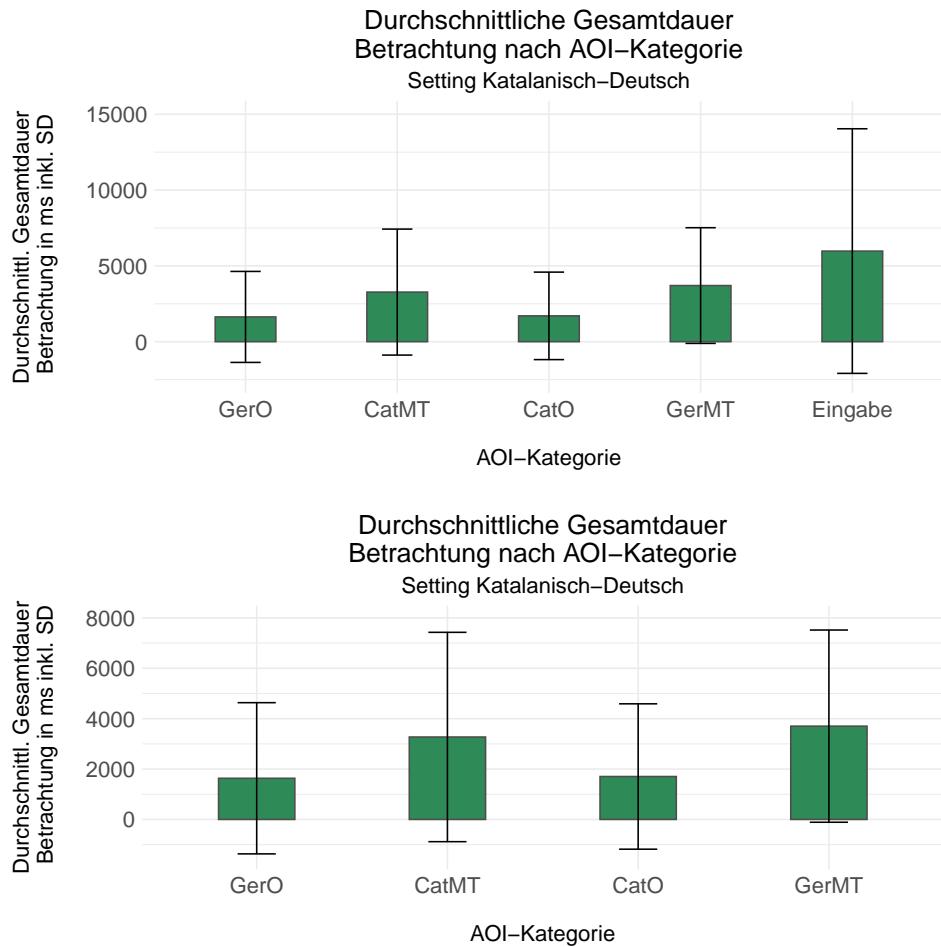


Abbildung 6.20: Durchschnittliche Betrachtungsdauer pro AOI-Kategorie im Setting Katalanisch-Deutsch, mit (oben) und ohne (unten) Eingabemaske

6 Analyse

Wie die graphische Inspektion des Datensatzes anhand von Abbildung 6.21 andeutet, kann nicht von einer Normalverteilung ausgegangen werden. Eine logarithmische Transformation (rechte Seite der Abbildung) ändert nichts an dieser Beobachtung. Auch die statistische Untersuchung des Datensatzes mittels Shapiro-Wilk-Test lehnt die Annahme einer Normalverteilung ab (roh: $W = 0,69, p < 0,001$ und \log_{10} : $W = 0,995, p < 0,001$).

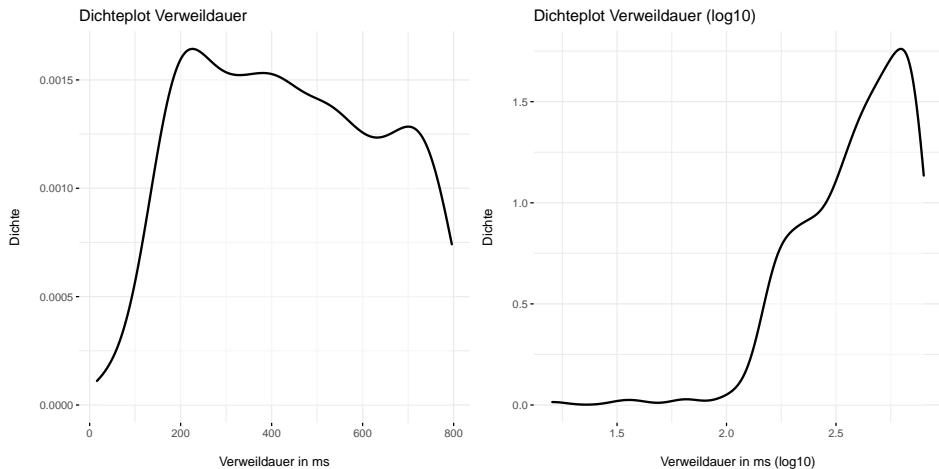


Abbildung 6.21: Verteilung der Betrachtungsdauer in ms: normal (links), logarithmisch transformiert (rechts)

Ein Kruskal-Wallis-Test belegt, dass Unterschiede in der zentralen Tendenz bei der Gesamtdauer der Betrachtung der AOI zwischen den einzelnen Teilnehmer*innen bestehen ($\chi^2(20) = 213,34, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 129 von 210 TN-Gruppen (61,43 %) signifikant unterscheiden. Es kann demnach gefolgert werden, dass die Betrachtungsdauer in gewissem Maße von der Individualität der TN abhängt.

Ein zweiter Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede bei der Gesamtbetrachtungsdauer pro AOI zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 170,7, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.18) zeigen, dass sich vor allem die Gruppierungen mit Beteiligung der maschinellen Übersetzung (*CatMT-CatO*, *CatMT-GerMT*, *CatO-GerMT*, *CatMT-GerO* sowie *GerMT-GerO*) unterscheiden, sodass gefolgert werden kann, dass die Betrachtungsdauer von der AOI-Kategorie, und speziell von der Richtung, abhängt.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.18: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Gesamtverweildauer nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	7,721200	0,0000 ***	0,27
CatMT-Eingabe	0,179601	0,4287	
CatO-Eingabe	-0,595598	0,3939	
CatMT-GerMT	-2,812706	0,0049 **	0,1
CatO-GerMT	-11,33659	0,0000 ***	0,37
Eingabe-GerMT	-0,453139	0,3614	
CatMT-GerO	6,331929	0,0000 ***	0,25
CatO-GerO	-0,730907	0,3874	
Eingabe-GerO	0,517520	0,3780	
GerMT-GerO	9,424509	0,0000 ***	0,33

Außerdem wurden die Unterschiede in der Gesamtbetrachtungsdauer nach progressiver erster Fixation untersucht. Da die progressive erste Fixation nur in zwei Ausprägungen vorliegt (0 und 1), wird nicht der Kruskal-Wallis-Test, sondern der Mann-Whitney-U-Test verwendet. Es besteht ein signifikanter Unterschied in der zentralen Tendenz der Betrachtungsdauer zwischen den AOI, bei denen zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, und denen, die chronologisch betreten wurden (s. Tabelle 6.19).

Tabelle 6.19: Ergebnisse des Mann-Whitney-U-Tests zur Gesamtverweildauer und progressiven ersten Fixation nach AOI-Kategorie

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	Effekt
GerO	21.197,0	0,019 *	0,14
CatMT	10.987,0	0,015 *	0,13
CatO	46.146,0	0,001 **	0,18
GerMT	22.032,5	0,092	
Eingabe	3,0	0,023 *	0,5
Global	364.732,5	0,001 **	0,23

Die Durchlaufdauer korreliert signifikant mit der Größe der AOI ($r_s = 0,45, p < 0,01, n = 374.884.161$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 20,33 %.

6.3.2.6 Regressive Durchlaufdauer

Tabelle 6.20 zeigt die Summe, Mittelwerte sowie die Standardabweichung der regressiven Durchlaufdauer in Millisekunden pro AOI-Kategorie. Während die Originalbeiträge im Deutschen insgesamt lediglich 397.000 ms fixiert werden, beträgt die aufsummierte regressive Durchlaufdauer bei der MÜ ins Deutsche 3.244.000 ms. Die MÜ wird in dieser Kategorie insgesamt also beinahe acht Mal so lange betrachtet. Auch die Durchlaufdauer des katalanischen Originals und der MÜ ins Katalanische unterscheidet sich etwa um den Faktor vier. Während das katalanische Original 695.000 ms lang betrachtet wird, liegt der Wert bei der MÜ in die Sprache bei etwa 2.533.000 ms. Insgesamt beträgt die regressive Durchlaufdauer 8.064.000 ms bzw. im Schnitt 4.929 ms.

Auch die Mittelwerte unterscheiden sich deutlich. So beträgt die durchschnittliche regressive Durchlaufdauer für die deutschen Originalbeiträge 1.318 ms, die der MÜ ins Deutsche hingegen 6.360 ms. ähnlich unterschiedlich sind die Werte für die katalanischsprachigen Beiträge. Die regressive Durchlaufdauer beträgt für die katalanischen Originale durchschnittlich 1.600 ms und für die MÜ in die Sprache 6.870 ms. Somit liegen die jeweiligen Originalbeiträge um etwa den Faktor fünf unter der Durchlaufdauer der MÜ. Während der absolute Höchstwert bei der MÜ ins Deutsche liegt, findet sich der höchste Schnitt bei der MÜ ins Katalanische.

Die Eingabemaske wurde als statisches AOI annotiert. Der hohe Mittelwert erklärt sich daher aus dem anfänglichen Warteverhalten der Versuchspersonen bis der katalanische Gegenüber die erste Nachricht gesendet hat. Bis zu diesem Zeitpunkt verweilten die Proband·innen meist sehr lange auf der Eingabemaske und damit innerhalb des AOI.

Eine graphische Inspektion der Daten (Abbildung 6.22) zeigt, dass weder die rohen noch die logarithmisch transformierten Werte normalverteilt sind. Auch die statistische Untersuchung unter Anwendung des Shapiro-Wilk-Tests deutet auf nicht normalverteilte Daten hin (roh: $W = 0,18$, $p < 0,001$, log10: $W = 0,99$, $p < 0,001$). Daher wird die regressive Durchlaufdauer mit nicht-parametrischen Tests untersucht.

Ein Kruskal-Wallis-Test belegt, dass Unterschiede in der zentralen Tendenz der regressiven Durchlaufdauer pro AOI zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(20) = 81,83$, $p < 0,05$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 61 von 210 TN-Gruppen (29,05 %) signifikant unterscheiden.

Ein zweiter Kruskal-Wallis-Test zeigt zudem, dass Unterschiede in der zentralen Tendenz der regressiven Durchlaufdauer pro AOI zwischen den einzelnen

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.20: Summe, Mittelwert, Median und SD der regressiven Durchlaufdauer in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
GerO	396.755	1.318,12	802,0	1.831,13
CatMT	2.532.654	6.863,56	3.292,0	11.960,89
CatO	694.504	1.596,56	710,0	3.244,24
GerMT	3.243.600	6.360,00	4.360,5	7.708,70
Eingabe	1.196.802	56.990,57	2053,0	128.990,12
Global	8.064.315	4.929,29	1.789,0	17.299,95

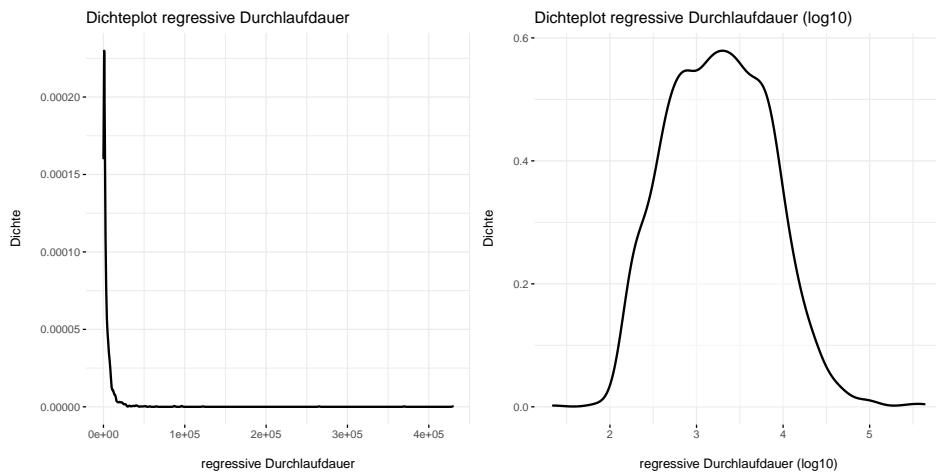


Abbildung 6.22: Verteilung regressive Durchlaufdauer: normal (links), logarithmisch transformiert (rechts)

6 Analyse

AOI-Kategorien bestehen ($\chi^2(4) = 498,67, p < 0,05$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, Tabelle 6.21) zeigen, dass sich alle außer die Gruppierungen *CatO-GerO* und *CatMT-Eingabe* signifikant unterscheiden, sodass gefolgert werden kann, dass die Durchlaufdauer von der AOI-Kategorie abhängt.

Ein Mann-Whitney-U-Test ergibt, dass ein signifikanter Unterschied in der zentralen Tendenz der regressiven Durchlaufdauer zwischen den AOI besteht, die von einem AOI mit höherer Ordnungszahl aus betreten wurden und jenen, die in chronologischer Reihenfolge eine Fixation erhalten. Dies gilt sowohl für den gesamten Datensatz als auch einzeln betrachtet für alle AOI-Kategorien (s. Tabelle 6.22).

Die regressive Durchlaufdauer korreliert signifikant mit der Größe der AOI ($r_s = 0,29, p < 0,01, n = 520.475.096$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 8,23 %.

6.3.2.7 Selektive regressive Durchlaufdauer

Tabelle 6.23 zeigt die Summe, Mittelwerte sowie die Standardabweichung der selektiven regressiven Durchlaufdauer in Millisekunden pro AOI-Kategorie. Während die Originalbeiträge im Deutschen insgesamt lediglich 261.970 ms fixiert werden, beträgt die aufsummierte selektive regressiven Durchlaufdauer bei der MÜ ins Deutsche 1.621.665 ms. Die MÜ wird in dieser Kategorie insgesamt also beinahe sechs Mal so lange betrachtet. Auch die Durchlaufdauer des katalanischen Originals und der MÜ ins Katalanische unterscheidet sich etwa um den Faktor vier. Während das katalanische Original 402.276 ms lang betrachtet wird, liegt der Wert bei der MÜ in die Sprache bei etwa 1.165.050 ms. Insgesamt beträgt die selektive regressiven Durchlaufdauer 3.718.921 ms bzw. im Schnitt 2.273 ms.

Auch die Mittelwerte unterscheiden sich deutlich. So beträgt die durchschnittliche selektive regressiven Durchlaufdauer für die deutschen Originalbeiträge 870 ms, die der MÜ ins Deutsche hingegen 3.180 ms. Ähnlich unterschiedlich sind die Werte für die katalanischsprachigen Beiträge. Die selektive regressiven Durchlaufdauer beträgt für die katalanischen Originale durchschnittlich 925 ms und für die MÜ in die Sprache 3.157 ms. Somit liegen die jeweiligen Originalbeiträge um etwa den Faktor vier unter der Durchlaufdauer der MÜ. Sowohl der absolute als auch durchschnittliche Höchstwert liegt bei der MÜ ins Deutsche.

Die Eingabemaske wurde als statisches AOI annotiert. Der hohe Mittelwert erklärt sich daher aus dem anfänglichen Warteverhalten der Versuchspersonen bis die katalanischen Gesprächspartner-innen die erste Nachricht gesendet haben. Bis zu diesem Zeitpunkt verweilten die Proband-innen meist sehr lange auf der Eingabemaske und damit innerhalb des AOI.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.21: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der regressiven Durchlaufdauer nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	14,06084	0,0000 ***	0,5
CatMT-Eingabe	1,447033	0,0822	
CatO-Eingabe	-3,001051	0,0022 **	0,14
CatMT-GerMT	-2,710149	0,0048 **	0,09
CatO-GerMT	-18,08532	0,0000 ***	0,59
Eingabe-GerMT	-2,289760	0,0138 *	0,1
CatMT-GerO	12,96185	0,0000 ***	0,5
CatO-GerO	0,154517	0,4386	
Eingabe-GerO	3,022081	0,0025 **	0,17
GerMT-GerO	16,39880	0,0000 ***	0,58

Tabelle 6.22: Ergebnisse des Mann-Whitney-U-Tests zur regressiven Durchlaufdauer und progressiven ersten Fixation nach AOI-Kategorie

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	Effekt
GerO	20.546,5	0,001 **	0,21
CatMT	8.762,0	0,001 **	0,31
CatO	48.470,0	0,001 **	0,1
GerMT	16.273,0	0,001 **	0,3
Eingabe	3,0	0,023 *	0,5
Global	319.377,0	0,001 **	0,39

Tabelle 6.23: Summe, Mittelwert, Median und SD der selektiven regressiven Durchlaufdauer in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
GerO	261.970	870,33	558,0	899,77
CatMT	1.165.050	3.157,32	1.364,0	5.388,72
CatO	402.276	924,77	534,0	1.469,86
GerMT	1.621.665	3.179,74	2.073,5	3.868,95
Eingabe	267.960	12.760,00	2.053,0	27.288,35
Global	3.718.921	2.273,18	959,0	4.870,12

6 Analyse

Eine graphische Inspektion der Daten (Abbildung 6.23) zeigt, dass weder die rohen noch die logarithmisch transformierten Werte normalverteilt sind. Auch die statistische Untersuchung unter Anwendung des Shapiro-Wilk-Tests deutet auf nicht normalverteilte Daten hin (roh: $W = 0,37, p < 0,001$, log10: $W = 0,99, p < 0,001$). Daher wird die selektive regressiven Durchlaufdauer mit nicht-parametrischen Tests untersucht.

Ein Kruskal-Wallis-Test belegt, dass Unterschiede in der zentralen Tendenz der selektiven regressiven Durchlaufdauer pro AOI zwischen den einzelnen Teilnehmer:innen bestehen ($\chi^2(20) = 82,77, p < 0,001$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 73 von 210 TN-Gruppen (34,76 %) signifikant unterscheiden. Ein zweiter Kruskal-Wallis-Test zeigt zudem, dass Unterschiede in der zentralen Tendenz der selektiven regressiven Durchlaufdauer pro AOI zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 368,49, p < 0,001$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, Tabelle 6.25) zeigen, dass sich alle außer die Gruppierungen *CatO-GerO*, *Eingabe-GerMT* und *CatMT-Eingabe* signifikant unterscheiden, sodass gefolgert werden kann, dass die Durchlaufdauer von der AOI-Kategorie abhängt.

Ein Mann-Whitney-U-Test ergibt, dass ein signifikanter Unterschied in der zentralen Tendenz der selektiven regressiven Durchlaufdauer zwischen den AOI besteht, die von einem AOI mit höherer Ordnungszahl aus betreten wurden und jenen, die in chronologischer Reihenfolge eine Fixation erhalten. Dies gilt sowohl für den gesamten Datensatz als auch für alle einzeln betrachtete AOI-Kategorien mit Ausnahme von *GerO* (s. Tabelle 6.24).

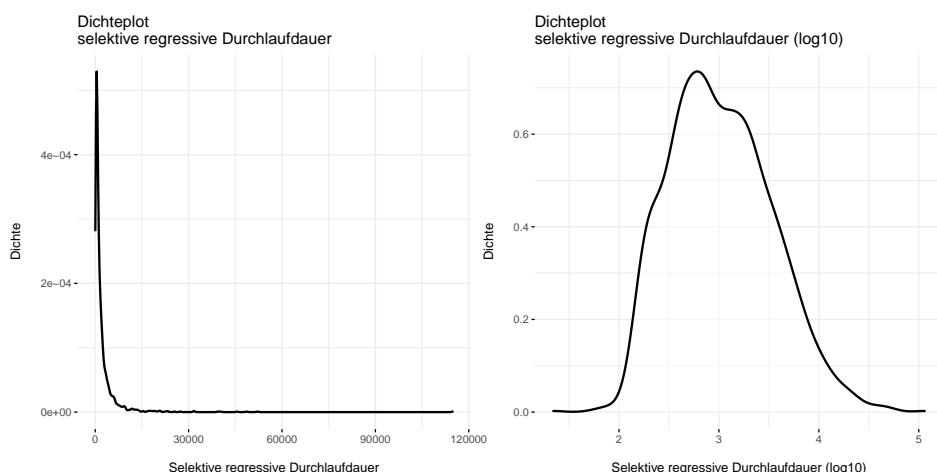


Abbildung 6.23: Verteilung selektive regressivie Durchlaufdauer: normal (links), logarithmisch transformiert (rechts)

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.24: Ergebnisse des Mann-Whitney-U-Tests zur selektiven regressiven Durchlaufdauer und progressiven ersten Fixation nach AOI-Kategorie

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	Effekt
GerO	21.874,0	0,068	
CatMT	9.027,5	0,001 **	0,3
CatO	47.553,5	0,005 **	0,14
GerMT	20.425,0	0,001 **	0,16
Eingabe	3,0	0,023 *	0,5
Global	345.901,0	0,001 **	0,31

Tabelle 6.25: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der selektiven regressiven Durchlaufdauer nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	10,97999	0,0000 ***	0,39
CatMT-Eingabe	-0,758305	0,2802	
CatO-Eingabe	-4,239537	0,0000 ***	0,2
CatMT-GerMT	-4,314935	0,0000 ***	0,15
CatO-GerMT	-16,42494	0,0000 ***	0,53
Eingabe-GerMT	-0,560382	0,2876	
CatMT-GerO	9,403701	0,0000 ***	0,36
CatO-GerO	-0,623239	0,2962	
Eingabe-GerO	3,989707	0,0000 ***	0,22
GerMT-GerO	14,10563	0,0000 ***	0,5

6 Analyse

Die selektive regressive Durchlaufdauer korreliert signifikant mit der Größe der AOI ($r_s = 0,34, p < 0,01, n = 482.868.254$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 11,45 %.

6.3.2.8 Pupillengröße

Tabelle 6.26 zeigt die Mittelwert und Standardabweichung der Pupillengröße in willkürlichen Einheiten pro AOI-Kategorie. Im katalanisch-deutschen Versuchsaufbau liegt die durchschnittliche Pupillengröße bei Betrachtung der Eingabemaske 92 Einheiten über dem Gesamtschnitt. Das entspricht einer Pupillenweitung um 17 %. Dies kann mitunter daran liegen, dass die Proband·innen den Bereich der Eingabe nutzen, um auf eingehende Nachrichten zu warten. Besonders dem Symbol „Schreibt gerade...“ kommt dabei eine zentrale Rolle zu, wie Schlösser u. a. (2018) in seinem Beitrag aufgreift. Dieses Symbol wird als zentraler Fokuspunkt auch im Angesicht der Rollenorganisation der Chatter·innen genutzt, um die Augen ruhen zu lassen und im Folgenden auf die Antwort des Gegenübers zu warten. Das Erfassen der Chatbeiträge des Gegenübers in der Muttersprache erhöht die kognitive Auslastung hingegen nicht signifikant, wie eine Testwiederholung unter Vernachlässigung der Eingabemaske belegen.

Tabelle 6.26: Mittelwert, Median und SD der Pupillengröße pro AOI-Kategorie in willkürlichen Einheiten

AOI-Kategorie	Mittelwert	Median	SD
GerO	505,39	504,40	82,84
CatMT	558,21	564,07	86,11
CatO	538,32	541,75	92,57
GerMT	546,93	554,83	87,66
Eingabe	632,93	640,69	61,44
Global	540,55	545,48	89,78

Im Setting Katalanisch-Deutsch sind es die Vergleichsgruppen mit Beteiligung der jeweiligen MÜ-Ausgabe, die signifikante Unterschiede in der Pupillengröße aufweisen. Die Pupillenweitung der MÜ-Ausgabe ins Deutsche liegt 1 % (*GerMT*: 546 Einheiten, Durchschnitt 540 Einheiten) über dem Durchschnitt. Die Pupillengröße beim deutschen Originalbeitrag ist 7 % kleiner als der Durchschnitt (*GerO*: 505 Einheiten, Durchschnitt: 540 Einheiten). Der Unterschied zwischen beiden Gruppen beträgt 8 %.

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Wie auch die MÜ-Ausgabe ins Deutsche liegt die Pupillengröße des katalanischen Originals nah am globalen Schnitt (*GerMT*: 546 Einheiten, *CatO*: 538 Einheiten, Unterschied 0,4 %). Die MÜ-Ausgabe ins Katalanische liegt 3 % über dem Schnitt (558 Einheiten). Der Unterschied zwischen beiden Kategorien beträgt 3,5 %. Die statistisch signifikanten Unterschiede der Pupillengröße beider katalanischen Kategorien könnte auf Priming-Effekte zurückzuführen sein. Diese Vermutung drängt sich durch die statistische Betrachtung der progressiven ersten Fixation auf. Die Pupillengröße unterscheidet sich zwischen AOI, die vorab bersprungen wurden, und solchen, die regulär in chronologischer Reihenfolge betrachtet wurden. Die Möglichkeit einer aufgabeninternen Varianz der Pupillengröße wurde hingegen nicht untersucht. Hyönä u. a. (1995: 6 f.) beschreiben in ihrer Studie die Veränderungen der Pupillengröße der Proband·innen über die gestellte Aufgabe hinweg und stellen dabei signifikante Unterschiede fest, deren Ursprung sie in einem Gewöhnungsprozess verorten.

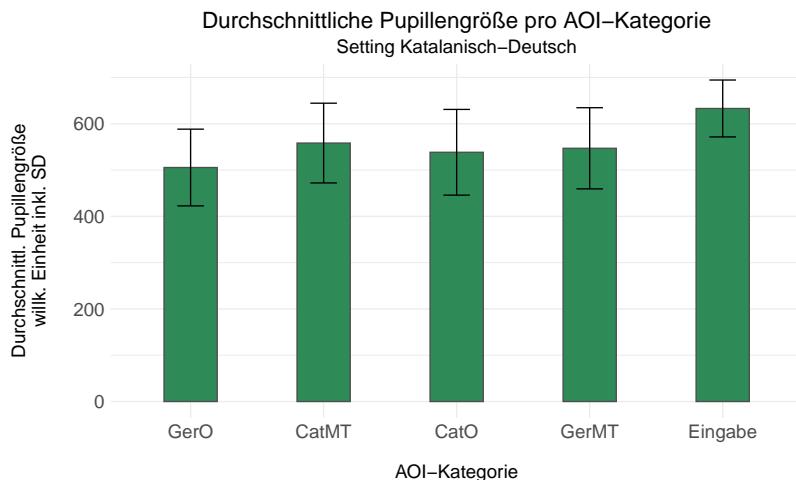


Abbildung 6.24: Durchschnittliche Pupillengröße pro AOI-Kategorie im Setting Katalanisch-Deutsch

Eine graphische Inspektion der Daten (Abbildung 6.25) zeigt, dass weder die rohen noch die logarithmisch transformierten Werte normalverteilt sind. Auch die statistische Untersuchung unter Anwendung des Shapiro-Wilk-Tests deutet auf nicht normalverteilte Daten hin (roh: $W = 0,99$, $p < 0,001$, log10: $W = 0,98$, $p < 0,001$). Daher wird die Pupillengröße mit nicht-parametrischen Tests untersucht.

Ein Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede in der zentralen Tendenz der Pupillengröße pro AOI zwischen den einzelnen Teilnehmer·innen

6 Analyse

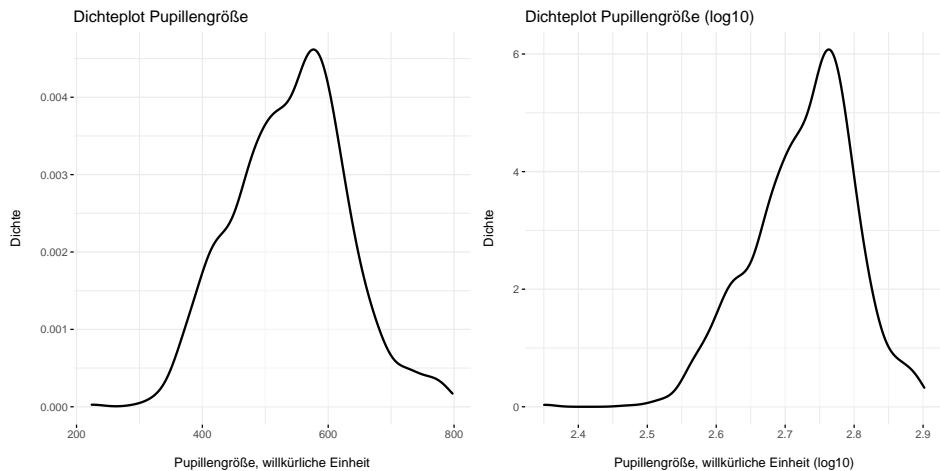


Abbildung 6.25: Verteilung der Pupillengröße in willkürlichen Einheiten: normal (links), logarithmisch transformiert (rechts)

bestehen ($\chi^2(20) = 937,34, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 168 von möglichen 210 TN-Paaren (80,00 %) signifikant unterscheiden, sodass gefolgert werden kann, dass Unterschiede bei der Pupillengröße in gewissem Maße zwischen den einzelnen Teilnehmer·innen bestehen. Auch eine Betrachtung der Mittelwerte (s. Tabelle 6.27) spiegelt diese Einschätzung wieder. Der globale Mittelwert liegt bei 540 Einheiten, wobei die mittlere Pupillengröße je nach Teilnehmer·in z. T. stark variiert.

Ein weiterer Kruskal-Wallis-Test ergibt, dass Unterschiede bei der Pupillengröße zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 91,45, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.28) zeigen, dass sich 8/10 Gruppierungen unterscheiden, sodass gefolgert werden kann, dass die Pupillengröße von der AOI-Kategorie abhängt. Die Gruppierungen *CatO-GerMT* sowie *CatMT-GerMT* unterscheiden sich hingegen nicht signifikant. Die kognitive Auslastung bei der Betrachtung der MÜ ins Deutsche ist demnach vergleichbar mit der bei der Erfassung der katalanischsprachigen Nachrichten.

Mit einer weiteren Testfolge sollen die Unterschiede zwischen progressiver erstmaliger Fixation und der Pupillengröße untersucht werden. Da diese Fixationsart nur in zwei Ausprägungen vorliegt (0 und 1), wird nicht der Kruskal-Wallis-Test, sondern der Mann-Whitney-U-Test verwendet. In Bezug auf den

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Tabelle 6.27: Mittelwert, Median und SD der Pupillengröße pro Studienteilnehmer·in

Pseudonym	Mittelwert	Median	SD
TN11	603,98	598,97	46,39
TN12	584,90	578,72	47,62
TN13	525,92	518,33	65,68
TN15	410,26	408,36	54,73
TN16	606,55	605,77	44,83
TN17	585,97	579,85	35,93
TN18	562,73	560,76	62,05
TN19	532,11	524,69	54,24
TN20	436,40	423,45	60,03
TN21	527,19	522,46	93,08
TN22	620,27	627,38	79,71
TN23	703,58	717,50	53,66
TN24	532,53	512,00	83,71
TN25	592,96	594,67	51,06
TN3	530,05	508,73	66,04
TN4	498,15	494,65	41,11
TN5	497,9	493,79	44,54
TN6	536,84	542,65	48,96
TN7	516,52	501,47	80,53
TN8	436,98	425,40	48,88
TN9	555,49	563,85	58,12
Global	540,55	545,48	89,78

6 Analyse

Tabelle 6.28: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Pupillengröße nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	3,135705	0,0011 **	0,11
CatMT-Eingabe	-3,934707	0,0001 ***	0,2
CatO-Eingabe	-4,929169	0,0000 ***	0,23
CatMT-GerMT	1,730853	0,0464	
CatO-GerMT	-1,592935	0,0556	
Eingabe-GerMT	4,487021	0,0000 ***	0,2
CatMT-GerO	7,743269	0,0000 ***	0,30
CatO-GerO	5,061620	0,0000 ***	0,19
Eingabe-GerO	6,533117	0,0000 ***	0,37
GerMT-GerO	6,659042	0,0000 ***	0,23

globalen Datensatz ($U = 370.818,0, p < 0,001$) besteht ein signifikanter Unterschied in der zentralen Tendenz der Pupillengröße zwischen den AOI, die von einem AOI mit höherer Ordnungszahl heraus betreten wurden, und denen, die chronologisch betrachtet wurden. Wie hierzu aus Tabelle 6.29 ersichtlich wird, beeinflusst die progressive erste Fixation auch pro AOI-Kategorie die Pupillengröße signifikant.

Tabelle 6.29: Ergebnisse des Mann-Whitney-U-Tests zur Pupillengröße und progressiven ersten Fixation nach AOI-Kategorie

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	Effekt
GerO	21.301,5	0,016 *	0,14
CatMT	9.557,5	0,0001 ***	0,23
CatO	46.275,5	0,0001 ***	0,14
GerMT	18.324,0	0,0001 ***	0,21
Eingabe	25,0	0,614	
Global	370.818,0	0,001 **	0,22

Die Pupillengröße korreliert zudem signifikant mit der Größe der AOI ($r_s = -0,20, p < 0,01, n = 839.458.138$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 3,83 %.

6.3.3 Das sakkadische Blickverhalten

Tabelle 6.30 stellt die absolute Anzahl an Sakkaden pro AOI im Setting Katalanisch-Deutsch dar. Beinahe drei Mal so viele Sakkaden werden innerhalb der MÜ-Ausgabe ins Deutsche getätigter wie im deutschen Original. Die wenigsten Sakkaden werden in den ausgehenden Beiträgen der Proband·innen getätigter, hier ist der Wert sogar noch niedriger als im AOI der Eingabemaske. Auf das katalanische Original entfallen weniger als halb so viele Sakkaden (2.198) wie auf die MÜ ins Deutsche (2.405). Ähnlich weit auseinander liegt die Anzahl an Sakkaden beim deutschen Original (1.835) und der MÜ ins Katalanische (3.229). Die entsprechenden Mittelwerte betragen für das deutsche Original 5,65, für die MÜ ins Katalanische 8,36, für das katalanische Original 11,98 und für die MÜ ins Deutsche 13,23. Im Durchschnitt werden also mehr Sakkaden innerhalb des katalanischen Originals und der MÜ ins Deutsche getätigter als im deutschen Original und der MÜ ins Katalanische. Pro AOI werden im gesamten Datensatz im Schnitt 55,30 Sakkaden erfasst. Die Standardabweichung ist mit Blick auf den gesamten Datensatz außergewöhnlich hoch. Es kann daher von einer starken Varianz innerhalb des Satzes ausgegangen werden. Da die Eingabemaske als statisches AOI annotiert wurde, ist eine Berechnung der durchschnittlichen Sakkadenanzahl nicht möglich.

Tabelle 6.30: Anzahl, Mittelwert und SD der Sakkaden pro AOI-Kategorie

AOI-Kategorie	Anzahl	Mittelwert	SD
GerO	1.835	9	5,65
CatMT	3.229	8,36	6,15
CatO	2.198	11,98	8,03
GerMT	5.320	13,23	8,96
Eingabe	2.405		
Global	14.987	55,30	7.108,98

Tabelle 6.31 zeigt die aufsummierte Sakkadenanzahl nach AOI und Richtung im Versuchssetting Katalanisch-Deutsch. Sakkaden, die außerhalb eines AOI gelandet sind, werden unter NA gezählt. Innerhalb aller AOI-Kategorien dominieren nach rechts gerichtete Sakkaden. In allen Kategorien beträgt der Anteil jeweils über 50 % mit Ausnahme der Eingabemaske. Der Anteil rechtsgerichteter Sakkaden ist im Vergleich zu den linksgerichteten Sakkaden auffällig gering. Die

6 Analyse

Tabelle 6.31: Sakkadenanzahl nach Richtung und AOI-Kategorie im Setting Katalanisch-Deutsch

Kategorie			GerO			Kategorie			CatMT		
Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent
Runter	26	1,42	Runter	150	4,65	Runter	327	6,15	Runter	327	6,15
Links	410	22,34	Links	811	25,12	Links	1.164	21,88	Links	1.164	21,88
Rechts	1.284	69,87	Rechts	1.876	58,10	Rechts	3.345	62,88	Rechts	3.345	62,88
Hoch	104	5,67	Hoch	362	11,21	Hoch	457	8,59	Hoch	457	8,59
NA	11	0,60	NA	30	0,93	NA	27	0,51	NA	27	0,51
Gesamt	1.835	100	Gesamt	3.229	100	Gesamt	5.320	100	Gesamt	5.320	100

Kategorie			CatO			Kategorie			GerMT		
Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent
Runter	122	5,55	Runter	327	6,15	Runter	327	6,15	Runter	327	6,15
Links	541	24,61	Links	1.164	21,88	Links	1.164	21,88	Links	1.164	21,88
Rechts	1.124	51,14	Rechts	3.345	62,88	Rechts	3.345	62,88	Rechts	3.345	62,88
Hoch	395	17,97	Hoch	457	8,59	Hoch	457	8,59	Hoch	457	8,59
NA	16	0,73	NA	27	0,51	NA	27	0,51	NA	27	0,51
Gesamt	2.198	100	Gesamt	5.320	100	Gesamt	5.320	100	Gesamt	5.320	100

Kategorie			Eingabe		
Richtung	Anzahl	Prozent	Richtung	Anzahl	Prozent
Runter	414	17,21	Runter	414	17,21
Links	797	33,14	Links	797	33,14
Rechts	1.052	43,74	Rechts	1.052	43,74
Hoch	107	4,45	Hoch	107	4,45
NA	35	1,46	NA	35	1,46
Gesamt	2.405	100	Gesamt	2.405	100

linksgerichteten Sakkaden in den übrigen vier Kategorien machen jeweils einen Anteil von 20–25 % aus. Auffälliger sind die unterschiedlich hohen Anteile an nach oben gerichteten Sakkaden. Im Falle des deutschen Originals und der MÜ ins Deutsche liegen die Anteile jeweils unter 10 %, bei der MÜ ins Katalanische bei etwas mehr als 11 % und beim katalanischen Original erreicht der Anteil sogar beinahe 18 %.

6.3.3.1 Sakkadenamplitude

Im Gegensatz zur statistischen Untersuchung der Normalverteilung der fixatotischen Indikatoren mit dem Shapiro-Wilk-Test wurde die Verteilung der Sakkadendaten mit dem Anderson-Darling-Test betrachtet. Dieser Test kann größere Datensätze erfassen als der Shapiro-Wilk-Test. Der Anderson-Darling-Test belegt, dass sowohl die Gesamtheit des Datensatzes als auch der Datensatz nach einzelnen AOI-Kategorien in Bezug auf die Sakkadenamplitude nicht normalverteilt ist. Es ergibt sich für die Gesamtheit ein Verteilungswert im AD-Test von ($A = 1.253,1, p < 0,01$) und für die Kategorien *GerO*: ($A = 162,84, p < 0,01$), *CatMT*: ($A = 305,34, p < 0,01$), *CatO*: ($A = 168,12, p < 0,01$), *GerMT*: ($A = 393,26, p < 0,01$) und *Eingabe*: ($A = 226,42, p < 0,01$). Auch eine logarithmische Transformation änderte die Verteilung nicht wesentlich (s. Abbildung 6.26). Auf die Kategorie *GerO* verfallen so noch 1.835 Sakkaden, 3.229 auf *CatMT*, 2.198 auf *CatO*, 5.320 auf *GerMT* und 2.405 auf die Eingabemaske. Insgesamt enthält der Datensatz somit noch 14.987 Sakkaden.

Es besteht eine lineare Abhängigkeit zwischen Sakkadenamplitude und -dauer, was Holmqvist (2011: 321) mit Verweis auf Carpenter (1988) herausstellt. Die genaue Art der Abhängigkeit kann mittels eines Korrelationstests mit Spearman's Rho untersucht werden. Sowohl der gesamte Datensatz als auch die Werte nach einzelnen AOI-Kategorien getrennt stehen demnach signifikant in einer positiven Beziehung zueinander: Gesamt: ($S = 2,9267 \times 10^{11}, p < 0,01, r = 0,65$), *GerO*: ($S = 368.854.171, p < 0,01, r = 0,76$), *CatMT*: ($S = 3.028.797.579, p < 0,01, r = 0,64$), *CatO*: ($S = 784.453.871, p < 0,01, r = 0,7$), *GerMT*: ($S = 1,0672 \times 10^{10}, p < 0,01, r = 0,69$) und *Eingabe*: ($S = 2.144.558.389, p < 0,01, r = 0,46$).

Ein Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede in der zentralen Tendenz der Sakkadenamplitude zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(20) = 286,86, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 134 von 210 möglichen Gruppierungen (63,81 %) signifikant voneinander unterscheiden, sodass gefolgt werden kann, dass die Sakkadenamplitude in gewissem Maße von der Individualität der TN abhängt. Dies ist nicht weiter verwunderlich, da Holmqvist (2011: 312) betont, dass die Sakkadenamplitude starken Variationen pro Teilnehmer·in unterliegt.

6 Analyse

Tabelle 6.32: Mittelwert, Median und SD der Sakkadenamplitude pro AOI-Kategorie

AOI-Kategorie	Mittelwert	Median	SD
GerO	1,94	1,42	1,51
CatMT	1,82	1,29	1,47
CatO	1,90	1,42	1,42
GerMT	1,79	1,37	1,30
Eingabe	1,86	1,29	1,56
Global	1,84	1,35	1,43

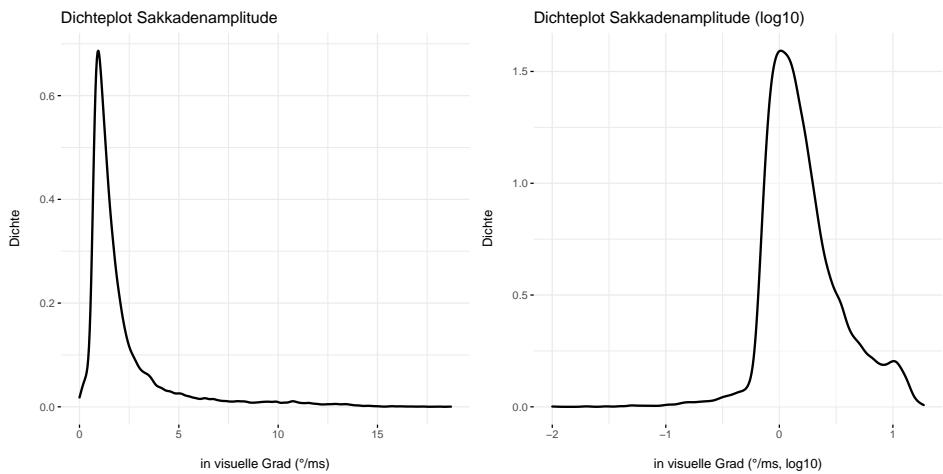


Abbildung 6.26: Verteilung der Sakkadenamplitude: normal (links), logarithmisch transformiert (rechts)

6.3 Der Datensatz im Setting Katalanisch-Deutsch

Ein zweiter Kruskal-Wallis-Test zeigt weiterhin, dass Unterschiede in der Sakkadenamplitude zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 56,56, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 8 von 10 möglichen Gruppierungen unterscheiden, sodass gefolgert werden kann, dass die Sakkadenamplitude von der betrachteten AOI-Kategorie beeinflusst wird (s. Tabelle 6.33, der Asterisk zeigt das Signifikanzlevel an: * $\alpha < 0,05$). Die Ausnahme bilden die Paare *CatMT-Eingabe* sowie *CatO-GerO*. Eine Effektstärke nach Cohen (1992) ist hier jedoch nicht nachweisbar ($r < 0,1$).

Tabelle 6.33: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Sakkadenamplitude nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	-4,946410	0,0000 ***	0,06
CatMT-Eingabe	0,869816	0,2136	
CatO-Eingabe	5,615444	0,0000 ***	0,08
CatMT-GerMT	-3,244902	0,001 **	0,03
CatO-GerMT	2,524951	0,0083 **	0,03
Eingabe-GerMT	-4,090050	0,0000 ***	0,04
CatMT-GerO	-4,789648	0,0000 ***	0,06
CatO-GerO	-0,151941	0,4396	
Eingabe-GerO	-5,427294	0,0000 ***	0,08
GerMT-GerO	-2,511196	0,0075 **	0,03

6.3.3.2 Sakkadendauer

Der Anderson-Darling-Test zeigt, dass sowohl die Gesamtheit des Datensatzes als auch der Datensatz nach einzelnen AOI-Kategorien in Bezug auf die Sakkadendauer nicht normalverteilt ist. Es ergibt sich für die Gesamtheit ein Verteilungswert im AD-Test von ($A = 3.442,2, p < 0,01$) und für die Kategorien von *GerO*: ($A = 374,99, p < 0,01$), *CatMT*: ($A = 749,75, p < 0,01$), *CatO*: ($A = 510,65, p < 0,01$), *GerMT*: ($A = 1.246,5, p < 0,01$) und *Eingabe*: ($A = 537,5, p < 0,01$). Auch eine logarithmische Transformation änderte die Verteilung nicht wesentlich. Auf die Kategorie *GerO* verfallen in absoluten Zahlen 2.099 Sakkaden, 3.700 auf *CatMT*, 2.491 auf *CatO*, 5.899 auf *GerMT* und 2.883 auf die Eingabemaske. Insgesamt enthält der Datensatz somit 17.072 Sakkaden.

6 Analyse

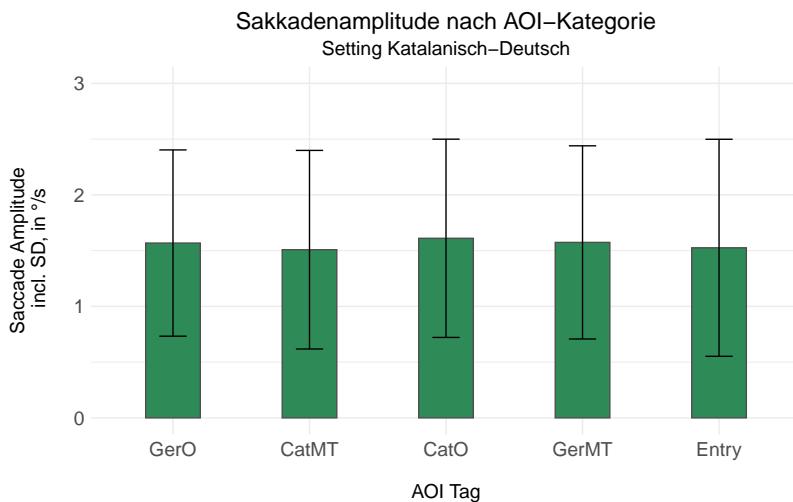


Abbildung 6.27: Sakkadenamplitude pro AOI

Tabelle 6.34: Mittelwert, Median und SD der Sakkadendauer nach AOI-Kategorie

AOI-Kategorie	Mittelwert	Median	SD
GerO	31,19	21	31,64
CatMT	34,25	20	39,59
CatO	33,44	21	37,71
GerMT	30,60	20	33,97
Eingabe	38,12	22	43,76
Global	33,15	21	37,41

6.3 Der Datensatz im Setting Katalanisch-Deutsch

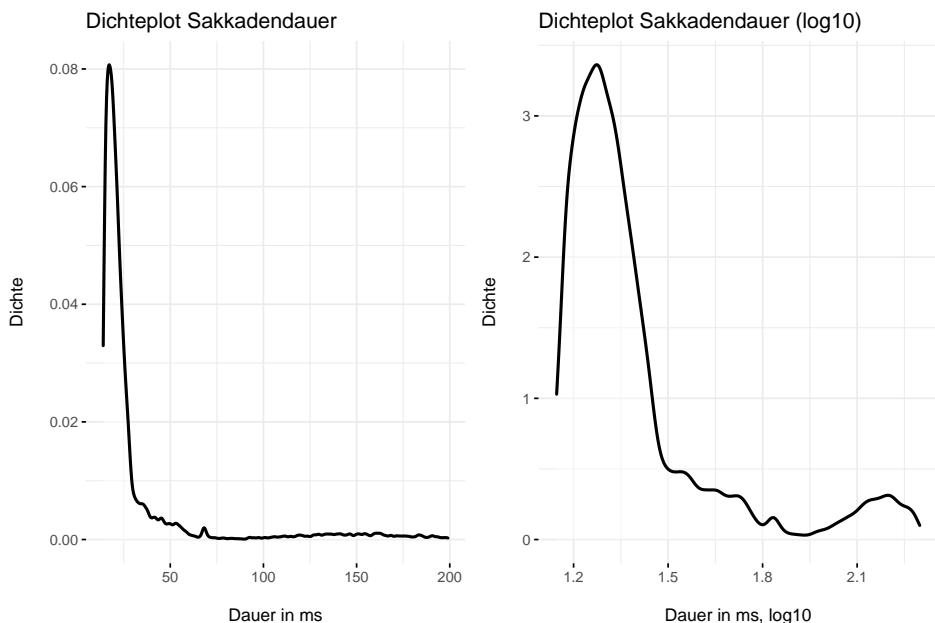


Abbildung 6.28: Verteilung der Sakkadendauer: normal (links), logarithmisch transformiert (rechts)

Ein Kruskal-Wallis-Test belegt, dass Unterschiede bei der Sakkadendauer zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(20) = 184,64, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 114 von 210 möglichen Gruppierungen (54,29 %) signifikant voneinander unterscheiden, sodass gefolgt werden kann, dass die Sakkadendauer in gewissem Maße von der Individualität der Teilnehmer·innen abhängt.

Ein weiterer Kruskal-Wallis-Test belegt, dass Unterschiede bei der Sakkadendauer zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(4) = 49,43, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tab. 6.35, der Asterisk zeigt das Signifikanzlevel an: * $\alpha < 0,05$) zeigen, dass sich 6 von 10 Gruppierungen signifikant voneinander unterscheiden, sodass gefolgt werden kann, dass die Sakkadendauer in gewissem Maße von der AOI-Kategorie abhängt. Die Effektgröße nach Cohen (1992) zeigt dabei keinen Effekt ($r < 0,1$) für alle Gruppierungen.

Für die Vergleichspaare *CatO-GerO* und *Eingabe-GerO* ergeben sich hingegen keine signifikanten Unterschiede. Der Vergleich der Sakkadendauer ist unauffällig, sodass angenommen werden kann, dass das Erfassen der deutschen Originalbeiträge etwa gleich andauernde Sakkaden hervorruft wie das Betrachten

6 Analyse

Tabelle 6.35: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Sakkadendauer nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
CatMT-CatO	-1,543763	0,0767	
CatMT-Eingabe	-4,700784	0,0000 ***	0,06
CatO-Eingabe	-2,806306	0,005 **	0,04
CatMT-GerMT	0,928215	0,1963	
CatO-GerMT	2,489077	0,0107 *	0,03
Eingabe-GerMT	5,995560	0,0000 ***	0,06
CatMT-GerO	-3,588229	0,0004 **	0,05
CatO-GerO	-1,958909	0,0358	
Eingabe-GerO	0,652672	0,2570	
GerMT-GerO	-4,623838	0,0000 ***	0,05

der Eingabemaske oder das Lesen des katalanischen Originals. Auch das Paar *CatMT-CatO* ist unauffällig, sowohl das katalanische Original als auch die MÜ in die Sprache werden also von den Versuchspersonen mit ähnlich langen Sakkaden betrachtet. Die gleiche Feststellung trifft ebenso auf das Paar *CatMT-GerMT* zu.

6.3.4 Zusammenfassung der Ergebnisse im Setting Katalanisch-Deutsch

Die Auswertung der beiden Fragebögen jeweils zu Beginn als auch zu Ende der Studie zeugt von einer variabel zusammengesetzten Kohorte. Die teilnehmenden Versuchspersonen sind, mit Ausnahme einer Person, allesamt Studierende der Universität Leipzig. Es handelt sich um junge Menschen Anfang 20, die einerseits mehrere Fremdsprachen sprechen und andererseits auch mehrheitlich bereits Auslandserfahrung aufweisen. Skype, und noch konkreter: der Skype Translator, ist für die Versuchspersonen nicht die präferierte Software für den Kontakt mit Freunden, Familie und anderen Menschen. Zugleich ist der Funktionsumfang von Skype nicht gänzlich unbekannt. Die Auswahl der Versuchsteilnehmer_innen entspricht der von vergleichbaren Studien im Umfeld der Translationsprozessforschung. So haben etwa auch Hyönä u. a. (1995), Kienle u. a. (2017) auf Studierende zur Durchführung der jeweiligen Studien zurückgegriffen.

Die Rückmeldungen zur Leistung und Qualität des Skype Translators und der Studiensituation sind schwankend. Einerseits werden mehrere Mängel bei der

6.3 Der Datensatz im Setting Katalanisch-Deutsch

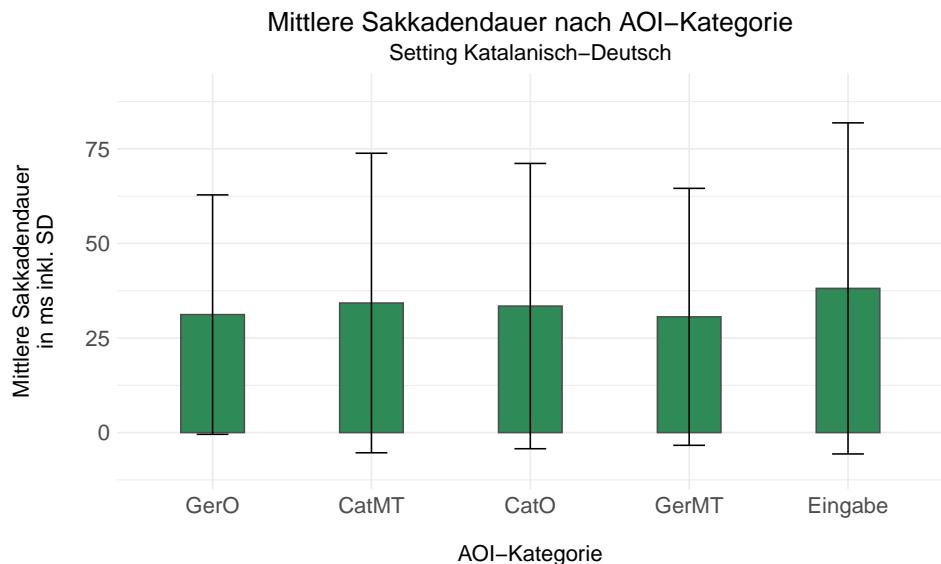


Abbildung 6.29: Sakkadendauer pro AOI

MÜ-Ausgabe benannt, andererseits geben viele Teilnehmer·innen an, dass der zwischengeschaltete Skype Translator kaum gestört habe bzw. kaum wahrgenommen wurde.

Die Analyse der untersuchten Indikatoren deutet auf Unterschiede bei der Wahrnehmung der einzelnen AOI-Kategorien hin. Sowohl die numerische als auch statistische Untersuchung weist in die Richtung eines Leseverhaltens, das sich wie folgt kennzeichnet:

Die Erfassung der Ausgabe der maschinellen Übersetzung ins Deutsche ist für die Versuchspersonen kognitiv fordernder als die Betrachtung der eigenen, deutschsprachigen Textnachrichten. Dafür sprechen sowohl numerische Indikatoren wie die Fixations- und Regressionsanzahl oder die Pupillengröße als auch temporäre Maße wie die Verweildauer und die Durchlaufdauer. Die Untersuchung der Sakkaden in Form der Amplitude und Dauer ist hingegen – unter Annahme einer natürlichen Korrelation beider Indikatoren – zum Teil widersprüchlich und bedarf einer genaueren Betrachtung. Als grundlegende Tendenz kann jedoch auch hier festgestellt werden, dass die Sakkaden sowohl mit Blick auf die Ausrichtung, die Anzahl, die Dauer als auch die Amplitude auf ein unterschiedliches Leseverhalten in Bezug auf die einzelnen AOI-Kategorien deuten.

Die Erkenntnisse, die aus den Daten der Eingabemaske gezogen werden können, sind nur begrenzt interpretierbar. Die Eingabemaske wurde als statisches

6 Analyse

AOI annotiert, somit sind Indikatoren, die auf ein erstmaliges Ereignis abzielen, auch tatsächlich immer nur als das erstmalige, singuläre Ereignis zu verstehen, selbst wenn die Eingabemaske im weiteren Verlauf der Studie erneut oder insgesamt mehrmals betrachtet wurde. Dazu zählt auch die Pupillengröße, da es mehrere Hinweise darauf gibt, dass die Proband·innen die Eingabemaske nutzen, um ihre Augen ruhen zu lassen. Somit kann es sein, dass die Pupillengröße, die auf der Eingabemaske erfasst wird, mehrfach variiert. Die Fixationsanzahl und die Gesamtverweildauer sind hingegen eindeutig messbar und können somit auch ohne Einschränkungen interpretiert werden.

6.4 Die Proband·innen im Setting Deutsch-Deutsch

Für die Vergleichsgruppe im Setting Deutsch-Deutsch wurden im Zeitraum zwischen Mai und Dezember 2019 acht regulär eingeschriebene Student·innen an der Universität Leipzig rekrutiert. Auch diese Gruppe erhielt eine Aufwandsentschädigung von jeweils 10 Euro. Die Gesprächspartner·innen waren in diesem Fall Studienfreunde des Verfassers dieser Arbeit. Analog zur Aufgabenstellung im Setting Katalanisch-Deutsch sollten die Proband·innen so tun, als würden sie vom IALT an den Fachbereich 06 der JGU Mainz, den FTSK in Germersheim, wechseln. Um Informationen zur Wohnungssuche (und zur Universität usw.) zu erhalten, kontaktierten sie deshalb die jeweilige Person.

Wie auch im Setting Katalanisch-Deutsch erhielten die Gesprächspartner·innen keinen Fragebogen und wurden ebenfalls nicht okkulometrisch untersucht. Anders als jedoch die Studie im Sprachenpaar Katalanisch-Deutsch erhielten die Proband·innen nur den Eingangs-, nicht jedoch den abschließenden Fragebogen (der die Erfahrung mit und Eindrücke vom Skype Translator erhebt). Inklusive Vor- und Nachbereitungszeit sollte die Studiendauer pro Person ca. 35 Minuten betragen, von denen auch hier etwa 15 Minuten auf die Gesprächssituation entfallen sollten.

Von den eingangs erwähnten acht Proband·innen in diesem Setting musste ein Datensatz ausgeschlossen werden, da dieser durch technische Probleme bei der Aufnahme unbrauchbar geworden war. Die verbleibende Kohorte besteht somit aus sieben Personen: fünf Frauen und zwei Männer im Alter von 19 bis 34 Jahren ($\varnothing 23,7$ Jahre, Median = 21 Jahre, SD = 4,7). Die Studienfächer der Proband·innen sind in Tabelle 6.36 aufgeschlüsselt. Im Gegensatz zum anderen Versuchssetting waren in diesem Fall auch Student·innen des IALT zugelassen. Dies ist mit den Erkenntnissen von Schaeffer u. a. (2017) zu rechtfertigen, die daraufhindeuten, dass das Leseverhalten zum Zwecke des Textverständnisses und beim Übersetzen

6.4 Die Proband·innen im Setting Deutsch-Deutsch

sich unterscheidet. Da es bei dieser Studie um das reine Textverständnis geht, ist davon auszugehen, dass die Proband·innen mit translatorischem Hintergrund die Ergebnisse nicht verfälschen.

Tabelle 6.36: Studienfächer der Teilnehmer·innen

	Bachelor	Master	Staatsexamen
Konferenzdolmetschen	-	1	-
Übersetzen	2	2	-
Romanische Studien	2	0	-
gesamt	4	3	0
		7	

Alle Studienteilnehmer·innen gaben bei den Fragen nach der Sprache, mit der sie primär aufgewachsen seien sowie der hauptsächlich im Alltag verwendeten Sprache Deutsch an. Eine Person fügte bei zweiter Kategorie noch Englisch und Katalanisch hinzu. Als Sprache im Beruf, ohne dass das Konzept im Fragebogen näher spezifiziert wurde, gaben zwei Personen nur Deutsch, eine Person Deutsch, Katalanisch, Französisch und Englisch und wiederum eine Person Deutsch, Englisch und Spanisch an. Eine weitere gab ausschließlich Englisch an. Die verbleibenden zwei enthielten sich der Antwort.

Die weiteren Fremdsprachen (s. Tabelle 6.30 und Tabelle 6.31, S. 168) der Proband·innen fokussieren sich auf den romanischen Sprachraum: Vier Mal wurde Spanisch angegeben, drei Mal jeweils Französisch und Katalanisch, jeweils einmal Latein, Schwedisch, Russisch, Portugiesisch und Russisch. Englisch wurde von vier Personen genannt, wobei eine darüber hinaus keine weiteren Sprachen erwähnte.

Der Eingangsfragebogen ging daraufhin im Weiteren auf die Sprachkompetenz entlang des Europäischen Referenzrahmens für eine Auswahl an romanischen Sprachen (Spanisch, Katalanisch, Galicisch, Italienisch, Französisch und Rumänisch) ein.

Alle Proband·innen gaben weiterhin an, bereits längere Zeit (mehr als vier Wochen Minimum, kein Urlaub) im Ausland verbracht zu haben. Dabei lagen die befreisten Länder der Teilnehmer·innen hauptsächlich in Europa. Mit Brasilien und den Vereinigten Staaten von Amerika gaben lediglich zwei Teilnehmer·innen Ziele außerhalb Europas an. Fünf der sieben Proband·innen haben hingegen bereits längere Zeit in Spanien verbracht. Die in den jeweiligen Ländern verbrachte Dauer reicht von fünf bis 24 Monate ($\bar{x} 11,1$ Monate, Median = 7 Monate, SD = 7,3).

6 Analyse

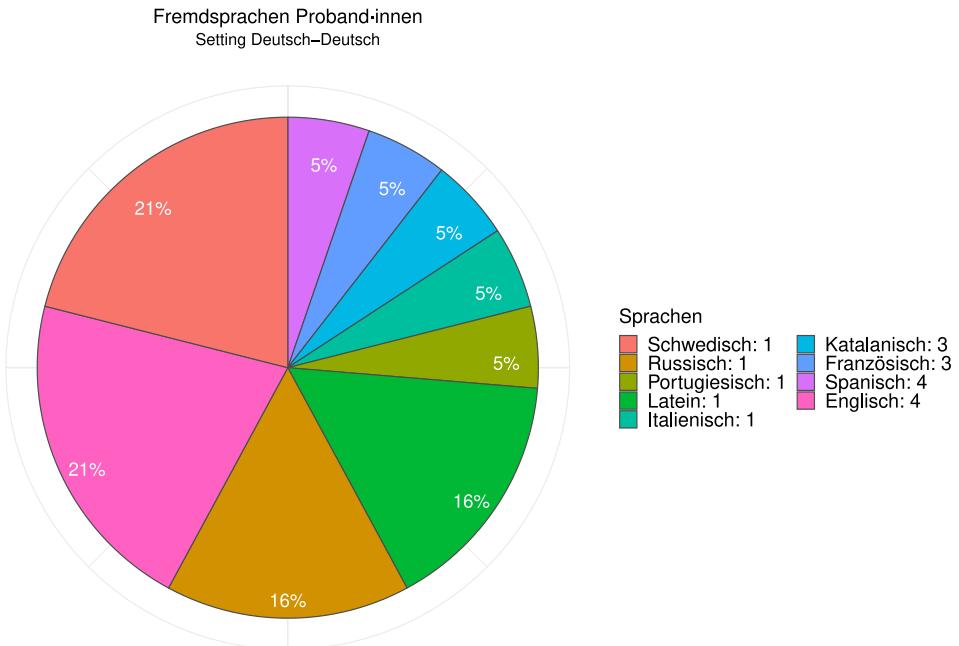


Abbildung 6.30: Fremdsprachen der Proband·innen

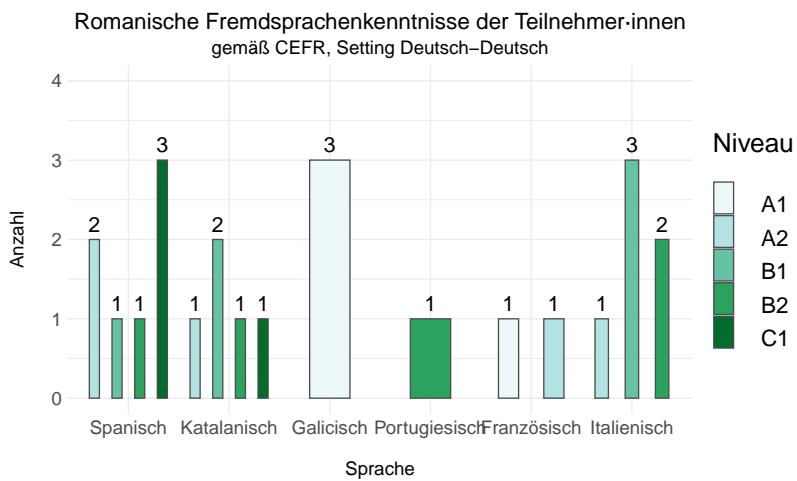


Abbildung 6.31: Kompetenz in ausgewählten romanischen Sprachen

6.4 Die Proband·innen im Setting Deutsch-Deutsch

Als Gründe wurden Spracherwerb (2/7), Sprachvertiefung (4/7), Kulturkontakt (3/7), ein Pflichtaufenthalt (1/7) sowie die (zeitweise) Verlagerung des Lebensmittelpunktes (1/7) genannt. Dabei lagen die Aufenthalte zwischen drei und 120 Monaten zurück (\varnothing 45,4 Monate, Median = 28 Monate, SD = 44,4).

6.4.1 Angaben der Proband·innen im Eingangsfragebogen

Von den sieben Proband·innen in der Vergleichsgruppe im Setting Deutsch-Deutsch gaben vier an, Skype überhaupt zu nutzen. Davon gaben jeweils zwei an, Skype monatlich bzw. seltener als monatlich zu verwenden. Die Nutzungs-dauer pro Sitzung belief sich dabei auf entweder bis zu 30 Minuten (2/4), bis zu einer Stunde (1/4) und länger als eine Stunde (1/4). Drei der vier Personen gaben an, Skype für Windows zu verwenden und eine gab an, die Plattform nicht zu kennen.

In Bezug auf die Nutzung der einzelnen Funktionen gab jeweils eine Person an, den Voice- bzw. den Textchat zu nutzen. Alle vier Teilnehmer·innen nutzten hingegen den Videochat. Aus dem Pool der weiteren Funktionen wurden nur die Bildschirmübertragung und der Konferenzmodus von jeweils einer Person verwendet.

Die Gesprächspartner·innen der Gruppe waren dabei sowohl Mitglieder der Familie im Inland (2/4) als auch im Ausland (2/4). Weiterhin verwendeten drei Teilnehmer·innen den Dienst für die Kommunikation mit Freunden im Inland und alle vier für den Kontakt mit Freunden im Ausland. Die weiteren Kategorien (berufliche Gesprächspartner·innen und Institutionen jeweils im In- und Ausland) wurden von keiner Person ausgewählt. Die verwendeten Sprachen in diesen Situationen waren dabei Deutsch (4/4) bzw. Katalanisch, Spanisch, Englisch und Portugiesisch mit jeweils einer Angabe.

Wie auch die Proband·innen im Setting Katalanisch-Deutsch gab auch hier keiner der Versuchsteilnehmer·innen Erfahrungswerte mit dem Skype Translator an. Die Erfahrung mit den Funktionen Video-, Voice- und Textchat wurde auf einer Skala von 1 - schlecht bis 4 - gut wie folgt bewertet: Zwei Personen haben eher gute Erfahrungen mit dem Textchat gemacht, eine eher schlechte, eine enthielt sich der Antwort. Alle vier Teilnehmer·innen haben eher gute Erfahrungen mit dem Videochat gemacht. Ebenfalls machten zwei eher gute Erfahrungen mit dem Voicechat, die verbleibenden zwei enthielten sich der Antwort.

6.4.2 Nutzung von Alternativen zu Skype

Die nachfolgenden Angaben beziehen sich nun wieder auf alle sieben Teilnehmer·innen im Setting Deutsch-Deutsch. Mit der Erhebung der verwendeten Alternativen zu Skype begann im Eingangsfragebogen ein neuer Block, zu dem die

6 Analyse

Teilnehmer·innen direkt weitergeleitet wurden, die die Frage nach der generellen Skype-Nutzung verneint hatten.

Sechs der sieben Proband·innen gaben demnach an, Alternativen zu Skype zu verwenden. Bei der Frage nach der generellen Nutzung (s. Abbildung 6.32, linke Abbildung) führt der Dienst WhatsApp als Alternative zu den angebotenen Kommunikationsmodi Voice, Video und Text mit jeweils fünf, vier und sechs Nennungen. Der am häufigsten genutzte Dienst, aufgeteilt nach Modus, ist bei der Kommunikation über Text und im Voicechat ebenfalls WhatsApp. Im Modus Videochat nutzten drei der sieben Versuchsteilnehmer·innen Apple Facetime (s. Abbildung 6.32, rechte Abbildung).

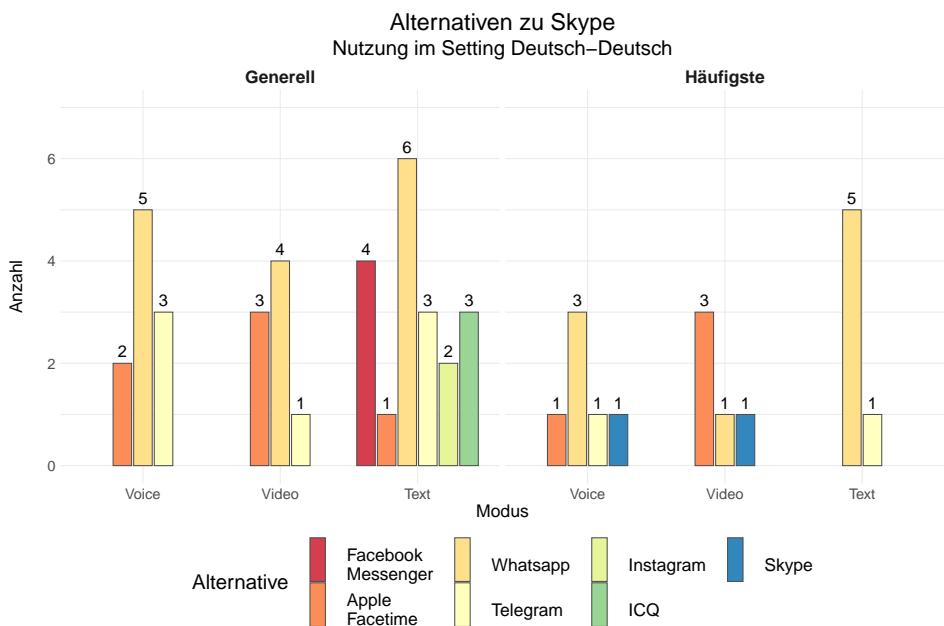


Abbildung 6.32: Vergleich: Generelle vs. häufigste Nutzung der Alternativen zu Skype nach Modus

6.5 Der Datensatz im Setting Deutsch–Deutsch

6.5.1 Visuelle Inspektion der Bildschirmmitschnitte

Die nachfolgenden Bildschirmfotos (Abbildung 6.33, 6.34 und 6.35) wurden exemplarisch einer Sitzung im Versuchsaufbau Deutsch–Deutsch entnommen und stellen die Fixationen, Sakkaden und Blinzler während der Kommunikationssituation dar.

6.5 Der Datensatz im Setting Deutsch-Deutsch

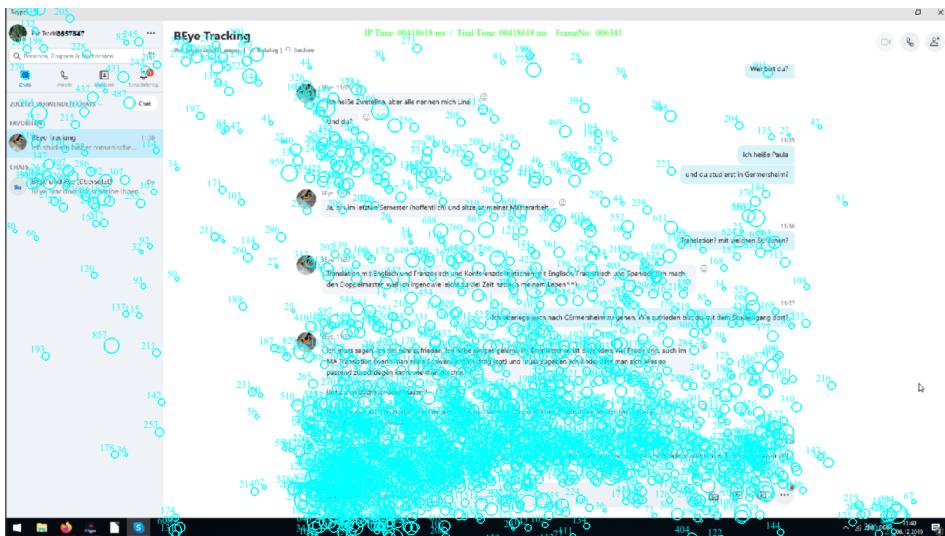


Abbildung 6.33: Fixationen während einer Bildschirmaufnahme

Die Fixationen sind am unteren Bildschirmrand oberhalb der Eingabemaske besonders dicht verteilt. Sie erstrecken sich über die gesamte Länge des Chat-Bereiches. Mehrere Fixationen sind auch entlang der Kontaktliste und am oberen Bildschirmrand erkennbar, jedoch deutlich weniger häufig und lang als jene im Bereich der eingehenden Nachrichten.

Generell sind über den gesamten Bildschirm, mit Ausnahme der oberen rechten Bildschirmecke, Sakkaden auszumachen. Die Verteilung ist jedoch nicht gleichmäßig, sondern weist eine Konzentration auf: Die visuelle Darstellung der Sakkaden bildet ein sehr dichtes Netz im unteren Bildschirmbereich, oberhalb der Eingabemaske und auf der gesamten Breite des Chatfensters. Die Mehrheit der Sakkaden bleibt in der unteren Bildschirmhälfte. Betrachtet man nur die horizontal ausgerichteten Sakkaden, spricht diese Beobachtung dafür, dass die beteiligten Personen vergleichsweise lange Nachrichten verfasst haben, die jeweils über die Mitte des Chatfensters hinausragen. Die Augenbewegungen im übrigen Bildschirmbereich sind möglicherweise Ausdruck dafür, dass die vielen Sprünge die Wartezeit der Versuchsperson darstellen, während die Person gegenüber einer neuen Nachricht verfasst.

Die vom Eye-Tracker erfassten Blinzler konzentrieren sich ebenfalls an der Eingabemaske, dort leicht oberhalb im Bereich mittig links. Wie auch im Falle des anderen Settings können die vermehrt auftretenden Blinzler dafür stehen, dass die Person in diesem Bereich eine stärkere kognitive Auslastung an den Tag

6 Analyse



Abbildung 6.34: Sakkaden während einer Bildschirmaufnahme

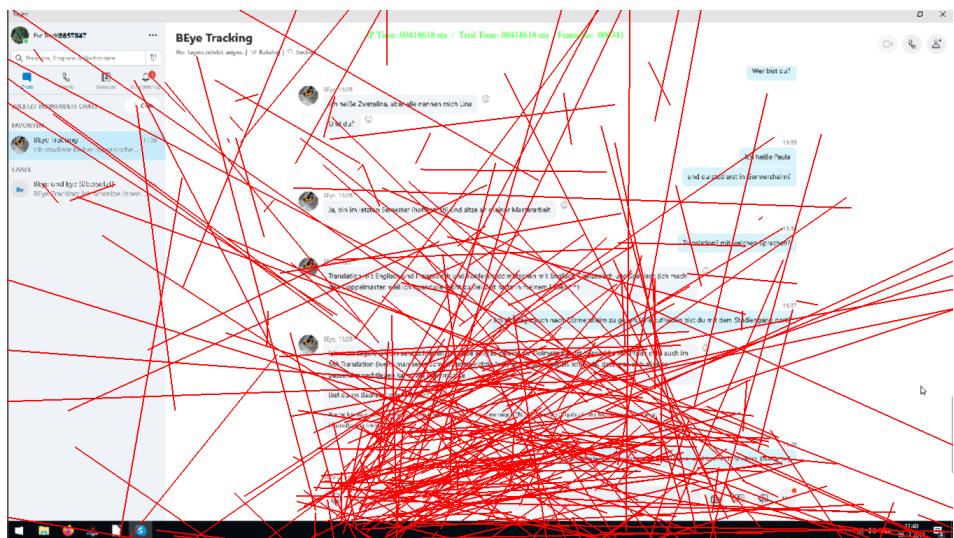


Abbildung 6.35: Blinzler während einer Bildschirmaufnahme

gelegt hat. Mehrere Blinzler führen außerdem aus dem erfassbaren Bildschirmbereich hinaus. In diesem Fall hat die Versuchsperson möglicherweise auf etwas geachtet, dass sich im Raum, in der die Studie durchgeführt wurde, befand.

6.5.2 Fixationen der Probanden

In Tabelle 6.37 sind grundlegende Informationen zu den einzelnen Bildschirmmitschnitten im Setting Deutsch-Deutsch aufgeführt. Wie auch im Setting Katalanisch-Deutsch hat die Mehrheit der Proband·innen länger als die ursprünglich veranschlagten 15 Minuten mit der Textchat-Kommunikation verbracht. Die Person mit dem Pseudonym DD_TN1 hat das Gespräch sogar erst nach etwas mehr als 26 Minuten beendet. Jeder einzelne Chatbeitrag wurde dabei – wie in Abschnitt 5.2.5 (S. 88) dargelegt – mit einem dynamischen AOI versehen und entsprechend mit einem Label versehen. Generell schwankt die Anzahl an Beiträgen auf Seiten der Proband·innen zwischen 12 und 23 ($\bar{x} 18$, Median = 19, SD = 3,84) und auf Seiten der Gesprächspartner·innen zwischen 13 und 36 ($\bar{x} 20$, Median = 18, SD = 7,19). Die starken Unterschiede und die hohe Standardabweichung zeugen von einer unterschiedlichen Produktivität und Aktivität während der Kommunikationssituation.

Wie auch im Falle der *fixation heatmaps* im katalanisch-deutschen Versuchsaufbau schauen die Proband·innen überwiegend in das untere linke Drittel des Chatfensters, knapp oberhalb der Eingabemaske. Auch in diesem Setting liegt der Fokus auf der Eingabemaske und dem unmittelbar darüber liegenden Bereich. Anders als die Darstellungen der katalanisch-deutschen Teilnehmer·innen sind die Fokuspunkte hier jedoch raumgreifender und unförmiger.

6.5.2.1 Fixationsanzahl

Tabelle 6.38 zeigt die aufsummierte Anzahl an Fixationen pro AOI-Kategorie. Insgesamt wurden 7.587 Fixationen in den mit AOI versehenen Bereichen erfasst. Der Gesamtschnitt beträgt 30,97 Fixationen pro AOI. Die Beiträge des Gegenübers (B) erhielten mit 4.329 Fixationen etwa drei Mal so viele wie die eigenen Nachrichten (1.481). Das Eingabefeld wurde 1.777 Mal fixiert. Der Durchschnittswert für die eigenen Beiträge ist mit 13,97 am kleinsten. Die durchschnittliche Fixationsanzahl auf die Nachrichten des Gegenübers liegt mit 31,8 über dem Gesamtschnitt. Auch die mittlere Anzahl an Fixationen, die auf die Eingabemaske verfallen, liegt über dem Schnitt (253,86).

Sowohl die visuelle Inspektion als auch die statistische Untersuchung deuten darauf hin, dass die Fixationsanzahl nicht normalverteilt ist. Ein Shapiro-

6 Analyse

Tabelle 6.37: Übersicht Bildschirmaufzeichnungen im Setting De-De

Pseudonym	Dauer in min	Anzahl AOI	Anzahl Chatbeiträge	
			Eigen	Fremd
DD_TN1	26,46	27	14	13
DD_TN2	20,92	59	23	36
DD_TN4	20,27	35	19	16
DD_TN5	17,43	30	15	15
DD_TN6	16,19	44	22	22
DD_TN7	18,93	42	19	23
DD_TN8	16,98	30	12	18
Gesamt	137,18	267	124	143
		\varnothing Anzahl Beiträge (gerundet)		
		18		
		Median Anzahl Beiträge		
		19		
		Standardabweichung (SD)		
		3,84		
		7,19		

Tabelle 6.38: Summe, Mittelwert, Median und SD der Fixationsanzahl pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
A	1.481	13,97	7,5	19,50
B	4.329	31,8	19,5	42,95
Eingabe	1.777	253,86	249,0	124,29
Global	7.587	30,97	12,0	55,55

6.5 Der Datensatz im Setting Deutsch-Deutsch

Wilk-Test ergibt für den gesamten Datensatz ($W = 0,51, p < 0,001$). Eine logarithmische Transformation ändert an dieser Beobachtung nichts (Shapiro-Wilk Test, \log_{10} : $W = 0,97, p < 0,001$). Deshalb wird die Fixationsanzahl mit nicht-parametrischen Tests analysiert.

Ein durchgeführter Kruskal-Wallis-Test deutet auf keine signifikanten Unterschiede bei der Fixationsanzahl nach Teilnehmer.in hin ($\chi^2(6) = 6,88, p = 0,33$). Ein weiterer Kruskal-Wallis-Test zeigt, dass Unterschiede bei der Fixationsanzahl zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 41,32, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.39) zeigen, dass sich alle Gruppierungen unterscheiden, sodass gefolgert werden kann, dass die Fixationsanzahl von der AOI-Kategorie abhängt. Dabei handelt es sich gemäß der Effektstärke nach Cohen (1992) um einen schwachen Effekt (*A-Eingabe*: $r = 0,27$) bzw. mittleren Effekt (*A-B*: 0,34, *B-Eingabe*: 0,39).

Tabelle 6.39: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Fixationsanzahl nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	-5,235690	0,0000 ***	0,34
A-Eingabe	-4,634883	0,0000 ***	0,39
B-Eingabe	-2,902860	0,0018 **	0,27

Ein exakter Mann-Whitney-U-Test wurde verwendet, um Unterschiede bei der Fixationsanzahl pro AOI zwischen beiden Arten der progressiven ersten Fixation zu untersuchen. Im Falle des gesamten Datensatzes ergibt der Test ein signifikantes Ergebnis ($U = 6.224,00, p < 0,001, r = 0,22$). Dabei handelt es sich gemäß der Effektstärke nach Cohen (1992) um einen schwachen Effekt. Der Median der Fixationsanzahl beträgt 16, wenn die AOI in chronologischer Reihenfolge betreten wurden, und 8, wenn zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde.

Unterteilt nach AOI-Kategorie ergibt der Mann-Whitney-U-Test ebenfalls für die eigenen Chatbeiträge ein signifikantes Ergebnis (AOI-Kategorie *A*: $U = 1.365,5, p < 0,05, r = 0,2$). Hierbei beträgt der Median der Fixationsanzahl 8, wenn die Reihenfolge eingehalten wurden, und 6, wenn zuvor ein AOI mit höherer Ordnungszahl betreten wurde. Auch die Testergebnisse für die Chatbeiträge des Gegenübers sind signifikant (AOI-Kategorie *B*: $U = 1.774,5, p < 0,05, r = 0,19$). Der Median der Fixationsanzahl bei konsekutiver Betrachtung der AOI beträgt 23, bei ungeordneter Betrachtung 12. Die Effekte sind nach Cohen (1992)

6 Analyse

als schwach einzustufen. Für die Eingabemaske konnten keine Werte berechnet werden, da für die Durchführung des Tests zu wenige Datenpunkte vorlagen.

Weiterhin korreliert die Fixationsanzahl signifikant mit der Größe der AOI ($r_s = 0,77, p < 0,01, n = 570.037$). Dabei handelt es sich nach Cohen (1992) um einen starken Effekt. Das Bestimmtheitsmaß beträgt 58,89 %.

Zur besseren Übersicht sind in Tabelle 6.40 die aufsummierte Größe in Pixel der AOI nach Kategorien aufgeführt sowie deren Mittelwert und Standardabweichung. Die mittlere Größe der AOI der eigenen Beiträge ist mit etwa 28.000 px deutlich geringer als die der Fremdbeiträge mit 40.000 px. Die Eingabemaske wurde statisch annotiert und ist in jedem einzelnen Datensatz etwa gleich groß. Hier gilt, wie weiter oben bereits problematisiert, die technisch bedingte Einschränkung der Aussagekraft der AOI-Größe (s. Abschnitt 5.2.2.1, S. 82).

Tabelle 6.40: Summe, Mittelwert und SD der AOI-Größe in px pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	SD
A	2.948.789	27.818,76	19.384,44
B	5.342.084	40.470,33	41.788,15
Eingabe	481.057	68.722,43	5.073,75
Gesamt	8.771.930	35.803,79	34.211,62

6.5.2.2 Dauer der ersten Fixation

Wie in Tabelle 6.41 ersichtlich wird, liegt die durchschnittliche Dauer der ersten Fixation auf den eingehenden Nachrichten 25 ms über dem Gesamtschnitt von 287,29 ms, wohingegen die Fixationsdauer auf den eigenen Nachrichten 14 ms unterhalb des Schnittes liegt. Der Wert bei Betrachtung der Eingabemaske liegt 50 ms unter dem globalen Durchschnitt.

Abbildung 6.37 (S. 178) stellt die Dauer der ersten Fixation pro AOI noch einmal als Boxplot dar. Der Mittelwert pro AOI-Kategorie beträgt für A: ($\bar{x} 273,56$ ms, Median: 209 ms, SD: 157,66), B: ($\bar{x} 302,22$ ms, Median: 274 ms, SD: 149,20) und Eingabe: ($\bar{x} 237,29$ ms, Median: 249 ms, SD: 62,25).

Sowohl die graphische als auch statistische Inspektion der Daten belegt, dass keine Normalverteilung vorliegt. Eine logarithmische Transformation nähert die Verteilung zwar der Gausschen Glockenkurve an, ist jedoch noch immer leicht verschoben (s. Abbildung 6.36). Auch eine Berechnung der Schiefe (normal: 1,67,

6.5 Der Datensatz im Setting Deutsch-Deutsch

log-transformiert: 0,46) und der Kurtosis (normal: 3,30, log-transformiert: -0,25) sowie Shapiro-Wilk-Tests (normal: $W = 0,85, p < 0,001$ und log-transformiert: $W = 0,98, p < 0,01$) deuten darauf hin, dass der Datensatz nicht normalverteilt ist. Deshalb werden für die statistische Analyse dieser Daten nicht-parametrische Tests verwendet.

Tabelle 6.41: Mittelwert, Median und SD der Dauer der ersten Fixation pro AOI-Kategorie

AOI-Kategorie	Mittelwert	Median	SD
A	273,56	209	157,66
B	302,22	274	149,25
Eingabe	237,29	240	62,25
Global	287,29	245	151,49

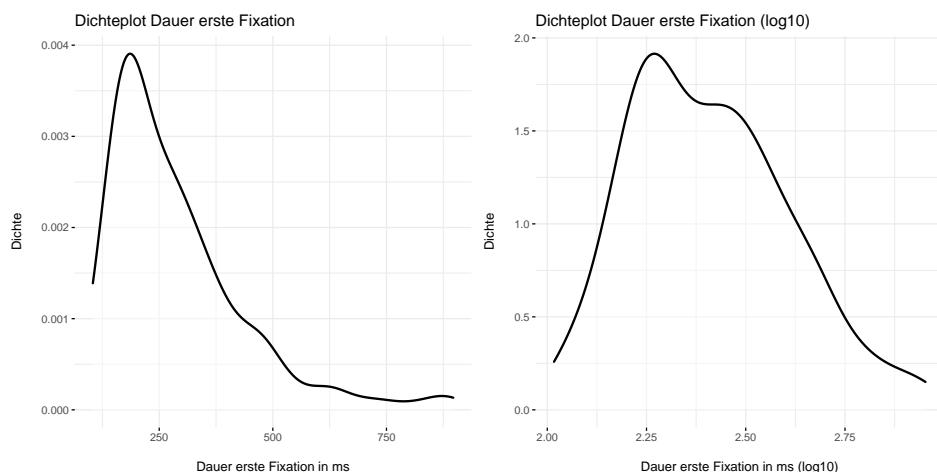


Abbildung 6.36: Verteilung der Dauer der ersten Fixation: normal (links), logarithmisch transformiert (rechts)

Ein Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede in der zentralen Tendenz der Dauer der ersten Fixation zwischen den einzelnen Teilnehmer*innen im Setting Deutsch-Deutsch bestehen ($\chi^2(6) = 17,28, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 7 von 21 möglichen Vergleichspaaren unterscheiden, sodass gefolgert werden kann, dass die Dauer der ersten Fixation teilweise von den Teil-

6 Analyse

nehmer:innen abhängt. Dies ist wenig verwunderlich, bedenkt man, dass die Individualität der Versuchspersonen im Rahmen naturalistisch angelegter Experimente häufig als Zufallseffekte betrachtet werden. Ein zweiter Kruskal-Wallis-Test lehnt die Vermutung ab, dass Unterschiede bei der Dauer der ersten Fixation zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 4,20, p = 0,12$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.42) belegen ebenfalls, dass keine signifikanten Unterschiede in der Fixationsdauer zwischen den einzelnen AOI bestehen. Allerdings ist der p-Wert in der Gruppe A-B niedrig ($z = -1,99, p = 0,07$). Generell kann gefolgert werden, dass die Fixationsdauer nicht von der AOI-Kategorie, also den Chatbeiträgen, abhängt.

Tabelle 6.42: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Dauer der ersten Fixation nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	-1,992319	0,07	-
A-Eingabe	0,080491	0,47	-
B-Eingabe	0,802203	0,32	-

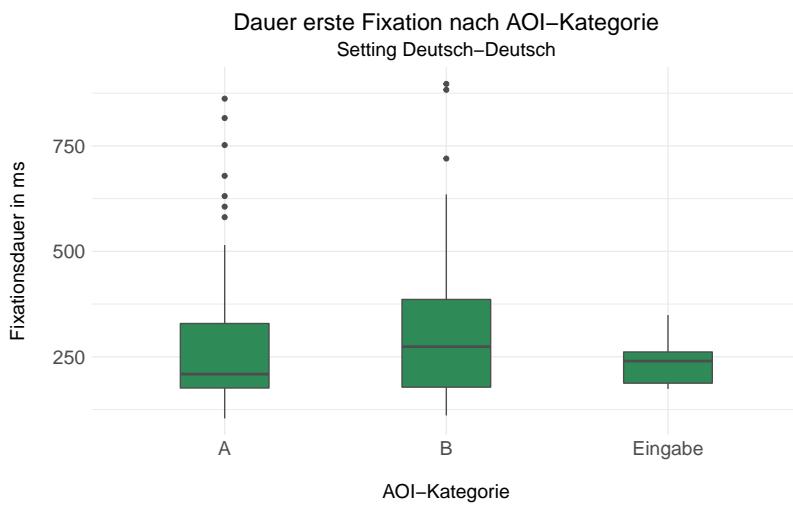


Abbildung 6.37: Dauer der ersten Fixation pro AOI-Kategorie

Da die progressive erste Fixation nur in zwei Ausprägungen vorliegt (0 und 1), wird nicht der Kruskal-Wallis-Test, sondern der Mann-Whitney-U-Test verwendet. Es besteht ein signifikanter Unterschied in der zentralen Tendenz der Dauer der erstmaligen Fixation zwischen den AOI, bei denen zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, und denen, die chronologisch betreten wurden. Die eigenen Chatbeiträge der Versuchspersonen werden kürzer fixiert, wenn zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, als wenn es sich um eine tatsächliche chronologische erstmalige Fixation handelt. Ein exakter Mann-Whitney-U-Test ($U = 1.034,5, p < 0,05, r = 0,24$) in Verbindung mit einer Untersuchung der Effektstärke nach Cohen (1992) zeigt dabei allerdings nur einen schwachen Effekt.

Tabelle 6.43: Ergebnisse des Mann-Whitney-U-Tests zur Dauer der ersten Fixation nach AOI-Kategorie und progressiver ersten Fixation

AOI-Kategorie	<i>U</i>	<i>p</i> (angepasst)	Effekt
A	1.034,5	0,02*	0,24
B	1.689,0	0,6	-
Eingabe	n/a	n/a	-
Global	5.596,5	0,08	-

Ein Korrelationstest (Spearman's Rho) belegt, dass keine signifikanten Zusammenhänge zwischen Dauer der ersten Fixation und der Größe der AOI bestehen ($r_s = -0,03, p = 0,68, n = 483.293$).

6.5.2.3 Regressionen

6.5.2.3.1 Eingehende Regressionen

Abbildung 6.38 und Tabelle 6.44 zeigen die aufsummierte Anzahl aller Regressionen in ein AOI nach Kategorie sowie deren Mittelwert und Standardabweichung. Die Kategorie *B*, also die Beiträge des Gegenübers, erhielten mit 221 ($\bar{x} 1,67, SD: 2,05$) die meisten Regressionen. In die Kategorien *A* und *Eingabe* fielen jeweils 81 ($\bar{x} 0,76, SD: 1,35$) bzw. 136 ($\bar{x} 19,43, SD: 16,47$) Regressionen. Auffällig sind die Durchschnittswerte, bei denen mindestens eine Regression in die Beiträge des Gegenübers ausgeführt wird und in die Eingabemaske sogar durchschnittlich mehr als 19.

Zur statistischen Untersuchung der Regressionen in ein AOI wurde mit einem binomialen logistischen Regressionsmodell mit gemischten Effekten gearbeitet,

6 Analyse

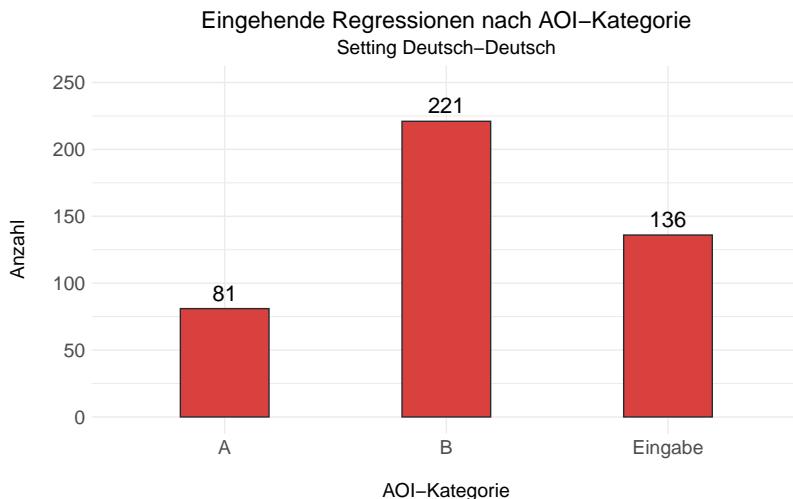


Abbildung 6.38: Anzahl Regressionen in ein AOI

Tabelle 6.44: Summe, Mittelwert und SD der eingehenden Regressio-nen pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	SD
A	81	0,76	1,35
B	221	1,67	2,05
Eingabe	136	19,43	16,47
Global	438	1,79	4,37

das die Teilnehmer-innen (*SessionLabel*) als Zufallseffekte beinhaltete (s. Tabel-le 6.45). Die Variablen wurden in einer bedingten Vorwärtsauswahl in das Mo-dell integriert. Das simpelste und zugleich genauste Modell erweist sich dabei, ebenso wie einzelne Koeffizienten der integrierten Variablen, als signifikant und liefert bessere Ergebnisse als ein einfaches lineares Regressionsmodell ($\chi^2(5) = 60,75, p < 0,001$). Zugleich zeigt sich, dass das Modell noch nicht optimal ange-passt (C: 0,79, Somers' D_{XY}: 0,58) ist. Während die Kategorien *Intercept* (in die-sem Fall A) bzw. *Eingabe* keinen Effekt auf die Wahrscheinlichkeit haben, dass eine Regression in ein AOI gemacht wird, steigt die relative Wahrscheinlichkeit bei B um 128 %. Weiterhin steigt die Wahrscheinlichkeit um 281 % mit der Größe der AOI. Hingegen nimmt die relative Wahrscheinlichkeit mit jeder Einheit der *IARegPD* um 35 % und mit jeder progressiven ersten Fixation um 66 % ab.

6.5 Der Datensatz im Setting Deutsch-Deutsch

Tabelle 6.45: Werte des Regressionsmodells für eingehende Regressio-
nen im Setting Deutsch-Deutsch

	Gruppen	Varianz	SD	L.R. X2	DF	Pr
Zufallseffekt(e)	SessionLabel	0	0	60,75	1	<0,05*
Fixe Effekt(e)	Schätzung	VIF	OR	SE	z	Pr(> z)
(Intercept)	0,4712		1,60	0,3257	1,447	0,14796
IATagB	0,8254	1,0329236	2,28	0,2995	2,755	0,00586**
IATagEingabe	3,1416	1,6200961	23,14	2,7880	1,127	0,25981
IAFFixPro1	-1,0749	1,0751683	0,34	0,3474	-3,094	0,00197**
IARegPD_z	-0,4271	1,7490922	0,65	0,2588	-1,650	0,09896.
IA_AREA_z	1,3367	1,0849144	3,81	0,2487	5,374	<0,001***

Model statistics	Wert
No. Groups	7
Number of cases in model	245
Observed misses	
Observed deviance	272,6
Residual deviance	
R2 (Nagelkerke)	0,295369
R2 (McFadden)	0,182236
R2 (Cox & Snell)	0,219609
C	0,79
Somers' D _{XY}	0,58
AIC	286,6
BIC	311,1
Prediction accuracy	
Model Likelihood Ratio Test	
x2:	60,75
df:	5
p	< 0,001***

6 Analyse

Wie auch im katalanisch-deutschen Versuchsaufbau wurde zur Überprüfung der Plausibilität mit nicht-parametrischen Tests gearbeitet. Bei eingehenden Regressionen im Setting Deutsch–Deutsch deutet ein Chi-Quadrat-Test auf ein signifikantes Verhältnis zwischen den jeweiligen AOI-Kategorien und den eingehenden Regressionen hin ($\chi^2(2) = 16,91, p < 0,001$). Weiterhin besteht ein enger Zusammenhang zwischen den eingehenden Regressionen und der progressiven ersten Fixation sowohl bei Betrachtung des gesamten Datensatzes ($\chi^2(1) = 6,19, p < 0,05$) als auch der AOI-Kategorie A ($\chi^2(1) = 10,84, p < 0,001$). Alle weiteren Variablen sind unauffällig.

6.5.2.3.2 Ausgehende Regressionen

Abbildung 6.39 und Tabelle 6.46 zeigen die aufsummierte Anzahl aller Regressionen aus einem AOI sowie deren Mittelwert und Standardabweichung. Aus der Kategorie B, also von den Chatbeiträgen des Gegenübers, wurden mit 118 ($\bar{x} 0,89, SD: 1,35$) die meisten Regressionen getätigt. Aus den Kategorien A und *Eingabe* wurden jeweils 79 ($\bar{x} 0,75, SD: 1,18$) bzw. 51 ($\bar{x} 7,29, SD: 19,28$) Regressionen vorgenommen. Aus den beiden Nachrichtenkategorien wird im Schnitt weniger als eine Regression ausgeführt. Anders verhält es sich bei der Eingabemaske, aus der durchschnittlich sieben Regressionen getätigt werden, auch wenn der absolute Wert der niedrigste in diesem Fall ist.

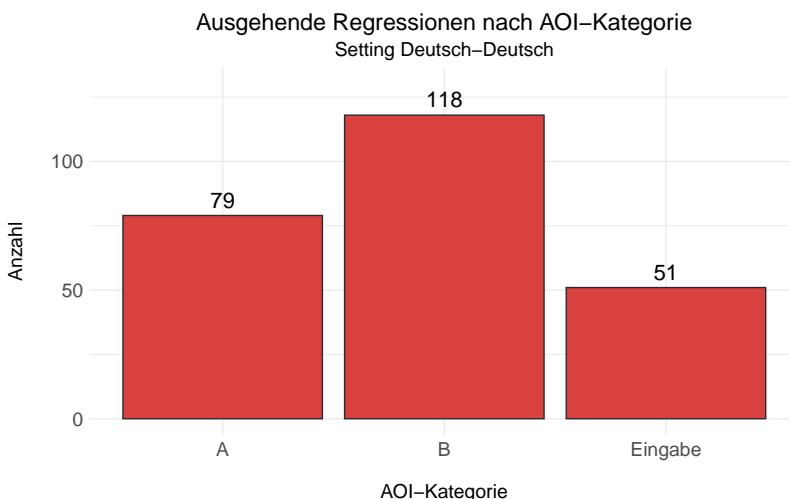


Abbildung 6.39: Anzahl Regressionen aus einem AOI

6.5 Der Datensatz im Setting Deutsch-Deutsch

Tabelle 6.46: Summe, Mittelwert und SD der ausgehenden Regressio-nen pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	SD
A	79	0,75	1,18
B	118	0,89	1,35
Eingabe	51	7,29	19,28
Global	248	1,01	3,45

Zur Untersuchung der Regressionen aus einem AOI wurde mit einem binomialen logistischen Regressionsmodell mit gemischten Effekten gearbeitet, das die Teilnehmer-innen (*SessionLabel*) als Zufallseffekte beinhaltet. Die Variablen wurden in einer bedingten Vorwärtsauswahl in das Modell integriert. Es gelang jedoch nicht, auf Grundlage der in den anderen Fällen verwendeten Variablen ein ausreichend stabiles, aussagekräftiges Modell zu erstellen. Eine Inspektion der Parameter des Modells belegte zu unterschiedlichen Zeitpunkten mangelnde Präzision (Signifikanz, AIC-Werte, R²).

Deshalb wurden in diesem Fall die in den anderen Modellen verwendeten Variablen einzig mittels Chi-Quadrat-Test paarweise untersucht. Während der Chi-Quadrat-Test für die Variablen *SessionLabel*, *IARegPD*, *IATag* und *FRundDwell* keine signifikanten Werte aufweist, deuten die Ergebnisse bei der Betrachtung von *IAFFixPro* auf eine signifikante Beziehung zwischen den ausgehenden Regressionen und der unabhängigen Variablen hin. Sowohl der gesamte Datensatz ($\chi^2(1) = 61,57, p < 0,001$) als auch die Unterteilung in beide AOI-Kategorien *A* ($\chi^2(1) = 28,72, p < 0,001$) und *B* ($\chi^2(1) = 33,19, p < 0,001$) sind dabei in Bezug auf die progressive erste Fixation statistisch auffällig.

6.5.2.4 Dauer des ersten Durchlaufs

Tabelle 6.47 (S. 184) zeigt die durchschnittliche Dauer des ersten Durchlaufs pro AOI (*first run dwell time*). Die aufsummierte Dauer der eigenen Beiträge beträgt in etwa nur ein Drittel der der Fremdbeiträge (63.114 ms zu 167.157 ms). Auch die Durchschnittswerte unterscheiden sich um mehr als das Doppelte (595,42 ms zu 1.266,34 ms). Die Durchlaufdauer der Eingabemaske beträgt 2.683 ms ($\varnothing 383,29$ ms).

6 Analyse

Tabelle 6.47: Mittelwert, Median und SD der Dauer des ersten Durchlaufs in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
A	63.114	595,42	331,0	661,81
B	167.157	1.266,34	635,5	1.498,92
Eingabe	2.683	383,29	349,0	250,56
Global	232.954	950,83	457,0	1.230,55

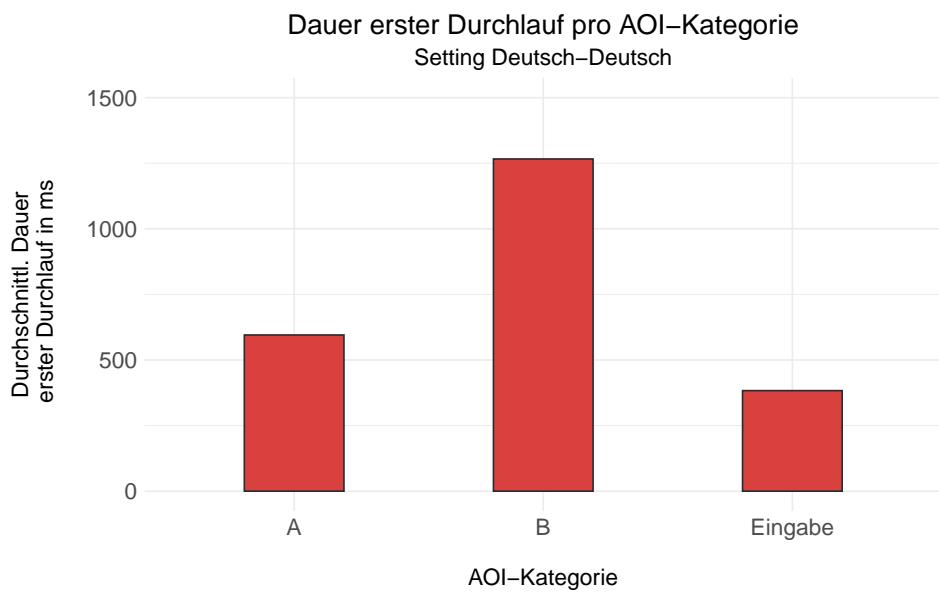


Abbildung 6.40: Durchschnittliche Betrachtungsdauer im ersten Durchlauf pro AOI-Kategorie im Setting Deutsch–Deutsch

6.5 Der Datensatz im Setting Deutsch-Deutsch

Wie die graphische Inspektion des Datensatzes anhand von Abbildung 6.41 belegt, kann nicht von einer Normalverteilung ausgegangen werden. Eine logarithmische Transformation (rechte Seite der Abbildung) ändert nichts an dieser Beobachtung. Auch die statistische Untersuchung des Datensatzes mittels Shapiro-Wilk-Test lehnt die Annahme einer Normalverteilung ab (roh: $W = 0,67$, $p < 0,001$ und log10: $W = 0,98$, $p < 0,01$).

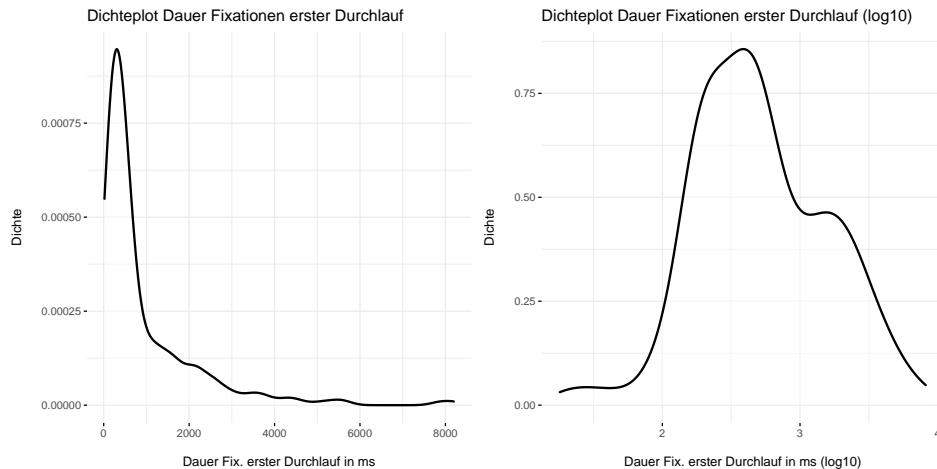


Abbildung 6.41: Verteilung der Dauer der Fixationen im ersten Durchlauf: normal (links), logarithmisch transformiert (rechts)

Keine Auffälligkeiten werden bei der Betrachtung der Unterschiede in der Dauer des ersten Durchlaufs pro AOI in Verbindung mit den einzelnen Teilnehmer*innen festgestellt. Ein Kruskal-Wallis-Test ist hierbei nicht signifikant ($\chi^2(6) = 9,73$, $p = 0,14$). Ein weiterer Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede bei der Dauer des ersten Durchlaufs pro AOI zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 19,08$, $p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, Tab. 6.48) zeigen, dass sich vor allem das Paar A-B ($z = -4,19$, $p < 0,01$, $r = 0,43$) unterscheidet, sodass gefolgert werden kann, dass die Durchlaufdauer von der AOI-Kategorie, und speziell von der Richtung (hereinkommende Nachrichten), abhängt. Dabei handelt es sich gemäß der Effektstärke nach Cohen (1992) um einen mittleren Effekt.

Da die unabhängige Variable der progressiven ersten Fixation nur zwei Ausprägungen aufweist, wird der exakte Mann-Whitney-U-Test angewendet. Dieses Testverfahren belegt, dass Unterschiede bei der Dauer der ersten Durchläufe pro AOI zwischen beiden Arten der erstmaligen Fixation bestehen ($U = 6.795,5$, $p < 0,05$, $r = 0,14$). Dabei handelt es sich gemäß der Effektstärke nach Cohen (1992) um einen schwachen Effekt.

6 Analyse

Tabelle 6.48: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Dauer des ersten Durchlaufs nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	-4,188020	0,0000***	0,43
A-Eingabe	0,444057	0,3285	-
B-Eingabe	1,855061	0,0477	-

Noch genauer zeigen sich signifikante Unterschiede bei der Durchlaufdauer der eigenen Chatbeiträge (*A*), wenn zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, als wenn es sich um eine tatsächliche chronologische erstmalige Fixation handelt. Ein exakter Mann-Whitney-U-Test ($U = 1.342,5, p < 0,05, r = 0,21$) in Verbindung mit einer Untersuchung der Effektstärke nach Cohen (1992) zeigt dabei einen schwachen Effekt. Ein Mann-Whitney-U-Test für die Be trachtung der progressiven ersten Fixation der fremden Chatbeiträge (*B*) erweist sich als nicht auffällig ($U = 1.976,0, p = 0,251, r = 0,1$).

Die Durchlaufdauer korreliert signifikant mit der Größe der AOI ($r_s = 0,28, p < 0,01, n = 1.775.506$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 7,59 %.

6.5.2.5 Gesamtverweildauer

Tabelle 6.49 gibt Summe, Mittelwert und Standardabweichung der Gesamtverweildauer in Millisekunden pro AOI-Kategorie wieder. Auf den Beiträgen des Gegenübers verweilen die Proband·innen mit 1.187.000 ms beinahe vier Mal so lang wie auf den eigenen Nachrichten, deren Gesamtdauer 378.000 ms beträgt. Die Eingabemaske wird mit den Nachrichten des Gegenübers vergleichbar lange fixiert, nämlich 1.089.000 ms. So ergibt sich in der Summe eine Gesamtdauer von 2.653.000 ms für alle AOI. Mit Blick auf die Mittelwerte ergibt sich ein ähnliches Bild. Auch hier macht die durchschnittliche Verweildauer für die eigenen Nachrichten (3.000 ms) etwa ein Drittel der Betrachtungsdauer der Nachrichten des Gegenübers (8.900 ms) aus.

Wie die graphische Inspektion des Datensatzes anhand von Abbildung 6.43 belegt, kann nicht von einer Normalverteilung ausgegangen werden. Eine logarithmische Transformation (rechte Seite der Abbildung) ändert nichts an dieser Beobachtung. Auch die statistische Untersuchung des Datensatzes mittels Shapiro-Wilk-Test lehnt die Annahme einer Normalverteilung ab (roh: $W = 0,27, p < 0,001$ und log10: $W = 0,99, p < 0,001$).

6.5 Der Datensatz im Setting Deutsch-Deutsch

Tabelle 6.49: Summe, Mittelwert, Median und SD der Gesamtverweildauer pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
A	377.527	3.069,33	1.330	4.908,75
B	1.186.758	8.856,40	5.368	12.609,55
Eingabe	1.088.618	155.516,86	102.914	126.036,86
Global	2.652.903	10.048,88	3.237	32.259,90

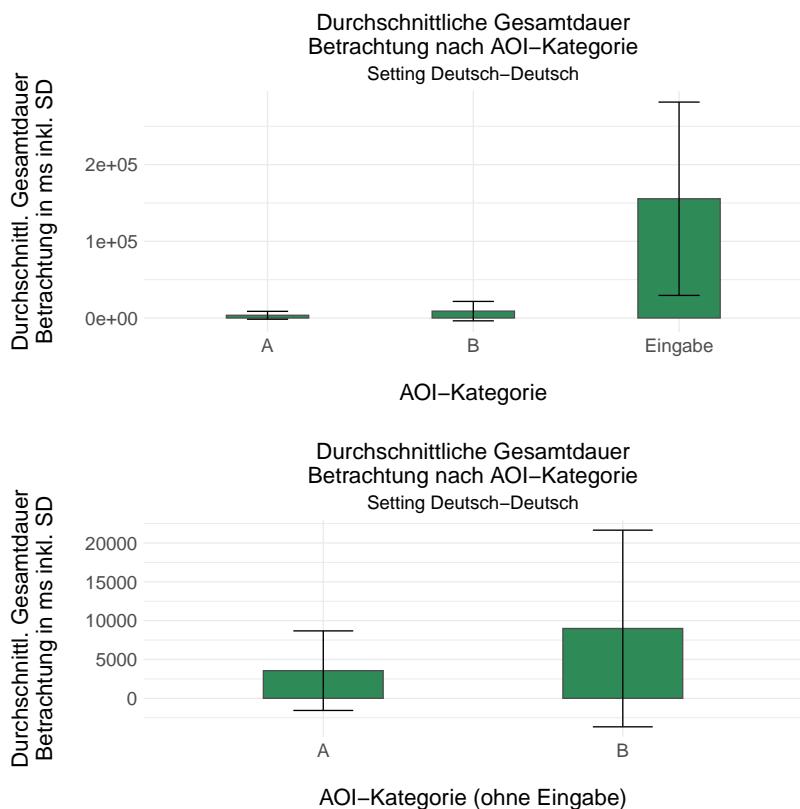


Abbildung 6.42: Durchschnittliche Betrachtungsdauer pro AOI-Kategorie im Setting Deutsch-Deutsch, mit (oben) und ohne (unten) Eingabemaske

6 Analyse

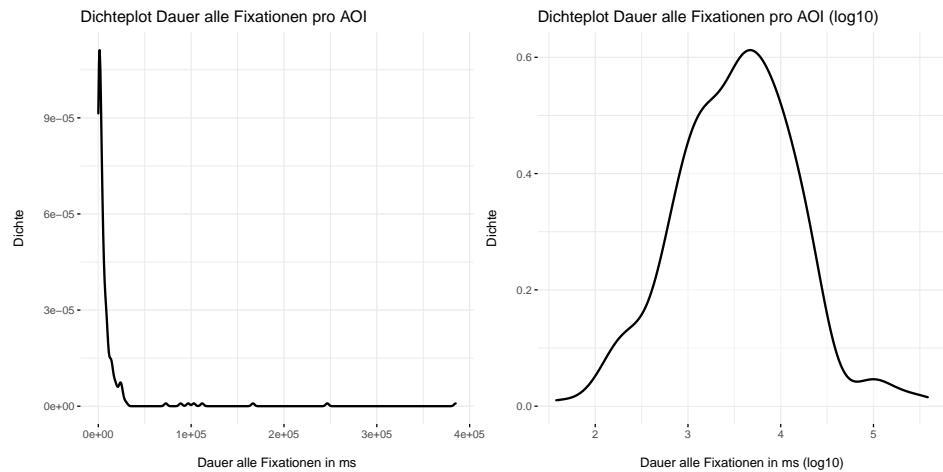


Abbildung 6.43: Verteilung der Gesamtdauer aller Durchläufe: normal (links), logarithmisch transformiert (rechts)

Keine Auffälligkeiten werden bei der Betrachtung der Unterschiede in der Gesamtverweildauer in Verbindung mit den einzelnen Teilnehmer·innen festgestellt. Ein Kruskal-Wallis-Test ist hierbei nicht signifikant ($\chi^2(6) = 4,66, p = 0,59$). Ein weiterer Kruskal-Wallis-Test zeigt allerdings, dass Unterschiede in der zentralen Tendenz der Gesamtverweildauer zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 48,55, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.50) zeigen, dass sich alle drei möglichen Gruppierungen signifikant unterscheiden, sodass gefolgert werden kann, dass die Verweildauer von der AOI-Kategorie abhängt. Unter Anwendung der Effektstärke nach Cohen (1992) handelt es sich bei *B-Eingabe* um einen schwachen Effekt ($r = 0,22$), bei *A-B* und *A-Eingabe* um jeweils mittlere Effekte ($r = 0,45$ und $r = 0,46$).

Tabelle 6.50: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Gesamtverweildauer nach AOI-Kategorie

1cAOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	-6,920374	0,0000 ***	0,45
A-Eingabe	-4,867397	0,0000 ***	0,46
B-Eingabe	-2,649329	0,0040 **	0,22

6.5 Der Datensatz im Setting Deutsch-Deutsch

Zur Untersuchung der progressiven ersten Fixation wird der exakte Mann-Whitney-U-Test angewendet. Dieses Testverfahren belegt, dass Unterschiede bei der Gesamtverweildauer pro AOI zwischen beiden Arten der erstmaligen Fixation bestehen ($U = 6.298,0, p < 0,001, r = 0,21$). Dabei handelt es sich gemäß der Effektstärke nach Cohen (1992) um einen schwachen Effekt. Ebenfalls zeigen sich signifikante Unterschiede bei der Verweildauer der Chatbeiträge des Gegenübers (B), wenn zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, als wenn es sich um eine tatsächliche chronologische erstmalige Fixation handelt. Ein exakter Mann-Whitney-U-Test ($U = 1.796,0, p < 0,05, r = 0,18$) in Verbindung mit einer Untersuchung der Effektstärke nach Cohen (1992) zeigt dabei einen schwachen Effekt. Ein Mann-Whitney-U-Test für die Betrachtung der progressiven ersten Fixation der eigenen Chatbeiträge (A) liegt auf der Grenze des Signifikanzniveaus ($U = 1.378,0, p = 0,051, r = 0,19$).

Die Gesamtverweildauer korreliert signifikant mit der Größe der AOI ($r_s = 0,75, p < 0,01, n = 764.790$). Dabei handelt es sich nach Cohen (1992) um einen starken Effekt. Das Bestimmtheitsmaß beträgt 56,34 %.

6.5.2.6 Regressive Durchlaufdauer

Tabelle 6.51 stellt Summe, Mittelwert und Standardabweichung der regressiven Durchlaufdauer in Millisekunden pro AOI-Kategorie dar. Der Gesamtwert beträgt 2.848.000 ms. Die Nachrichten der Proband-innen werden insgesamt 1.173.000 ms betrachtet, wohingegen die Gesamtdauer der regressiven Durchlaufdauer der Beiträge des Gegenübers nur 1.093.330 ms beträgt. Die Mittelwerte betragen jeweils 11.067 ms (A) und 8.282 ms (B).

Tabelle 6.51: Summe, Mittelwert, Median und SD der regressiven Durchlaufdauer in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
A	1.173.109	11.067,07	4.460,5	16.635,21
B	1.093.330	8.282,80	5.061,5	10.272,36
Eingabe	581.094	83.013,43	1.678,0	212.477,43
Global	2.847.533	11.622,58	4.797,0	37.924,73

6 Analyse

Wie auch schon im anderen Setting erklärt, wurde die Eingabemaske als statisches AOI annotiert. Der hohe Mittelwert erklärt sich daher möglicherweise aus dem anfänglichen Warteverhalten der Versuchspersonen bis die Gesprächspartner·innen die erste Nachricht gesendet hat. Bis zu diesem Zeitpunkt verweilten die Proband·innen meist sehr lange auf der Eingabemaske und damit innerhalb des AOI. Wie die graphische Inspektion des Datensatzes anhand von Abbildung 6.44 belegt, kann nicht von einer Normalverteilung ausgegangen werden. Eine logarithmische Transformation (rechte Seite der Abbildung) ändert nichts an dieser Beobachtung. Die statistische Untersuchung des Datensatzes mittels Shapiro-Wilk-Test lehnt die Annahme einer Normalverteilung ab (roh: $W = 0,21, p < 0,001$ bzw. log10: $W = 0,99, p < 0,05$). Daher wird mit nicht-parametrischen Tests gearbeitet.

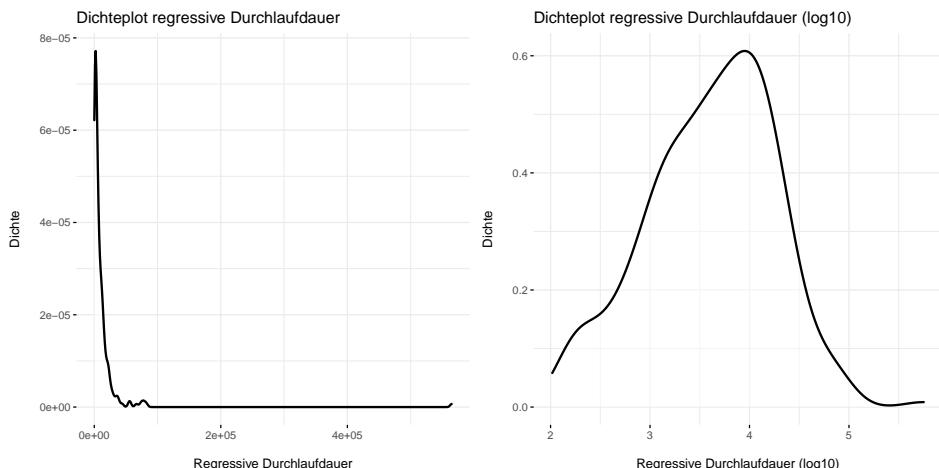


Abbildung 6.44: Verteilung der regressiven Durchlaufdauer: normal (links), logarithmisch transformiert (rechts)

Ein Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede bei der regressiven Durchlaufdauer zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(6) = 33,54, p < 0,001$). Ein zweiter Kruskal-Wallis-Test belegt, dass keine Unterschiede bei der regressiven Durchlaufdauer zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 0,9, p = 0,64$).

Es besteht wiederum ein signifikanter Unterschied bei der regressiven Durchlaufdauer zwischen AOI, die in chronologischer Reihenfolge betreten wurden und jenen, die aus einem AOI mit höherer Ordnungszahl besucht wurden. Da die Variable der progressiven ersten Fixation nur in zwei Ausprägungen vorliegt, wurde hier der Mann-Whitney-U-Test angewendet ($U = 3.899,5, p < 0,001$,

$r = 0,52$). Dieser Befund gilt ebenfalls für die Unterteilung nach AOI-Kategorie ($A: U = 743,0, p < 0,001, r = 0,62, B: U = 1.177,0, p < 0,001, r = 0,47$). Für die Durchlaufdauer durch beide Nachrichtenarten macht es demnach einen Unterschied, ob die Proband·innen vorher bereits nachfolgende Nachrichten(-AOI) betrachtet haben oder nicht.

Die regressive Durchlaufdauer korreliert außerdem mit der Größe der AOI ($r_s = 0,35, p < 0,01, n = 1.593.303$). Dabei handelt es sich nach Cohen (1992) um einen mittleren Effekt. Das Bestimmtheitsmaß beträgt 12,24 %.

6.5.2.7 Selektive Regressive Durchlaufdauer

Tabelle 6.52 stellt Summe, Mittelwert und Standardabweichung der regressiven Durchlaufdauer in Millisekunden pro AOI-Kategorie dar. Der Gesamtwert beträgt 1.281.202 ms. Die Nachrichten der Versuchspersonen werden insgesamt 247.249 ms betrachtet, wohingegen die Gesamtdauer der regressiven Durchlaufdauer der Beiträge des Gegenübers 632.744 ms beträgt. Die Mittelwerte betragen jeweils 2.332,54 ms (A) und 4.793,52 ms (B). Die Annotation der Eingabemaske als statisches AOI und die damit verbundenen Konsequenzen für die Auswertung wurden bereits mehrfach erklärt (s. o.).

Tabelle 6.52: Summe, Mittelwert, Median und SD der selektiven regressiven Durchlaufdauer in ms pro AOI-Kategorie

AOI-Kategorie	Summe	Mittelwert	Median	SD
A	247.249	2.332,54	994,0	3.883,06
B	632.744	4.793,52	2.600,5	7.454,11
Eingabe	401.209	57.315,57	1.678,0	144.493,5
Global	1.281.202	5.229,4	1.509,0	25.125,72

Wie die graphische Inspektion des Datensatzes anhand von Abbildung 6.45 belegt, kann nicht von einer Normalverteilung ausgegangen werden. Eine logarithmische Transformation (rechte Seite der Abbildung) ändert nichts an dieser Beobachtung. Die statistische Untersuchung des Datensatzes mittels Shapiro-Wilk-Test lehnt die Annahme einer Normalverteilung ab (roh: $W = 0,13, p < 0,001$ bzw. $\log_{10}: W = 0,99, p < 0,05$). Daher wird mit nicht-parametrischen Tests gearbeitet.

Ein Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede bei der selektiven regressiven Durchlaufdauer zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(6) = 15,99, p < 0,01$). Ein zweiter Kruskal-Wallis-Test belegt, dass

6 Analyse

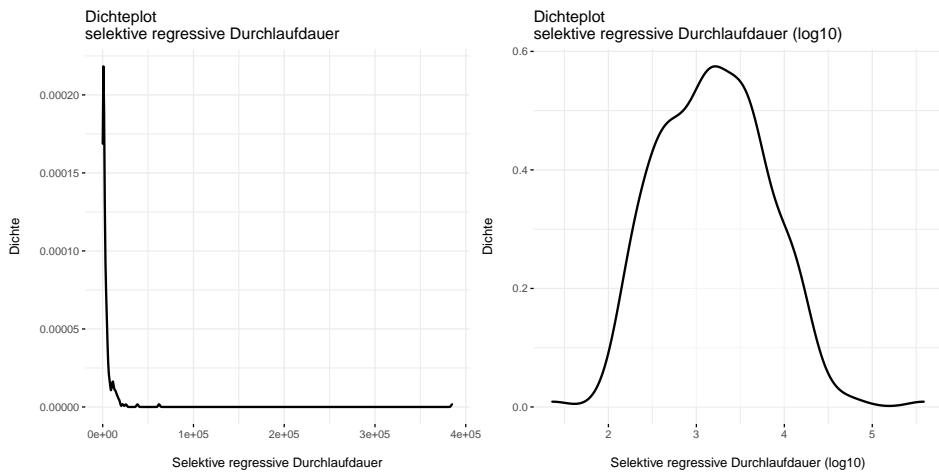


Abbildung 6.45: Verteilung der selektiven regressiven Durchlaufdauer: normal (links), logarithmisch transformiert (rechts)

Unterschiede bei der selektiven regressiven Durchlaufdauer zwischen den AOI-Kategorien bestehen ($\chi^2(2) = 18,36, p < 0,001$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.56, S. 195) zeigen, dass sich das Paar A-B signifikant unterscheidet. Die Effekstärke ist mit 0,25 allerdings als schwach anzusehen.

Tabelle 6.53: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der selektiven regressiven Durchlaufdauer nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	-4,166639	0,0000***	0,25
A-Eingabe	-1,756923	0,0593	-
B-Eingabe	-0,365754	0,3573	-

Es besteht wiederum ein signifikanter Unterschied bei der selektiven regressiven Durchlaufdauer zwischen AOI, die in chronologischer Reihenfolge betreten wurden und jenen, die aus einem AOI mit höherer Ordnungszahl besucht wurden (progressive erste Fixation). Hierzu wurde der Mann-Whitney-U-Test angewendet ($U = 5.115,0, p < 0,001, r = 0,36$). Dieser Befund gilt ebenfalls für die Unterteilung nach AOI-Kategorie (A: $U = 967,0, p < 0,001, r = 0,47$, B: $U = 1.536,5, p < 0,001, r = 0,30$). Für die Durchlaufdauer durch beide Nachrichtarten macht es demnach einen Unterschied, ob die Proband·innen vorher bereits nachfolgende Nachrichten(-AOI) betrachtet haben oder nicht.

Die selektive regressive Durchlaufdauer korreliert außerdem mit der Größe der AOI ($r_s = 0,56, p < 0,01, n = 1.072.123$). Dabei handelt es sich nach Cohen (1992) um einen starken Effekt. Das Bestimmtheitsmaß beträgt 31,64 %.

6.5.2.8 Pupillengröße

Tabelle 6.54 (S. 193) zeigt die durchschnittliche Pupillengröße sowie die Standardabweichungen. Bei der Betrachtung der Eingabemaske liegt die Pupillengröße 122 Einheiten (126 %) über dem Gesamtschnitt. Die Pupillenweitung bei Betrachtung der Chatbeiträge des Gegenübers gleicht beinahe dem Durchschnitt (*B*: 473 Einheiten, Durchschnitt 470 Einheiten). Die Pupillengröße beim eigenen Beitrag ist 2,4 % kleiner als der Durchschnitt (*A*: 459 Einheiten, Durchschnitt: 470 Einheiten). Der Unterschied zwischen beiden Gruppen (*A*: 459, *B*: 473) beträgt 2,96 %.

Tabelle 6.54: Mittelwert, Median und SD der Pupillengröße pro AOI-Kategorie in willkürlichen Einheiten

AOI-Kategorie	Mittelwert	Median	SD
A	459,19	420,57	121,15
B	473,07	439,29	112,05
Eingabe	592,34	602,74	128,63
Global	470,64	438,50	118,09

Mit Blick auf Tabelle 6.55 (S. 194) kann zudem festgestellt werden, dass die durchschnittliche Pupillengröße pro Versuchsteilnehmer:in stark variiert: Das Minimum beträgt 358,58 Einheiten und liegt somit 120 Einheiten unter dem Durchschnitt. Das Maximum hingegen beträgt 693,82 Einheiten und liegt 220 Einheiten über dem Schnitt.

Eine graphische Inspektion der Daten (Abbildung 6.47) zeigt, dass weder die tatsächlichen noch die logarithmisch transformierten Werte normalverteilt sind. Auch die statistische Untersuchung unter Anwendung des Shapiro-Wilk-Tests deutet auf nicht normalverteilte Daten hin (roh: $W = 0,88, p < 0,001$, log10: $W = 0,93, p < 0,001$). Daher wird die Pupillengröße mit nicht-parametrischen Tests untersucht.

Ein Kruskal-Wallis-Test belegt, dass Unterschiede bei der Pupillengröße zwischen den einzelnen Teilnehmer:innen bestehen ($\chi^2(6) = 152,19, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass die gruppierte Betrachtung von 15/21 (71,42 %) möglichen Paaren signifikante Unterschiede aufweist, sodass gefolgert werden kann, dass die Pupillengröße zwischen den Versuchspersonen variiert.

6 Analyse

Tabelle 6.55: Mittelwert, Median und SD der Pupillengröße pro Studienteilnehmer·in

Pseudonym	Mittelwert	Median	SD
DD_TN1	420,39	418,52	49,70
DD_TN2	422,16	419,62	35,34
DD_TN4	693,82	702,04	41,09
DD_TN5	518,64	501,83	104,51
DD_TN6	353,58	349,84	23,14
DD_TN7	483,22	468,00	97,69
DD_TN8	447,12	456,67	54,22
Global	470,64	438,50	118,09

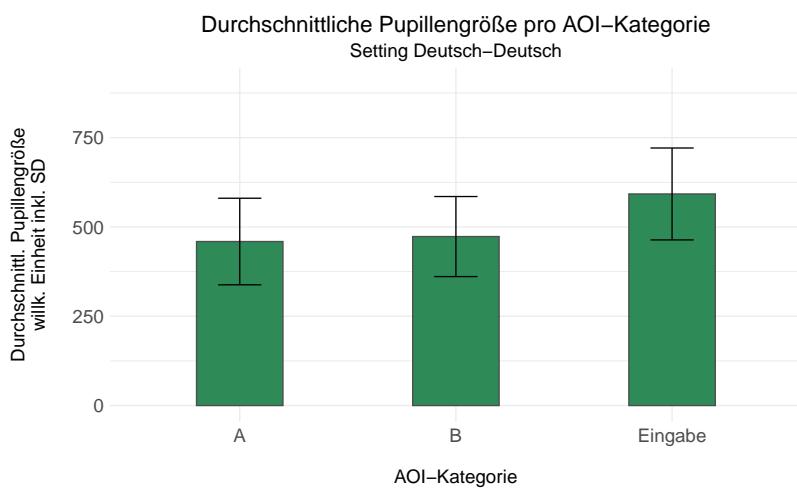


Abbildung 6.46: Durchschnittliche Pupillengröße pro AOI–Kategorie im Setting Deutsch–Deutsch

6.5 Der Datensatz im Setting Deutsch-Deutsch

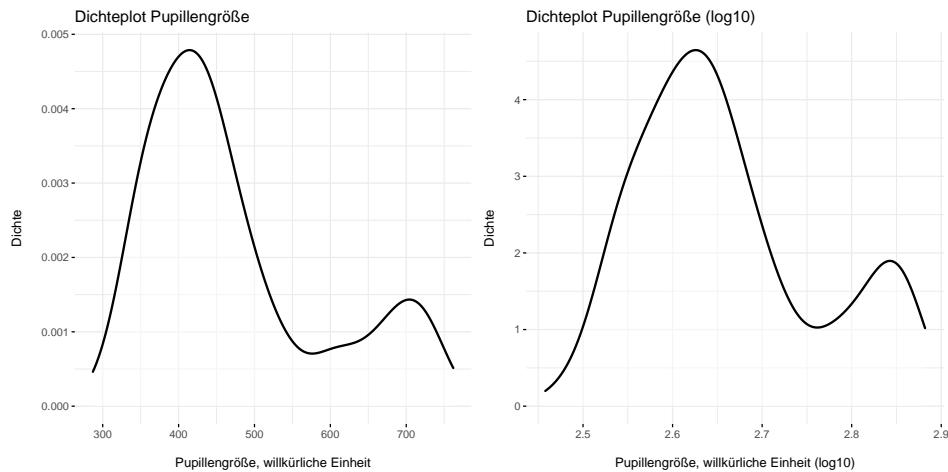


Abbildung 6.47: Verteilung der Pupillengröße in willkürlichen Einheiten: normal (links), logarithmisch transformiert (rechts)

Ein zweiter Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede bei der Pupillengröße zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(6) = 8,68$, $p < 0,05$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.56) zeigen, dass sich die Kategoriegruppen *A-Eingabe* ($z = -2,74$, $p < 0,01$, $r = -0,2621229$) und *B-Eingabe* ($z = -2,21$, $p < 0,05$, $r = -0,189165$) unterscheiden, sodass gefolgert werden kann, dass die Betrachtung der Eingabemaske einen signifikanten Einfluss auf die Pupillengröße hat, wobei die Effektgröße nur schwach ist.

Tabelle 6.56: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Pupillengröße nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	-1,588856	0,0560	
A-Eingabe	-2,736643	0,0093**	0,26
B-Eingabe	-2,214119	0,0201*	0,19

Es besteht hingegen ein signifikanter Unterschied in der zentralen Tendenz der Pupillengröße in Bezug auf die progressive erste Fixation zwischen den AOI, bei denen zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde, und denen, die chronologisch betreten wurden (global: $U = 4.280,5$, $p < 0,001$, $r = 0,45$).

6 Analyse

Sowohl bei eigenen Chatbeiträgen als auch bei fremden zeigen sich signifikante Unterschiede bei der Pupillengröße, wenn zuvor ein AOI mit höherer Ordnungszahl betrachtet wurde als wenn es sich um eine tatsächliche chronologische erstmalige Fixation handelt. Der exakte Mann-Whitney-U-Test für die eigenen Chatbeiträge ($U = 967,5, p < 0,01, r = 0,42$) in Verbindung mit einer Untersuchung der Effektstärke nach Cohen (1992) zeigt dabei einen mittleren Effekt. Bei fremden Chatbeiträgen weist der exakte Mann-Whitney-U-Test ($U = 1.154,00, p < 0,001, r = 0,473$) ebenfalls einen mittleren Effekt auf (s. hierzu Tabelle 6.57).

Tabelle 6.57: Ergebnisse des Mann-Whitney-U-Tests zur Pupillengröße nach AOI-Kategorie und progressiver ersten Fixation

AOI-Kategorie	U	p (angepasst)	Effekt
A	967,5	0,001**	0,42
B	1.154,0	0,001**	0,47
Eingabe	n/a	n/a	-
Global	4.280,5	0,001**	0,45

Die Pupillengröße korreliert signifikant mit der Größe der AOI ($r_s = -0,18, p < 0,01, n = 2.675.871$). Dabei handelt es sich nach Cohen (1992) um einen schwachen Effekt. Das Bestimmtheitsmaß beträgt 3,1 %.

6.5.3 Das sakkadische Blickverhalten

Im Setting Deutsch-Deutsch werden generell mehr als drei Mal so viele Sakkaden im Bereich der Beiträge des Gegenübers getätigten wie jeweils auf den eigenen bzw. auf der Eingabemaske (s. Tabelle 6.58), auf die zugleich die absolut wenigsten Sakkaden verfallen. Durchschnittlich werden hingegen in den eigenen und in den Beiträgen des Gegenübers etwa gleich viele Sakkaden gezählt. Die Mittelwerte betragen für ausgehende, eigene Nachrichten 9,76 Sakkaden und für eingehende, fremde Nachrichten 10,27 Sakkaden. Pro AOI im gesamten Datensatz werden im Schnitt 40,34 Sakkaden erfasst. Eine Berechnung der Werte für die Eingabemaske war nicht möglich, da es sich um ein statisches AOI handelt.

Tabelle 6.59 stellt die aufsummierte Sakkadenanzahl nach Richtung und AOI-Kategorie dar. Sakkaden, die außerhalb eines AOI gelandet sind, werden unter NA gezählt. In absoluten Zahlen fällt auf, dass innerhalb der AOI-Kategorie B mehr Sakkaden erfolgen als in den anderen beiden Kategorien. Rechtsgerichtete

Tabelle 6.58: Anzahl, Mittelwert und SD der Sakkaden pro AOI-Kategorie

AOI-Kategorie	Anzahl	Mittelwert	SD
A	781	9,76	4,36
B	2.557	10,27	5,08
Eingabe	615	n/a	n/a
Global	3.953	40,34	75.381,78

Sakkaden dominieren in jeder Kategorie. Das ist nicht weiter verwunderlich, da das Deutsche der Leserichtung von links nach rechts folgt. Auffällig sind hingegen die vergleichsweise vielen Sakkaden nach links und nach oben in der Kategorie *B*.

6.5.3.1 Sakkadenamplitude

Die durchschnittliche Sakkadenamplitude (s. Tabelle 6.60) bei Betrachtung der eingehenden Nachrichten liegt mit $1,95^\circ/\text{ms}$ unterhalb des Gesamtschnittes, die der Eingabemaske ($1,85^\circ/\text{ms}$) und der ausgehenden Nachrichten ($1,81^\circ/\text{ms}$) oberhalb des Gesamtschnittes ($1,84^\circ/\text{ms}$).

Der Anderson-Darling-Test zeigt, dass sowohl die Gesamtheit des Datensatzes als auch der Datensatz nach einzelnen AOI-Kategorien in Bezug auf die Sakkadenamplitude nicht normalverteilt ist. Der AD-Test weist ein Ergebnis von ($A = 429,07, p < 0,01$) für den gesamten Datensatz auf und für die Kategorien A: ($A = 85,64, p < 0,01$), B: ($A = 262,37, p < 0,01$) und Eingabe: ($A = 80,89, p < 0,01$). Auch eine logarithmische Transformation änderte die Verteilung nicht wesentlich (s. Abbildung 6.48). Auf die Kategorie A verfallen 781 Sakkaden, 2.557 auf B und 615 auf die Eingabemaske. Insgesamt enthält der Datensatz somit 3.953 Sakkaden.

Wie auch im Falle der Sakkadenamplitude und -dauer im Setting Katalanisch-Deutsch (s. S. 157) wurde die Korrelation beider Variablen untersucht. Für das Setting Deutsch-Deutsch ergeben sich folgende Werte aus dem Korrelationstest nach Spearman's Rho: Sowohl der gesamte Datensatz als auch die Werte nach einzelnen AOI-Kategorien getrennt stehen in einer signifikant positiven Beziehung zueinander – Gesamt: ($S = 8.037.112.694, p < 0,01, r = 0,63$), A: ($S = 48.444.315, p < 0,01, r = 0,66$) und B: ($S = 1.979.782.735, p < 0,01, r = 0,66$). Für die Eingabemaske lagen zu wenig Datenpunkte vor, als dass eine Berechnung möglich gewesen wäre.

6 Analyse

Tabelle 6.59: Sakkadenanzahl nach Richtung und AOI-Kategorie im Setting Deutsch-Deutsch

Kategorie		A		Kategorie		B	
Richtung		Anzahl	Prozent	Richtung		Anzahl	Prozent
Runter		33	4,23	Runter		62	2,42
Links		168	21,51	Links		611	24,90
Rechts		531	67,99	Rechts		1.661	64,96
Hoch		13	4,61	Hoch		177	6,92
NA		13	1,66	NA		46	1,80
Gesamt		781	100	Gesamt		2.557	100

Kategorie		Eingabe	
Richtung		Anzahl	Prozent
Runter		102	16,59
Links		205	33,33
Rechts		268	43,58
Hoch		25	4,07
NA		15	2,44
Gesamt		615	100

Tabelle 6.60: Mittelwert, Median und SD der Sakkadenamplitude pro AOI-Kategorie

AOI-Kategorie	Mittelwert	Median	SD
A	1,95	1,42	1,6
B	1,81	1,38	1,39
Eingabe	1,85	1,20	1,70
Global	1,84	1,36	1,49

6.5 Der Datensatz im Setting Deutsch-Deutsch

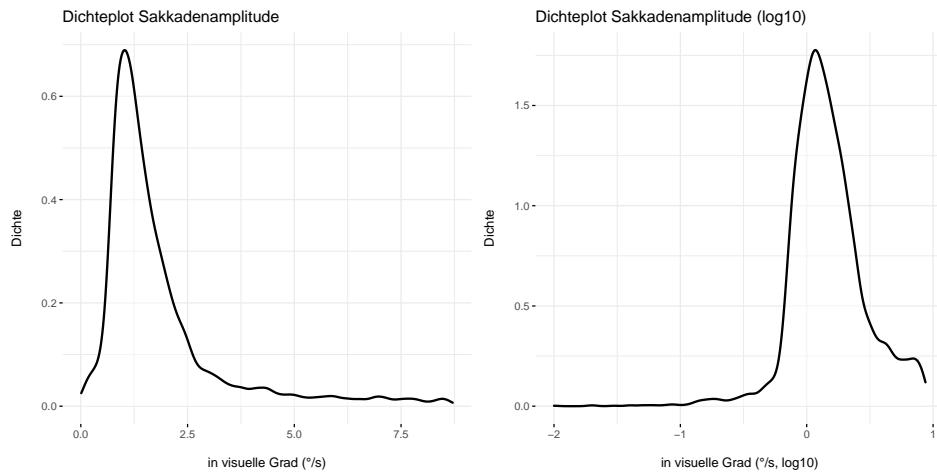


Abbildung 6.48: Verteilung der Sakkadenamplitude: normal (links), logarithmisch transformiert (rechts)

Ein Kruskal-Wallis-Test belegt, dass Unterschiede bei der Sakkadenamplitude zwischen den einzelnen Teilnehmer·innen bestehen ($\chi^2(6) = 81,54, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 14 von 21 möglichen Paaren (66,67 %) signifikant voneinander unterscheiden, sodass gefolgert werden kann, dass die Sakkadenamplitude in gewissem Maße von der Individualität der TN abhängt.

Ein weiterer Kruskal-Wallis-Test zeigt, dass Unterschiede bei der Sakkadenamplitude zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 30,63, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich die AOI-Kategorien *A-Eingabe* und *B-Eingabe* signifikant unterscheiden. Im Falle des Paares *A-Eingabe* gilt ($z = 17, p < 0,01, r = 0,12$) und der Kategorie *B-Eingabe* ($z = 4,83, p < 0,01, r = 0,08$), sodass gefolgert werden kann, dass die Sakkadenamplitude von der AOI-Kategorie abhängt (s. Tabelle 6.61), der Asterisk zeigt das Signifikanzlevel an: $^*\alpha < 0,05$).

Tabelle 6.61: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Sakkadenamplitude nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	1,591889	0,0557	-
A-Eingabe	5,174487	0,0000***	0,12
B-Eingabe	4,829444	0,0000***	0,08

6 Analyse

Die Tests im Setting Deutsch-Deutsch belegen signifikante Unterschiede in der Sakkadenamplitude für die Vergleichspaare mit Beteiligung der Eingabemaske. Die vergleichende Untersuchung der eigenen vs. fremden Beiträge bleibt unauffällig.

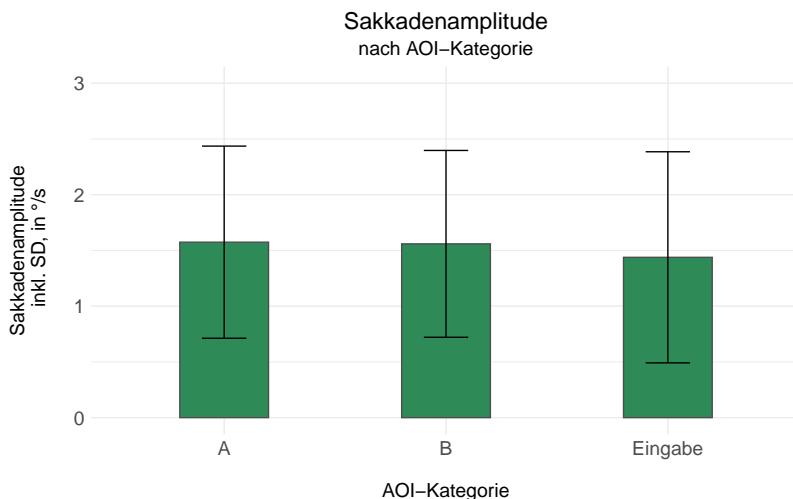


Abbildung 6.49: Sakkadenamplitude pro AOI

6.5.3.2 Sakkadendauer

Tabelle 6.62 stellt Mittelwert und Standardabweichung der Sakkadendauer pro AOI-Kategorie dar. Die Sakkaden im Bereich der ausgehenden Nachrichten dauern im Schnitt 1,8 ms länger als im Bereich der Nachrichten des Gegenübers. Auf den gesamten Datensatz bezogen beträgt die durchschnittliche Sakkadendauer 31,83 ms und liegt damit zwischen den Mittelwerten der beiden Beitragsarten.

Der Anderson-Darling-Test zeigt, dass sowohl die Gesamtheit des Datensatzes als auch der Datensatz nach einzelnen AOI-Kategorien in Bezug auf die Sakkadendauer nicht normalverteilt ist. Es ergibt sich für die Gesamtheit ein Verteilungswert im AD-Test von ($A = 935, p < 0,01$) und für die Kategorien A: ($A = 165,76, p < 0,01$), B: ($A = 634,64, p < 0,01$) und Eingabe: ($A = 138,21, p < 0,01$). Auch eine logarithmische Transformation änderte die Verteilung nicht wesentlich (s. Abbildung 6.50). Auf die Kategorie A verfallen somit noch 952 Sakkaden, 3.279 auf B und 847 auf die Eingabemaske. Insgesamt enthält der Datensatz 5.078 Sakkaden. Die durchschnittliche Sakkadendauer bei der Betrachtung eingehender Nachrichten liegt unterhalb des Gesamtschnittes, die der ausgehenden Nachrichten und der Eingabemaske überhalb des Schnittes.

6.5 Der Datensatz im Setting Deutsch-Deutsch

Tabelle 6.62: Mittelwert, Median und SD der Sakkadendauer pro AOI-Kategorie

AOI-Kategorie	Mittelwert	Median	SD
A	32,90	21	33,00
B	31,10	20	32,01
Eingabe	33,44	22	33,69
Global	31,83	21	32,49

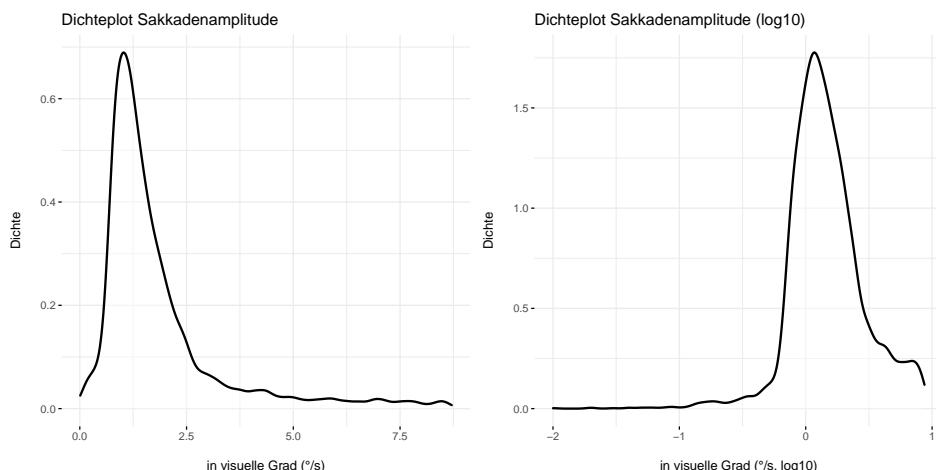


Abbildung 6.50: Verteilung der Sakkadendauer: normal (links), logarithmisch transformiert (rechts)

Ein Kruskal-Wallis-Test deutet darauf hin, dass Unterschiede bei der Sakkadendauer zwischen den einzelnen Teilnehmer.innen bestehen ($\chi^2(6) = 53,66, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests) zeigen, dass sich 12 von 21 (57,14 %) TN-Gruppen signifikant unterscheiden, sodass gefolgert werden kann, dass die Sakkadendauer in gewissem Maße von der Individualität der TN abhängt.

Ein weiterer Kruskal-Wallis-Test belegt, dass Unterschiede bei der Sakkadendauer zwischen den einzelnen AOI-Kategorien bestehen ($\chi^2(2) = 10,81, p < 0,01$). Anschließend durchgeführte Post-hoc-Tests (Dunn-Benjamini-Hochberg-Tests, s. Tabelle 6.63, der Asterisk zeigt das Signifikanzlevel an: * $\alpha < 0,05$) zeigen, dass sich die AOI-Kategorien *A-B* und *B-Eingabe* signifikant unterscheiden. Somit kann gefolgert werden, dass die Sakkadendauer von der AOI-Kategorie abhängt.

6 Analyse

Tabelle 6.63: Ergebnisse des Dunn-Tests: Gruppierte Vergleiche der Sakkadendauer nach AOI-Kategorie

AOI-Kategoriepaar	<i>z</i>	<i>p</i> (angepasst)	Effekt
A-B	2,800713	0,0076**	0,04
A-Eingabe	0,318257	0,3751	-
B-Eingabe	-2,285139	0,0167*	0,04

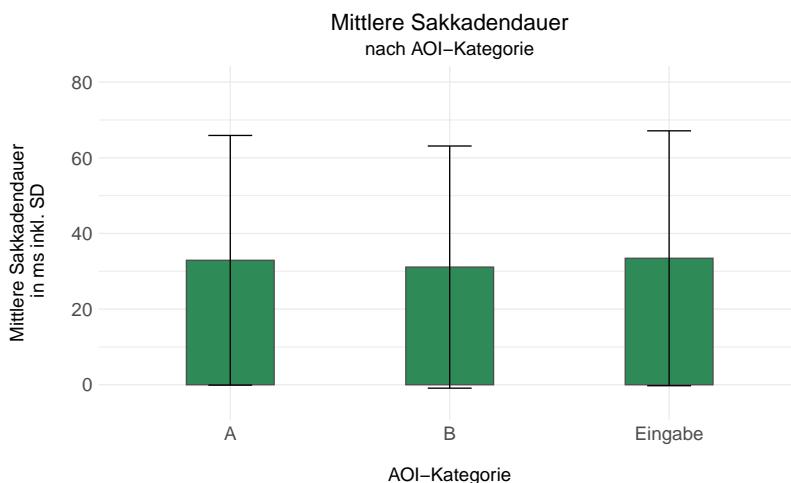


Abbildung 6.51: Sakkadendauer pro AOI

6.5.4 Zusammenfassung der Ergebnisse der einsprachigen Vergleichsstudie

Für die Vergleichsstudie wurden weniger Versuchspersonen rekrutiert als für die Untersuchung des Skype Translators. Damit einher geht auch ein verkürzter Versuchsaufbau, der nicht mit einem Abschlussfragebogen endet, in dem die Proband·innen ihre Eindrücke zu der maschinell übersetzten Kommunikations-situation darlegen sollen. Ziel dieser Vergleichsstudie ist es vielmehr, okulometrische Referenzdaten zu generieren, die die Wahrnehmung der Versuchsteilnehmer·innen im Setting Katalanisch-Deutsch kontrastieren.

Die Auswertung des Eingangsfragebogens zu Beginn der Studie zeugt von einer variabel zusammengesetzten Kohorte. Alle sieben Personen sind Studierende der Universität Leipzig. Es handelt sich um junge Menschen Anfang 20, die einerseits mehrere Fremdsprachen sprechen und andererseits auch mehrheitlich bereits Auslandserfahrung aufweisen können. Wie auch im anderen Versuch ist

6.5 Der Datensatz im Setting Deutsch-Deutsch

Skype für die Versuchspersonen nicht die präferierte Software für den Kontakt mit Freunden, Familie und anderen Menschen. Zugleich ist der Funktionsumfang von Skype nicht gänzlich unbekannt.

Mit Blick auf alle untersuchten Indikatoren wird ein Leseverhalten charakterisiert, das besonders die Eingabemaske und die eingehenden Textnachrichten des Gegenübers in den Fokus rückt. Dies wird bereits durch die Betrachtung der Bildschirmaufnahmen visuell ersichtlich. Es wirkt allerdings so, als wiesen die Indikatoren in diesem einsprachigen Setting eine breitere Streuung über den Bildschirm auf als im anderen Versuchsaufbau. Diese Beobachtung ist vorläufig und daher mit Vorsicht auszusprechen. Eine eingehende Untersuchung sollte vorgenommen werden. Die fremden Beiträge werden sowohl absolut-numerisch als auch statistisch mehr betrachtet. Wie auch im katalanisch-deutschen Versuchsaufbau sind die Erkenntnisse, die aus den Daten der Eingabemaske gezogen werden können, nur begrenzt interpretierbar. Die Eingabemaske wurde auch hier als statisches AOI annotiert, somit sind Indikatoren, die auf ein erstmaliges Ereignis abzielen, auch tatsächlich immer nur als das erstmalige, singuläre Ereignis zu verstehen, selbst wenn die Eingabemaske im weiteren Verlauf der Studie erneut oder insgesamt mehrmals betrachtet wurde. Die in der Zusammenfassung zur katalanisch-deutschen Studie erwähnten interpretierbaren Indikatoren (Fixationszahl und Gesamtverweildauer, s. Abschnitt 6.3.4, S. 164) gelten auch hier.

Im Gegensatz zur Analyse der Regressionen im anderen Studienteil war es in diesem Fall nur begrenzt möglich, entsprechende Regressionsmodelle zur Untersuchung zu erstellen. Es lassen sich dennoch wenige, begrenzte Annahmen aufstellen. Einerseits liegt die Vermutung nahe, dass die Versuchspersonen im deutsch-deutschen Setting noch einmal zu den eingehenden Nachrichten des Gegenübers zurückspringen, um die erhaltenen Informationen mit den eigenen Beiträgen abzugleichen. Dies kann sowohl für die bereits gesendeten Nachrichten (*A*) gelten als auch für Nachrichten, die die Proband·innen gerade im Begriff sind zu verfassen und sich daher noch im AOI der Eingabemaske befinden.

Im Gegensatz zur Sakkadenamplitude weisen die Tests der Sakkadendauer auf signifikante Unterschiede zwischen den eigenen und fremden Beiträgen sowie den fremden Beiträgen und der Eingabemaske hin. Hieraus lässt sich die Annahme ableiten, dass die Beiträge des Gegenübers anders wahrgenommen und verarbeitet werden als einerseits der Text während der Produktion in der Eingabemaske als auch das Rezipieren der eigenen Beiträge.

7 Diskussion

Das vorangegangene Kapitel war der Analyse der verschiedenen Studiendaten gewidmet. Nach der Betrachtung der subjektiven Angaben im Rahmen der Umfragen und der Eye-Tracking-Daten werden im Folgenden nun die Ergebnisse der einzelnen Studienteile kontrastiert und in Bezug sowohl zu bestehenden Forschungsergebnissen als auch zu der aufbereiteten Theorie gesetzt. Die Diskussion folgt dabei der bereits bestehenden linearen Struktur des vorangegangenen Analyseteils und verweist innerhalb der jeweiligen Unterkapitel (Abschnitt 7.3.1–7.4.3, S. 210 ff.) auf vergleichbare Studien aus dem aktuellen Forschungsumfeld. Entlang dieses Weges werden auch immer wieder die zu Beginn dieser Arbeit präsentierten theoretischen Grundlagen aufgegriffen. Das Kapitel endet mit einem Überblick über weiterführende Denkanstöße, praktische Schlussfolgerungen sowie der kritischen Auseinandersetzung mit dem Studienaufbau (Abschnitt 7.5, S. 228).

7.1 Merkmale der CvK an Skype und dem Translator

Im vorausgehenden Kapitel wurde die Chatkommunikation über Skype jeweils einmal mit und einmal ohne Vermittlung durch den Skype Translator analysiert. Die Studienteilnehmer·innen beider Teile stammen aus einer Zielgruppe, die nativ mit den Vorzügen dieser Kommunikationstechnologie aufgewachsen ist. Das Durchschnittsalter der Teilnehmer·innen in den drei Erhebungen liegt bei etwa 24 Jahren und somit in einer Generation, die von der Sozialforschung als *Generation Y* eingegrenzt wird. Diese Personengruppe umfasst in etwa die Jahrgänge 1985–1999 und zeichnet sich u. a. durch einen intuitiven Umgang mit und zugleich intensiven bis exzessiven Konsum von Medien (in diesem Fall sind besonders die im Verlauf dieser Arbeit thematisierten Kommunikationsmedien wie Skype, WhatsApp u. ä. gemeint) aus.

Weiterhin wird dieser Generation nachgesagt, mehr Zeit und Geld in Bildung zu investieren als alle anderen. Interpretiert man hierbei die Fremdsprachenkenntnisse als möglichen Indikator, so zeichnen sich die Personen in den Erhebungen nicht nur formell über die offiziellen Abschlüsse als gebildet aus. Damit

7 Diskussion

einher gehen auch ein oder mehrere Auslandsaufenthalte von durchschnittlich etwa 12 Monaten, die häufig im Rahmen eines Studiums oder bereits nach dem Schulabschluss absolviert werden. Die Mehrheit gibt Deutsch als Muttersprache an – ein Kriterium, das für die Rekrutierung der Proband·innen zwingend erfüllt werden musste –, jedoch beherrschen die meisten Personen mindestens eine, wenn nicht sogar mehrere weitere Fremdsprachen. Dies ist zunächst das Englische, aber auch die in der deutschen Schullandschaft weit verbreiteten Französisch und Spanisch sind häufig vertreten.

In allen Umfrageergebnissen fällt auf, dass Skype nicht mehr die erste Wahl für die Kommunikation über das Internet ist. Die überwiegende Mehrheit der Teilnehmer·innen an der Online-Umfrage (144 Personen) nutzt Skype demnach lediglich weniger als monatlich. Die gleichen Verhältnisse sind in den Eingangsfragebögen beider Studienteile wiederzufinden. Über die vergangenen Jahre hat die Anwendung harte Konkurrenz zu jeder Funktion aus dem Portfolio erhalten. Hiervon zeugen sowohl die Angaben in der Online-Umfrage als auch die der Proband·innen in beiden Versuchsaufbauten. Auch wenn Skype lange Zeit eine Ikone für die Videokommunikation über das Internet darstellte, sind es heute Dienste wie WhatsApp, Facetime oder der Facebook Messenger, die größere Präsenz auf der mentalen Landkarte der Endnutzer·innen aufweisen.

Zugleich scheint es, als wäre dem Dienst nicht gänzlich der Rang abgelaufen worden. Davon zeugt die nicht unbedeutliche Anzahl an Umfrageteilnehmer·innen an der Online-Erhebung, die besonders für Videochats auf Skype zurückgreifen. Als starker Kontrast wirken allerdings die Angaben der Studienteilnehmer·innen in beiden Studienvarianten, bei denen Skype sowohl generell als auch in puncto am häufigsten genutzte Anwendung hinter WhatsApp, Apple Facetime und Telegram zurückfällt.

Sowohl die Auswertung des deutschlandweit verteilten Online-Fragebogens als auch die Antworten der Versuchsteilnehmer·innen zeugt davon, dass der Skype Translator bislang keinerlei Reichweite besitzt. Die Funktion ist keiner Person bekannt gewesen. Auch die Angaben des Abschlussfragebogens deuten darauf hin, dass dem Skype Translator an sich wenig Aufmerksamkeit zuteil wurde, auch wenn die Proband·innen sich in Ermanglung von Katalanischkenntnissen auf die MÜ-Ausgabe verlassen mussten. Zugleich fielen den Personen grundlegende syntaktische, semantische und pragmatische Fehler auf. Die Chatsituation hat dies offensichtlich nicht sonderlich stark beeinflusst.

Weiterhin ist auf Grundlage der Durchsicht der Bildschirmaufzeichnungen festzustellen, dass die Versuchspersonen zu keinem Zeitpunkt willentlich innerhalb des Chats zurückscrollten, um ältere Nachrichten noch einmal erneut zu lesen. Die Aufmerksamkeit war die gesamte Chatsituation über immer auf die

7.1 Merkmale der CvK an Skype und dem Translator

jüngsten Nachrichten gerichtet. Die Informationsintegration verläuft daher offenbar strikt linear, selbst wenn ein MÜ-System in die Kommunikation involviert ist.

Auf Seite der Entwickler drängt sich somit allerdings die Frage auf, welchen Vorteil große, multinationale Konzerne wie Microsoft oder Google von dem Angebot haben. Wenn die Software einerseits nicht (mehr) Marktführer ist und zugleich die eingeschaltete Funktion offenbar kaum bewusst von den Nutzer·innen wahrgenommen wird, stehen sich an dieser Stelle Zugewinn für die Unternehmen und Aufwand auf den ersten Blick in einem Missverhältnis gegenüber. Als naheliegender Erklärungsansatz könnte vermutet werden, dass die Echtzeit-Übersetzung als Funktion eines der Softwareprodukte lediglich ein Nebenerzeugnis eines anderen Unternehmensbereiches ist. Microsoft und Google sind qua Unternehmensausrichtung auf natürliche Sprachdaten angewiesen und generieren durch ihre Produkte zugleich Unmengen an Daten. Nicht zuletzt die Betriebssystemsparte (Windows, Android uvm.) bedarf der Lokalisierung. Somit kann die Aufbereitung der Sprachdaten in Form des Skype Translators womöglich als Entgegenkommen der Konzerne auf die Endnutzer·innen angesehen werden. Zumal die omnipräsente Verfügbarkeit von übersetzten Inhalten bislang noch immer ein Wunschtraum für Unternehmen und Privatpersonen darstellt und das Angebot eines Echtzeit-Übersetzungssystems somit gute Werbung darstellt.

Hieraus ergibt sich dann jedoch eine weitere Frage: Welchen Nutzen bringt das so breit gefächerte Sprachenportfolio? Das bloße Vorhandensein der Sprachdaten bei den Konzernen vereinfacht sicherlich die Entwicklung solcher Technologie auch für „kleinere“ Sprachen. Das wiederum stärkt einerseits das Bewusstsein für diese Sprachen. Zugleich sind häufig dennoch zu wenige Daten vorhanden, als dass von Beginn an qualitativ zumindest akzeptable Resultate beobachtet werden können.

In diesem Zusammenhang macht ein Beitrag im Handelsblatt aus dem Jahre 2020 deutlich, welchen Stellenwert kleinere Sprachen für die Modellentwickler·innen bei den großen Technologiekonzernen haben. So zitiert Holzki (2020) den Chefentwickler von Google, „dass es viel einfacher ist, eine elfte Sprache zu lernen als die ersten zehn“. Da Google gleichzeitig das Ziel ausweist, „Modelle trainieren [zu] können, die das Erlernen so generalisieren, dass sie nur kleine Datenmengen brauchen“ (ebd.), ergibt sich hieraus das Bild eines Konzerns, der vermeintlich kleinere Sprachen als gute Sparringspartner für die Entwicklung von NLP-Systemen ansieht, die nur mit wenigen Ressourcen viel leisten und gute Ergebnisse liefern. Die Möglichkeit, Inhalte für kleinere Sprachen im digitalen Raum verfügbar zu machen, scheint demnach nicht zwangsläufig ein erklärtes Ziel. Vielmehr geht es darum, die Modellentwicklung bei der Datenverarbeitung

7 Diskussion

von natürlicher Sprache soweit zu verallgemeinern, dass mit möglichst wenig Aufwand möglichst viele Sprachen und Sprachdaten verarbeitet werden können.

Mit Blick auf die Konzeption sowohl der Online-Umfrage als auch des inhaltlich gleichen Eingangsfragebogens der beiden Eye-Tracking-Studien ist als kritikwürdig festzuhalten, dass einzelne, im Rückblick als durchaus aufschlussreich erscheinende Fragen nicht aufgegriffen wurden. So hätte es sich im Rückblick als hilfreich erwiesen, das genauere Nutzungsverhalten von Chatanwendungen, die eine Alternative zu Skype darstellen, zu erheben. Hiermit wäre es möglich gewesen, eine genauere Aussage über die Gründe, weshalb diese Alternativen Skype gegenüber bevorzugt werden, und generell über die Usability (s. Abschnitt 3.7, S. 68) zu treffen. Auch eine detaillierter Erhebung der Nutzung der einzelnen Funktionen von Skype hätte in diesem Zusammenhang sicherlich dazu beigetragen, das Nutzungsverhalten der Anwendung genauer zu umschreiben.

Nicht zuletzt muss sich jedoch auch stets vergegenwärtigt werden, dass innerhalb des Angebots an Kommunikationsanwendungen Skype die einzige ist, die bislang eine kostenfreie maschinelle Übersetzung in Echtzeit für die Chatkommunikation anbietet.

7.2 Visuelle Inspektion

Die visuelle Inspektion der Eye-Tracking-Daten wurde exemplarisch an den Bildschirmfotos jeweils einer Versuchsperson für das katalanisch-deutsche und einer Person für das monolinguale Setting durchgeführt. Hierzu wurde eine Visualisierung der Indikatoren *Fixationen*, *Sakkaden* und *Blinzler* verwendet. Einen vergleichbaren Ansatz verfolgt Burridge (2014: 224 f.) bei der strukturellen Untersuchung von Social-Media-Webseiten. Auch hier liefert die visuelle Inspektion des *gaze path* sowie der Fixationen wertvolle Hinweise zur grundlegenden Wahrnehmung des Designs sowie zur Orientierung der Nutzer:innen der jeweiligen Seite und kann darüber hinaus zur Bewertung der *Usability* verwendet werden. Everdell (2014: 167 f.) hingegen orientiert sich bei der Untersuchung der Informationsverteilung bei Webinhalten an Heatmaps und nutzt diese zur Identifizierung relevanter Bereiche auf dem Bildschirm.

Generell lässt sich deshalb mit Blick auf die in dieser Arbeit durchgeführte Untersuchung feststellen, dass die Konzentration der erfassten Indikatoren zwischen den beiden Versuchsaufbauten schwankt. Die Wolke, die die Fixationen, die Sakkaden und Blinzler jeweils im Bereich der Eingabemaske bilden, wirkt im Falle des katalanisch-deutschen Versuchsaufbaus dichter als im monolingualen Setting. Das Auftreten der drei Indikatoren im zweisprachigen Versuch ist be-

sonders im Bereich der eingehenden bzw. maschinell übersetzten Nachrichten auszumachen.

Diese Beobachtungen können mit höchster Vorsicht und nur vorläufig mit Blick auf die weitere Diskussion auf folgenden Erklärungsansätzen basieren: Zum einen verlagern die Proband·innen im einsprachigen Versuchssetting ihre Aufmerksamkeit auf die eingehenden Nachrichten, da beide Nachrichtenarten in der Muttersprache abgefasst sind. Eine erneute Betrachtung der eigenen Beiträge zwecks Überprüfung des Inhalts ist also nicht bzw. nur in Ausnahmefällen nötig. Eine sprachenbedingte Konzentration der Aufmerksamkeit findet nicht hier, sondern nur im zweisprachigen Versuch statt, wo zudem im linken Bereich des Chats gleich drei Nachrichtenarten ($2 \times$ MÜ, $1 \times$ katalanisches Original) eingehen. Das entspricht den Beobachtungen von Everdell (2014: 169), wonach Inhalte die größte Aufmerksamkeit der/des Leser·in an der Stelle gewinnen, auf die sich die Person am meisten fokussiert. Ferner deuten die Durchschnittswerte der AOI-Größe im monolingualen Versuch darauf hin, dass die Chatpartner·innen womöglich längere Nachrichten am Stück geschrieben haben. Auch das sorgt in der Folge für eine Konzentration auf den in Frage stehenden Bereich.

Zum anderen kann die Konzentration in beiden Settings auf den Bereich oberhalb der Eingabemaske dahingehend gedeutet werden, dass die Versuchspersonen selten über das untere Drittel nach oben hinausspringen, um ältere Nachrichten noch einmal zu lesen. Dieses Verhalten kommt zwar gelegentlich vor, das Hauptaugenmerk liegt jedoch eindeutig auf den neusten Beiträgen. Grundsätzlich jedoch verbringen die Studienteilnehmer·innen einen substantiellen Teil mit der Betrachtung der linksbündigen, eingehenden Nachrichten, so etwa der MÜ-Ausgabe. Burridge (2014: 229) hebt hervor, dass ein zweispaltiger Aufbau in eben diesem Fall eine strukturierende Wirkung haben soll. Ursprünglich in Bezug auf soziale Medien gedacht, gilt die Beobachtung, wonach ein zweispaltiges Design mit der Präsentation von Inhalten in kleinen Blöcken (analog zu den Chatbeiträgen) als Hilfsmittel zur Informationsverarbeitung beiträgt (Burridge 2014: 229 f.). Verstärkt wird diese Konzentration sicherlich auch durch die Einblendung des Hinweises, dass die Person gegenüber gerade schreibt. Diese Nachricht wird unmittelbar zwischen Eingabemaske und neu eingehenden Nachrichten, ebenfalls linksbündig, angezeigt.

7.3 Fixatorische Augenbewegung

7.3.1 Fixationsanzahl

In absoluten Zahlen fallen in beiden Versuchssettings Parallelen auf: Die eigenen Beiträge erhalten jeweils deutlich weniger Fixationen, als auf die eingehenden Nachrichten in deutscher Sprache verfallen. Die durchschnittliche Fixationsanzahl im monolingualen Versuchsaufbau ist bei beiden Nachrichtenkategorien jedoch höher als im Setting Katalanisch-Deutsch. Eine mögliche Erklärung für diese Werte ist, dass die Versuchspersonen im Setting Katalanisch-Deutsch ihre Aufmerksamkeit auf mehrere Reize verteilen müssen. Dort sind vier verschiedene Nachrichtenarten zu erfassen, im einsprachigen Setting nur zwei. Daher lesen die Personen die vier Nachrichtenarten unterschiedlich, zumal nur zwei von diesen in ihrer Muttersprache abgefasst sind bzw. zwei von einem MÜ-System stammen. Andererseits besteht die Möglichkeit, dass die Studienteilnehmer-innen im katalanisch-deutschen Versuch mehrere kürzere Nachrichten versenden, da sie im Vergleich zum einsprachigen Setting – ob bewusst oder unbewusst – weniger komplexe Nachrichten versenden, um die Kommunikation aufrechtzuerhalten.

Beide Versuchssettings weisen unter Beachtung der Fixationsanzahl signifikante Merkmale auf. Die Ergebnisse von Schaeffer u. a. zum übersetzen und verstehenden Lesen deuten in diese Richtung. Das übersetzende Lesen erhält im Schnitt mehr Fixationen als das reine auf Verständnis ausgerichtete Lesen (Schaeffer u. a. 2017: 36). Eingehende bzw. Fremdbeiträge erhalten mehr Fixationen als eigene bzw. ausgehende Nachrichten. Wird die Thema-Rhema-Progression missachtet, beispielsweise durch die Betrachtung von AOI mit höherer Ordnungszahl (progressive erste Fixation), so erhalten die einzelnen Beiträge weniger Fixationen. Da dieses Phänomen statistisch sowohl bei beiden Datensätzen als auch bei den einzelnen Nachrichtenkategorien auftritt, kann vermutet werden, dass eine lineare Thema-Rhema-Progression innerhalb des Chats als wichtige Einflussgröße auf die Fixationsanzahl, und dadurch ausgedrückt die Lesetiefe, wirkt.

Auch wenn bei der statistischen Untersuchung mittels nicht-parametrischer Tests neun von zehn Vergleichspaare in Bezug auf die Fixationsanzahl pro AOI im Setting Katalanisch-Deutsch signifikant unterschiedlich sind, fällt besonders das eine unauffällige Vergleichspaar *CatO-GerO* auf. Auf Grundlage dieses Testergebnisses ist deshalb anzunehmen, dass sich die zentrale Tendenz der Fixationsanzahl zwischen den beiden Kategorien nicht signifikant unterscheidet. Wenn nun weiterhin angenommen wird, dass die Versuchspersonen die eigenen verfassten Beiträge bereits kennen und deshalb weniger Fixationen auf die Kategorie *GerO* entfallen, so bedeutet dies im Umkehrschluss für die Kategorie *CatO*,

dass das katalanische Original nur oberflächlich gelesen wird. Diese Möglichkeit wird durch Abgleich mit den Testergebnissen im einsprachigen Versuchsaufbau unterstrichen, in dem sich alle Kategoriepaare – also die eigenen und die Fremdbeiträge – in puncto Fixationsanzahl signifikant unterscheiden, was wiederum auf eine unterschiedliche Lesetiefe hindeutet.

Das Verhältnis der Fixationsanzahl nach Reiz in der Mutter- und Fremdsprache deckt sich mit den Erkenntnissen von Hvelplund (2017: 70), der in Bezug auf vier verschiedene Arten des Lesens feststellt, dass das Lesen von einem bestehenden Zieltext kognitivfordernder ist als die Erfassung eines Ausgangstextes. Während in der Untersuchung von Hvelplund die Gründe für dieses Verhältnis in der Vertrautheit des durch die/den Übersetzer.in produzierten Zieltext bzw. der Fremdheit des Ausgangstextes gesucht werden, gelten für die Studie in dieser Arbeit vertauschte Rollen: Der Ausgangstext ist der produzierte, vertraute Text, wohingegen die Ausgabe der maschinellen Übersetzung sowie die Beiträge des Gegenübers die unbekannten Texte sind.

In beiden Versuchen ist zudem eine positive Korrelation von Größe der AOI und Anzahl an Fixationen festzustellen, die in die jeweiligen Bereiche fallen. Die naheliegende Erklärung hierfür ist, dass mit zunehmender Länge des zu lesenden Textes bzw. der damit einhergehenden Vergrößerung der Textbox mehr Fixationen getätigter werden müssen, um die gesamte Nachricht zu erfassen. Das gilt in beiden Versuchsanordnungen ungeachtet der beteiligten Sprachen.

7.3.2 Dauer der ersten Fixation

Die Dauer der ersten Fixation wird als Indikator für den anfänglichen Verarbeitungsprozess eines Wortes angesehen wird, während dem auf das mentale Lexikon zugegriffen wird. Es fällt auf, dass eingehende Nachrichten in beiden Versuchssettings durchschnittlich länger fixiert werden als eigene Beiträge. Die Ausnahme stellt dabei der Durchschnittswert der MÜ-Ausgabe ins Deutsche dar, der geringer ist als der des deutschen Originals (*GerO*: 296,46, *GerMT*: 278,75) und somit nahe bei dem Wert des eigenen Beitrags im Setting Deutsch-Deutsch (*A*: 273,56) liegt. Die im Schnitt längste Dauer der ersten Fixation wird mit 304 ms bei den katalanischsprachigen Originalnachrichten festgestellt. Generell sind alle Durchschnittswerte vergleichbar mit denen von Holmqvist (2011: 382).

In beiden Versuchsaufbauten besteht eine hohe Standardabweichung der Werte. Nach Dormann (2013: 7) deutet die hohe Standardabweichung auf eine hohe Streuung der Werte um den Median hin, da generell davon ausgegangen werden kann, dass unter Annahme einer Normalverteilung des Datensatzes „etwa 68%

7 Diskussion

der Datenwerte ± 1 Standardabweichung und etwa $95\% \pm 2$ Standardabweichungen um den Mittelwert“ (Dormann 2013: 7) herumliegen. Da die Dauer der ersten Fixation allerdings mehreren nur begrenzt kontrollierbaren, subjektiven Faktoren wie Müdigkeit, Konzentration, Lichtverhältnisse oder auch Koffeinkonsum unterliegt, ist diese Abweichung nicht weiter verwunderlich (Rayner 1998: 376, Holmqvist 2011: 378 f.).

Die inferenzstatistische Betrachtung der einzelnen AOI-Kategorien deutet weder im Setting Deutsch-Deutsch noch im Setting Katalanisch-Deutsch auf signifikante Unterschiede bei der Dauer der ersten Fixation hin. Ausgehend von den Erkenntnissen von Hvelplund (2017) könnte die durchschnittliche Dauer der ersten Fixation pro AOI-Kategorie mit der Vertrautheit und der Wahrnehmung durch die Proband·innen erklärt werden. Ergeben sich keine statistisch signifikanten Unterschiede bei der Dauer der ersten Fixation der einzelnen AOI-Kategorien, ist anzunehmen, dass weder bei der Betrachtung der MÜ-Ausgaben ins Katalanische oder Deutsche noch bei dem katalanischen Original das mentale Lexikon, sprich: der kognitive Verarbeitungsprozess, in vergleichbarem Maße angeregt wird. Doherty u. a. (2010: 8 f.) berichten in ihrer Untersucht zur MÜ-Evaluierung mittels Eye-Tracking von einem ähnlichen Effekt: Vermehrt leichter zu verarbeitende, von menschlichen Revisoren als gut eingestufte Sätze der MÜ-Ausgabe werden von den Proband·innen teilweise länger fixiert als Testsätze, die als schlecht markiert sind. Außerdem wurde in der referenzierten Studie ein Gewöhnungsprozess beobachtet: Mit Ausschluss der ersten fünf Sätze verschiebt sich die durchschnittliche Fixationsdauer. Gute Test-Sätze werden unter dieser Bedingung durchschnittlich kürzer fixiert als schlechte, auch wenn der Effekt minimal ist. Da im Gegensatz zu diesen ähnlichen Lesestudien die AOI auf Nachrichtenebene und nicht auf Wortebene festgelegt wurden, stellt sich hier die Frage nach der Interpretierbarkeit der Werte.

Die Dauer der ersten Fixation bezieht sich auf das gesamte AOI, das die ganze Nachricht umschließt, und nicht auf ein einzelnes Wort. Die erste Fixation erfolgt am linken Zeilenanfang jedes Beitrages gemäß der konventionellen Leserichtung des Katalanischen und Deutschen von links nach rechts. Es liegt so die Vermutung nahe, dass die Dauer der ersten Fixation mit einer anfänglichen Verarbeitung des ersten Wortes im fokalen Sichtbereich der Teilnehmer·innen in Verbindung steht und nicht mit dem gesamten Textblock.

Auch die statistische Betrachtung der progressiven ersten Fixation in Bezug auf die Dauer der ersten Fixation ergibt signifikante Unterschiede bei zwei untersuchten Datensätzen: der Gesamtdatensatz im Setting Katalanisch-Deutsch sowie die MÜ ins Deutsche. Hierzu ergeben sich zwei Erklärungen: Zum einen ist

es möglich, dass entweder die Versuchsperson oder ihr Gegenüber mehrere aufeinanderfolgende, zusammenhängende Nachrichten gesendet hat. Da diese vom Skype Translator getrennt voneinander maschinell übersetzt werden, erhält jede gesendete Nachricht automatisch ihre jeweilige maschinelle Übersetzung. Das wiederum bietet der Versuchsperson die Möglichkeit, mit der jeweils jüngsten Nachricht zu beginnen, die jedoch zugleich die höchste aktuelle Ordnungszahl besitzt. Wenn die Person sich nun vom unteren Bildschirmrand nach oben, zu den älteren eingegangenen Nachrichten hocharbeitet, kann dies die Dauer der ersten Fixation beeinträchtigen. Zum anderen folgen in der Annotationsreihenfolge aller AOI auf die MÜ ins Deutsche wieder deutsche Originalbeiträge.

Die Studie von Inhoff & Rayner (1986: 433) deutet zudem darauf hin, dass die Fenstergröße mit der Fixationsdauer korreliert. Allerdings ergibt sich ein entsprechendes Ergebnis nur im katalanisch-deutschen, nicht jedoch im deutsch-deutschen Versuchsaufbau. Eine mögliche Erklärung ist, dass die Versuchspersonen durch die vier unterschiedlichen Nachrichtarten und deren räumlicher Anordnung auf dem Bildschirm stärker kognitiv gefordert werden, als die Personen, die an der einsprachigen Studie mit nur zwei Nachrichtarten teilgenommen haben. Wie auch schon bei der Fixationsanzahl angemerkt, ist weiterhin zu untersuchen, ob auch ein sprachenabhängiges Kommunikationsverhalten als Faktor für den Umfang der Nachrichten in Erwägung zu ziehen ist.

7.3.3 Regressionen

Die Analyse der Regressionen erfolgte unterteilt in zwei Arten: Regressionen in ein und solchen aus einem AOI. In absoluten Zahlen gehen auf den fremden Beiträgen in beiden Settings jeweils mehr Regressionen ein als auf den eigenen Nachrichten der Versuchspersonen. Die Mittelwerte der Kategorien *GerO*, *CatO* und *B* liegen allesamt über 1. Somit wird auf jeden Beitrag dieser Kategorien jeweils im Schnitt mindestens einmal zurückgekehrt. Die beiden Nachrichtarten der MÜ ins Katalanische sowie ins Deutsche weisen einen Schnitt von 0,65 bzw. 0,67, die eigenen Nachrichten *A* im einsprachigen Setting von 0,76 auf. In diese AOI wird im Schnitt demnach weniger als einmal zurückgekehrt.

Umgekehrt gehen von beiden MÜ-Ausgaben mehr Regressionen aus als von den von einem Menschen verfassten Nachrichten. Ähnlich gestaltet sich das Verhältnis für das deutsch-deutsche Setting, auch hier werden mehr Regressionen aus den Fremdbeiträgen heraus getätigt. Auffällig ist dabei, dass die Mittelwerte für die Kategorien *GerO* und *CatO* mit jeweils 0,15 und 0,18 deutlich unter dem Schnitt von *A* und *B* mit 0,75 bzw. 0,89 liegt. Aus allen aufgezählten Beitragskategorien wird folglich im Schnitt weniger als eine Regression ausgeführt. Die

7 Diskussion

Eingabemaske weist in beiden Versuchsaufbauten und in Bezug auf beide Arten von Regressionen vergleichsweise hohe Werte auf. Dies kann als Indiz für das Leseverhalten der Proband·innen genommen werden, die zur Eingabemaske zurückkehren, um eine neue eigene Nachricht zu verfassen, oder dort verweilen, bis eine neue Nachricht auf dem Bildschirm eingeht.

Die Regressionen stellen zudem den einzigen Indikator dar, der inferenzstatistisch mit komplexen Modellen untersucht wurde. Da die Betrachtung der abhängigen Variablen in diesem Fall nur zwei Zustände möglich waren (Regression: ja/nein), wurde mit logistischen Regressionsmodellen gearbeitet, die nicht notwendigerweise eine Normalverteilung der Daten voraussetzen. Im Versuchsaufbau Katalanisch-Deutsch war eine Modellierung sowohl für die eingehenden als auch die ausgehenden Regressionen möglich. Aufgrund nicht hinreichender Modellwerte sind die Ergebnisse jedoch nur begrenzt interpretierbar. Deshalb wurden die Ergebnisse der Modellierung mit nicht-parametrischen Tests erneut überprüft. Im deutsch-deutschen Setting war eine Modellierung nur für eingehende Regressionen möglich, scheiterte allerdings komplett für die abhängige Variable der ausgehenden Regressionen. Die verwendeten vier unabhängigen Variablen wurden für beide Regressionsarten jeweils mit nicht-parametrischen Tests auf Plausibilität überprüft.

Es ist festzustellen, dass die eingehenden Regressionen in beiden Studienteilen von der AOI-Kategorie abhängen. Genauer noch deuten die Ergebnisse der Regressionsmodelle für die eingehenden Regressionen darauf hin, dass besonders die Beiträge des Gegenübers eingehende Regressionen provozieren. Im Setting Katalanisch-Deutsch kann dies als Indiz genommen werden, dass die Proband·innen die MÜ-Ausgabe ins Deutsche mit der katalanischen Originalnachricht vergleichen und sich dabei möglicherweise an markanten Wörtern oder Zahlen orientieren. Im einsprachigen Studienteil ist es denkbar, dass die Versuchspersonen noch einmal auf die Nachricht des Gegenübers zurückkehren, um zu überprüfen, ob die eigene Nachricht auch an alle wichtigen Informationen anknüpft.

Bezieht man nun die chronologische Reihenfolge der AOI mit ein, so ergibt sich ein vergleichendes Leseverhalten: Die Wahrscheinlichkeitswerte des Modells für die eingehenden Regressionen deuten primär auf Augenbewegungen hin, die zu einem zuvor bereits betrachteten, mit einer niedrigeren Ordnungszahl versehenen, AOI zurückspringen. Wenn also eine erhöhte Wahrscheinlichkeit besteht, dass Regressionen in das deutsche oder katalanische Original getätigter werden, so erfolgen diese Rücksprünge in der Regel von dem jeweilig darauffolgenden AOI mit nächsthöherer Ordnungszahl, also den deutschsprachigen *GerMT*

im Falle des katalanischen Originals bzw. der katalanischen MÜ (*CatMT*) im Falle des deutschen Originals. Der auffallend hohe Wahrscheinlichkeitswert der Regressionen in das AOI der Eingabemaske ist demnach schlichtweg der zwangsläufigen Nutzung zum Verfassen von Nachrichten geschuldet.

Während die Regressionen in das katalanischen Originals mutmaßlich von der MÜ ins Deutsche zurück einfallen, ist es verwunderlich, wieso vom katalanischen Original Regressionen in die vorausgegangene und zum deutschen Original gehörende MÜ ins Katalanische fallen (sollten). Ein erster Erklärungsansatz könnte in der kontinuierlichen Nummerierung der AOI liegen und somit lediglich technisch bedingt sein.

Mit Blick auf die Wahrscheinlichkeitswerte des Modells für die ausgehenden Regressionen deuten die Ergebnisse im Falle der MÜ ins Deutsche darauf hin, dass die Proband·innen zurückspringen zu vorausgehenden Nachrichten (der Kategorie *CatO*), um diese womöglich erneut zu betrachten. Bei der gesunkenen Wahrscheinlichkeit im Falle des deutschen und katalanischen Originals liegt der Erklärungsansatz nahe, dass die Originalbeiträge schlichtweg keine notwendige Vergleichsmöglichkeit zum Rücksprung bieten. Dem deutschen Original geht innerhalb der Sequenz der AOI-Annotation die MÜ-Ausgabe ins Deutsche, dem katalanischen Original die MÜ ins Katalanische voraus. Außerdem wurde das deutsche Original von den Testpersonen selbst verfasst und beinhaltet somit keinerlei neue Information.

Für Lesestudien auf Wortebene heben Inhoff u. a. die Bedeutung von Regressionen für das Textverständnis hervor und gehen dabei auf verschiedene Erklärungsansätze ein, weshalb es zu größeren Regressionen kommt. Wenn das gerade betrachtete Element nicht im semantischen oder grammatischen Einklang mit dem zuvor erfassten linguistischen Kontext steht, so wird mit großer Wahrscheinlichkeit eine Regression ausgeführt (Inhoff u. a. 2019: 3). Die Regression wird folglich genutzt, um eine linguistische Fehlinterpretation zu korrigieren. Weiterhin ist es möglich, dass Regressionen nicht zwingend zielgerichtet stattfinden, sondern lediglich unbewusst dazu verwendet werden, um mehr Verarbeitungszeit für die aufgenommenen Reize zu gewinnen (Inhoff u. a. 2019: 3 f.). Als dritte Hypothese verweisen Inhoff u. a. auf die Möglichkeit, dass Regressionen dem besseren Zugriff auf das Arbeitsgedächtnis dienen (Inhoff u. a. 2019: 4). Überträgt man diese Erklärungsansätze von der Wortebene auf ganze Chatbeiträge, so lassen sich ebenfalls drei Annahmen treffen.

Einerseits können die Studienergebnisse dahingehend interpretiert werden, dass die Proband·innen besonders bei der MÜ häufiger Regressionen ausführen, weil die MÜ fehlerbehaftet ist und sie im Sinne der ersten Hypothese von Inhoff u. a. eben diese linguistischen Fehlinterpretationen ausgleichen. In diesem Falle

7 Diskussion

müsste untersucht werden, ob die Regressionen unmittelbar an fehlerbehafteten Stellen der Chatbeiträge stattfinden.

Andererseits könnten die Regressionen aus der und in die MÜ-Ausgabe jedoch auch dafür sprechen, dass die Versuchspersonen die Beiträge zunächst komplett gelesen haben und dann, gemäß der zweiten und dritten Annahme, durch die Regressionen die Zeit zur kognitiven Verarbeitung ausweiten. In jedem Fall kann angenommen werden, dass die Regressionen an bestimmten Punkten innerhalb des Chatbeitrags geschehen. Entweder werden die Regressionen genutzt, um die fehlerhafte MÜ auszugleichen, somit wäre das Leseverhalten sprunghaft und punktuell, oder die regressiven Augenbewegungen dienen der tiefergehenden Verarbeitung des Gelesenen und spiegeln somit ein eher vergleichendes, umfassendes Leseverhalten wider. Es zeichnet sich in jedem Fall ab, dass die MÜ-Ausgabe anders referenziert wird als die Originalbeiträge.

7.3.4 Dauer des ersten Durchlaufs

Die Dauer des ersten Durchlaufs unterscheidet sich in beiden Versuchsaufbauten jeweils nach AOI-Kategorie. Im monolingualen Setting sind es die eingehenden Nachrichten des Gegenübers, die in absoluten Zahlen eine längere Dauer des ersten Durchlaufs aufweisen. Im katalanisch-deutschen Versuchsaufbau fallen in absoluten Zahlen vor allem die beiden Arten der MÜ auf. Die durchschnittliche Dauer des ersten Durchlaufs durch die original deutschsprachigen Beiträge liegt in beiden Versuchen bei etwa 600 ms. Für den ersten Durchlauf durch die MÜ ins Deutsche benötigen die Proband·innen hingegen im Schnitt 816 ms. Noch auffälliger ist dann der Mittelwert bei den eingehenden, deutschsprachigen Nachrichten im monolingualen Versuch, der 1.266 ms beträgt. Die hohe Standardabweichung in beiden Versuchsaufbauten zeugt davon, dass eine hohe intersubjektive Varianz besteht. So ist es möglich, dass die zu verarbeitenden Textboxen im monolingualen Setting generell größer sind als die in der vom Skype Translator vermittelten Kommunikationssituation.

Diese Vermutung drängt sich auch mit Blick auf die Durchschnittswerte der AOI-Größe in Pixel auf (s. Tabelle 6.6, S. 129 und 6.40, S. 176). Während alle AOI-Kategorien in beiden Settings einen Mittelwert um 28.000 px aufweisen, stellen die eingehenden Nachrichten des Gegenübers im einsprachigen Setting mit durchschnittlich 40.000 px einen auffällig hohen Wert dar. Somit kann vermutet werden, dass die Gegenüber im Setting Deutsch-Deutsch vergleichsweise umfangreiche Nachrichten am Stück verfasst haben, während die Beiträge im katalanisch-deutschen Versuch hingegen kürzer gefasst waren.

Unter Anwendung statistischer Tests sind es alle Vergleichspaare mit Beteiligung der MÜ ins Deutsche, die signifikante Unterschiede in der zentralen Tendenz der Durchlaufdauer aufweisen. Wenn davon auszugehen ist, dass neue Informationen in Form von eingehenden Nachrichten die längere Durchlaufdauer hervorrufen, lässt sich dies auch auf das Setting Katalanisch-Deutsch, und konkret auf die MÜ ins Deutsche, anwenden. Es handelt sich in jedem Fall um neue Informationen in der Muttersprache der Versuchsperson. Weiterhin kann davon ausgegangen werden, dass die MÜ bislang in irgendeiner Form fehlerbehaftet ist. Die Versuchspersonen sind deshalb veranlasst, langsamer zu lesen, um die Schwachstellen auszugleichen. Holmqvist (2011: 388) beschreibt dies mit Schwierigkeiten von Versuchspersonen, allgemeine oder auch lexikalische Informationen aus den präsentierten Reizen zu extrahieren.

So erklärt sich auch die signifikant unterschiedliche Dauer der Betrachtung der MÜ ins Deutsche im Vergleich mit allen anderen AOI-Kategorien. Die Testergebnisse der übrigen Vergleichspaare sind dahingehend zu interpretieren, dass keine signifikanten Unterschiede in der zentralen Tendenz der Dauer des ersten Durchlaufs bestehen. Die Betrachtungsdauer der katalanischen MÜ und Originalbeiträge unterscheidet sich demnach nicht wesentlich von der der deutschen Originalbeiträge. Ausgehend vom deutschen Original, das den Versuchspersonen gegenüber als inhaltlich bereits bekannt vorausgesetzt werden kann, ergibt sich für die MÜ ins Katalanische und das katalanische Original die Schlussfolgerung, dass diese beiden Beitragsarten mit vergleichsweise geringerer Tiefe gelesen und verarbeitet werden. Die nicht auffälligen Testergebnisse in diesem Fall deuten im Umkehrschluss darauf hin, dass die Versuchspersonen nicht gänzlich über diese Nachrichten hinweggehen, da die Dauer ansonsten ebenfalls signifikant auffällig wäre.

Bemerkenswert ist zudem das Ergebnis der Signifikanztests, die die Dauer des ersten Durchlaufs unter Beachtung der progressiven ersten Fixation untersuchen. Im katalanisch-deutschen Setting ergaben sich weder auf den gesamten Datensatz bezogen noch unterteilt nach den einzelnen AOI-Kategorien signifikante Unterschiede. Es scheint für die Dauer des ersten Durchlaufs nicht von Belang, ob vorab ein AOI mit höherer Ordnungszahl betreten wurde und somit bereits eine Vorwegnahme an noch aktuelleren Informationen im Chat stattgefunden hat. In der einsprachigen Variante sind die Werte jedoch sowohl in Bezug auf den Gesamtdatensatz als auch auf die ausgehenden, eigenen Nachrichten signifikant.

Die Größe der AOI korreliert in beiden Studienvarianten signifikant mit der Dauer des ersten Durchlaufs. Dabei kann generell davon ausgegangen werden, dass eine größere Textbox mehr Informationen beinhaltet und somit auch eine

7 Diskussion

längere Lesezeit benötigt. Das kann zunächst ungeachtet der beteiligten Sprachen angenommen werden.

Eine Diskussion der Werte der Eingabemaske erscheint nicht sinnvoll. Da die Eingabemaske als statisches AOI annotiert wurde, ergibt sich der Mittelwert aus der Gesamtsumme der Dauer des ersten Durchlaufs, geteilt durch die Anzahl an Versuchsteilnehmer-innen. Da die Gesamtsumme sich jedoch tatsächlich nur aus den Zeiten zusammensetzt, die zwischen dem erstmaligen Fixieren dieses AOI und dem erstmaligen Verlassen stehen, ist die Aussagekraft begrenzt .

7.3.5 Gesamtverweildauer

Die durchschnittliche Gesamtverweildauer für die eingehenden Nachrichten (*B*) im monolingualen Setting hebt sich mit etwa 8.900 ms deutlich von den Werten der anderen AOI-Kategorien ab. Während die eigenen Beiträge (*A*) im einsprachigen Versuch durchschnittlich etwa 3.000 ms fixiert werden, sind es im katalanisch-deutschen Versuchsaufbau 2.200 ms (*GerO*). Die anderen Beitragsarten werden jeweils etwa 2.100 ms (*CatO*), 3.600 ms (*CatMT*) und 3.900 ms (*GerMT*) fixiert. Auffällig ist dabei, dass beide MÜ-Ausgaben jeweils etwa 1.000 ms länger als die Originalbeiträge betrachtet werden.

Die unterschiedlichen Durchschnittswerte der Beitragsarten im einsprachigen Versuch sind insofern auffällig, als nur sieben Versuchspersonen beteiligt waren. Eine Erklärung dieser Werte ist daher, dass die Gegenüber in der Studienvariante deutlich längere Nachrichten verfasst haben, die die Proband·innen lesen mussten. Die Werte der Eingabemaske wirken im Falle der einsprachigen Studie nur dann plausibel, wenn angenommen wird, dass der Bereich der Eingabemaske zugleich auch immer als Ruheort für die Augen dient, während die Chat-Partner·innen ihre Nachrichten verfassen. Die Werte der Eingabemaske im Setting unter Beteiligung des Skype Translators wurden bereits im Zuge der Analyse problematisiert. Hier scheint ein Messfehler oder eine ungenaue Erfassung vorzuliegen.

Mit Blick auf die statistische Untersuchung fallen im Setting Katalanisch-Deutsch die Vergleichspaare auf, bei denen die MÜ-Ausgabe ins Deutsche beteiligt ist. Die Verweildauer der MÜ ins Deutsche unterscheidet sich signifikant von der der katalanischen MÜ-Ausgabe, des katalanischen Originals sowie des deutschen Originals. Gemessen an den absoluten Zahlen liegt somit die Vermutung nahe, dass die MÜ ins Deutsche eine längere Verarbeitungszeit bei den Proband·innen hervorruft. Als mögliche Erklärung sind hier sprachliche Fehler in der Ausgabe denkbar. Auch ist es, mit Blick auf die anderen Indikatoren denkbar, dass die Versuchspersonen häufiger zu der MÜ ins Deutsche zurückkehren,

um den Inhalt erneut zu lesen. Umgekehrt erweisen sich die gepaarten Vergleiche von Eingabe, den beiden katalanischen Nachrichtenarten und dem deutschen Original als statistisch nicht auffällig. Dies kann erneut mit dem vermuteten Messfehler auf dem AOI der Eingabemaske in Verbindung gebracht werden. Im Setting Deutsch-Deutsch fallen die Unterschiede in der zentralen Tendenz zwischen allen Vergleichspaaren statistisch signifikant aus. Die Verweildauer in den AOI der eigenen Nachrichten unterscheidet sich also von der in den AOI des Gegenübers sowie in der Eingabemaske. Allerdings sind, wie oben angemerkt, die Werte der statisch annotierten Eingabemaske nur bedingt plausibel.

Die Erkenntnisse von Hvelplund (2017) bieten hierzu einen Erklärungsansatz. Bei den eigenen Beiträgen handelt es sich um bekannte, vertraute Texte, die für die Proband·innen leichter zu erfassen sind. Bei den Beiträgen des Gegenübers handelt es sich hingegen um unbekannte Texte, die zunächst verarbeitet werden müssen. Die längere Verweildauer auf dem AOI der Eingabemaske ist demnach auf die Erkenntnisse von Hvelplund (2017) zurückzuführen, nachdem die Lesewahrnehmung und die Textproduktion gleichermaßen kognitiven Aufwand erfordern.

Durchgeführte Korrelationstest zeigen außerdem, dass in beiden Settings die Größe des AOI und die Gesamtverweildauer in einem positiven Verhältnis zueinander stehen. Je größer das AOI, sprich: die Textbox im Chat, ist, desto länger ist die Gesamtverweildauer. Eine Erklärung hierzu ist in der begrenzten Distanz von 7–9 Zeichen pro Sakkade zu finden, die die Augen nur springen können und der Lesefortschritt somit begrenzt ist.

7.3.6 Regressive Durchlaufdauer

Die regressive Durchlaufdauer drückt die Dauer aller Fixationen aus, die zwischen dem erstmaligen Betreten eines AOI und dem Verlassen hin zu einem AOI mit höherer Ordnungszahl gemessen wird. Es wird auch die Fixationsdauer hinzugaddiert, die nach dem erstmaligen Betreten des AOI in vorausgehenden AOI mit niedrigerer Ordnungszahl erfasst wird. Die Ergebnisse können deshalb zur Untersuchung des Leseverhaltens genutzt werden. Je kürzer die regressive Durchlaufdauer ist, desto früher haben die Versuchspersonen das jeweilige AOI in Richtung eines AOI mit höherer Ordnungszahl verlassen und somit möglicherweise nicht alle Informationen aus der Chatnachricht und dem angebotenen Kontext in Form der vorausgehenden Nachrichten aufgenommen. Im Setting Deutsch-Deutsch ist die regressive Durchlaufdauer der Chatbeiträge der Versuchspersonen sowohl im Schnitt als auch absolut länger als auf denen des Gegenübers. Im Gegensatz dazu ist die Betrachtungsdauer der eigenen Beiträge und die des katalanischen

7 Diskussion

Originals im Setting Katalanisch-Deutsch sowohl absolut als auch im Schnitt kürzer als die anderen beiden Arten von Textbeiträgen. Auf Grundlage dieser Werte kann angenommen werden, dass die Originalnachrichten (*GerO* und *CatO*) im Setting Katalanisch-Deutsch sowie die Nachrichten des Gegenübers (*B*) im einsprachigen Versuchsaufbau wesentlich früher in Richtung eines anderen AOI verlassen wurden als jeweils im Falle der MÜ-Ausgabe (*CatMT* und *GerMT*) und der eigenen Beiträge (*A*). Auch bei dieser Untersuchung sind die Werte der statisch annotierten Eingabemaske nur bedingt zu interpretieren, da die regressive Durchlaufdauer das erstmalige Betreten bis zum erstmaligen Verlassen des AOI erfasst.

In beiden Versuchsaufbauten waren die Daten nicht normal verteilt und wiesen zudem eine starke inter-individuelle Varianz auf. Das ist für naturalistisch orientierte Untersuchungen dieser Art nicht ungewöhnlich. Die unterschiedlichen Ergebnisse der Signifikanztests in Bezug auf die einzelnen AOI-Kategorien sind hingegen beachtenswert. Im einsprachigen Versuchsaufbau finden sich keine signifikanten Unterschiede bei der regressiven Durchlaufdauer zwischen den beiden Beitragsarten. Im Gegensatz dazu weisen bis auf zwei Ausnahmen (*CatMT-Eingabe* und *CatO-GerO*) alle möglichen Vergleichspaare im katalanisch-deutschen Versuch signifikante Unterschiede in der zentralen Tendenz auf. Auch hat die progressive erste Fixation in beiden Versuchen einen Einfluss auf die regressive Durchlaufdauer. Die AOI, die von einem anderen AOI mit höherer Ordnungszahl betreten wurden, unterscheiden sich in der regressiven Durchlaufdauer signifikant von jenen, die chronologisch betreten wurden.

Bei der Beteiligung der MÜ im Falle des Skype Translators unterscheidet sich die Lesedauer nicht nur zwischen Beiträgen in den beteiligten Sprachen, sondern auch zwischen Nachrichten, die zwar in der gleichen Sprache abgefasst, jedoch entweder von einem Menschen oder einer Maschine generiert wurden. Demnach können die niedrigen absoluten und durchschnittlichen Werte im katalanisch-deutschen Setting auf ein sprunghaftes Leseverhalten zurückgeführt werden. Die MÜ-Ausgabe der jeweiligen Textbeiträge stellt einen neuen Reiz auf dem Bildschirm dar.

Die Proband·innen, die so zunächst ihre eigenen Nachrichten (*GerO*) lesen, springen zu der MÜ ins Katalanische, um diese mit dem Original zu vergleichen. Das gleiche Muster kann auch den Werten des katalanischen Originals zugrunde gelegt werden. Die Versuchspersonen sind im Begriff, die katalanischsprachige Nachricht zu lesen, wenn dann plötzlich die maschinelle Übersetzung auf dem Bildschirm eingeblendet wird. Entweder erneut aus dem Antrieb heraus, nun einen Vergleich zur Verfügung stehen zu haben, der in diesem Fall sogar in der Muttersprache abgefasst ist, oder auf Grundlage der Eye-Mind-Hypothese von Just &

Carpenter (1980) springen sie in die MÜ ins Deutsche. Umgekehrt wiederum verbleiben die Personen offensichtlich lange auf der Ausgabe der MÜ. Hierzu kann angenommen werden, dass die Nachricht zuerst vollständig gelesen wird, bevor sich die Versuchspersonen entscheiden, ob sie zum nächsten AOI springen oder noch einmal zum entsprechenden Original zwecks Vergleich zurückkehren.

Die Annahme wird durch einen Vergleich der Werte der regressiven Durchlaufdauer mit denen der selektiven regressiven Durchlaufdauer sowie der Gesamtverweildauer bestärkt, deren jeweilige Werte deutlich niedriger sind. Die absoluten und durchschnittlichen Werte der beiden MÜ-Kategorien sind dort beispielsweise etwa nur halb so groß. Daraus lässt sich schließen, dass die zeitliche Differenz, um die die regressiven Durchlaufdauer größer ist, in vorausgehenden AOI mit niedrigerer Ordnungszahl verbracht wurde. Es zeigt sich, dass die Versuchspersonen zunächst die ins deutsche maschinell übersetzte Nachricht erstmalig fixieren, wodurch die Messung der regressiven Durchlaufdauer beginnt. Danach springen sie zu Vergleichszwecken zu dem katalanischsprachigen Original zurück, wobei die Messung der regressiven Durchlaufdauer weiterläuft, jedoch nicht die der Gesamtverweildauer oder der selektiven regressiven Durchlaufdauer.

7.3.7 Pupillengröße

Die Pupillengröße auf den Fremdbeiträgen mit Ausnahme des katalanischen Originals liegt in beiden Settings jeweils über dem Schnitt. Besonders auffällig ist dabei die durchschnittliche Pupillengröße bei Betrachtung der maschinell übersetzten Nachrichten ins Katalanische. Diese liegt 18 Einheiten über dem Schnitt. Doch auch die Pupillengröße bei Betrachtung der MÜ ins Deutsche liegt über dem Gesamtschnitt. Auf Grundlage der Erkenntnisse von Hess & Polt (1964), Beatty (1982) können die Ergebnisse dahingehend interpretiert werden, dass die Proband*innen bei der Verarbeitung der einzelnen Chatbeiträge unterschiedlich stark kognitiv ausgelastet sind. Krejtz u. a. (2018: 17 f.) kommen in ihrer Studie zu dem Ergebnis, dass die Pupillengröße von dem Schwierigkeitsgrad der gestellten Aufgabe abhängt. Dabei stellen sich sowohl Unterschiede zwischen den Versuchspersonen als auch innerhalb jeder Versuchsteilnahme heraus. Iqbal u. a. (2005: 317 f.) stellen hierzu in einer anderen Studie fest, dass sich die Pupille bei kognitiv anspruchsvollen Aufgaben weitet. Proband*innen, die ein Dokument bearbeiten, weisen eine höhere Auslastung während der Sprachverarbeitung als während des Leseverständnisses auf. Ein vergleichbares Bild zeichnen die Daten dieser Arbeit sowohl für das einsprachige Setting als auch für die Untersuchung mit Skype Translator ab.

7 Diskussion

Die statistische Untersuchung der Pupillengröße ergibt zunächst, dass die Werte signifikant mit der Größe des AOI korrelieren. Dies gilt in beiden Versuchssettings. In Hinblick auf die einzelnen Kategorien ist in beiden Versuchsanordnungen weiterhin die Eingabemaske hervorzuheben. Die Pupillengröße ist bei beiden Versuchen im Falle der Eingabemaske am größten. Ein Grund für die durch diese Werte ausgedrückte kognitive Auslastung könnte einerseits das Verfassen von neuen Nachrichten sein, während die Proband:innen die Eingabemaske fixieren. Andererseits drücken diese Werte möglicherweise das Warten auf neue einge hende Nachrichten des Gegenübers aus. Die Werte der eigenen Beiträge sind hingegen jeweils am geringsten. So scheint es, als stelle die eigene Nachricht in beiden Settings die geringsten kognitiven Anforderungen an die Versuchspersonen.

Weiterhin deuten die Testergebnisse in beiden Versuchsanordnungen darauf hin, dass inter-individuelle Variabilität besteht, da die durchgeführten Kruskal-Wallis-Tests jeweils signifikante Unterschiede zwischen den Proband:innen ergeben. Die Ergebnisse in Bezug auf die beteiligte Sprache werden auch durch die Studie von Hyönä u. a. (1995: 608 f.) unterstützt, in der sich signifikante Unterschiede bei der Pupillengröße zwischen zu lesenden Wörtern in der Fremd- und Muttersprache ergaben. Die Verarbeitung in der Fremdsprache führt zu einer stärkeren Pupillenweitung als in der Muttersprache. Die Betrachtung der einzelnen AOI ergibt im Setting Deutsch-Deutsch keine signifikanten Unterschiede zwischen den einzelnen Beitragsarten, sondern nur im Vergleich mit der Eingabemaske. Im katalanisch-deutschen Setting ergeben sich hingegen bei 8 von 10 möglichen Vergleichspaaren signifikante Unterschiede. Zwei Vergleichspaare unter Beteiligung der MÜ ins Deutsche (jeweils *CatMT-GerMT* und *CatO-GerMT*) sind unauffällig. Aus zweierlei Gründen ist das beachtenswert. Erstens wird die MÜ-Ausgabe ins Deutsche (*GerMT*) auf Grundlage des katalanischen Originals (*CatO*) erzeugt. Zweitens steht dieses Testergebnis somit in einem Widerspruch zu den Erkenntnissen von Hyönä u. a., wonach die Betrachtung von Wörtern in der Fremdsprache kognitiv fordernder sind als in der Muttersprache (Hyönä u. a. 1995: 610).

Offen bleibt noch die Einordnung der progressiven ersten Fixation. In beiden Versuchsanordnungen ergeben sich in Bezug auf den gesamten Datensatz und auf die jeweiligen AOI signifikante Unterschiede in der Pupillengröße zwischen AOI, die aus einem anderen AOI mit höherer Ordnungszahl heraus betreten wurden, und jenen, die in chronologischer Reihenfolge besucht wurden. Es liegt die Vermutung nahe, dass dies der Verarbeitung von neuen Informationen im Rahmen der Thema-Rhema-Progression geschuldet ist. Möglicherweise ist diese

Auffälligkeit mit dem Konzeptionsprozess von Nachrichten zu erklären. Während die Proband:innen die Eingabemaske betrachten, verfassen sie die neuen Nachrichten und sind demnach kognitiv anders ausgelastet als bei einer Leseaufgabe wie der Betrachtung der eingehenden und ausgehenden Nachrichten.

Nichtsdestotrotz bleibt die Aussagekraft der Unterschiede bei der Pupillengröße begrenzt. Wie auch Doherty u. a. (2010: 10) abwägen, sind einerseits zwar Veränderungen der Durchschnittswerte erkennbar, andererseits kann dies auch auf die kleine Fontgröße, einen Gewöhnungsprozess oder schlichtweg darauf zurückzuführen sein, dass der Verarbeitungsprozess der MÜ-Ausgabe nicht wesentlich kognitivfordernder ist. Auch Krejtz u. a. weisen auf genau dieses Problem bei der Untersuchung der Pupillengröße hin. Die alleinige Analyse der Pupillengröße besitzt nur eine limitierte Aussagekraft, da die Erfassung vielen Ungenauigkeiten und Einflussfaktoren unterworfen ist (Krejtz u. a. 2018: 19). Die in dem Artikel als Alternative vorgeschlagene Untersuchung der Microsakkaden wurde im Rahmen dieser Arbeit allerdings nicht durchgeführt.

7.4 Sakkadische Augenbewegungen

7.4.1 Sakkadenanzahl

Das sakkadische Blickverhalten der Proband:innen wurde unter Verwendung von drei Indikatoren untersucht: die Sakkadenamplitude und die -dauer sowie die absolute Sakkadenanzahl. Genauer noch wurde die Anzahl an Sakkaden unterteilt nach Richtung und AOI, sodass auch hierüber Rückschlüsse auf das Leseverhalten möglich sind. Bereits an dieser Stelle ist auf die unterschiedliche Anzahl an Sakkaden pro AOI zu verweisen, die sich innerhalb der Untersuchung der Sakkadendauer und -amplitude ergibt. Die divergierenden Zahlen sind mutmaßlich auf die vergleichsweise grobe Annotation mit dynamischen AOI zurückzuführen, die die genaue Erfassung aller Sakkaden an ihrem Ansatz- und Absatzpunkt nicht durchweg mit voller Präzision zulassen.

Generell ist in beiden Versuchsanordnungen festzustellen, dass etwa drei bis vier Mal so viele Sakkaden auf die eingehenden, bzw. maschinell in die deutsche Sprache übersetzten Nachrichten (*GerO*: 9 und *A*: 9,76) verfallen wie auf die eigenen Beiträge. Die Mittelwerte der deutschen Originalbeiträge im zweisprachigen Versuch sowie deren MÜ ins Katalanische sind hingegen kleiner als der Mittelwert der eigenen Nachrichten im einsprachigen Setting, auch wenn der Unterschied nur 0,76 beträgt. Die durchschnittliche Sakkadenanzahl innerhalb der MÜ-Beiträge ins Katalanische (*CatMT*: 8:36) ist allgemein am niedrigsten. Die durchschnittliche Sakkadenanzahl der katalanischen Originalbeiträge (*CatO*:

7 Diskussion

11,98) im zweisprachigen Versuchsaufbau sowie deren MÜ ins Deutsche (*GerMT*: 13,23) liegt über der der beiden Beitragsarten im einsprachigen Setting (A: 9,76 und B: 10,27). Gemessen an der Sakkadenanzahl bedarf die MÜ ins Deutsche offenbar eine intensivere Betrachtung – und damit einhergehende Verarbeitung – als von deutschen Muttersprachler-innen verfasste Nachrichten auf Deutsch.

Mit Blick auf die Sakkadenrichtung dominieren die nach rechts gerichteten Sprünge der Augen ungeachtet der AOI-Kategorie. Das entspricht der zu erwartenden Leserichtung in den beiden beteiligten Sprachen. Auch die Bewegungen nach links haben einen untereinander vergleichbar großen Anteil an der Gesamtzahl, der allerdings über dem in anderen Studien beobachteten Wert liegt. Buscher u. a. (2012: 6) beispielsweise gehen von einem üblichen Anteil an Regressionen bei Leseaufgaben von 10–15 % aus¹. Der Anteil an Sakkaden, die im zweisprachigen Setting innerhalb der katalanischen Originalbeiträge nach rechts getätigten werden, ist zugleich auffällig. Während der Anteil der nach rechts gerichteten Sakkaden in den AOI-Kategorien *CatMT*, *GerO* sowie *GerMT* zwischen 58 % und 70 % liegt, macht er bei der Eingabemaske und beim katalanischsprachigen Original nur 43 % bzw. 51 % aus. Möglicherweise erklärt sich dieser geringere Anteil damit, dass die Versuchspersonen in beiden Fällen eher kurSORisch springen als intensiv lesen. Diese Möglichkeit wird auch durch die Durchschnittswerte der Sakkadenamplitude und -dauer gestützt. Biedert, Dengel u. a. (2012: 124) unterscheiden in ihrer Studie ähnlich zwischen Sakkaden, die tiefgehendes und solchen, die überfliegendes Leseverhalten (*skimming*) ausdrücken. In der Studie wird die Unterscheidung allerdings auf Grundlage der über eine Sakkade hinweg zurückgelegten Zeichen getroffen, wie sie auch Rayner (1998: 379 f.) als Maß anwendet.

Wenn ein vergleichendes Leseverhalten der Versuchsteilnehmer-innen zu grunde gelegt wird, kann angenommen werden, dass die Personen innerhalb desselben katalanischsprachigen AOI zurückspringen, um markante Stellen zu identifizieren. Wenn das AOI über mehrere Zeilen geht oder mehrere Chatbeiträge derselben Person aufeinander folgen, erklärt sich so der höhere Anteil an nach oben gerichteten Sakkaden. Einschränkend stellen jedoch Biedert, Hees u. a. (2012: 129) in ihrer Studie fest, dass die Fenstergröße der präsentierten Leseaufgabe einen starken Einfluss auf die Genauigkeit der Sakkaden hat. So führt ein

¹Hierbei handelt es sich um sakkadische Regressionen. Sie werden auf Grundlage des Ansatz- und Absatzpunktes der jeweiligen Sakkade (bzw. des Winkels den diese beiden Punkte relativ zur horizontalen Ebene auf dem Bildschirm beschreiben) identifiziert und stehen damit immer in Verbindung mit der Leserichtung. Daher dürfen diese Regressionen auch nicht mit dem Indikator verwechselt werden, der im Methodikteil beschrieben und in der Analyse verwendet wird. Bei jenen Regressionen handelt es sich um ein proprietäres Maß der Software von SR Research Ltd. auf Grundlage der Ordnungszahl der fixierten AOI.

zu kleines Fenster dazu, dass die Zeilenabstände eines Beispieltexes zu nah aneinanderstehen, wodurch der Eye-Tracker nicht mehr präzise in der Lage sei, die Sakkaden den einzelnen Zeilen zuzuordnen. Umgekehrt führe ein zu großes Fenster dazu, dass die Sakkaden von zu viel *Noise*, also Ungenauigkeit, begleitet würden. Eine ähnliche Abwägung wurde auch mit Bezug auf die Nachrichtengröße bei Skype im Methodikteil dieser Arbeit problematisiert und ist bei der Bewertung der Befunde an dieser Stelle im Hinterkopf zu behalten.

Außerdem ist es möglich, dass die Versuchsteilnehmer-innen zwischen den eigenen und fremden Beiträgen wechseln. In diesem Fall deutet der Anteil an nach links gerichteten Sakkaden auf ein abgleichendes Leseverhalten zwischen den beiden katalanischen Nachrichtenarten und den eigenen Nachrichten hin.

Eine andere Erklärung ist der Redaktionsprozess einer Nachricht an sich, der vermehrte Korrekturen und Abwandlungen der Nachricht notwendig macht. Hierbei ist die Reihenfolge zu beachten, in der die Nachrichten auf dem Bildschirm eingeblendet werden. So ist es denkbar, dass der hohe Anteil nach oben gerichteter Sakkaden in den AOI des katalanischen Originals durch ein abgleichendes Leseverhalten der Versuchspersonen mit den jeweils vorausgehenden eigenen Nachrichten und deren MÜ ins Katalanische hervorgerufen wird. Das ist allerdings noch tiefgreifender zu überprüfen. Bislang mutet es seltsam an, dass vergleichsweise viele nach oben gerichtete Sakkaden von dem katalanischen Original ausgehen.

7.4.2 Sakkadenamplitude

Da keine Normalverteilung vorliegt, war eine Betrachtung der Sakkadenamplitude in beiden Settings unter inferenzstatistischen Gesichtspunkten nur mit nicht-parametrischen Tests möglich. Beide Datensätze weisen jeweils die gleiche durchschnittliche Amplitude von 1,84 vor. Auch die durchschnittliche Amplitude der jeweils eigenen Beiträge ist vergleichbar (*GerO*: 1,94 und *A*: 1,95). Obwohl sich die Amplitude der beiden Beitragskategorien im Setting Deutsch-Deutsch um 0,14 in absoluten Zahlen unterscheidet, ergeben sich keine statistischen Aufälligkeiten aus dem Vergleich der beiden AOI-Kategorien (*A-B*). Die zentrale Tendenz der Amplitude im katalanisch-deutschen Versuchsaufbau hingegen unterscheidet sich bei 8 von 10 Vergleichspaaren signifikant. Dieses Ergebnis kann dahingehend interpretiert werden, dass eine unterschiedliche Lesetiefe zwischen den einzelnen AOI-Kategorien vorliegt. Statistisch unauffällig sind nur die Paare *CatMT-Eingabe* und *CatO-GerO*.

Die höhere Amplitude kann als Indiz für eine geringere Lesetiefe genommen werden, da sie die zurückgelegte Strecke einer Sakkade zwischen An- und Absatz

abbildet. Je größer also der Amplitudenwert ist, desto länger sind die einzelnen Sakkaden, die innerhalb der AOI erfasst wurden. Eine größere Distanz zwischen Ansatzpunkt und Absatz stellt also die Spanne dar, über die das Auge springt und in der möglicherweise mehrere Zeichen liegen. Das gilt offenbar für alle Vergleichspaare mit Ausnahme der oben genannten zwei Paare *CatMT-Eingabe* und *CatO-GerO*. Die Produktionsaktivität wurde hingegen in dieser Studie nicht erfasst, stellt möglicherweise jedoch einen Einflussfaktor dar. In anderen Studien wurde der Faktor der Produktion in der Fremd- vs. in der Muttersprache in das Studiendesign integriert. Das ist hier nicht der Fall. Die Proband·innen müssen nicht selbst auf einer oder aus einer Fremdsprache produzieren.

7.4.3 Sakkadendauer

Wie auch die Amplitude drückt die Sakkadendauer die Lesetiefe aus. Höhere Werte bedeuten längere Sakkaden, was auf eine geringere Lesetiefe hindeutet. In Hinblick auf die Sakkadendauer wurden nur die Durchschnittswerte erfasst. Eine Darstellung der aufsummierten Werte scheint für diesen Indikator nicht sinnvoll. Die hohe Standardabweichung zeugt nebenher von einer hohen Varianz innerhalb der beiden Datensätze. Dies kann mehrere Gründe haben: Individuelle Varianz der Proband·innen, hohe Fehlerraten, bislang noch nicht beachtete Variablen und falsch-positive Beobachtungen. Auch die visuelle Untersuchung, exemplarisch anhand von Abbildung 6.12 (s. S. 125), deutet darauf hin, dass die meisten Sakkaden in das untere Drittel des Bildschirms – und dort in die linke Hälfte – fallen. Im Versuchssetting Katalanisch-Deutsch werden dort die eingehenden Nachrichten des Gegenübers sowie alle MÜ-Ausgaben angezeigt. Nur wenige Sakkaden gehen über diesen Bereich hinaus.

In beiden Versuchsaufbauten liegt die Sakkadendauer innerhalb der Eingabemaske über dem Durchschnitt. Die Werte innerhalb der Beiträge des Gegenübers im monolingualen Setting sowie innerhalb der MÜ ins Deutsche im Versuchsaufbau Katalanisch-Deutsch liegen hingegen unter dem Gesamtdurchschnitt. Die Durchschnittsdauer der Sakkaden innerhalb der eigenen Nachrichten ist im monolingualen Setting größer als der Durchschnittswert, im Versuchsaufbau Katalanisch-Deutsch liegt sie unter dem Schnitt. Im Falle der eigenen Beiträge im Setting Deutsch-Deutsch sowie der Eingabemaske in beiden Versuchen ist dies möglicherweise mit der Vertrautheit der Nachrichteninhalte zu erklären, die zudem auch noch in der Muttersprache der Versuchspersonen abgefasst sind. Die niedrigeren Werte innerhalb der eigenen Beiträge sowie der MÜ ins Deutsche im Setting Katalanisch-Deutsch deuten darauf hin, dass die Lesetiefe in diesen beiden Fällen hoch ist. Das wiederum kann in Verbindung mit der Leserichtung

und der Sakkadenamplitude als vergleichendes Leseverhalten der Proband·innen jeweils zwischen Original und maschineller Übersetzung gewertet werden. Die Personen wechseln zwischen *GerO* und *CatMT* bzw. zwischen *CatO* und *GerMT*, um die beiden Paare zu vergleichen und so markante Stellen zu identifizieren.

Zur statistischen Untersuchung der Sakkadendauer wurden ebenfalls nicht-parametrische Tests angewendet, da keine Normalverteilung bei beiden Datensätzen vorlag. In beiden Versuchsaufbauten unterscheidet sich die Sakkadendauer innerhalb der ausgehenden, eigenen Nachrichten und der eingehenden, fremden Beiträgen signifikant. Im Setting Katalanisch-Deutsch gilt dies für die deutschen Originalbeiträge, die bedeutsame Unterschiede im Vergleich zur MÜ-Ausgabe jeweils ins Deutsche und Katalanische aufweisen. Im einsprachigen Versuchsaufbau unterscheidet sich die Sakkadendauer im Vergleichspaar *A-B* ebenfalls signifikant. Im Gegensatz dazu ergeben sich für die MÜ-Ausgabe ins Deutsche signifikante Unterschiede zu den Werten des katalanischen Originals und der Eingabemaske. Auch im anderen Versuchsaufbau weist der Vergleich der Sakkadendauer zwischen Fremdbeiträgen und Eingabemaske auf signifikante Unterschiede hin. Mit Rückgriff auf die Feststellung von Holmqvist (2011: 322), wonach die Sakkadendauer mit dem Schwierigkeitsgrad der präsentierten Aufgabe steigt, kann angenommen werden, dass die Betrachtung der einzelnen Beitragsarten eine unterschiedlich schwierige Aufgabe für die Proband·innen darstellt. Das wiederum bietet insbesondere eine Erklärung für das statistisch unauffällige Paar *CatMT-GerMT*, wonach die Betrachtung der MÜ-Ausgabe ungeachtet der Sprache für die Personen gleichermaßen kognitivfordernd ist.

Das Vergleichspaar *A-Eingabe* im einsprachigen Setting bleibt ebenso statistisch unauffällig wie auch das Paar *GerO-Eingabe* im anderen Versuch. Dies lässt sich mit der Vertrautheit der Proband·innen mit dem erfassten Informationen erklären. Die Nachricht wird in der Eingabemaske verfasst und ist deckungsgleich mit der Anzeige, die kurz darauf als ausgehende Nachrichten der AOI-Kategorie *A* bzw. *GerO* erfasst wird. Die Schlussfolgerung liegt somit nahe, dass die unterschiedlichen Beitragsarten eine unterschiedliche Lesetiefe, ausgedrückt durch die Sakkadendauer und -amplitude, im monolingualen Chat hervorrufen. Unter Bezugnahme auf die absoluten Zahlen der Amplitude und Sakkadendauer kann weiterhin angenommen werden, dass die Lesetiefe der MÜ ins Deutsche (*GerMT*) bzw. der deutschsprachigen Fremdbeiträge (*B*) weitaus höher ist und eine stärkere kognitive Auslastung hervorruft als das Lesen der deutschen Originalbeiträge.

Wie in den Tabellen 6.32 (S. 160) und 6.34 (S. 162) erkenntlich wird, sind sowohl die durchschnittliche Sakkadenamplitude als auch die durchschnittliche Sakkadendauer vergleichbar mit früheren Studien (Rayner 1998, Gangl u. a. 2018, Nikolova u. a. 2018).

Ausgehend von der bestehenden Literatur (Beatty 1982) ist einschränkend festzustellen, dass die alleinige Betrachtung der sakkadischen Augenbewegung nur begrenzt umfassende Erkenntnisse im Rahmen dieser Studie liefert. Zwar ist eine generelle Untersuchung des sakkadischen Blickverhaltens dahingehend sinnvoll, als es einen Teil der wissenschaftlichen Kartierung des Skype Translators darstellt. Es bestehen offenkundige Unterschiede im sakkadischen Blickverhalten der Proband·innen zwischen den beiden Versuchssettings. Innerhalb der beiden Settings zeichnen die ein- und ausgehenden Nachrichten jeweils markante sakkadische Augenbewegungen ab. Jedoch wird ebenso sichtbar, dass die Betrachtung der Sakkaden allein nicht fruchtbar ist, sondern in jedem Fall die Notwendigkeit einer kombinierten Untersuchung mit den fixatorischen Augenbewegungen besteht. Die alleinige Betrachtung der Sakkaden ist schlichtweg in diesem Rahmen, besonders aufgrund der im Vergleich zu anderen Studien vergleichsweise kleinen Schriftgröße, zu ungenau.

7.5 Blinde Flecken

An dieser Stelle soll noch auf verschiedene denkbare Ausweitungen bzw. Vertiefungen des Forschungsthemas eingegangen werden.

7.5.1 Controlled World

Der Bereich der kontrollierten Umwelt (*controlled world*) lässt sich gut mit den Angaben in den Ausgangsfragebögen verknüpfen. Die Studienteilnehmer·innen wurden dort gefragt, unter welchen Umständen sie eine Technologie wie den Skype Translator noch einmal nutzen würden. Die Rückmeldungen waren zweiseitig: Mehrere Proband·innen befürworteten den Einsatz des Skype Translators im Rahmen der Arbeit oder für berufliche Kontakte. Diese Meinung zielt also unmittelbar auf Bereiche ab, die im Bereich der Qualitäts- und Risikobewertung von Sprachdienstleistern möglicherweise als Hoch-Risiko-Texte qualifiziert werden. Als Gegenpol hierzu lehnt ein anderer Teil der Versuchspersonen genau dieses Einsatzfeld ab und grenzt die Nutzbarkeit des Skype Translators auf Gespräche mit Freund·innen und Bekannt·innen ein. Zu beiden Standpunkten wird Forschung betrieben. So untersucht Spinuzzi (2017) etwa den Einsatz von Chat-Kommunikation am Arbeitsplatz. Die Charakteristika von Chat-Kommunikation in einem informellen Umfeld werden hingegen beispielsweise von Pappert (2017) oder Verheijen (2017) betrachtet.

Implizit werfen diese Aussagen also die Frage auf, wie stark eine Kommunikationssituation kontrolliert werden kann – durch und mit Technologien wie dem Skype Translator.

7.5.2 Visual World Paradigm

Die Erkenntnisse aus der Studie lassen in Bezug auf das Visual World Paradigm² noch Lücken offen. Auf Grundlage der statistischen und der visuellen Inspektion der Daten bleibt bisweilen ungeklärt, ob die Proband·innen willentlich und bewusst zwischen Original und maschineller Übersetzung wechseln, oder ob die Betrachtung der fremdsprachlichen Textboxen lediglich dem Reflex folgt, auf neue, eingehende Reize zu reagieren. Im ersten Fall könnten daher zwei Annahmen aufgestellt werden: Entweder suchen die Proband·innen lediglich nach vergleichbaren Anhaltspunkten wie Eigennamen, Zahlen oder Wörtern, die dieselbe Herkunft und Zeichenform in beiden Sprachen aufweisen, und springen deshalb punktuell zwischen Original und MÜ, oder die Versuchspersonen vergleichen die beiden Einblendungen akribisch und versuchen tatsächlich den gesamten Chatbeitrag abzulegen.

7.5.3 Sichtbarkeit der Übersetzer·innen

Auch wenn es nur informell und zuweilen höchst interpretativ zwischen den Antwortzeilen der Fragebögen durchschimmert, wirkt es so, als spiele die maschinelle Übersetzung in der Wahrnehmung der Proband·innen allenfalls eine untergeordnete Rolle. Der/die Übersetzer·in bleibt in dieser Technologie also weitestgehend unsichtbar und tritt nur dann ins Bewusstsein der Versuchsteilnehmer·innen, wenn einer der Gesprächsabschnitte unzulänglich übersetzt wird. Der Status Quo jedoch scheint die Annahme zu sein, dass die Kommunikation gelingt und daher keine weitere Aufmerksamkeit auf dieses System fallen muss. Offenbar sind sich die Proband·innen zwar bewusst, dass sie an einer von einer Maschine vermittelten Kommunikation teilnehmen (s. Beispiele 1 und 2, S. 112). Dass es sich explizit um ein maschinelles Übersetzungssystem handelt, wird ihnen jedoch erst auf Rückfrage zur Qualität klar (s. Beispiele 8 sowie 9, S. 114). Damit würde die Auseinandersetzung mit dem Skype Translator in die Richtung der bereits von Venuti (1986) problematisierten Sichtbarkeit der Übersetzer·innen gehen.

²S.a. <https://likan.info/en/paradigm/language/visual-world-paradigm/>, letzter Aufruf am 31. Januar 2022.

7 Diskussion

Über diesen Punkt sollte auch noch einmal eine Rückanbindung an die Verwendung des Katalanischen als Studiensprache (s. u. Abschnitt 2.6.1, S. 32) stattfinden. So könnte die geringe Wahrnehmung des MÜ-Systems auch der relativen Nähe der beteiligten Sprachen geschuldet sein. Wie bereits mehrfach als Denkanstoß angeboten, verfügen das Katalanische und das Deutsche über dasselbe Zeichensystem. Besonders Versuchspersonen, die eine andere romanische Sprache beherrschen, fühlen sich deshalb möglicherweise auch mit dem Katalanischen vertraut, wodurch im Bewusstsein die Tatsache abgeschwächt wird, dass eine MÜ in der Kommunikation zwischengeschaltet ist. Mit zunehmender Distanz zwischen den beteiligten Sprachen – etwa bei einer Studienwiederholung mit Beteiligung des Russischen, Griechischen oder Chinesischen (s. u.) – könnte daher vielleicht auch die Aufmerksamkeit für die maschinelle Übersetzung stärker werden.

7.5.4 Pilotierung

Diese Studie ist nicht als Erweiterung einer bestehenden Forschungslinie gedacht. Vielmehr ist es so, dass diese Studie eine erste Pilotierung eines bislang wenig betrachteten, aber immer mehr in den Vordergrund drängenden Feldes darstellt. Um die maschinell übersetzte, computervermittelte Kommunikation eingehender zu untersuchen, bieten sich mehrere Studien zur Erweiterung dieses Bereichs an:

Über die Betrachtung des Skype Translators unter Beteiligung von Sprachen desselben Zeichensystems hinaus von Interesse und sicherlich eine zu schließen-de Forschungslücke ist die Untersuchung des Kommunikationsverhaltens von Personen, denen eines der beteiligten Schriftsystemen gänzlich unbekannt ist (s. o.). Eine auf die multi-subjektive und womöglich auch kollaborative Interaktion in Chat-Gesprächen ausgelegte Studie böte außerdem einen gänzlich anderen Blick auf internationale, synchrone CvK.

Eine weitere – bislang nur vorläufige – Beobachtung ist, dass die Probandinnen im katalanisch-deutschen Versuch kürzere Nachrichten schreiben als die Personen im monolingualen Setting. Das kann daran liegen, dass die Personen unbewusst auf die MÜ bzw. den fremdsprachigen Gegenüber Rücksicht nehmen, während sie sich im Falle des ein- und zugleich muttersprachlichen Versuchs der Aussagen sicherer sind. Einschränkend ist hierbei der Einsatz mehrerer Gesprächspartner·innen in beiden Versuchsaufbauten zu beachten. Das individuelle Schreibverhalten der verschiedenen Personen ist ohne Zweifel ein Faktor, der bei der Untersuchung der Länge der einzelnen Chatbeiträge miteinbezogen werden muss.

Damit einher geht die Betrachtung des nunmehr umstrukturierten Layouts des Skype Translators. Da nun nicht mehr jeder einzelne Beitrag mit der einhergehenden MÜ direkt unterhalb der Nachricht angezeigt wird, sondern alle Nachrichten automatisch maschinell übersetzt in der jeweiligen Systemsprache des verwendeten Endgerätes dargestellt werden, besteht auch hier Raum für mögliche Hypothesen. Ein Vorteil dieses Layouts für naturalistisch ausgerichtete Studien mit Eye-Tracking-Einsatz ist sicherlich die bessere Nachvollziehbarkeit der Augenbewegungen. Jeder beteiligten Person werden nicht mehr vier verschiedene Beitragsarten (bei zwei Personen jeweils Original und MÜ in den beteiligten Sprachen) angezeigt, sondern lediglich zwei. Und diese sind darüber hinaus auch im üblichen zweispaltigen Chat-Aufbau eindeutig voneinander getrennt. Erst mit Klick auf die jeweilige Nachricht wird dann das Original bzw. die MÜ an selbiger Stelle angezeigt.

Aus korpuslinguistischer Sicht sind zudem die im Rahmen dieses Projektes entstandenen natürlichen Sprachdaten aufzubereiten und zu untersuchen. Neu an diesem kleinen Korpus ist die Tatsache, dass es sich nicht nur um ein bilanguales Korpus aus deutschen und katalanischen Textchat-Beiträgen handelt, sondern dass darüber hinaus auch noch deren jeweilige maschinelle Übersetzung zur Verfügung steht. In diesem Zusammenhang können auch die Datensätze des Key-Loggings zur Untersuchung hinzugezogen werden.

Die Studie sollte weiterhin in einzelnen Bereichen detaillierter wiederholt werden. So stellt O'Brien (2009) in ihrer Untersuchung fest, dass es zu einem stetigen Gewöhnungsprozess bei der Verarbeitung der präsentierten Reize kommt. Daraufhin schlossen sie die ersten Sekunden der Reizexponierung aus der Analyse aus, woraufhin sich auch die Resultate änderten. Eine erneute Durchführung der hier beschriebenen Studie könnte genau diese Prozesse erforschen: Wie verhalten sich die Werte der Durchlaufzeit, der Verweildauer oder die Fixationsanzahl, wenn jede Nachrichtenkategorie auch zeitlich unterteilt wird? Angenommen, man betrachtete sowohl den Chatverlauf als Ganzen als auch die einzelnen Nachrichten zeitlich unterteilt nach Beginn, Mitte und Ende. Die Ergebnisse erlaubten sicherlich Aufschluss über Gewöhnungsprozesse im Umgang mit der MÜ sowie mit der Technologie an sich.

Weiterhin wäre es wünschenswert, die Studie auch mit Personen durchzuführen, die keinem derart global vernetztem Umfeld wie das der Universität entstammen. Die hier rekrutierten Versuchspersonen sind allesamt Studierende in einem Alter und aus einer Generation, für die es weitaus selbstverständlicher ist, mit verschiedenen Sprachen im Alltag und im Beruf konfrontiert zu sein, als es noch vor wenigen Jahren der Fall war. Die Annahmen über die Problemlösungsstrategien der Teilnehmer·innen sind also ganz andere als bei Personen, die nicht derart

7 Diskussion

häufig und tief mit Fremdsprachen in Kontakt stehen. Die Rekrutierung von Personen, die nicht zur Generation Y gehören, ist daher womöglich als ertragreich zu bewerten. Es wäre demnach spannend zu beobachten, wie diese Zielgruppe mit derartiger Technologie umgeht. Gleiches gilt auch für Personen, die einen formell niedrigeren Bildungsgrad aufweisen als auf akademischem Niveau.

Ebenso bleibt ein soziologisches Spannungsfeld bestehen, auf dem in den nächsten Jahren sicherlich vielseitige und tiefgreifende Diskussionen geführt sowie stetige Neuerungen bekanntgegeben werden. Bislang stellt die Nutzeridentität ein mittel- bis schwerwiegendes Problem für die Maschine dar. Nicht nur der Umgang mit Eigennamen ist damit gemeint, sondern auch die Referenzierung über Personalpronomen und damit die Zuschreibung des eigenen Geschlechts. In den gesammelten Sprachdaten dieser Arbeit kam es beispielsweise zur Übersetzung des katalanischen Eigennamens *Alba*. Die MÜ hatte diese Einheit nicht als solche erkannt, sondern das gemeinsprachliche Lexem identifiziert, dem im Deutschen „der Morgengrauen“ entspricht. Gleichermassen problematisiert Roser (2018) den Umgang mit der Geschlechtsidentität sowohl zwischen Nutzer·innen als auch zwischen Nutzer·in und Maschine. Konkret weist der Autor darauf hin, dass die Maschinen bislang eher weiblich geprägt sind, was sich bereits in der von Werk eingestellten Stimme und dem Anredenamen wie *Siri*, *Alexa* oder *Cortana* wider-spiegelt.

7.6 Zusammenfassung

Die triangulierten Ergebnisse sowohl der Online-Umfrage als auch den beiden Versuchsaufbauten der Eye-Tracking-Studie zeichnen das Bild einer Sprachtechnologie, die in einem bemerkenswerten Zwiespalt steht. Einerseits hinkt Skype, sowohl was den Funktionsumfang als auch die Nutzungsintensität angeht, offenbar geläufigeren Diensten hinterher. Zugleich führt die Marktmacht Microsofts jedoch dazu, dass der Dienst nicht ganz abgehängt wird. Dafür ist die Nutzung der Software besonders für Unternehmen nach wie vor zu groß. Aber auch der Kontakt zwischen Privatpersonen in verschiedenen Ländern scheint ein Bereich zu sein, der bevorzugt die Kommunikation über Skype verlangt. In diesem Kontext scheint es beinahe verwunderlich, dass ausgerechnet Skype im europäischen Kontext der bislang einzige Dienst ist, der eine maschinelle Übersetzung für die Chat-Kommunikation anbietet.

Die folgenden wesentlichen Erkenntnisse lassen sich aus den Studien entnehmen: Einerseits ist die maschinelle Übersetzung ein starker Einflussfaktor bei der Kommunikation zwischen verschiedensprachigen Personen, die der Sprache des

7.6 Zusammenfassung

Gegenübers nicht mächtig sind. Die Vergleichsstudien belegen, dass das Vorhandensein der MÜ-Ausgabe einen Einfluss auf das Leseverhalten und die Wahrnehmung der Situation haben.

Aus berufspraktischer Sicht weist Spinuzzi (2017: 214 f.) darauf hin, dass die CvK in vielen beruflichen Situation heutzutage einen gewöhnlichen Bestandteil darstellt. Auch wenn es ungewöhnlich anmuten mag, nutzen selbst Kolleginnen, die nur wenige Zentimeter voneinander entfernt sitzen, vermehrt CvK, um sich nicht aus dem Arbeitsfluss zu reißen und wenden sich erst dann persönlich zu, wenn eine grundlegende Klärung des Anliegens erfolgt ist. Mit dieser Perspektive bietet sich auch für den Skype Translator eine Chance. So könnte die Technologie folglich zur Kontaktaufnahme und anfänglichen Erörterung von Problemen zwischen Personen mit einem gewissen Anliegen verwendet werden (z. B. tatsächlich zur Vorbereitung auf einen Auslandsaufenthalt). Aufbauend auf diesem Erstkontakt kann dann festgelegt werden, wie die weitere Begleitung oder gar Beratung auszusehen hat. Die Kommunikation von Angesicht zu Angesicht ist dabei jedoch unerlässlich, sodass der Technologie eher die Rolle einer Ergänzung als denn einer Ersetzung zukommt.

8 Fazit und Ausblick

Die vorliegende Arbeit hat gezeigt, dass der Einsatz von maschineller Übersetzung (in Echtzeit) im Rahmen der CvK einen Einfluss auf die Wahrnehmung der Kommunikation hat. Hierzu wurde eine nach bestem Wissen und Gewissen bislang so zuvor noch nicht durchgeführte, explorative Fallstudie entwickelt, die das Chatverhalten von Versuchspersonen in einem Setting mit maschineller Übersetzung mit dem Verhalten in einem monolingualen Chat vergleicht. Die Ergebnisse der vorgenommenen Analysen in dieser Arbeit lassen sich mit denen anderer, in jüngster Vergangenheit durchgeföhrter naturalistisch orientierter Studien vergleichen und reihen sich damit in den aktuellen Forschungskontext mit kommunikations- und MÜ-bezogener Ausrichtung ein. Einen theoretischen Mehrwert liefert die Arbeit dadurch, dass sie aus kommunikationswissenschaftlicher Sicht die Sphäre der Gruppenchats verlässt und eine Konstellation von lediglich zwei am Chat beteiligten Personen aufgreift und diese wiederum mit dem bestehenden Forschungsstand der Translationstechnologien verknüpft. Während die Mensch-Maschine-Interaktion sowie die Informationsbeschaffung mittels Chatbots mittlerweile einen eigenen, erschlossenen Forschungsbereich darstellen, bestehen mehrere blinde Flecken auf dem Gebiet der CvK in Verbindung mit Echtzeit-MÜ. Die vorliegende Arbeit weist auch durch die kritische Be trachtung der eigenen Methoden und Ergebnisse auf diese Punkte hin und trägt ein wenig zu diesem Brückenschluss bei. Für die methodologische Wegbereitung wurden Elemente der naturalistisch orientierten Forschungskonzeption auf den Gebieten des Eye-Trackings sowie der CvK auf das kleinste erdenkliche Setting mit lediglich zwei beteiligten Personen übertragen. Die besondere Herausforderung lag in der Anpassung bestehender Methoden auf den Einsatz von maschineller Übersetzung innerhalb einer Chatanwendung wie – in diesem konkreten Fallbeispiel – dem Skype Translator. Dabei haben sich die einzelnen für die Analyse verwendeten Indikatoren als unterschiedlich aussagekräftig erwiesen. Der praktische Wert der Arbeit besteht in der Untersuchung des ungeleiteten, am Alltag orientierten Textchats zwischen Personen, die der Muttersprache der anderen beteiligten Person nicht mächtig sind. Aus dieser Betrachtung entstammen Erkenntnisse sowohl in Bezug auf das Chatverhalten als auch auf die Anforderungen an eine möglichst selbsterklärende und intuitiv konzipierte Kommuni-

8 Fazit und Ausblick

kationssoftware, damit eine solche Kommunikationssituation überhaupt erfolgreich verläuft. Besonders hervorzuheben ist die Erkenntnis, dass der Echtzeit-MÜ einerseits substanziell hohe Aufmerksamkeit zukommt, andererseits im Bewusstsein der Chatteilnehmer·innen weniger präsent zu sein scheint als angenommen. Nicht zuletzt bietet die Arbeit einen praktischen Ausblick auf die Anwendbarkeit der Ergebnisse. Kommunikationsanwendungen unterliegen einem stetigen Wandel, was auch über den Verlauf dieses gesamten Projektes spürbar wurde. Eine kontinuierliche Untersuchung ist gerade deshalb unerlässlich.

Anhang A: Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

Dies ist der Hinweis zum Datenschutz des Online-Umfragetools *LimeSurvey*:

Eine Bemerkung zum Datenschutz

Dies ist eine anonyme Umfrage. In den Umfrageantworten werden keine persönlichen Informationen über Sie gespeichert, es sei denn, in einer Frage wird explizit danach gefragt. Wenn Sie für diese Umfrage einen Zugangsschlüssel benutzt haben, so können Sie sicher sein, dass der Zugangsschlüssel nicht zusammen mit den Daten abgespeichert wurde. Er wird in einer getrennten Datenbank aufbewahrt und nur aktualisiert, um zu speichern, ob Sie diese Umfrage abgeschlossen haben oder nicht. Es gibt keinen Weg, die Zugangsschlüssel mit den Umfrageergebnissen zusammenzuführen.

Der individuell ausgearbeitete Hinweis zum Fragebogen:

Liebe Teilnehmerin,

lieber Teilnehmer,

diese Umfrage findet im Rahmen des Forschungsvorhabens von *Felix Hoberg* mit dem Titel *Informationsintegration in mehrsprachigen Textchats* unter der Leitung von *Prof. Dr. Oliver Czulo* am *Institut für Angewandte Linguistik und Translatologie* der *Universität Leipzig* statt.

Die Umfrage erhebt Daten zu zwei Themengebieten. Zunächst sollen Angaben zu Ihrem Kenntnisstand und Nutzungsverhalten im Umgang mit dem Messaging-Dienst Skype erhoben werden. Ein Teil wird hierbei genauer auf

A Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

Ihre Erfahrungen mit dem Skype Translator eingehen. Das zweite Themengebiet umfasst Fragen zu Ihren Auslandsaufenthalten und Sprachkenntnissen.

Ziel der Umfrage ist es, Angaben von möglichst vielen unterschiedlichen Nutzergruppen von Skype zu erhalten. Diese Daten möchte ich mit den Ergebnissen eines Experiments am Skype Translator vergleichen. Abschließend bitte ich Sie noch, ein paar Angaben zu Ihrer Person zu machen.

Das **Ausfüllen wird ca. 10 Minuten** in Anspruch nehmen. Das Ausfüllen kann jederzeit zwischengespeichert und zu einem späteren Zeitpunkt fortgesetzt werden. Die Teilnahme ist freiwillig. Die Daten werden nur zum Zwecke des o. g. Forschungsvorhabens in anonymisierter Form verwendet. Es werden keine Namen oder sonstige personenspezifische Daten erhoben, die einen Rückschluss auf Teilnehmer erlauben könnten. Die Angaben beim Zwischenspeichern der Umfrage dienen ausschließlich der Wiederaufnahme zu einem späteren Zeitpunkt und können vom Umfrageleiter nicht eingesenkt werden.

Die Mailadresse wird unabhängig vom jeweiligen Fragebogen für die Dauer des Forschungsvorhabens aufbewahrt und auf Wunsch gelöscht. Das Umfragetool erhebt weder IP-Adresse noch Datumsangaben, die einen Rückschluss auf den Probanden erlauben. Somit ist zu keiner Zeit eine Zuordnung von Mailadresse zu Fragebogen möglich. Die Verwendung und der Betrieb dieses Online-Tools unterliegt gängigen Datenschutzbestimmungen (s. Datenschutzrichtlinien dieses Umfragetools). Sie haben jederzeit das Recht, ohne Angabe von Gründen die Erhebung zu unterbrechen oder gar komplett abzubrechen. Weiterhin steht Ihnen das Recht zu, Ihre Eingaben unwiderbringlich löschen zu lassen. Hierzu bedarf es der schriftlichen Mitteilung an den Umfrageleiter, Felix Hoberg.

Sollten Sie Rückfragen haben oder von den genannten Rechten Gebrauch machen wollen, kontaktieren Sie mich über felix.hoberg@uni-leipzig.de oder auf anderem Wege unter der hier angegebenen Kontaktadresse. Für Anliegen, die die Umfrage unmittelbar betreffen, steht Ihnen am Ende des Fragebogens ein Kommentarfeld zur Verfügung.

Vielen Dank für Ihre Teilnahme!

Am Ende der Umfrage wird weiterhin folgender Hinweis eingeblendet:

Sie sind nun am Ende der Umfrage angelangt. An dieser Stelle möchte ich mich noch einmal herzlich bei Ihnen für die Teilnahme und die investierte Zeit bedanken. Ihre Teilnahme ist für mich ein wichtiger Beitrag zu meinem Forschungsvorhaben.

Sollten Sie weitere Informationen wünschen, oder über die weitere Entwicklung dieses Vorhabens informiert werden möchten, hinterlassen Sie bitte Ihre Kontaktdaten über das folgende Eingabefenster (Klicken sie unten auf Weiter). Diese werden selbstverständlich gemäß der eingangs angesprochenen Datenschutzrichtlinien verwaltet und getrennt von den Angaben der Umfrage aufbewahrt. Bitte beachten Sie, dass es mehrere Monate dauern kann, bis aussagekräftige, veröffentlichte Ergebnisse vorliegen.

Andernfalls können Sie das Browserfenster jetzt schließen.

Beste Grüße

Felix Hoberg

Nutzung von Skype

1. Nutzen Sie Skype?
 - Nein → *weiter mit Frage 17*
 - Ja → *weiter mit Frage 2*
2. Wie häufig nutzen Sie Skype?
 - Nie → *weiter mit Frage 17*
 - weniger als monatlich
 - monatlich
 - wöchentlich
 - täglich

A Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

Nutzung von Skype

3. Wie lange nutzen Sie Skype durchschnittlich pro Sitzung?
 - bis 5 Minuten
 - bis 15 Minuten
 - bis 30 Minuten
 - bis 1 Stunde
 - länger als 1 Stunde
4. Welche Skype-Version verwenden Sie?
 - Skype für Windows
 - Skype für MacOS
 - Skype für Linux
 - Skype für Mobiltelephone
 - Browser
5. Welche Funktionen nutzen Sie bei Skype?
 - Voice-Chat
 - Video-Chat
 - Text-Chat
 - Festnetzanruf
 - Translator
 - Bildschirmübertragung
 - Konferenzmodus
 - Dateifreigabe
6. Wer sind Ihre Gesprächspartner?
 - Familie im selben Land
 - Familie im Ausland
 - Freunde im selben Land
 - Freunde im Ausland
 - Berufliche, im selben Land
 - Berufliche, im Ausland
 - Institutionen im selben Land
 - Institutionen im Ausland

Nutzungsbewertung

7. Bewerten Sie ihre Erfahrung mit der Text-Chat-Funktion.

	1	2	3	4
schlecht	◦	◦	◦	◦
				gut

8. Bewerten Sie ihre Erfahrung mit der Video-Chat-Funktion.

	1	2	3	4
schlecht	◦	◦	◦	◦
				gut

9. Bewerten Sie ihre Erfahrung mit der Voice-Chat-Funktion.

	1	2	3	4
schlecht	◦	◦	◦	◦
				gut

10. Haben Sie bereits Erfahrungen mit dem Skype Translator gemacht?

- Ja → *weiter mit Frage 11*
- Nein → *weiter mit Frage 17*

Nutzungsbewertung Skype Translator

11. Bewerten Sie ihre Erfahrung mit dem Skype Translator.

	1	2	3	4
schlecht	◦	◦	◦	◦
				gut

12. Ihr Gesprächspartner hat die gleiche Muttersprache wie Sie. Wie tolerant sind Sie bei sprachlichen Fehlern ihm gegenüber?

	1	2	3	4
nicht tolerant	◦	◦	◦	◦
				tolerant

A Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

13. Ihr Gesprächspartner hat NICHT die gleiche Muttersprache wie Sie. Wie tolerant sind Sie bei sprachlichen Fehlern ihm gegenüber?

1	2	3	4	
nicht tolerant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> tolerant

14. Wie bewerten Sie die Leistung des Skype Translators hinsichtlich...?

	1	2	3	4	
allgemeine Leistung	schlecht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> gut
Präzision	unpräzise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> präzise
Verständlichkeit	unverständlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> verständlich
Brauchbarkeit des Outputs	unbrauchbar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> brauchbar
Vollständigkeit des Outputs	unvollständig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> vollständig

15. Wie sehen Sie das Verhältnis von Skype Translator zu menschlichen Übersetzern bzw. Dolmetschern?

	1	2	3	4
Konkurrenz	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/> verschiedene Welten

16. Würden Sie die Funktion auch nutzen, wenn Sie kostenpflichtig wäre?

- Ja
- Nein

Alternativen

17. Nutzen Sie Alternativen zu Skype, die ebenfalls einen Voice- oder Video-Chat bieten?

- Ja → *weiter mit Frage 18*
- Nein → *weiter mit Frage 24*

Alternativen

18. Welche Alternativen zum Video-Chat nutzen Sie?

- Appear.in

- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

19. Welche Alternativen zum Voice-Chat nutzen Sie?

- Apear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

20. Welche Alternativen zum Text-Chat nutzen Sie?

- Apear.in
- Facebook Video-Chat
- Google Hangouts

A Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

21. Welchen Dienst nutzen Sie für Video-Chats am Häufigsten?

- Skype
- Apear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

22. Welchen Dienst nutzen Sie für Voice-Chats am Häufigsten?

- Skype
- Apear.in
- Facebook Video-Chat
- Google Hangouts

- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

23. Welchen Dienst nutzen Sie für Text-Chats am Häufigsten?

- Skype
- Apear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

Auslandserfahrung

24. Befinden Sie sich derzeit längerfristig (Urlaub ausgenommen, länger als 4 Wochen) außerhalb Deutschlands?

- Ja → *weiter mit Frage 26*
- Nein → *weiter mit Frage 25*

A Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

Auslandserfahrung

25. Haben Sie bereits längere Zeit (Urlaub ausgenommen, länger als 4 Wochen) außerhalb Deutschlands verbracht?
- Ja → *weiter mit Frage 30*
 - Nein → *weiter mit Frage 35*

Auslandserfahrung – gegenwärtig

26. In welchem Land befinden Sie Sich?
-

27. In welcher Stadt befinden Sie Sich?
-

28. Wie lange dauert der Aufenthalt in Monaten?
-

29. Was ist der Grund des Aufenthaltes?

- Spracherwerb
- Sprachvertiefung
- Kulturkontakt
- Pflichtaufenthalt
- Fortbildung
- Beruflich
- Sonstiges: _____
Weiter mit Frage 25

Auslandserfahrung – in der Vergangenheit

30. In welchen Ländern haben Sie längere Zeit (Urlaub ausgeschlossen, länger als 4 Wochen) verbracht?
-
-
-
-

31. In welchen Städten haben Sie sich primär aufgehalten?

32. Wie lange dauerte der Aufenthalt in Monaten?

33. Was war der Grund des Aufenthaltes?

- Spracherwerb
- Sprachvertiefung
- Kulturkontakt
- Pflichtaufenthalt
- Fortbildung
- Beruflich
- Sonstiges: _____

34. Wie lange liegt der Aufenthalt in Monaten zurück?

Individuelle Angaben

35. Bitte geben Sie ihr Alter in Jahren an.

36. Bitte geben Sie ihr Geschlecht an.

- Weiblich
- Männlich
- Sonstiges: _____

37. Bitte geben Sie Ihren höchsten Abschluss an.

- keinen
- bis mittlere Reife (z. B. Hauptschule, Realschule, o. ä.)
- Hochschulzugangsberechtigung (z. B. Fachabitur, allgemeine Hochschulreife, o. ä.)

A Fragebogen 1: Allgemeine Umfrage zur Nutzung von Skype

- abgeschlossene Ausbildung
- höherer Berufsabschluss (z. B. Meister, Techniker, o. ä.)
- Hochschulabschluss (z. B. Bachelor, Master, Diplom, o. ä.)
- Sonstiges: _____

38. Mit welcher Sprache sind Sie primär aufgewachsen?

39. Welche Sprache verwenden Sie hauptsächlich im Alltag?

40. Welche Sprache verwenden Sie hauptsächlich im Beruf

41. Welche Sprachen beherrschen Sie weiterhin?¹

42. Wie schätzen Sie Ihre Kompetenz entlang des Europäischen Referenzrahmens für die folgenden Sprachen ein?

	A1	A2	B1	B2	C1	C2
Spanisch	◦	◦	◦	◦	◦	◦
Katalanisch	◦	◦	◦	◦	◦	◦
Galizisch	◦	◦	◦	◦	◦	◦
Portugiesisch	◦	◦	◦	◦	◦	◦
Italienisch	◦	◦	◦	◦	◦	◦
Französisch	◦	◦	◦	◦	◦	◦
Rumänisch	◦	◦	◦	◦	◦	◦

43. Hier haben Sie abschließend noch die Möglichkeit, mir bei Bedarf Rückmeldung zu geben, Fragen zu stellen, oder auf Auffälligkeiten hinzuweisen. Bitte nutzen Sie dieses Feld nur in unmittelbaren Zusammenhang mit der

¹Unter *beherrschen* ist hier eine nachweisliche Qualifikation zu verstehen, etwa ein Zeugnis, ein Diplom oder eine Urkunde. Zu vernachlässigen ist hingegen das Jahr, aus dem diese Qualifikation stammt.

Umfrage. Für alle weiteren Anliegen nutzen Sie bitte die eingangs aufgeführten Kontaktmöglichkeiten.

Anhang B: Fragebogen 2: Fragebogen Pre-Experiment

Dies ist der Einweisungstext für die Teilnehmer des Experiments:

Liebe Teilnehmerin, lieber Teilnehmer,

diese Beobachtung findet im Rahmen des Forschungsvorhabens von *Felix Hoberg* mit dem Titel *Informationsintegration in mehrsprachigen Textchats* unter der Leitung von *Prof. Dr. Oliver Czulo* am *Institut für Angewandte Linguistik und Translatologie* der *Universität Leipzig* statt.

Die Beobachtung ist aus drei Teilen aufgebaut:

1. Zunächst bitte ich Sie, diesen Fragebogen auszufüllen. Es werden Daten zu Ihrer Nutzungserfahrung mit Skype – und spezieller: dem Skype Translator – sowie zu (möglicher) Auslandserfahrung erhoben. Der Fragebogen enthält Multiple-Choice-Fragen, Single-Choice-Fragen, Bewertungsskalen und Fragen mit offenem Antwortbereich.
2. Nach Ausfüllen des Bogens bitte ich Sie an einer Kommunikationssituation über Skype teilzunehmen, bei der der Skype Translator aktiviert ist. Ihr Gesprächspartner wird ein katalanischer Muttersprachler sein. Für das Gespräch erhalten Sie ein bestimmtes Themengebiet, über das Sie sich in Ihrer eigenen Muttersprache informieren sollen. Der Bildschirm an dem Sie arbeiten sowie Ihre Pupillenbewegung wird mittels Eye Tracker aufgezeichnet.
3. Zum Abschluss der Beobachtung erhalten Sie einen weiteren Fragebogen, der Ihre jüngste Erfahrung – mit besonderem Fokus auf dem Skype Translator – dokumentiert.

B Fragebogen 2: Fragebogen Pre-Experiment

Insgesamt ist die Beobachtungssituation auf eine Dauer von ca. 30 Minuten ausgelegt. Die Teilnahme ist freiwillig. Die Daten werden nur zum Zwecke des o. g. Forschungsvorhabens in anonymisierter Form verwendet. In Bezug auf den Eye Tracker ist zu betonen, dass lediglich Ihre Pupillenbewegung am verwendeten Bildschirm aufgezeichnet wird. Ihr Gesicht hingegen wird zu keinem Zeitpunkt gespeichert.

Es werden weiterhin keine Namen oder sonstige personenspezifische Daten erhoben, die einen Rückschluss auf den Teilnehmer erlauben könnten. Dies gilt besonders für die Video-Chat-Funktion von Skype.

Die Angabe der Mailadresse ist freiwillig und dient der Kontaktaufnahme. Die Mailadresse wird unabhängig vom jeweiligen Datensatz aufbewahrt und auf Wunsch gelöscht. Die Gesprächssituation erfolgt über das Skype-Konto von Felix Hoberg und macht so keine Veröffentlichung privater Probanden-nutzerkonten notwendig. Der Proband hat jederzeit das Recht, ohne Angabe von Gründen die Erhebung zu unterbrechen oder gar komplett abzubrechen. Weiterhin steht dem Teilnehmer das Recht zu, seine Eingaben unwie-derbringlich löschen zu lassen. Hierzu bedarf es der schriftlichen Mitteilung an den Versuchsleiter.

Sollten Sie Rückfragen haben oder von den o.g. Rechten Gebrauch machen wollen, so melden Sie sich bitte per Mail unter felix.hoberg@uni-leipzig.de oder auf anderem Wege unter hier angegebenen Kontaktadresse.

Vielen Dank für Ihre Teilnahme!

Nutzung von Skype

1. Haben Sie bereits an der Online-Umfrage zum Nutzungsverhalten von Skype im Zeitraum von 05/18-07/18 teilgenommen?
 - Nein
 - Ja
2. Nutzen Sie Skype?
 - Nein → *weiter mit Frage 19*
 - Ja → *weiter mit Frage 3*

3. Wie häufig nutzen Sie Skype?

- Nie → *weiter mit Frage 19*
- weniger als monatlich
- monatlich
- wöchentlich
- täglich

Nutzung von Skype

4. Wie lange nutzen Sie Skype durchschnittlich pro Sitzung?

- bis 5 Minuten
- bis 15 Minuten
- bis 30 Minuten
- bis 1 Stunde
- länger als 1 Stunde

5. Welche Skype-Version verwenden Sie?

- Skype für Windows
- Skype für MacOS
- Skype für Linux
- Skype für Mobiltelephone
- Browser

6. Welche Funktionen nutzen Sie bei Skype?

- Voice-Chat
- Video-Chat
- Text-Chat
- Festnetzanruf
- Translator
- Bildschirmübertragung
- Konferenzmodus
- Dateifreigabe

B Fragebogen 2: Fragebogen Pre-Experiment

7. Wer sind Ihre Gesprächspartner?

- Familie im selben Land
- Familie im Ausland
- Freunde im selben Land
- Freunde im Ausland
- Berufliche, im selben Land
- Berufliche, im Ausland
- Institutionen im selben Land
- Institutionen im Ausland

8. Welche Sprache verwenden Sie im Kontakt mit diesen Gesprächspartnern?

Nutzungsbewertung

9. Bewerten Sie ihre Erfahrung mit der Text-Chat-Funktion.

1 2 3 4

schlecht ◦ ◦ ◦ ◦ gut

10. Bewerten Sie ihre Erfahrung mit der Video-Chat-Funktion.

1 2 3 4

schlecht ◦ ◦ ◦ ◦ gut

11. Bewerten Sie ihre Erfahrung mit der Voice-Chat-Funktion.

1 2 3 4

schlecht ◦ ◦ ◦ ◦ gut

12. Haben Sie bereits Erfahrungen mit dem Skype Translator gemacht?

- Ja → *weiter mit Frage 13*
- Nein → *weiter mit Frage 19*

Nutzungsbewertung Skype Translator

13. Bewerten Sie ihre Erfahrung mit dem Skype Translator.

	1	2	3	4	
schlecht	◦	◦	◦	◦	gut

14. Ihr Gesprächspartner hat die gleiche Muttersprache wie Sie. Wie tolerant sind Sie bei sprachlichen Fehlern ihm gegenüber?

	1	2	3	4	
nicht tolerant	◦	◦	◦	◦	tolerant

15. Ihr Gesprächspartner hat NICHT die gleiche Muttersprache wie Sie. Wie tolerant sind Sie bei sprachlichen Fehlern ihm gegenüber?

	1	2	3	4	
nicht tolerant	◦	◦	◦	◦	tolerant

16. Wie bewerten Sie die Leistung des Skype Translators hinsichtlich...?

		1	2	3	4	
allgemeine Leistung	schlecht	◦	◦	◦	◦	gut
Präzision	unpräzise	◦	◦	◦	◦	präzise
Verständlichkeit	unverständlich	◦	◦	◦	◦	verständlich
Brauchbarkeit des Outputs	unbrauchbar	◦	◦	◦	◦	brauchbar
Vollständigkeit des Outputs	unvollständig	◦	◦	◦	◦	vollständig

17. Wie sehen Sie das Verhältnis von Skype Translator zu menschlichen übersetzern bzw. Dolmetschern?

	1	2	3	4	
Konkurrenz	◦	◦	◦	◦	verschiedene Welten

18. Würden Sie die Funktion auch nutzen, wenn Sie kostenpflichtig wäre?

- Ja
- Nein

B Fragebogen 2: Fragebogen Pre-Experiment

Alternativen

19. Nutzen Sie Alternativen zu Skype, die ebenfalls einen Voice- oder Video-Chat bieten?

- Ja → *weiter mit Frage 20*
- Nein → *weiter mit Frage 26*

Alternativen

20. Welche Alternativen zum Video-Chat nutzen Sie?

- Appear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

21. Welche Alternativen zum Voice-Chat nutzen Sie?

- Appear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter

- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

22. Welche Alternativen zum Text-Chat nutzen Sie?

- Apear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

23. Welchen Dienst nutzen Sie für Video-Chats am Häufigsten?

- Skype
- Apear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat

B Fragebogen 2: Fragebogen Pre-Experiment

- Instagram-Chat
- ICQ
- Sonstige: _____

24. Welchen Dienst nutzen Sie für Voice-Chats am Häufigsten?

- Skype
- Appear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat
- Instagram-Chat
- ICQ
- Sonstige: _____

25. Welchen Dienst nutzen Sie für Text-Chats am Häufigsten?

- Skype
- Appear.in
- Facebook Video-Chat
- Google Hangouts
- Apple Facetime
- WhatsApp
- Telegram
- Viber
- Twitter
- Snapchat

- Instagram-Chat
- ICQ
- Sonstige: _____

Auslandserfahrung – in der Vergangenheit

26. Haben Sie bereits längere Zeit (Urlaub ausgenommen, länger als 4 Wochen) außerhalb Deutschlands verbracht?

- Ja → *weiter mit Frage 27*
- Nein → *weiter mit Frage 32*

27. In welchen Ländern haben Sie längere Zeit (Urlaub ausgeschlossen, länger als 4 Wochen) verbracht?

28. In welchen Städten haben Sie sich primär aufgehalten?

29. Wie lange dauerte der Aufenthalt in Monaten?

30. Was war der Grund des Aufenthaltes?

- Spracherwerb
- Sprachvertiefung
- Kulturkontakt
- Pflichtaufenthalt
- Fortbildung
- Beruflich
- Sonstiges: _____

31. Wie lange liegt der Aufenthalt in Monaten zurück?

Individuelle Angaben

32. Bitte geben Sie ihr Alter in Jahren an.

33. Bitte geben Sie ihr Geschlecht an.

◦ Weiblich

◦ Männlich

◦ Sonstiges: _____

34. Bitte geben Sie Ihren höchsten Abschluss an.

◦ keinen

◦ bis mittlere Reife (z. B. Hauptschule, Realschule, o. ä.)

◦ Hochschulzugangsberechtigung (z. B. Fachabitur, allgemeine Hochschulreife, o. ä.)

◦ abgeschlossene Ausbildung

◦ höherer Berufsabschluss (z. B. Meister, Techniker, o. ä.)

◦ Hochschulabschluss (z. B. Bachelor, Master, Diplom, o. ä.)

◦ Sonstiges: _____

35. Mit welcher Sprache sind Sie primär aufgewachsen?

36. Welche Sprache verwenden Sie hauptsächlich im Alltag?

37. Welche Sprache verwenden Sie hauptsächlich im Beruf

38. Welche Sprachen beherrschen Sie weiterhin?¹

¹Unter *beherrschen* ist hier eine nachweisliche Qualifikation zu verstehen, etwa ein Zeugnis, ein Diplom oder eine Urkunde. Zu vernachlässigen ist hingegen das Jahr, aus dem diese Qualifikation stammt.

39. Hier haben Sie abschließend noch die Möglichkeit, mir bei Bedarf Rückmeldung zu geben, Fragen zu stellen, oder auf Auffälligkeiten hinzuweisen. Bitte nutzen Sie dieses Feld nur in unmittelbaren Zusammenhang mit der Umfrage. Für alle weiteren Anliegen nutzen Sie bitte die eingangs aufgeführten Kontaktmöglichkeiten.

Anhang C: Fragebogen 3: Fragebogen Post-Experiment

Am Ende der Beobachtungssituation, nach Ausfüllen des Ausgangsfragebogens, wird weiterhin folgender Hinweis eingeblendet:

Sie sind nun am Ende der Umfrage angelangt. An dieser Stelle möchte ich mich noch einmal herzlich bei Ihnen für die Teilnahme und die investierte Zeit bedanken. Ihre Teilnahme ist für mich ein wichtiger Beitrag zu meinem Forschungsvorhaben.

Sollten Sie weitere Informationen wünschen, oder über die weitere Entwicklung dieses Vorhabens informiert werden möchten, hinterlassen Sie bitte Ihre Kontaktdaten über das folgende Eingabefenster (Klicken sie unten auf Weiter). Diese werden selbstverständlich gemäß der eingangs angesprochenen Datenschutzrichtlinien verwaltet und getrennt von den Angaben der Umfrage aufbewahrt. Bitte beachten Sie, dass es mehrere Monate dauern kann, bis aussagekräftige, veröffentlichbare Ergebnisse vorliegen.

Andernfalls können Sie das Browserfenster jetzt schließen.

Beste Grüße
Felix Hoberg

Grundsätzliche Beurteilung

1. Welche der folgenden Aspekte können Sie nun beurteilen?
 - Gesprächsqualität
 - übersetzungsqualität
 - Tonqualität

C Fragebogen 3: Fragebogen Post-Experiment

- Videoqualität
 - Versuchssetting
2. Wie beurteilen Sie...?¹

	schlecht		gut	
	1	2	3	4
die Gesprächsqualität	◦	◦	◦	◦
die Übersetzungsqualität	◦	◦	◦	◦
die Tonqualität	◦	◦	◦	◦
die Videoqualität	◦	◦	◦	◦
das Versuchssetting	◦	◦	◦	◦

3. Wie haben Sie sich in dieser Kommunikationssituation gefühlt?

Wahrnehmung des Probanden

4. Wie beurteilen Sie...?

	schlecht		gut	
	1	2	3	4
die Qualität der Kommunikationssituation	◦	◦	◦	◦
die Informationsdichte	◦	◦	◦	◦
die Leistung des Skype Translators	◦	◦	◦	◦

5. Wie beurteilen Sie...?

	zu kurz		angemessen		zu lang	
	1	2	3	4	5	
die Dauer der Kommunikationssituation	◦	◦	◦	◦	◦	

¹Nur gemäß vorausgehender Frage beantworten.

6. Ist Ihnen etwas in semantischer Hinsicht aufgefallen?
 - Ja → *Blende ein: Frage 12*
 - Nein → *Blende aus: Frage 12*
7. Ist Ihnen etwas in syntaktischer Hinsicht aufgefallen?
 - Ja → *Blende ein: Frage 13*
 - Nein → *Blende aus: Frage 13*
8. Ist Ihnen etwas in terminologischer Hinsicht aufgefallen?
 - Ja → *Blende ein: Frage 14*
 - Nein → *Blende aus: Frage 14*
9. Ist Ihnen etwas in puncto Stil aufgefallen?
 - Ja → *Blende ein: Frage 15*
 - Nein → *Blende aus: Frage 15*
10. Ist Ihnen etwas in puncto Register aufgefallen?
 - Ja → *Blende ein: Frage 16*
 - Nein → *Blende aus: Frage 16*
11. Ist Ihnen weiterhin etwas aufgefallen?
 - Ja → *Blende ein: Frage 17*
 - Nein → *Blende aus: Frage 17*

Auffälligkeiten in der Leistung des Skype Translators

12. Was ist Ihnen in semantischer Hinsicht aufgefallen?

13. Was ist Ihnen in syntaktischer Hinsicht aufgefallen?

C Fragebogen 3: Fragebogen Post-Experiment

14. Was ist Ihnen in terminologischer Hinsicht aufgefallen?

15. Was ist Ihnen in puncto Stil aufgefallen?

16. Was ist Ihnen in puncto Register aufgefallen?

17. Was ist Ihnen weiterhin aufgefallen?

18. Welchen Einfluss hatten die ... Auffälligkeiten auf die Kommunikation?

	keinen	großen		
	1	2	3	4
semantischen	○	○	○	○
syntaktischen	○	○	○	○
terminologischen	○	○	○	○
stilistischen	○	○	○	○
das Register betreffenden	○	○	○	○
sonstigen	○	○	○	○

Informationsgehalt der Kommunikationssituation

19. Haben Sie das Gefühl, dass Ihnen im Nachhinein noch immer Informationen fehlen?

- Ja → *weiter mit Frage 20*
- Nein → *weiter mit Frage 21*

20. Woran machen Sie das Fehlen der Informationen fest?

- individuelle, persönliche Gründe:
-

- individuelle, den Kommunikationspartner betreffende Gründe:
-

- technische:
-

- sonstige:
-

21. Haben Sie durch diese Kommunikation Anregungen bekommen, über die Sie vorher nicht Bescheid wussten?

- Ja → *weiter mit Frage 22*
- Nein → *weiter mit Frage 23*
- Unsicher → *weiter mit Frage 23*

Weiterführung des Informationsbedarfs

22. Wie werden Sie diese weiteren Informationen einholen?

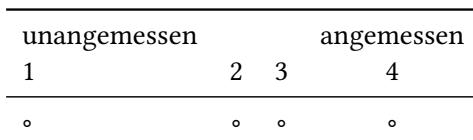
- Skype Translator
- Freunde
- Heimatuniversität
- Internet

C Fragebogen 3: Fragebogen Post-Experiment

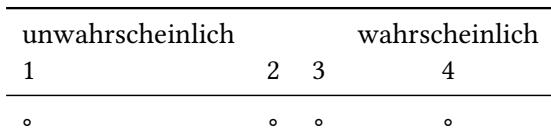
- Gastuniversität
- bis zum Beginn des Auslandsaufenthaltes warten
- Sonstige: _____

Weitere Verwendung des Skype Translators

23. Wie angemessen fanden Sie die Verwendung des Skype Translators in dieser Situation?



24. Wie wahrscheinlich ist es, dass Sie den Skype Translator erneut für eine solche Situation verwenden?



25. Würden Sie den Skype Translator für eine andere Situation verwenden?

- Ja → *weiter mit Frage 26*
- Nein → *weiter mit Frage 27*

26. Wie sähe die Situation aus, in der Sie den Skype Translator verwenden würden?

Vergleich Pre-Experiment-Fragebogen

27. Wie beurteilen Sie nun das Verhältnis von Skype Translator zu übersetzern und Dolmetschern?

	1	2	3	4	
Konkurrenz	◦	◦	◦	◦	verschiedene Welten

28. Weshalb sehen Sie das Verhältnis zwischen Skype Translator und übersetzern so?

29. Wie beurteilen Sie eine Kombination von menschlicher übersetzung/Verdolmetschung und Skype Translator?

	1	2	3	4	
unsinnig	◦	◦	◦	◦	sinnvoll

30. Würden Sie auch Geld für diese Technologie bezahlen?

- Ja → *weiter mit Frage 31*
- Nein

31. Wie viel Geld (Euro) wären Sie bereit dafür auszugeben...?

- in einem Monatsabo

- in einem Jahresabo

- einmalig

Literaturverzeichnis

- Aeppli, Jürg, Luciano Gasser, Eveline Gutzwiller & Annette Tettenborn. 2016. *Empirisches wissenschaftliches Arbeiten: Ein Studienbuch für die Bildungswissenschaften*. 4. Auflage. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Albert, Ruth & Nicole Marx. 2014. *Empirisches Arbeiten in Linguistik und Sprachlehrforschung: Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht*. 2. Auflage. Tübingen: Narr Verlag.
- Altena, Andreas & Frank Moritz. 2014. *Expertenwissen für DGQ-Mitglieder: Risikomanagement – Eine Einführung*. <https://www.dgq.de/wp-content/uploads/2014/03/Risikomanagement.pdf>.
- Assmann, Jan. 2002. Das kulturelle Gedächtnis. *Erwägen, Wissen, Ethik* 194(13). 239–247. DOI: <https://doi.org/10.11588/propylaeumdok.00004040>.
- Atteslander, Peter, Jürgen Cromm, Busso Grabow, Harald Klein, Andrea Maurer & Gabriele Siegert. 2010. *Methoden der empirischen Sozialforschung*. 13. Auflage. Berlin: Erich Schmidt Verlag.
- Automatic Language Processing Advisory Committee. 1966. *Language and Machines: Computers in Translation and Linguistics*. Washington, D.C.: National Academy of Sciences.
- Baechler, Coline, Eva Martha Eckkrammer, Johannes Müller-Lancé & Verena Thaler (Hrsg.). 2016. *Medienlinguistik 3.0: Formen und Wirkung von Textsorten im Zeitalter des Social Web* (Sprachwissenschaft Band 34). Berlin: Frank & Timme.
- Bahdanau, Dzmitry, KyungHyun Cho & Yoshua Bengio. 2017. Neural machine translation by jointly learning to align and translate. In San Diego, CA, USA. <https://arxiv.org/pdf/1409.0473.pdf>.
- Baker, Mona & Gabriela Saldanha (Hrsg.). 2011. *Routledge encyclopedia of translation studies*. 2. Auflage. London: Routledge, Taylor & Francis Group.
- Balling, Laura Winther, Kristian Tangsgaard Hvelplund & Annette C. Sjørup. 2014. Evidence of parallel processing during translation. *Meta* 59(2). 234–259. DOI: [10.7202/1027474ar](https://doi.org/10.7202/1027474ar).
- Beatty, Jackson. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91(2). 276–292. DOI: [10.1037/0033-2909.91.2.276](https://doi.org/10.1037/0033-2909.91.2.276).

Literaturverzeichnis

- Beatty, Jackson & Daniel Kahneman. 1966. Pupillary changes in two memory tasks. *Psychonomic Science* 5(10). 371–372.
- Beck, Klaus. 2006. *Computervermittelte Kommunikation im Internet*. München: Oldenbourg Wissenschaftsverlag. DOI: 10.1524/9783486839203.
- Beißwenger, Michael. 2007. *Sprachhandlungskoordination in der Chat-Kommunikation* (Linguistik, Impulse & Tendenzen 26). Berlin: W. de Gruyter. DOI: 10.1515/9783110953121.
- Beißwenger, Michael. 2010. Chattern unter die Finger geschaut: Formulieren und Revidieren bei der schriftlichen Verbalisierung in synchroner internetbasierter Kommunikation. *Nähe und Distanz im Kontext variationslinguistischer Forschung* 35. 247–294.
- Berché, Mercè Pujol. 2014. Política lingüística: Lengua, cultura e identidad, el ejemplo de Cataluña. *Amnis*. DOI: 10.4000/amnis.2061.
- Bergstrom, Jennifer Romano & Andrew Jonathan Schall (Hrsg.). 2014. *Eye tracking in user experience design*. Amsterdam: Elsevier.
- Bernardini, Silvia, Pierrette Bouillon, Dragos Ciobanu, Josef van Genabith, Silvia Hansen-Schirra, Sharon O'Brien, Erich Steiner & Elke Teich. 2020. Language service provision in the 21st century: Challenges, opportunities and educational perspectives for translation studies. Bologna Process beyond 2020: Fundamental values of the EHEA. 297–303.
- Biedert, Ralf, Andreas Dengel, Georg Buscher & Arman Vartan. 2012. Reading and estimating gaze on smart phones. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA'12), 385–388. Santa Barbara, California: Association for Computing Machinery. DOI: 10.1145/2168556.2168643.
- Biedert, Ralf, Jörn Hees, Andreas Dengel & Georg Buscher. 2012. A robust real-time reading-skimming classifier. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (ETRA'12), 123–130. Santa Barbara, California: Association for Computing Machinery. DOI: 10.1145/2168556.2168575.
- Bisson, Marie-Josée, Walter J. B. Van Heuven, Kathy Conklin & Richard J. Tunney. 2014. Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics* 35(2). 399–418. DOI: 10.1017/S0142716412000434.
- Bochmann, Klaus, Jenny Brumme, Gerlinde Ebert, Jürgen Erfurt, Leipziger Forschungsgruppe, Ralf Müller & Bärbel Plötner. 2011. *Sprachpolitik in der Romania: Zur Geschichte sprachpolitischen Denkens und Handelns von der Französischen Revolution bis zur Gegenwart*. Berlin: De Gruyter. DOI: 10.1515 / 9783110859416.

- Booth, Andrew D. 1958. The history and recent progress of machine translation. In Andrew D. Booth u. a. (Hrsg.), *Aspects of translation* (Studies in Communication 2), 88–104. London: Secker & Warburg.
- Bowker, Lynne. 2002. *Computer-aided translation technology: A practical introduction* (Didactics of Translation Series). Ottawa: University of Ottawa Press.
- Bowker, Lynne & Jairo Buitrago Ciro. 2019. *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Bingley, UK: Emerald Publishing.
- Brinker, Klaus. 2001. *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, Bd. 2. Berlin; New York: De Gruyter.
- Burchardt, Aljoscha, Markus Egg, Kathrin Eichler, Brigitte Krenn, Jörn Kreutel, Annette Leßmöllmann, Georg Rehm, Manfred Stede, Hans Uszkoreit & Martin Volk. 2012. *Die Deutsche Sprache im Digitalen Zeitalter – The German Language in the Digital Age* (META-NET White Paper Series.). Wiesbaden: Springer.
- Burridge, Lorenzo. 2014. Social Media. In Jennifer Romano Bergstrom & Andrew Jonathan Schall (Hrsg.), *Eye tracking in user experience design*, 217–254. Amsterdam: Elsevier.
- Buscher, Georg, Andreas Dengel, Ralf Biedert & Ludger V. Elst. 2012. Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems* 1(2). 1–30. DOI: [10.1145/2070719.2070722](https://doi.org/10.1145/2070719.2070722).
- Bußmann, Hadumod, Gregory Trauth & Kerstin Kazzazi. 1998. *Routledge dictionary of language and linguistics*. London: Routledge.
- Canfora, Carmen & Angelika Ottmann. 2015a. Im grünen Bereich. *Technische Kommunikation* 37(5). 34–39.
- Canfora, Carmen & Angelika Ottmann. 2015b. Risikomanagement für Übersetzungen. *Trans-Kom* 2(8). 314–346.
- Carpenter, Roger H. S. 1988. *Movements of the eyes*. 2. Auflage. London, England: Pion Limited.
- Carstensen, Kai-Uwe, Christian Ebert, Cornelia Ebert, Susanne J. Jekat, Ralf Klabbunde & Hagen Langer (Hrsg.). 2010. *Computerlinguistik und Sprachtechnologie: Eine Einführung*. 3. Auflage. Heidelberg: Spektrum, Akad. Verl.
- Casesnoves, Raquel, Josep-Àngel Mas & Anna Tudela. 2019. Primary and secondary factors in language maintenance in a medium-sized community language: Catalan in Spain. *International Journal of Bilingualism* 23(2). 525–552. DOI: [10.1177/1367006917745697](https://doi.org/10.1177/1367006917745697).
- Castilho, Sheila, Stephen Doherty, Federico Gaspari & Joss Moorkens. 2018. Approaches to human and machine translation quality assessment. In Joss Moorkens, Sheila Castilho, Federico Gaspari & Stephen Doherty (Hrsg.), *Translati-*

Literaturverzeichnis

- on Quality Assessment (Machine Translation: Technologies and Applications 1), 9–38. Cham: Springer. DOI: 10.1007/978-3-319-91241-7_2.
- Castilho, Sheila & Ana Guerberof Arenas. 2018. Reading comprehension of machine translation output: What makes for a better read? In Juan Antonio Pérez-Ortiz (Hrsg.), *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, 79–88. Alacant, Spain: European Association for Machine Translation.
- Castilho, Sheila & Sharon O'Brien. 2016. Evaluating the impact of light post-editing on usability. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 310–316. Portorož, Slovenia: European Language Resources Association (ELRA). <https://aclanthology.org/L16-1048>.
- Cattelan, Alessandro & Marcello Federico. 2017. An introduction to Neural Machine Translation for linguists. Webinar. <https://www.matecat.com/webinar/introduction-neural-machine-translation-linguists/>.
- Cohen, Jacob. 1992. A power primer. *Psychological Bulletin* 112(1). 155–159. DOI: 10.1037/0033-2909.112.1.155.
- Coseriu, Eugenio. 1988. Kontrastive Linguistik und Übersetzung: Ihr Verhältnis zueinander. In Eugenio Coseriu, Jörn Albrecht, Harald Thun & Jens Lüdtke (Hrsg.), *Energeia und Ergon: Sprachliche Variation, Sprachgeschichte, Sprachtypologie* (Tübinger Beiträge zur Linguistik B300), 311–326. Tübingen: G. Narr.
- Daems, Joke & Lieve Macken. 2019. Interactive adaptive SMT versus interactive adaptive NMT: A user experience evaluation. *Machine Translation* 33(1). 117–134. DOI: 10.1007/s10590-019-09230-z.
- Dalla-Zuanna, Jean-Marc. 2010. Direkte Qualitätsmessung: Evaluierung fremdsprachiger technischer Dokumentation im Bereich Vertrieb After Sales der Marke Volkswagen Pkw. *MDÜ: Mitteilungen für Dolmetscher und Übersetzer* 2. 20–25.
- Delavenay, Émile. 1960. *An introduction to Machine Translation*. London: Thames & Hudson.
- Deloitte. 2017. Nutzung von Messaging-Diensten nach Altersgruppen 2017. *Global Mobile Consumer Survey* 2017. 3.
- Deloitte. 2019. Top 4 Messaging-Dienste nach Nutzung 2013, 2018. *Global Mobile Consumer Survey* 2019. 3.
- Doherty, Stephen & Sharon O'Brien. 2012. A user-based usability assessment of raw machine translated technical instructions. In *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas*, 1–10. San Diego, CA, USA: Association for Machine Translation in the Americas. <http://doras.dcu.ie/19470/1/AMTA-2012-Doherty-2.pdf>.

- Doherty, Stephen & Sharon O'Brien. 2014. Assessing the usability of raw machine translated output: A user-centered study using eye tracking. *International Journal of Human-Computer Interaction* 30(1). 40–51. DOI: 10.1080/10447318.2013.802199.
- Doherty, Stephen, Sharon O'Brien & Michael Carl. 2010. Eye tracking as an MT evaluation technique. *Machine Translation* 24(1). 1–13. DOI: 10.1007/s10590-010-9070-9.
- Döring, Nicola. 2003. *Sozialpsychologie des Internet: Die Bedeutung des Internet für Kommunikationsprozesse, Identitäten, soziale Beziehungen und Gruppen*. 2. Auflage (Internet und Psychologie 2). Göttingen: Hogrefe.
- Döring, Nicola. 2019. Sozialkontakte online: Identitäten, Beziehungen, Gemeinschaften. In Wolfgang Schweiger & Klaus Beck (Hrsg.), *Handbuch Online-Kommunikation*, 167–194. Wiesbaden: Springer. DOI: 10.1007/978-3-658-18016-4_8.
- Döring, Nicola. 2013. Modelle der Computervermittelten Kommunikation. In Rainer Kuhlen, Wolfgang Semar & Dietmar Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation*, 425–430. Berlin, Boston: De Gruyter. DOI: 10.1515/9783110258264.424.
- Döring, Nicola & Sandra Pöschl. 2007. Wissenskommunikation in Chats. In Kai Lehmann & Michael Schetsche (Hrsg.), *Die Google-Gesellschaft: Vom digitalen Wandel des Wissens*, 273–280. Bielefeld: Transcript. DOI: 10.25969/mediarep/11892.
- Dormann, Carsten F. 2013. *Parametrische Statistik*. Berlin: Springer. DOI: 10.1007/978-3-642-34786-3.
- Drugan, Joanna. 2013. *Quality in professional translation: Assessment and improvement* (Bloomsbury Advances in Translation). London: Bloomsbury.
- Duchowski, Andrew T. 2017. *Eye tracking methodology*. Cham: Springer. DOI: 10.1007/978-3-319-57883-5.
- Dürscheid, Christa. 2016. Neue Dialoge – alte Konzepte? *Zeitschrift für germanistische Linguistik* 44(3). DOI: 10.1515/zgl-2016-0023.
- Dürscheid, Christa. 2018. Die personale Kommunikation im Internet: (K)ein Dialog? In Emmanuel J. Bauer (Hrsg.), *Das dialogische Prinzip: Aktualität über 100 Jahre*, 29–53. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Eckkrammer, Eva Martha. 2019. Textlinguistik und Digitalität: Eine Diskussion. In Nina Janich (Hrsg.), *Textlinguistik: 15 Einführungen und eine Diskussion*, 2. Auflage (Narr Studienbücher), 341–366. Tübingen: Narr Francke Attempto.
- Ehlich, Konrad. 2011. *Sprache und sprachliches Handeln*. Berlin: De Gruyter. DOI: 10.1515/9783110922721.

Literaturverzeichnis

- Eskenazi, Michael A. & Jocelyn R. Folk. 2017. Regressions during reading: The cost depends on the cause. *Psychonomic Bulletin & Review* 24(4). 1211–1216. DOI: 10.3758/s13423-016-1200-9.
- Everdell, Ian. 2014. Web Content. In Jennifer Romano Bergstrom & Andrew Jonathan Schall (Hrsg.), *Eye tracking in user experience design*, 163–186. Amsterdam; Boston: Elsevier.
- Fantinuoli, Claudio. 2009. InterpretBank: Ein Tool zum Wissens- und Terminologienmanagement für Simultandolmetscher. In *Übersetzen in die Zukunft. Herausforderungen der Globalisierung für Dolmetscher und Übersetzer*, 411–417. Berlin: BDÜ. https://www.staff.uni-mainz.de/fantinuo/download/publications/Fantinuoli_InterpretBank.pdf.
- Fox, Wendy. 2018. *Can integrated titles improve the viewing experience?* (Translation and Multilingual Natural Language Processing 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.1180721.
- Gangl, Melanie, Kristina Moll, Manon W. Jones, Chiara Banfi, Gerd Schulte-Körne & Karin Landerl. 2018. Lexical reading in dysfluent readers of German. *Scientific Studies of Reading* 22(1). 24–40. DOI: 10.1080/10888438.2017.1339709.
- Garcia, Ignacio. 2012. A brief history of postediting and of research on post-editing. *Anglo Saxonica* 3(3). 291–312.
- Habra, Abdul. 2017. *Neural networks: An introduction*. Technology Exponent. <http://www.tek271.com/documents/others/intro-to-neural-networks>.
- Hartmann, Maren & Friedrich Krotz. 2019. Onlinekommunikation als Kultur. In Wolfgang Schweiger & Klaus Beck (Hrsg.), *Handbuch Online-Kommunikation*, 257–281. Wiesbaden: Springer. DOI: 10.1007/978-3-658-18016-4_11.
- Henisz-Dostert, Bozena, R. Ross Macdonald & Michael Zarechnak. 1979. *Machine translation* (Trends in linguistics 11). New York: Mouton.
- Hess, E. H. & J. M. Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143(3611). 1190–1192. DOI: 10.1126/science.143.3611.1190.
- Himmelein, Gerald. 2019. Die smarte Wortfabrik. *Heise Magazine* 25. 138–144. <https://www.heise.de/select/ct/2019/25/1575050419063249>.
- Holmqvist, Kenneth (Hrsg.). 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Holzki, Larissa. 2020. *Digitale Revolution: Das Ende der Sprachgrenzen – wenn der Algorithmus alles übersetzt*. <https://www.handelsblatt.com/technik/digitale-revolution/digitale-revolution-das-endе-der-sprachgrenzen-wenn-der-algorithmus-alles-uebersetzt/25480752.html>.

- Hutchins, William John. 1995. Machine Translation: A Brief History. In William John Hutchins (Hrsg.), *Concise history of the language sciences: From the Sumerians to the cognitivists*, 431–445. New York: Pergamon.
- Hutchins, William John & Harold L. Somers. 1997. *An introduction to machine translation*. 2. Auflage. London: Academic Press.
- Hvelplund, Kristian Tangsgaard. 2017. Chapter 2. Four fundamental types of reading during translation. In Arnt Lykke Jakobsen & Bartolomé Mesa-Lao (Hrsg.), *Translation in Transition: Between cognition, computing and technology*, Bd. 133, 56–78. Amsterdam: John Benjamins. DOI: 10.1075/btl.133.02hve.
- Hyönä, Jukka, Jorma Tommola & Anna-Mari Alaja. 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A* 48(3). 598–612. DOI: 10.1080/14640749508401407.
- Inhoff, Albrecht W., Andrew Kim & Ralph Radach. 2019. Regressions during Reading. *Vision* 3(3). 35. DOI: 10.3390/vision3030035.
- Inhoff, Albrecht W. & Keith Rayner. 1986. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics* 40(6). 431–439. DOI: 10.3758/BF03208203.
- Iqbal, Shamsi T., Piotr D. Adamczyk, Xianjun Sam Zheng & Brian P. Bailey. 2005. Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05) (Hrsg.). Portland, Oregon, USA: Association for Computing Machinery. DOI: 10.1145/1054972.1055016.
- Joachim Thommes. 2018. Die große Lust am Informellen: Whatsapp und Co mauern sich zu Service-Tools. *DDV dialog* September 2018. 3.
- Just, Marcel A. & Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4). 329–354. DOI: 10.1037/0033-295X.87.4.329.
- Kade, Otto. 1968. *Zufall und Gesetzmäßigkeit in der Übersetzung*. Leipzig: Verl. Enzyklopädie.
- Kahneman, Daniel & Jackson Beatty. 1966. Pupil Diameter and Load on Memory. *Science* 154(3756). 1583–1585. DOI: 10.1126/science.154.3756.1583.
- Khayrallah, Huda, Brian Thompson, Matt Post & Philipp Koehn. 2020. Simulated multiple reference training improves low-resource machine translation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. DOI: 10.18653/v1/2020.emnlp-main.7.
- Kienle, Andrea, Michael Beißwenger, Linda Cedli, Torsten Holmer, Philipp Schlieker-Steens & Christian Schlösser. 2017. Eyetracking als Ressource zur Unterstützung des Interaktionsmanagements in synchroner Schriftkommunikation.

Literaturverzeichnis

- nikation. In Michael Beißwenger (Hrsg.), *Empirische Erforschung internetbasierter Kommunikation*, 143–174. Berlin, Boston: De Gruyter. DOI: 10.1515/9783110567786-006.
- King, Margaret. 1997. Evaluating translation. In Christina Hauenschild (Hrsg.), *Machine translation and translation theory*, 251–263. New York: Mouton de Gruyter.
- Kirchhoff, Sabine, Sonja Kuhnt, Peter Lipp & Siegfried Schlawin. 2010. *Der Fragebogen: Datenbasis, Konstruktion und Auswertung*. 5. Auflage. Wiesbaden: VS Verlag.
- Klerke, Sigrid, Sheila Castilho, Maria Barrett & Anders Søgaard. 2015. Reading metrics for estimating task efficiency with MT output. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, 6–13. Lisbon, Portugal: Association for Computational Linguistics. DOI: <http://dx.doi.org/10.18653/v1/W15-2402>.
- Koch, Peter & Wulf Oesterreicher. 1994. Schriftlichkeit und Sprache. In Hartmut Günther & Otto Ludwig (Hrsg.), *Schrift und Schriftlichkeit*, Bd. 1, 587–604. Berlin: De Gruyter.
- Koch, Peter & Wulf Oesterreicher. 2011. *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. 2. Auflage (Romanistische Arbeitshefte 31). Berlin: De Gruyter.
- Koehn, Philipp. 2009. *Statistical Machine Translation*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511815829.
- Koehn, Philipp. 2020. *Neural Machine Translation*. Cambridge: Cambridge University Press. DOI: 10.1017/9781108608480.
- Krejtz, Krzysztof, Andrew T. Duchowski, Anna Niedzielska, Cezary Biele & Izabela Krejtz. 2018. Eye tracking cognitive load using pupil diameter and micro-saccades with fixed gaze. *PLOS ONE* 13(9). e0203629. DOI: 10.1371/journal.pone.0203629.
- Krenz, Michael, Markus Ramlow & Uta Seewald-Heeg (Hrsg.). 2008. *Maschinelle Übersetzung und XML im Übersetzungsprozess: Prozesse der Translation und Lokalisierung im Wandel* (TransÜD 19). Berlin: Frank & Timme.
- Kriesel, David. 2005. *Ein kleiner Überblick über Neuronale Netze*. http://www.dkriesel.com/_media/science/neuronalenetze-de-zeta2-2col-dkrieselcom.pdf.
- Kruger, Jan-Louis, Stephen Doherty, Wendy Fox & Peter de Lissa. 2018. Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign-language viewers. In Isabel Lacruz & Riitta Jääskeläinen (Hrsg.), *American Translators Association Scholarly Monograph Series*, Bd. 18, 267–294. Amsterdam: John Benjamins. DOI: 10.1075/ata.18.12kru.

- Krüger, Ralph. 2017. Von Netzen und Vektoren: Neuronale maschinelle Übersetzung. *MDÜ: Mitteilungen für Dolmetscher und Übersetzer* 1. 38–44.
- Laisiepen, Klaus, Ernst Lutterbeck, Karl-Heinrich Meyer-Uhlenried, Marianne Buder, Werner Rehfeld, Thomas Seeger & Dietmar Strauch (Hrsg.). 1996. *Grundlagen der praktischen Information und Dokumentation: Ein Handbuch zur Einführung in die fachliche Informationsarbeit*. Berlin, Boston: De Gruyter. DOI: 10.1515/9783110961355.
- Läubli, Samuel, Rico Sennrich & Martin Volk. 2018. Has machine translation achieved human parity? A case for document-level evaluation. In *Proceedings of the 2018 conference on empirical methods in Natural Language Processing*, 4791–4796. Brussels, Belgium: Association for Computational Linguistics. DOI: 10.18653/v1/D18-1512.
- Le Meur, Olivier & Zhi Liu. 2015. Saccadic model of eye movements for free-viewing condition. *Vision Research* 116. 152–164. DOI: 10.1016/j.visres.2014.12.026.
- Lehner, Franz. 2014. *Wissensmanagement: Grundlagen, Methoden und technische Unterstützung*. 5. Auflage. München: Hanser.
- Ljubešić, Nikola & Antonio Toral. 2014. *caWaC: A web corpus of Catalan and its application to language modeling and machine translation*. Dept. of Information & Communication Sciences, University of Zagreb Ivana Lučića 3, HR-10000 Zagreb, Croatia. <http://www.mt-archive.info/10/LREC-2014-Ljubesic.pdf>.
- Locke, William N. 1956. Translation by machine. *Scientific American* 194(1). 29–33. <http://www.jstor.org/stable/24943825>.
- Lommel, Arle, Aljoscha Burchardt, Maja Popović, Kim Harris, Eleftherios Avramidis & Hans Uszkoreit. 2014. Using a new analytic measure for the annotation and analysis of MT errors on real data. In *Proceedings of the 17th Annual conference of the European Association for Machine Translation*, 165–172. Dubrovnik, Croatia: European Association for Machine Translation. <https://aclanthology.org/2014.eamt-1.38>.
- Lommel, Arle, Hans Uszkoreit & Aljoscha Burchardt. 2014. *Multidimensional Quality Metrics (MQM) Definition*. Arle Lommel, Attila Görög, Alan Melby, Hans Uszkoreit, Aljoscha Burchardt, Leonid Glazychev, Serge Gladkoff & Kim Harris (Hrsg.). <http://www.qt21.eu/mqm-definition/>.
- Lykke Jakobsen, Arnt & Kristian Tangsgaard Hvelplund Jensen. 2008. Eye movement behaviour across four different types of reading task. *Copenhagen Studies in Language* 36. 103–124.
- Macken, Lieve, Daniel Prou & Arda Tezcan. 2020. Quantifying the effect of machine translation in a high-quality human translation production process. *Informatics* 7(2). DOI: 10.3390/informatics7020012.

Literaturverzeichnis

- Martín Martín, José Miguel. 2010. Sobre la evaluación de traducciones en el ámbito académico. *Revista española de lingüística aplicada* 23. 229–246.
- Massardo, Isabella, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri & Katrin Drescher. 2016. *MT Post-Editing Guidelines*. Richtlinie. Amsterdam.
- McCaffrey, James. 2012. Test Run - Neural Network Back: Propagation for Programmers. *MSDN Magazine* 12(10). <https://msdn.microsoft.com/en-us/magazine/jj658979.aspx>.
- Megill, Kenneth A. 2005. *Thinking for a living: The coming age of knowledge work*. Berlin; New York: De Gruyter. DOI: 10.1515/9783598440090.
- Messerer, Thomas & Beate Eickhoff. 2013. *Einsatz von Skype im Unternehmen: Chancen, Risiken und Policy-Empfehlungen*. Empfehlung. München: Fraunhofer ESK. 24. <http://publica.fraunhofer.de/documents/N-374881.html>.
- Misoch, Sabina. 2006. *Online-Kommunikation*. Konstanz: UVK Verlag.
- Moorkens, Joss, Sheila Castilho, Federico Gaspari & Stephen Doherty (Hrsg.). 2018. *Translation quality assessment: From principles to practice*, Bd. 1 (Machine Translation: Technologies and Applications). Cham: Springer. DOI: 10.1007/978-3-319-91241-7.
- Moreno, Asunción, Núria Bel, Eva Revilla, Emilia Garcia, Sisco Vallverdú, Georg Rehm & Hans Uszkoreit. 2012. *La llengua catalana a l'era digital – The Catalan language in the digital age* (META-NET White Paper Series.). Wiesbaden: Springer.
- Nikolova, Mirela, Stephanie Jainta, Hazel I. Blythe & Simon P. Liversedge. 2018. Binocular advantages for parafoveal processing in reading. *Vision Research* 145. 56–63. DOI: 10.1016/j.visres.2018.02.005.
- O'Brien, Sharon. 2009. Eye tracking in translation process research: Methodological challenges and solutions. *Methodology, technology and innovation in translation process research* 38. 251–266.
- O'Brien, Sharon, Laura Winther Balling, Michael Carl, Michel Simard & Lucia Specia (Hrsg.). 2014. *Post-editing of machine translation: Processes and applications*. Newcastle upon Tyne: Cambridge Scholars Press.
- Ott, Myle, Michael Auli, David Grangier & Marc'Aurelio Ranzato. 2018. Analyzing uncertainty in neural machine translation. In Jennifer Dy & Andreas Krause (Hrsg.), *Proceedings of the 35th International Conference on Machine Learning*, Bd. 80 (Proceedings of Machine Learning Research), 3956–3965. Stockholm, Sweden: PMLR. <http://proceedings.mlr.press/v80/ott18a.html>.

- Pappert, Steffen. 2017. Zu kommunikativen Funktionen von Emojis in der WhatsApp-Kommunikation. In Michael Beißwenger (Hrsg.), *Empirische Erforschung internetbasierter Kommunikation*, 175–212. Berlin: De Gruyter. DOI: 10.1515/9783110567786-007.
- Popović, Maja. 2018. Language-related issues for NMT and PBMT for English-German and English-Serbian. *Machine Translation* 32(3). 237–253. DOI: 10.1007/s10590-018-9219-5.
- Porsiel, Jörg (Hrsg.). 2017. *Maschinelle Übersetzung: Grundlagen für den professionellen Einsatz*. Berlin: Bundesverband der Dolmetscher und Übersetzer e.V. (BDÜ).
- Pym, Anthony. 1992. Translation error analysis and the interface with language teaching. In Cay Dollerup & Anne Loddegaard (Hrsg.), *The teaching of translation*, 279–288. Amsterdam: John Benjamins.
- Ramlow, Markus. 2009. *Die maschinelle Simulierbarkeit des Humanübersetzens: Evaluation von Mensch-Maschine-Interaktion und der Translatqualität der Technik* (TransÜD 27). Berlin: Frank & Timme.
- Raschke, Michael, Tanja Blascheck & Michael Burch. 2014. Visual analysis of eye tracking data. In Weidong Huang (Hrsg.), *Handbook of Human Centric Visualization*, 391–409. New York, NY: Springer. DOI: 10.1007/978-1-4614-7485-2_15.
- Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3). 372–422. DOI: 10.1037/0033-2909.124.3.372.
- Rayner, Keith, Alexander Pollatsek & Erik D. Reichle. 2003. Eye movements in reading: Models and data. *Behavioral and Brain Sciences* 26(4). 507–518. DOI: 10.1017/S0140525X03520106.
- Roser, Andreas. 2018. Warum sprechen Menschen mit Maschinen? *Information: Wissenschaft & Praxis* 69(5). 249–256. DOI: 10.1515/iwp-2018-0041.
- Sajjad, Hassan, Francisco Guzmán, Nadir Durrani, Ahmed Abdelali, Houda Bouamor, Irina Temnikova & Stephan Vogel. 2016. Eyes don't lie: Predicting machine translation quality using eye movement. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1082–1088. San Diego, California: Association for Computational Linguistics. DOI: 10.18653/v1/N16-1125.
- Schaeffer, Moritz, Kevin B. Paterson, Victoria A. McGowan, Sarah J. White & Kirsten Malmkjær. 2017. Reading for translation. In Arnt Lykke Jakobsen & Bartolomé Mesa-Lao (Hrsg.), *Translation in transition: Between cognition, computing and technology* (133), 17–53. Amsterdam: John Benjamins. DOI: 10.1075/btl.133.

Literaturverzeichnis

- Schlösser, Christian, Benedikt Schröder, Linda Cedli & Andrea Kienle. 2018. Beyond gaze cursor: Exploring information-based gaze sharing in chat. In *Proceedings of the Workshop on Communication by Gaze Interaction* (COGAIN'18). New York, NY, USA: Association for Computing Machinery. DOI: 10.1145/3206343.3206345.
- Schmitt, Peter A. & Heike Elisabeth Jüngst (Hrsg.). 2007. *Translationsqualität* (Leipziger Studien zur angewandten Linguistik und Translatologie 5). Frankfurt: P. Lang.
- Schopflin, Katharine & Matt Walsh. 2018. *Practical Knowledge and Information Management*. London: Facet. DOI: 10.29085/9781783303373.
- Schweiger, Wolfgang & Klaus Beck. 2019. *Handbuch Online-Kommunikation*. Cham: Springer. DOI: 10.1007/978-3-658-18016-4.
- Seewald-Heeg, Uta. 2005. Der Einsatz von Translation-Memory-Systemen am Übersetzerarbeitsplatz. *MDÜ: Mitteilungen für Dolmetscher und Übersetzer* 52(4). 8–38.
- Seidel, Christian. 2005. *Matrizen und Vektoren in der Computerlinguistik*. http://www.cis.uni-muenchen.de/~micha/kurse/mathel-WS0607/begleitmaterial/matrizen_und_vektoren.pdf.
- Spinuzzi, Clay. 2017. Analyse computervermittelter Kommunikation in Arbeitsumgebungen: Ein tätigkeitstheoretischer Ansatz. In Michael Beißwenger (Hrsg.), *Empirische Erforschung internetbasierter Kommunikation*, 213–238. Berlin: De Gruyter. DOI: 10.1515/9783110567786-008.
- SR Research Ltd. 2019. *EyeLink Data Viewer*. Version 4.1.1. Mississauga, Ontario, Canada.
- Stapelkamp, Torsten. 2007. *Screen- und Interfacedesign: Gestaltung und Usability für Hard- und Software*. Berlin: Springer.
- Statista. 2017. *Facebook Messenger – Anzahl der monatlich aktiven Nutzer 2017*. Statista. <https://de.statista.com/statistik/daten/studie/419453/umfrage/anzahl-der-monatlich-aktiven-nutzer-des-facebook-messengers-weltweit/>.
- Stein, Daniel. 2009. Maschinelle Übersetzung: Ein Überblick. *Maschinelle Übersetzung: Von der Theorie zur Anwendung* 24(3). 5–18. <http://www.gscl.info/>.
- Stein, Daniel. 2013. Machine Translation: Past, Present and Future. *Translation: Computation, Corpora, Cognition*. 3(1). 5–13. <http://www.blogs.uni-mainz.de/fb06-tc3/files/2015/11/28-143-1-PB.pdf>.
- Storrer, Angelika. 2001a. Getippte Gespräche oder dialogische Texte? Zur kommunikationstheoretischen Einordnung der Chat-Kommunikation. In Andrea Lehr, Matthias Kammerer, Klaus-Peter Konderding, Angelika Storrer, Caja Thimm & Werner Wolski (Hrsg.), *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik*. 439–465. Berlin: De Gruyter.

- Storrer, Angelika. 2001b. Sprachliche Besonderheiten getippter Gespräche: Sprecherwechsel und sprachliches Zeigen in der Chat-Kommunikation. In Michael Beißwenger (Hrsg.), *Chat-Kommunikation. Spache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld*. 3–24. Stuttgart: Ibidem.
- Streitz, Norbert A. 1990. Hypertext: Ein innovatives Medium zur Kommunikation von Wissen. In Peter A. Gloor & Norbert A. Streitz (Hrsg.). Bearb. von W. Brauer, *Hypertext und Hypermedia*, Bd. 249, 10–27. Berlin: Springer. DOI: [10.1007/978-3-642-84282-5_2](https://doi.org/10.1007/978-3-642-84282-5_2).
- Tesseur, Wine. 2017. Incorporating translation into sociolinguistic research: Translation policy in an international non-governmental organisation. *Journal of Sociolinguistics* 21(5). 629–649. DOI: [10.1111/josl.12245](https://doi.org/10.1111/josl.12245).
- Translation Automation User Society (TAUS). 2017a. *Category: Evaluation type - knowledgebase*. https://www.taus.net/knowledgebase/index.php?title=Category:Evaluation_type.
- Translation Automation User Society (TAUS). 2017b. *Category: Metrics - knowledgebase*. <https://www.taus.net/knowledgebase/index.php?title=Category:Metrics>.
- Trepte, Sabine & Leonard Reinecke. 2012. *Medienpsychologie*. Stuttgart: Kohlhammer Verlag.
- Vardaro, Jennifer, Moritz Schaeffer & Silvia Hansen-Schirra. 2019. Translation quality and error recognition in professional neural machine translation post-editing. *Informatics* 6(3). 41. DOI: [10.3390/informatics6030041](https://doi.org/10.3390/informatics6030041).
- Venuti, Lawrence. 1986. The translator's invisibility. *Criticism* 28(2). 179–212. <http://www.jstor.org/stable/23110425>.
- Verheijen, Lieke. 2017. WhatsApp with social media slang? Youth language use in Dutch written computer-mediated communication. In Darja Fišer & Michael Beißwenger (Hrsg.), *Investigating computer-mediated communication: Corpus-based approaches to language in the digital world* (Book series translation studies and applied Linguistics), 72–101. Ljubljana: Ljubljana University Press.
- Wahlster, Wolfgang, Saul Amarel, Alan W. Biermann, Leonard Bolc, John P. Hayes, Ameet Joshi, Douglas Lenat, Donald W. Loveland, Alan Mackworth, Dana Nau, Raymond Reiter, Erik Sandewall, Steven Shafer, Yoav Shoham & Jörg Siekmann (Hrsg.). 2000. *Verbmobilität: Foundations of speech-to-speech translation* (Artificial Intelligence). Berlin: Springer. DOI: [10.1007/978-3-662-04230-4](https://doi.org/10.1007/978-3-662-04230-4).
- Walther, Joseph B., Yuhua Jake Liang, David C. DeAndrea, Stephanie Tom Tong, Caleb T. Carr, Erin L. Spottswood & Yair Amichai-Hamburger. 2011. The effect

Literaturverzeichnis

- of feedback on identity shift in computer-mediated communication. *Media Psychology* 14(1). 1–26. DOI: 10.1080/15213269.2010.547832.
- Weingarten, Rüdiger. 2001. Voraussetzungen und Formen technisch realisierter Kommunikation. In *Text- und Gesprächslinguistik: Ein internationales Handbuch zeitgenössischer Forschung*, Bd. 2, 1132–1140. Berlin: De Gruyter.
- Wilks, Yorick. 2009. *Machine translation: Its scope and limits*. New York: Springer.
- Williams, Malcolm. 2009. Translation quality assessment: A question for the holy grail. *Mutatis Mutandis* 2(1). 3–23.
- Zech, Rainer. 2015. *Qualitätsmanagement und gute Arbeit*. Wiesbaden: Springer. DOI: 10.1007/978-3-658-07504-0.
- Zimmermann, Harald H. 2012. Maschinelle Übersetzung. In Klaus Laisiepen, Ernst Lutterbeck, Karl-Heinrich Meyer-Uhlenried, Marianne Buder, Werner Rehfeld, Thomas Seeger & Dietmar Strauch (Hrsg.), *Grundlagen der praktischen Information und Dokumentation: Ein Handbuch zur Einführung in die fachliche Informationsarbeit*, 244–254. Berlin: De Gruyter. DOI: doi:10.1515 / 9783110961355.244.

Autorenregister

- Aeppli, Jürg, 76–78, 86, 93
Albert, Ruth, 76, 93
Altena, Andreas, 62
Artsrouni, George, 38
Assmann, Jan, 28
Atteslander, Peter, 77, 78, 86
Automatic Language Processing Advisory Committee, 40
- Baechler, Coline, 8
Bahdanau, Dzmitry, 50, 51
Baker, Mona, 60, 67
Balling, Laura Winther, 89
Bar-Hillel, Yehoshua, 39, 40
Beatty, Jackson, 81, 221, 228
Becher, Johann Joachim, 37
Beck, Klaus, 8, 10, 11, 16–18, 25, 26
Beißwenger, Michael, 8, 11–14, 17, 24, 25
Berché, Mercè Pujol, 33
Bergstrom, Jennifer Romano, 78, 79
Bernardini, Silvia, 30, 32
Biedert, Ralf, 224
Bisson, Marie-Josée, 26, 79, 122
Bochmann, Klaus, 4
Booth, Andrew D., 38, 39
Bowker, Lynne, 41, 42, 52, 53
Brinker, Klaus, 25
Brown, Peter F., 41
Burchardt, Aljoscha, 28, 31, 32, 41, 66
Burridge, Lorenzo, 208, 209
Buscher, Georg, 224
- Bußmann, Hadumod, 25
Canfora, Carmen, 62, 65
Carpenter, Patricia A., 79, 220
Carpenter, Roger H. S., 159
Carstensen, Kai-Uwe, 44, 45, 53, 64, 66
Casesnoves, Raquel, 33
Castilho, Sheila, 61, 67, 69, 84
Cattelan, Alessandro, 48, 51
Ciro, Jairo Buitrago, 41, 42, 52
Cohen, Jacob, 127, 132, 141, 145, 148, 152, 156, 161, 163, 175, 176, 179, 185, 186, 188, 189, 191, 193, 196
Coseriu, Eugenio, 24, 25
- Daems, Joke, 69
Dalla-Zuanna, Jean-Marc, 63, 64
Delavenay, Émile, 39
Deloitte, 104, 105
Dengel, Andreas, 224
Doherty, Stephen, 69, 70, 80, 83, 84, 212, 223
Döring, Nicola, 8, 16, 18–21, 29, 30
Dormann, Carsten F., 211, 212
Drugan, Joanna, 64
Duchowski, Andrew T., 80, 93
Dürscheid, Christa, 8, 10, 11, 13
- Eckkrammer, Eva Martha, 28
Ehlich, Konrad, 24

Autorenregister

- Eickhoff, Beate, 56
Eskenazi, Michael A., 80
Everdell, Ian, 208, 209

Fantinuoli, Claudio, 54
Federico, Marcello, 48, 51
Folk, Jocelyn R., 80
Fox, Wendy, 26

Gangl, Melanie, 227
Garcia, Ignacio, 68
Guerberof Arenas, Ana, 67, 69, 84

Habra, Abdul, 49
Hartmann, Maren, 13, 26
Hees, Jörn, 224
Henisz-Dostert, Bozena, 40
Hess, E. H., 81, 221
Himmelein, Gerald, 52
Holmqvist, Kenneth, 79–82, 159, 211, 212, 217, 227
Holzki, Larissa, 52, 207
Hutchins, William John, 37, 38, 40–43, 53
Hvelplund, Kristian Tangsgaard, 93, 211, 212, 219
Hyönä, Jukka, 81, 153, 164, 222

Inhoff, Albrecht W., 213, 215
Iqbal, Shamsi T., 80, 221

Jensen, Kristian Tangsgaard Hvelplund, 25
Joachim Thommes, 105
Jüngst, Heike Elisabeth, 62
Just, Marcel A., 79, 220

Kade, Otto, 59
Kahneman, Daniel, 81
Khayrallah, Huda, 66

Kienle, Andrea, 8, 10, 12–14, 164
King, Margaret, 64
Kirchhoff, Sabine, 77
Klerke, Sigrid, 84
Koch, Peter, 8–11, 24, 29
Koehn, Philipp, 39, 41, 43, 45, 46, 51, 52, 54, 64, 66
Krejtz, Krzysztof, 81, 221, 223
Krenz, Michael, 41, 42
Kriesel, David, 47, 50
Krotz, Friedrich, 13, 26
Kruger, Jan-Louis, 26
Krüger, Ralph, 46–51

Laisiepen, Klaus, 43
Läubli, Samuel, 66
Le Meur, Olivier, 81
Lehner, Franz, 27, 29
Leibniz, Gottfried W., 37
Liu, Zhi, 81
Ljubešić, Nikola, 34
Llull, Ramon, 33, 37
Locke, William N., 38, 39
Lommel, Arle, 66, 67
Lykke Jakobsen, Arnt, 25

Macken, Lieve, 67, 69
Martín Martín, José Miguel, 61, 63, 64
Marx, Nicole, 76, 93
Massardo, Isabella, 68
McCaffrey, James, 49
Megill, Kenneth A., 27
Messerer, Thomas, 56
Misoch, Sabina, 8, 18–20, 22
Moorkens, Joss, 61
Moreno, Asunción, 33–35
Moritz, Frank, 62

Nikolova, Mirela, 227

- O'Brien, Sharon, 67, 69, 82–84, 88, 91, 231
Oesterreicher, Wulf, 8–11, 24, 29
Ott, Myle, 66
Ottmann, Angelika, 62, 65
- Pappert, Steffen, 17, 228
Polt, J. M., 81, 221
Popović, Maja, 66
Porsiel, Jörg, 43
Pöschl, Sandra, 29, 30
Pym, Anthony, 65
- Ramlow, Markus, 43
Raschke, Michael, 79, 93
Rayner, Keith, 25, 79, 212, 213, 224, 227
Reinecke, Leonard, 7, 10, 17–19, 21, 22
Roser, Andreas, 232
- Sajjad, Hassan, 84
Saldanha, Gabriela, 60, 67
Schaeffer, Moritz, 25, 80, 92, 93, 166, 210
Schall, Andrew Jonathan, 78, 79
Schlösser, Christian, 122, 152
Schmitt, Peter A., 62
Schopflin, Katharine, 27
Schweiger, Wolfgang, 10
Seewald-Heeg, Uta, 53
Seidel, Christian, 50
Smirnov-Troyanskii, Petr Petrovich, 38
Somers, Harold L., 43
Spinuzzi, Clay, 228, 233
SR Research Ltd., 94, 224
Stapelkamp, Torsten, 69, 70
Statista, 55
Stein, Daniel, 37, 38, 40–46
- Storrer, Angelika, 8, 11–15
Streitz, Norbert A., 27, 28
- Tesseur, Wine, 31
Toral, Antonio, 34
Translation Automation User Society (TAUS), 66
Trepte, Sabine, 7, 10, 17–19, 21, 22
Turing, Alan, 38
- Uszkoreit, Hans, 67
- Vardaro, Jennifer, 25, 85
Venuti, Lawrence, 229
Verheijen, Lieke, 228
- Wahlster, Wolfgang, 17, 42
Walsh, Matt, 27
Walther, Joseph B., 22
Weaver, Warren, 38, 39
Weingarten, Rüdiger, 16–18
Wilks, Yorick, 43
Williams, Malcolm, 64, 65
- Yngve, Victor, 39
- Zech, Rainer, 60
Zimmermann, Harald H., 53, 54

Sachregister

- AIC, *siehe* Akaikes Informationskriterium
Aktualität, 28
Alexa, *siehe* Amazon
Alignierung, 45
ALPAC
Automatic Language Processing Advisory Committee, 40
Amazon, 55
 Alexa, 55
Anderson-Darling-Test, *siehe* Statistik, 159, 161, 197, 200
Annotationsreihenfolge, 213
Anonymität, 22, 30
ANOVA, *siehe* Varianzanalyse
AOI-Label, *siehe* Area of Interest
Apertium, 34
Appear.in, 87
Apple, 55
 Facetime, 87
 Siri, 55
Apple Facetime, 87, 170, 206
Area of Interest, 82
 -Label, 95
 Anteil an, 94
 Anzahl an, 94
 Anzahl an Fixationen pro, 94
 dynamisches, 82, 90, 173
 Größe des, 95, 128, 132, 137, 189, 191, 193, 196, 211, 216, 217, 219
Kategorie, 126, 133, 137
statisches, 82, 90, 190, 203, 218
Verweildauer im, 94
AS-TRANSAC, 42
ATLAS, 42
Autonome Gemeinschaft Katalanen, 32, 33
Bi-Lingual Evaluation Understudy-Score, *siehe* BLEU
Bing Translate, *siehe* Microsoft
binocular recording, *siehe* Eye-Tracking
BLEU, 66
blink, *siehe* Blinzler
Blinzler, 82, 90, 91, 125, 171, 208
CAT-Tool, 25
Chat, 92
 -anwendung, 7, 10, 12, 208
 -beitrag, 12–14, 23, 75, 76, 82, 90, 173, 178, 179, 182, 183, 193, 215, 216
 -kommunikation, 10, 11, 13, 18, 23, 25, 108, 173
 -nachricht, 22, 82
 -partner·in, 218
 -rooms, 10
 -technologie, 19
 -verlauf, 12, 14¹
moderierter, 23

Sachregister

- Text-, 8, 10, 11, 26, 82, 87, 91, 92, 104, 108–111, 170
Video-, 8, 87, 104, 109–111, 170, 206
Voice-, 8, 87, 104, 109–111, 170
- Chat-
beitrag, 221
Interaktion, 13
kommunikation, 13
- Chatnachricht, *siehe* Chatbeitrag
Chatprotokoll, *siehe* Chatverlauf
Chi-Quadrat-Test, *siehe* Statistik
Cohen
Effektstärke nach, 163, 175, 176, 179, 185, 186, 188, 189, 191, 193, 196
- Cornea, 78
Creative Commons, 28
CULT, 41
Cyberspace, 26
- De-Individualisierung, 21
Debriefing, 106
DeepL, 35
Deutsches Forschungszentrum für Künstliche Intelligenz, 35, 67
DFKI, *siehe* Deutsches Forschungszentrum für Künstliche Intelligenz
Dialog, 12, 26
Dialogizität, 9
Digitalisierungs-Modell, 19
Diskurs, 11
Drift, *siehe* Fixation
Dunn-Benjamini-Hochberg-Test,
siehe Statistik, 141, 163, 175, 177, 178, 185, 188, 192, 193, 195, 199, 201
- Durchlauf
-dauer, 165, 188
regressive, 80, 94, 137, 146–148, 189–191
selektive regressive, 95, 148–152, 191–193
Dauer des ersten, 137, 139–141, 183–186, 216–218
- Echtzeit-Übersetzung, 207
-ssystem, 207
- Effekt
gemischter, 133, 136, 179, 183
Zufalls-, 133, 137, 178, 180, 183
- Effektstärke nach Cohen, 127, 132, *siehe* Statistik, 156
- ENIGMA, 38
eTranslation, 35
EU Council Presidency Translator, 35
EUROTRA, 41
Evaluation, 84
Eye-Mind-Hypothese, 79
Eye-Tracking, 78–85, 92, 106, 111
-Indikatoren, 93
-Labor, 88
-Modus, 88
- Facebook
Instagram, 3, 87
Messenger, 10, 54, 55, 87, 104, 105
- Facebook Messenger, 87, *siehe* Facebook, 206
- Fachbereich Sprache, Kultur und Translation, 166
- Filter-Modell, 19
first fixation progressive, 141

- first run dwell time, *siehe* Verweildauer im ersten Durchlauf, 183
- Fixation, 79–81, 90, 91, 93, 171, 208
-sanzahl, 79, 93, 94, 126–128, 173–176, 210–211
-sdauer, 79, 93, 129, 189
-srate, 93
-ssequenz, 93
Anzahl der, 165, 166
- Dauer der, 93
- Dauer der ersten, 79, 94, 129–132, 176–179, 211–213
- Drift, 79
- Durchlaufdauer, 80
- erstmalige, 218
- progressive erste, 95, 127, 133, 137, 141, 145, 154, 156, 175, 179, 180, 182, 183, 185, 189, 190, 195, 212, 217, 220, 222
- Pupillengröße, 80, 94, 95
- Tremor, 79
- Formen der Wissensrepräsentation, *siehe* Wissen
- Fragebogen, 76–78, 86, 92, 106
Ausgangs-, 97, 111, 166, 202
- Eingangs, 169
- Eingangs-, 97, 108, 166, 169, 202, 208
- FTSK, *siehe* Fachbereich Sprache, Kultur und Translation
- Fujitsu, 42
- gaze path, 208
- gaze point, 79
- General Text Matcher, 66
- Generalitat de Catalunya, 34
- Georgetown-Experiment
- Maschinelle Übersetzung, 39
- Google, 46, 55, 207
Google Translate, 34
- Hangouts, 54, 87
- Google Hangouts, *siehe* Google, 87
- Gruppenzugehörigkeit, 21
- Heatmap, 90, 91, 173
- hyperpersonale Kommunikation
- Modell der, 21–22
- IALT, *siehe* Institut für Angewandte Linguistik und Translatologie
- IBM, 41
- ICQ, 87
- Idealisierung, 22
-sprozess, 22
- Identität
- Gruppen-, 22
- personale, 21
- Persönlichkeit
-sstereotypen, 22
- Selbstbild, 22
- soziale, 21
- individuumsbzogenes Kommunikationsmodell, *siehe* Kommunikation
- Information
-smanagement, 27
- Inspektion
- graphische, 129, 176, 185, 186, 190, 191, 193
- statistische, 129, 176, 212
- visuelle, 203, 208
- Instagram, *siehe* Facebook, 87
- Institut für Angewandte Linguistik und Translatologie, 166
- Interaktion, 13, 20, 25, 31
Chat-, 13

Sachregister

- ergebnisorientierte, 12
- Internet Relay Chat, 7
- interpersonale Medienwahl, 20
- Johannes Gutenberg-Universität Mainz, 166
- kanalbezogenes Kommunikationsmodell, *siehe* Kommunikation
- Kanalreduktionsmodell, 18–20
- Key Logging, 89
- kollaborative Projekte, 28
- Kommunikation, 42, 169
 - sbedingungen, 11
 - sbedinungen, 9
 - smedium, 19
 - smodell
 - Digitalisierungs-Modell, 19
 - Filter-Modell, 19
 - individuumsbezogenes, 18
 - interpersonale Medienwahl, 20
 - kanalbezogenes, 18
 - Kanalreduktionsmodell, 18
 - medienbezogenes, 18
 - Modell der hyperpersonalen Kommunikation, 21–22
 - Modell der medialen Reichhaltigkeit, 19
 - Modell der rationalen Medienvwahl, 19
 - normative Medienwahl, 20
 - SIDE-Modell, 21
 - Theorie der sozialen Informationsverarbeitung, 20
 - smodus, 170
 - spartner-in, 109
 - splattform, 7
- ssituation, 21, 75, 76, 106, 111, 112, 170, 173, 202, 216
- stechnologie, 205
- sthema, 9
- sverhalten, 18, 21, 22, 103, 104
- Aggression in der, 19
- audio-visuelle, 8
- Chat-, 10–13, 18, 23, 25, 108, 173
- computervermittelte, 7, 10, 18, 20, 22, 31
- Distanz-, 9–14, 16
- Enthemmung in der, 19
- mündliche, 14
- Netz-, 25
- Nähe, 12
- Nähe-, 9–11, 13
- quasi-synchrone, 13, 22
- simultane, 13
- spontane, 9
- synchrone, 30
- Telefon, 11
- textbasierte, 8, 11
- Video-, 206
- von Angesicht zu Angesicht, 11, 13, 18, 20
- Wissens-, 30
- Kommunikationstechnologie
- Zoom, 3
- Kooperation
- intensive, 9
- Korrelationstest nach Spearman, 128, 132, *siehe* Statistik, 145, 148, 152, 156, 159, 176, 179, 186, 191, 193, 196, 197
- Kruskal-Wallis-Test, *siehe* Statistik, 163, 175, 177, 178, 185, 188, 190, 191, 193, 195, 199, 201

- Kurtosis, 129, 177
Künstliche Intelligenz, 42
- Libre Office, 28
Linux, 28
Logos, 41
- Mann-Whitney-U-Test, *siehe* Statistik, 175, 179, 185, 186, 189, 190, 192, 196
maschinelle Übersetzung, 84, 92, 126, 213, *siehe auch* Übersetzung Evaluation, 212
neuronale, 46–48
Qualität der, 164
regelbasierte, 41–45
statistische, 41, 45
- Massenmedium, 23
media richness theory, *siehe* mediale Reichhaltigkeit
mediale Reichhaltigkeit Modell der, 19
- medienbezogenes Kommunikationsmodell, *siehe* Kommunikation
- Medium Wahl des, 20
- META NET, 31
- METAL, 41
- METEOR, 66
- Microsoft, 28, 207, 232
Bing Translate, 34
Conversations, 59
Skype, 3
Windows Live Messenger, 55
- Mitsubishi, *siehe* MELTRAN
- MSN, *siehe* Windows
- Multidimensionale Qualitätsmetrik, 66
MÉTÉO, 41
Mündlichkeit, 8, 11, 24
- Nachricht eingehende, 171
- Netzkommunikation, 25
- Netzwerk künstliches neuronales, 47
- nicht-parametrischer Test, *siehe* Statistik, 191
- NIST, *siehe* US National Institute of Standard and Technology
- non-verbale Elemente, 18
- normative Medienwahl, 20
- Nutzerfreundlichkeit, 78, 81
- Ominium Cultural, 33
- Parallelkorpus, *siehe* Korpus
paraverbale Elemente, 18
- Persönlichkeit -sstereotypen, *siehe* Identität
phrasenbasierter Ansatz, *siehe* maschinelle Übersetzung
- Pilotierung, 75, 105
- Plausibilität, 182, 214
- Post-Editing, 25
- Post-Studien-Interview, 84
- Postediting, 67
- probabilistischer Ansatz, *siehe* maschinelle Übersetzung
- Progression, 84
- Pupillengröße, 80–81, 152–156, 165, 166, 193–196, 221–223
- Qualität, 164
- Qualitätsbewertung, *siehe auch* Qualitätsmanagement

Sachregister

- rationale Medienwahl, 19
Raum
 digitaler, 26
 Informations-, 26
Rederecht, 14
reduced social cues approach, *siehe*
 Filter-Modell
Regionalstatut Kataloniens, 33
Regression, 79, 80, 84, 93, 94
 ausgehende, 80, 94, 136–137,
 182–183, 213–216
 eingehende, 80, 94, 133–136,
 179–182, 213–216
 nicht-zurückkehrende, 80
 Summe aller, 133, 136
 Summe der, 165, 182
 zurückkehrende, 80
Regressionsmodell, *siehe* Statistik
 binominales logistisches, 133,
 136, 179, 183, 214
 lineares, 133, 137, 180
regressive Durchlaufdauer, *siehe*
 Durchlauf, 189
Reliabilität, 28
Restriktionsmodell, *siehe* Kanalre-
 duktionsmodell
Risikomanagement, 61
Rollenorganisation, *siehe* Spre-
 cher.innenorganisation

Sakkade, 79, 81, 84, 90, 91, 94, 208
 -namplitude, 95, 159–161, 165,
 197–200, 227
 -ndauer, 95, 161–165, 200–201,
 227
 Anzahl an, 157–159, 196–197
 gerichtete, 157–159, 171
 Glissade, 81
 Micro-, 79, 81

Sakkadenlänge, 84
Sakkadenrichtung, 84
Sanyo, 42
Scan Path
 Sakkade, 125
scan sath, 81
Schriftlichkeit, 8, 24
Selbstbild, *siehe* Identität
Selbstdarstellung
 positive, *siehe* Idealisierung
selektive regressive Durchlaufdauer,
 siehe Durchlauf, 191
Shapiro-Wilk-Test, *siehe* Statistik,
 185, 186, 190, 191, 193
SIDE-Modell, 21
Siri, *siehe* Apple
Skype, 14¹, 86–87, 91, 97, 100–105, 110,
 164, 169, 203
 -Version, 89
Alternativen zu, 103, 110–111,
 169, 170, 208
Nutzung von, 108
Skype Translator, 10, 26, 75, 85,
 92, 101, 109–111, 164, 169,
 207, 213, 216
social identity and de-individuation,
 siehe SIDE-Modell
social information processing theo-
 ry, *siehe* soziale Informati-
 onsverarbeitung
Somers' D_{XY}, *siehe* Statistik
soziale Informationsverarbeitung
 Theorie der, 20
Spearman's Rho, 128, *siehe* Statistik,
 132, 141, 145, 148, 152, 156,
 159, 176, 179, 186, 191, 193,
 196, 197
Sprachdienstleistung, 32
Sprachhandlung, 24

- Sprachkompetenzen, 167
Sprecher·innenorganisation, 13–15
Sprecher·innenwechsel, 13–15
Statistik, 92, 93
 Testverfahren
 Anderson-Darling-Test, 159, 161, 197, 200
 Chi-Quadrat-Test, 133, 137, 182, 183
 Dunn-Benjamini-Hochberg-Test, 127, 131, 139, 141, 144, 146, 148, 150, 154, 159, 161, 163, 175, 177, 178, 185, 188, 192, 193, 195, 199, 201
 Effektstärke nach Cohen, 127, 132, 141, 145, 156, 163, 175, 176, 179, 185, 186, 188, 189, 191, 193, 196
 Korrelationstest nach Spearman, 128, 132, 141, 145, 148, 152, 156, 159, 176, 179, 186, 191, 193, 196, 197
 Kruskal-Wallis-Test, 127, 131, 139, 141, 144, 146, 150, 153, 159, 161, 163, 175, 177, 178, 185, 188, 190, 191, 193, 195, 199, 201
 Mann-Whitney-U-Test, 127, 132, 141, 145, 148, 150, 154, 175, 179, 185, 186, 189, 190, 192, 196
 nicht-parametrisches, 93, 131, 133, 146, 150, 177, 182, 190, 191, 193, 210, 214
 Regressionsmodell, 92, 133, 136, 137, 179, 180, 183, 214
Shapiro-Wilk-Test, 139, 144, 146, 150, 153, 154, 185, 186, 190, 191, 193
- Somers' D_{XY} , 133, 137
Varianzanalyse, 92
Transformation
 logarithmische, 144, 146, 150, 153, 159, 161, 176, 185, 186, 190, 191, 193, 197, 200
Statistisches Bundesamt, 97
Stimulus, 78, 80, 90, 92
Systran, 41
- TAUS, *siehe* Translation Automation User Society
Telegram, 10, 87
TERMCAT, 33
Thema-Rhema-Progression, 210, 222
Think-Aloud-Protokoll, 84
Tilde, 35
TITUS, 41
Toshiba, 42
Trainingskorpus, *siehe* Korpus
Transferübersetzung, 44, *siehe auch* Übersetzung
Transformation
 logarithmische, *siehe* Statistik
Translation Automation User Society, 66
Translation Error Rate, 66
Translationsprozessforschung, 93, 164
Tremor, *siehe* Fixation
Turn-Taking, 13–15
Twitter, 87
- Umfrage, 76–78, 97
 mündliche, 78
 online, 76, 86, 97, 103, 110, 208
Pilotierung, 77
Pretest, 77

Sachregister

- Telefon-, 76
- Universität Leipzig, 166
- US National Institute of Standard and Technology, 66
- Usability, 208
- Valencianische Gemeinschaft, 32
- Validität, 28
- Varianzanalyse, *siehe auch* Statistik
- Verbmobil, 17, 42
- Versprachlichungsstrategien, 9, 11
- Verweildauer, 84
 - durchschnittliche, 186
 - Gesamt-, 142–145, 186–189, 218–219
- Verweildauer im ersten Durchlauf, 94
- Viber, 87
- Weaver Memorandum, 39
- WhatsApp, 10, 14¹, 87, 104, 105, 110, 170, 206
- Wikipedia, 28
- Windows Live Messenger, *siehe* Microsoft
- Wissen
 - saustausch
 - Möglichkeiten des, 30
 - sbegriff, 27
 - sinstanz, 26
 - skommunikation, 30
 - skuratierung, 28
 - smanagement, 27
 - sproduktion, 31
 - srepräsentation, 26
 - Formen der, 27
 - sstruktur, 26
 - sverteilung, 28, 29
 - szugang, 31

Informationsintegration in mehrsprachigen Textchats

Die vorliegende Arbeit widmet sich der Informationsintegration in maschinell übersetzten, mehrsprachigen Textchats am Beispiel des Skype Translators im Sprachenpaar Katalanisch-Deutsch. Der Untersuchung von Textchats dieser Konfiguration wurde sich bislang nur wenig zugewendet. Deshalb wird der zunächst grundlegend explorativ ausgerichteten Forschungsfrage nachgegangen, wie Personen eine maschinell übersetzte Textchat-Kommunikation wahrnehmen, wenn sie nicht der Sprache des Gegenübers mächtig sind. Damit einher geht auch die Untersuchung der Informationsextraktion und -verarbeitung zwischen Nachrichten, die in der eigenen Sprache verfasst wurden, und der Ausgabe der Maschinellen Übersetzung.

Zur Erfassung des Nutzungsverhalten im Umgang mit Skype und dem Skype Translator wurde mit einer deutschlandweit an Studierende gesendeten Online-Umfrage gearbeitet. In einer zweiteiligen, naturalistisch orientierten Pilotstudie unter Einsatz des Eye-Trackers wurde das Kommunikationsverhalten von Studierenden mit deutscher Muttersprache einerseits in maschinell vom Skype Translator übersetzten Chats mit katalanischen Muttersprachler-innen und andererseits, als Referenz, in monolingualen, rein deutschsprachigen Chats ohne Skype Translator untersucht. Bei den Teilnehmer-innen an diesen Studien handelt es sich um zwei unabhängige Gruppen. Beide wurden ebenfalls mit Fragebögen zum Nutzungsverhalten und zu den Eindrücken des Skype Translators erfasst.

Das sicher überraschendste Ergebnis der Studie ist, dass die Versuchspersonen einen substanzialen Teil der Chatkommunikation auf der MÜ-Ausgabe in beiden beteiligten Sprachen verbringen. Die Untersuchung der Sakkaden und Regressionen deutet auf einen sprunghaften Wechsel zwischen Originalnachricht und MÜ hin. Der Schwerpunkt der Aufmerksamkeit liegt dabei konsequent auf den neusten Nachrichten. Es ist daher anzunehmen, dass die Versuchspersonen die MÜ-Ausgabe aktiv in die Kommunikation miteinbeziehen und wesentliche Informationen zwischen Original und MÜ abzugleichen versuchen.