

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Qualität beim Simultandolmetschen: Der Einfluss der Intonation auf die
Bewertung der Dolmetschleistung

verfasst von | submitted by

Lorenza Truffa

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

Master of Arts (MA)

Wien | Vienna, 2026

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

UA 070 348 331

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

Masterstudium Translation Italienisch Deutsch

Betreut von | Supervisor:

Univ.-Prof. Mag. Dr. Franz Pöchhacker

Danksagung

Ich möchte mich bei allen bedanken, die direkt oder indirekt zur Verfassung dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt meinem Betreuer, Univ.-Prof. Mag. Dr. Franz Pöchlhammer, der mich bei der Erstellung dieser Arbeit begleitet und mir stets interessantes, zeitnahes und konstruktives Feedback gegeben hat.

Ich danke auch allen, die sich die Zeit genommen haben, an der Studie teilzunehmen und den Fragebogen auszufüllen.

Ringrazio i miei genitori, Patrizia e Giovanni e mio fratello Federico. Grazie per essere i miei fan numero uno e per non avermi mai stressato ai pranzi di Natale chiedendomi quando finalmente avessi completato la tesi. Ringrazio le mie zie Gabriella e Monica, che hanno sempre creduto in me. Ringrazio i miei cuoricini Ilaria, Sabrina, Karen, Yasser, Cristina, Giada e Cristiana, che hanno reso Vienna la mia casa. Ringrazio Luki e ringrazio Raphi. E ringrazio le mie amiche che ho lasciato in Italia Ana, Dajana, Sara e Francesca che anche da lontano mi hanno sempre sostenuto.

Vi voglio bene, grazie.

Inhaltsverzeichnis

Abbildungsverzeichnis	5
Einleitung	12
Grundlegendes zum Simultandolmetschen	15
1.1 Simultandolmetschen: Definition und Anforderungen	16
1.2 Kommunikative Aspekte des Dolmetschens	17
1.3 Kompetenzen von Simultandolmetscher*innen	17
1.4 Kognitive Anforderungen und das Effort-Modell	18
1.5 Einsatzbereiche des Simultandolmetschens: Konferenz- und Mediendolmetschen	19
2 Qualität beim Simultandolmetschen	23
2.1 Qualitätsstandards und Parameter	24
2.1.1 Die Kriterien von Bühler (1986)	25
2.1.2 Die Bedeutung nonverbaler Parameter	26
2.2 Qualitätsperspektive	28
2.2.1 Die Perspektive der Dolmetscher*innen: Studien zu den Erwartungen	30
2.2.2 Die Perspektive der Nutzer*innen	32
2.2.2.1 Umfrage-Studien zu Erwartungen	32
2.2.2.2 Umfrage-Studien zu Erwartungen und Qualitätsbewertung	35
3 Intonation	38
3.1 Intonatorische Gestaltung	40
3.2 Intonationsmodelle	41
3.3 Akustische Korrelate der Intonation	43
3.4 Die Wahrnehmung der Intonation	45
3.5 Funktionen der Intonation	46
4 Intonation beim Simultandolmetschen	48
4.1 Merkmale der Intonation beim Simultandolmetschen	48

4.2 Forschungsstand zur Intonation beim Simultandolmetschen	49
4.2.1. Studien von Collados Aís zur Qualitätsbewertung	51
4.2.2 Folgestudien zum Experiment von Collados Aís (1998)	53
4.3 Die Rolle von Dolmetscher*innen bei der Wiedergabe der Intonation	56
5 Methodik	58
5.1 Zielsetzung und Hypothesen	58
5.2 Versuchsdesign.....	59
5.2.1 Material und Stimuli.....	60
5.2.1.1 Ausgangstext	60
5.2.1.2 Textgestaltung & Präsentationsweise.....	61
5.2.2 Die Vorbereitung der Dolmetscherin	62
5.2.3 Dolmetschversionen	62
5. 3 Proband*innen.....	64
5.4 Versuchsablauf.....	66
5.5 Erhebungsinstrument: Fragebogen.....	67
5.5.1 Gestaltung und Layout	67
5.5.2 Eingangsfragebogen	68
5.5.3 Fragebogen zur Qualitätsbewertung.....	69
5.5.4 Verwendete Skalen.....	71
5.6 Datenanalyse	72
5.6.1 Datenüberprüfung für nichtparametrische Tests	72
5.6.2 Nichtparametrische Tests zur Untersuchung von Bewertungskriterien und Störvariablen	74
5.6.3 Deskriptive Analyse: Intra- und Interstudien-Vergleich mit der Studie von Collados Aís (1998).....	75
5.6.3.1 Vorgehen der deskriptiven Analysen	76
6 Ergebnisse	78
6.1 Vergleich der Ergebnisse von Test A und Test B	78
6.2 Analyse der Störvariable „Teilnahme an Konferenzen“ für Test A und Test B	85
6.3 Analyse der Störvariable „Deutschkenntnisse“ für Test A und Test B.....	96

6.4 Ergebnisse des Intra- und Interstudien-Vergleichs mit der Studie von Collados Aís (1998)	104
7 Diskussion und Schlussfolgerungen.....	114
7.1 Ergebnisse des Vergleichs zwischen Test A und Test B sowie der Störvariablen	114
7.2 Ergebnisse des Intra- und Interstudienvergleichs.....	116
7.3 Limitationen	118
7.4 Schlussfolgerungen	120
Literaturverzeichnis.....	123
Anhang	131
Abstract	143

Abbildungsverzeichnis

Abb. 1: Darstellung des Spektrums des Dolmetschens (Pöchhacker 2016: 17)	16
Abb. 2: Grafische Darstellung des Modells der Qualitätsstandards (Pöchhacker 2002: 97) ...	25
Abb. 3: Grafische Darstellung der Qualitätsperspektiven (Pöchhacker 2001: 42)	29
Abb. 4: Darstellung der Schallsignalen in Bezug auf die Tonhöhe, adaptiert nach Peters (2021: 13).....	44
Abb. 5: Beispiel von harmonischen Klängen mit Grundfrequenzen (100 Hz, 120 Hz und 80 Hz) und Darstellung des Grundtons und der ersten vier Obertöne, adaptiert nach Peters (2021: 14)	45
Abb. 6: Grafische Darstellung der Pitch-Konturen – links die Kontrollversion, rechts die Monotonversion	64
Abb. 7: Links: Verteilungsdarstellung der Bewertungen des Kriteriums Intonation für Test A und B; Rechts: Beispiel einer Normalverteilung der Daten (Field et al. 2025: 251)	74
Abb. 8: Verteilung der Bewertungen für das Kriterium Gesamtbewertung	79
Abb. 9: Verteilung der Bewertungen für das Kriterium Flüssigkeit	80
Abb. 10: Verteilung der Bewertungen für das Kriterium Intonation	81
Abb. 11: Verteilung der Bewertungen für das Kriterium Kohärenz	83
Abb. 12: Verteilung der Bewertungen für das Kriterium Professionalität.....	84
Abb. 13: Vergleich der Mittelwertdifferenzen zwischen der lebhaften und der monotonen Version nach Studie über alle Kriterien	112
Abb. 14: Vergleich der Bewertungen über alle Kriterien – Video 1 (Collados Aís) und Test B	113

Tabellenverzeichnis

Tab. 1. Grundfrequenzwerte der Kontroll- und Monotonversion im Vergleich	64
Tab. 2. Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung	79
Tab. 3: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung	79
Tab. 4: Deskriptive Statistiken für das Kriterium Gesamtbewertung	79
Tab. 5: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit.....	80
Tab. 6: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit.....	80
Tab. 7: Deskriptive Statistiken für das Kriterium Flüssigkeit.....	81
Tab. 8: Häufigkeiten der Bewertungen für das Kriterium Intonation	82
Tab. 9: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation	82
Tab. 10: Deskriptive Statistiken für das Kriterium Intonation.....	82
Tab. 11: Häufigkeiten der Bewertungen für das Kriterium Kohärenz.....	83
Tab. 12: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz.....	83
Tab. 13: Deskriptive Statistiken für das Kriterium Kohärenz.....	83
Tab. 14: Häufigkeiten der Bewertungen für das Kriterium Professionalität.....	84
Tab. 15: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität.....	84
Tab. 16: Deskriptive Statistiken für das Kriterium Professionalität	84
Tab. 17: Zusammenfassung der Mann-Whitney-U-Testergebnisse für Test A und Test B.....	85
Tab. 18: Zusammenfassung deskriptiver Statistiken für Test A und Test B.....	85
Tab. 19: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme	86
Tab. 20: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme.....	86
Tab. 21: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme	86
Tab. 22: Ergebnisse des Kendall's Tau-Tests für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme.....	86
Tab. 23: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme	87
Tab. 24: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme	87
Tab. 25: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme	87

Tab. 26: Ergebnisse des Kendall's Tau-Tests für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme.....	87
Tab. 27: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme	88
Tab. 28: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme	88
Tab. 29: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme	88
Tab. 30: Ergebnisse des Kendall's Tau-Tests für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme	88
Tab. 31: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme	89
Tab. 32: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme	89
Tab. 33: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme	89
Tab. 34: Ergebnisse des Kendall's Tau-Tests für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme	89
Tab. 35: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test A nach Konferenzteilnahme	90
Tab. 36: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test A nach Konferenzteilnahme	90
Tab. 37: Deskriptive Statistiken für das Kriterium Intonation für Test A nach Konferenzteilnahme	90
Tab. 38: Ergebnisse des Kendall's Tau-Tests für das Kriterium Intonation für Test A nach Konferenzteilnahme	91
Tab. 39: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test B nach Konferenzteilnahme	91
Tab. 40: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test B nach Konferenzteilnahme	91
Tab. 41: Deskriptive Statistiken für das Kriterium Intonation für Test B nach Konferenzteilnahme	92
Tab. 42: Ergebnisse des Kendall's Tau-Tests für das Kriterium Intonation für Test B nach Konferenzteilnahme	92

Tab. 43: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test A nach Konferenzteilnahme	92
Tab. 44: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test A nach Konferenzteilnahme	93
Tab. 45: Deskriptive Statistiken für das Kriterium Kohärenz für Test A nach Konferenzteilnahme	93
Tab. 46: Ergebnisse des Kendall's Tau-Tests für das Kriterium Kohärenz für Test A nach Konferenzteilnahme	93
Tab. 47: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test B nach Konferenzteilnahme	93
Tab. 48: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test B nach Konferenzteilnahme	94
Tab. 49: Deskriptive Statistiken für das Kriterium Kohärenz für Test B nach Konferenzteilnahme	94
Tab. 50: Ergebnisse des Kendall's Tau-Tests für das Kriterium Kohärenz für Test B nach Konferenzteilnahme	94
Tab. 51: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test A nach Konferenzteilnahme	95
Tab. 52: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test A nach Konferenzteilnahme.....	95
Tab. 53: Deskriptive Statistiken für das Kriterium Professionalität für Test A nach Konferenzteilnahme	95
Tab. 54: Ergebnisse des Kendall's Tau-Tests für das Kriterium Professionalität für Test A nach Konferenzteilnahme	95
Tab. 55: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test B nach Konferenzteilnahme	96
Tab. 56: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test B nach Konferenzteilnahme.....	96
Tab. 57: Deskriptive Statistiken für das Kriterium Professionalität für Test B nach Konferenzteilnahme	96
Tab. 58: Ergebnisse des Kendall's Tau-Tests für das Kriterium Professionalität für Test B nach Konferenzteilnahme	96
Tab. 59: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test A nach Deutschkenntnissen.....	97

Tab. 60: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test A nach Deutschkenntnissen	97
Tab. 61: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test A nach Deutschkenntnissen.....	97
Tab. 62: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test B nach Deutschkenntnissen.....	98
Tab. 63: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test B nach Deutschkenntnissen.....	98
Tab. 64: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test B nach Deutschkenntnissen.....	98
Tab. 65: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test A nach Deutschkenntnissen.....	98
Tab. 66: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test A nach Deutschkenntnissen.....	99
Tab. 67: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test A nach Deutschkenntnissen.....	99
Tab. 68: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test B nach Deutschkenntnissen.....	99
Tab. 69: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test B nach Deutschkenntnissen.....	99
Tab. 70: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test B nach Deutschkenntnissen.....	99
Tab. 71: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test A nach Deutschkenntnissen.....	100
Tab. 72: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test A nach Deutschkenntnissen.....	100
Tab. 73: Deskriptive Statistiken für das Kriterium Intonation für Test A nach Deutschkenntnissen.....	100
Tab. 74: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test B nach Deutschkenntnissen.....	101
Tab. 75: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test B nach Deutschkenntnissen.....	101
Tab. 76: Deskriptive Statistiken für das Kriterium Intonation für Test B nach Deutschkenntnissen.....	101

Tab. 77: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test A nach Deutschkenntnissen.....	102
Tab. 78: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test A nach Deutschkenntnissen.....	102
Tab. 79: Deskriptive Statistiken für das Kriterium Kohärenz für Test A nach Deutschkenntnissen.....	102
Tab. 80: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test B nach Deutschkenntnissen.....	102
Tab. 81: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test B nach Deutschkenntnissen.....	103
Tab. 82: Deskriptive Statistiken für das Kriterium Kohärenz für Test B nach Deutschkenntnissen.....	103
Tab. 83: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test A nach Deutschkenntnissen.....	103
Tab. 84: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test A nach Deutschkenntnissen	103
Tab. 85: Deskriptive Statistiken für das Kriterium Professionalität für Test A nach Deutschkenntnissen.....	103
Tab. 86: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test B nach Deutschkenntnissen.....	104
Tab. 87: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test B nach Deutschkenntnissen	104
Tab. 88: Deskriptive Statistiken für das Kriterium Professionalität für Test B nach Deutschkenntnissen.....	104
Tab. 89: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Gesamtbewertung.....	105
Tab. 90: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Video 1 (Collados Aís) und Test B.....	106
Tab. 91: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Video 1 (Collados Aís) und Test B	106
Tab. 92: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Flüssigkeit.....	107
Tab. 93: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Video 1 (Collados Aís) und Test B	107

Tab. 94: Deskriptive Statistiken für das Kriterium Flüssigkeit für Video 1 (Collados Aís) und Test B	108
Tab. 95: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Intonation.....	108
Tab. 96: Häufigkeiten der Bewertungen für das Kriterium Intonation für Video 1 (Collados Aís) und Test B	109
Tab. 97: Deskriptive Statistiken für das Kriterium Intonation für Video 1 (Collados Aís) und Test B	109
Tab. 98: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Kohärenz.....	109
Tab. 99: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Video 1 (Collados Aís) und Test B	110
Tab. 100: Deskriptive Statistiken für das Kriterium Kohärenz für Video 1 (Collados Aís) und Test B	110
Tab. 101: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Professionalität	111
Tab. 102: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Video 1 (Collados Aís) und Test B	111
Tab. 103: Deskriptive Statistiken für das Kriterium Professionalität für Video 1 (Collados Aís) und Test B	112

Einleitung

Die vorliegende Arbeit widmet sich der Qualitätsbewertung beim Simultandolmetschen aus der Perspektive der Nutzer*innen, wobei das Augenmerk hauptsächlich auf den Parameter der Intonation gelegt wird.

Im Bereich des Konferenzdolmetschens sind die Kriterien, denen bei der Bewertung einer Verdolmetschung traditionell die größte Bedeutung beigemessen wird, inhaltsbezogene Kriterien wie die Sinnübereinstimmung mit dem Original und die Kohärenz. Die Kriterien hingegen, die sich auf die Prosodie und somit auf die Darbietung beziehen, darunter Flüssigkeit und Intonation, werden als weniger relevant angesehen (vgl. Zwischenberger 2010: 127). Dies wurde durch verschiedene Studien zu den Qualitätserwartungen der Nutzer*innen bestätigt, darunter jene von Kurz (1989, 1993), Kopczynski (1994) und Moser (1996). Es hat sich jedoch gezeigt, dass, wenn Nutzer*innen eine konkrete Dolmetschleistung bewerten müssen, die darbietungsbezogenen Kriterien eine weitaus größere Bedeutung erlangen.

Im Jahr 1998 veröffentlichte Collados Aís eine Studie, die sich mit der Untersuchung der Auswirkungen einer monotonen Intonation auf die Bewertung einer Dolmetschleistung durch spezialisierte Nutzer*innen und professionelle Dolmetscher*innen befasste. Um dies zu überprüfen, erstellte sie drei verschiedene Dolmetschversionen auf der Grundlage eines Ausgangstextes, der mit monotoner Intonation vorgetragen wurde, wobei sie die Intonation und die Bedeutung des Ausgangstextes manipuliert: Eine Version wies eine monotone Intonation auf und enthielt keine inhaltlichen Abweichungen vom Ausgangstext, eine weitere war nicht monoton vorgetragen, wies jedoch inhaltliche Fehler auf, und eine war die Kontrollversion, die nicht monoton war und keine inhaltlichen Fehler enthielt (vgl. Collados Aís 1998: 90). Die Ergebnisse zeigten, dass die inhaltlich korrekte Version, die jedoch eine monotone Intonation aufwies, schlechtere Ergebnisse erzielte als die nicht monotone Version, die inhaltlichen Fehler enthielt (vgl. Collados Aís 1998: 251). Es wurde somit nachgewiesen, dass eine monotone Intonation die Qualitätsbewertung beim Simultandolmetschen beeinträchtigt, was die Bedeutung der darbietungsbezogenen Kriterien für die konkrete Bewertung einer Dolmetschleistung deutlich macht.

In ihrer Studie forderte Collados Aís (1998) dazu auf, in zukünftigen Forschungen zu überprüfen, ob eine monotone Intonation der Dolmetschleistung bei einem lebhaft intonierten Ausgangstext als noch stärker nachteilig wahrgenommen wird (vgl. 1998: 90). Auf der Grundlage dieser Feststellung wurde diese Masterarbeit verfasst, die eine partielle Replikation der Studie von Collados Aís darstellt (1998). Ausgehend von einem lebhaften Ausgangstext

wurden zwei Dolmetschversionen erstellt: eine mit monotoner und eine mit lebhafter Intonation, die von zwei verschiedenen Gruppen nicht spezialisierter Nutzer*innen anhand eines online vorgelegten Fragebogens bewertet wurden. Ziel ist es, die in der Studie von Collados Aís (1998) aufgestellte Hypothese erneut zu bestätigen, wonach eine monotone Intonation die Bewertung der Qualität der Verdolmetschung sowie die Wahrnehmung der Nutzer*innen hinsichtlich des subjektiven Kriteriums der Professionalität der Dolmetscherin negativ beeinflusst (vgl. Collados Aís 1998: 81f). Darüber hinaus zielt diese Arbeit darauf ab, zu überprüfen, ob die monotone Version unter Verwendung eines lebhaften Ausgangstextes noch stärker nachteilig wahrgenommen wird als in der Studie von Collados Aís (1998).

Strukturell besteht diese Masterarbeit aus acht Hauptkapiteln. Die ersten vier Kapitel sind dem theoretischen Teil gewidmet. Das erste Kapitel erläutert die Grundlagen des Simultandolmetschens. Das zweite Kapitel konzentriert sich auf die Qualität des Simultandolmetschens, analysiert die Standards und Parameter sowie die verschiedenen Perspektiven, aus denen eine Dolmetschleistung bewertet werden kann. Dabei wird insbesondere auf Studien Bezug genommen, die sich mit der Perspektive der Nutzer*innen befassen – die Perspektive, auf die sich diese Arbeit stützt. Kapitel 3 befasst sich mit der Intonation als Tonhöhenverlauf einer sprachlichen Äußerung und analysiert deren Merkmale und Funktionen. Kapitel 4 behandelt die Intonation im Bereich des Dolmetschens. Hier wird die Studie von Collados Aís (1998) ausführlich beschrieben und es werden die Merkmale der Intonation beim Simultandolmetschen behandelt. Die Kapitel 5 bis 8 widmen sich dem experimentellen Teil dieser Arbeit. Kapitel 5 befasst sich mit der Methodik: Es beschreibt die Erstellung des Audio-Video-Materials, stellt die Proband*innen sowie die Teilnahmebedingungen vor, führt den Fragebogen als Instrument der Datenerhebung ein und widmet sich der Datenanalyse.

Die Analyse wurde sowohl mittels inferenzstatistischer Methoden unter Verwendung nichtparametrischer Tests als auch auf deskriptive Weise durchgeführt. Zunächst wurde ein Vergleich zwischen den beiden Dolmetschversionen der vorliegenden Studie mittels des Mann-Whitney-U-Tests durchgeführt. Damit sollte überprüft werden, ob die monotone Version schlechtere Ergebnisse erzielte als die lebhafte Version. Anschließend wurde ein intrastudienbezogener deskriptiver Vergleich durchgeführt, bei dem die Mittelwertdifferenzen der Bewertungen in der Studie von Collados Aís (1998) und in der vorliegenden Studie gemessen wurden. Dies dient dazu, zu untersuchen, inwieweit die monotone Version in beiden Studien negativer bewertet wurde als die lebhafte Version. Auf diese Weise konnte die Hypothese überprüft werden, ob eine monotone Intonation der Dolmetschleistung bei einem

Ausgangstext mit lebhafter Intonation als noch nachteiliger empfunden wird. Schließlich wurde ein direkter Vergleich zwischen den monotonen Versionen der beiden Studien durchgeführt, um die Ergebnisse der vorliegenden Studie in einen vergleichenden Kontext zu stellen. Die Ergebnisse sind in Kapitel 6 ausführlich dargestellt, und in Kapitel 7 werden Schlussfolgerungen gezogen. Es folgen die Anhänge mit der Transkription des Ausgangstextes, des Zieltextes und dem Fragebogen.

Grundlegendes zum Simultandolmetschen

Das Dolmetschen ermöglicht die Kommunikation zwischen Menschen, die nicht dieselbe Sprache sprechen (vgl. Ahrens 2004: 5). Diese Tätigkeit wird allgemein als mündliche Übersetzung bezeichnet (vgl. Pöchhacker 2010a: 154). Beim Dolmetschen wird in der Regel ein mündlich vorgetragener Ausgangstext (AT) in der Zielsprache ebenfalls mündlich als Zieltext (ZT) wiedergegeben (vgl. Herbert 1952: 1). Der Zieltext kann im Gegensatz zur schriftlichen Übersetzung nach seiner Äußerung nicht korrigiert werden (vgl. Kade 1968: 35). Dolmetschen ist somit eine Echtzeitleistung, bei der die Ausgangsbotschaft nicht wiederholt werden kann und der von den Dolmetscher*innen ausgedrückte Inhalt weder korrigiert noch überarbeitet werden kann (vgl. Pöchhacker 2010a: 154).

Die wichtigsten Einsatzbereiche des Dolmetschens sind das Konferenzdolmetschen und das *Community Interpreting* (Kommunaldolmetschen) (vgl. Pöchhacker 2010a: 155). Auf der Grundlage verschiedener Kriterien erfolgt eine weitere Klassifizierung. Betrachtet man beispielsweise die situative Konstellation der Interaktion, so findet die Interaktion beim *Community Interpreting* in den meisten Fällen zwischen drei kommunikativen Akteuren statt, wobei die Dolmetscher*innen als Vermittler*innen zwischen zwei Personen fungieren, die unterschiedliche Sprachen sprechen. In diesem Fall spricht man vom Dialogdolmetschen (vgl. Pöchhacker 2016: 16). Die Interaktion beim Konferenzdolmetschen ist hingegen in der Regel multilateral und nicht dialogisch (vgl. Pöchhacker 2010a: 155). Eine weitere wichtige Unterscheidung wird nach dem institutionellen Rahmen getroffen, in dem das Dolmetschen stattfindet. Konferenzdolmetschen findet in der Regel in einem internationalen Rahmen statt, während *Community Interpreting* in einer Institution einer bestimmten Gesellschaft, also in einem innergesellschaftlichen Rahmen, stattfindet (2010a: 155a). Diese Unterscheidungen sind in Abbildung 1 schematisch dargestellt.

Es existieren jedoch auch weitere „hybride“ Formen wie das Dialogdolmetschen im internationalen Rahmen, beispielsweise bei diplomatischen Gesprächen, und das Konferenzdolmetschen in innergesellschaftlichen Settings, das beispielsweise gehörlose Menschen betrifft (2010a: 155). Darüber hinaus gibt es noch einen weiteren Einsatzbereich des Dolmetschens, das Mediendolmetschen, das den Mediennutzer*innen einer soziokulturellen

Gemeinschaft die Inhalte fremdsprachiger Massenmedien zugänglich macht (vgl. Pöchhacker 2016: 15).

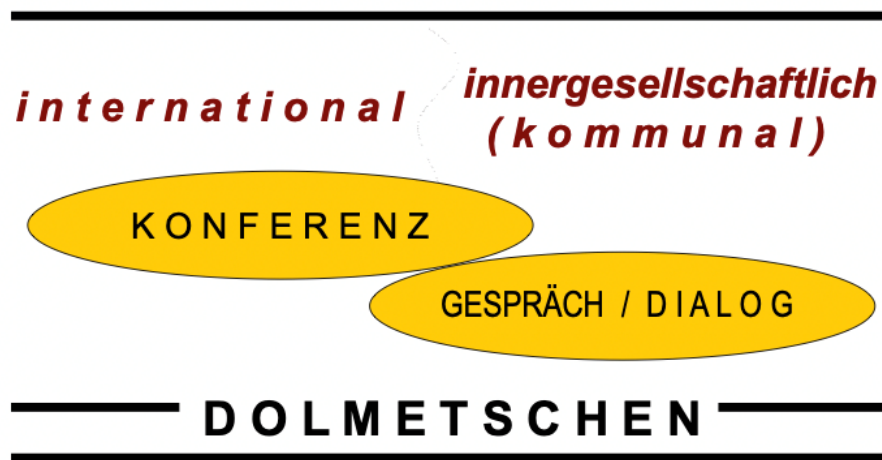


Abb. 1: Darstellung des Spektrums des Dolmetschens (Pöchhacker 2016: 17)

Es lassen sich zudem drei verschiedene Modi des Dolmetschens unterscheiden. Die wichtigsten Modi sind das Simultandolmetschen, gewöhnlich mit SI abgekürzt, das Konsekutivdolmetschen und das Flüsterdolmetschen, auch „Chuchotage“ genannt (vgl. Pöchhacker 2016). In diesem Kapitel wird das Simultandolmetschen näher beschrieben, da sich die vorliegende Masterarbeit mit diesem Modus befasst.

1.1 Simultandolmetschen: Definition und Anforderungen

“Simultaneous interpreting is a mode in which the interpreter reformulates the source speech as it unfolds, generally with a lag of a few seconds at most” (Gile 2004: 11). Dieser Modus wird bei 90% der Konferenzen angewendet und entwickelte sich aufgrund der Anforderungen dieser formellen Veranstaltungen, die hohe Standards in Bezug auf Geschwindigkeit und Genauigkeit verlangen. Diese Standards konnten nur durch die Einführung geeigneter Geräte wie Mikrofone, schalldichte Kabinen, Kopfhörer und spezielle Kanäle für verschiedene Sprachen erfüllt werden (vgl. Setton 2010: 69). Die Entwicklung der ersten SI-Geräte erfolgte in den 1920er Jahren und ermöglichte den erstmaligen Einsatz dieses Verfahrens bei der Konferenz der Internationalen Arbeitsorganisation in Genf im Jahr 1927. Allerdings traten technische Schwierigkeiten auf und erst während der Nürnberger Prozesse (1945–1946) konnten im Rahmen des ersten großen internationalen Gerichtsverfahrens Simultandolmetschungen in Englisch, Französisch, Deutsch und Russisch angeboten werden (vgl. Phelan 2019: 2).

1.2 Kommunikative Aspekte des Dolmetschens

Im Kommunikationsprozess des Simultan- und Konsektivdolmetschens übertragen die Dolmetscher*innen einen Ausgangstext (AT) der Sprecher*innen an die Empfänger*innen, und erzeugen dabei einen Zieltext (ZT). Der Kommunikationsakt des Simultandolmetschens findet in einer multidirektionalen dreiseitigen Kommunikationsbeziehung statt, an der *speaker* (Sprecher*in/Redner*in), *translator* (hier Dolmetscher*in) und *receiver* (Empfänger*in) beteiligt sind. Auch der *client* (Kund*in) ist Teil dieses Kommunikationsprozesses, obwohl er/sie nicht direkt am Informationsfluss teilnimmt. Der/die Sprecher*in formuliert eine Botschaft in der Ausgangssprache, welche der/die Dolmetscher*in in die Zielsprache überträgt und an die Empfänger*innen vermittelt. Der/die Kund*in ist schließlich die Person oder Einrichtung, die den Dolmetschdienst in Auftrag gibt (vgl. Gile 2009b: 27f.).

In dieser Masterarbeit wird die Bewertung der Qualität einer Simultandolmetschleistung durch die Empfänger*innen untersucht. Diese sind für den Kommunikationsprozess des Simultandolmetschens von grundlegender Bedeutung, da sie das Publikum darstellen, an das sich die Rede des AT richtet. Laut Kurz (2001) hängt die Qualität der Dolmetschleistung direkt von der Zufriedenheit der Empfänger*innen ab, d. h. der Nutzer*innen, die den Dolmetschdienst in Anspruch nehmen (vgl. 2001: 405). Sie nehmen sowohl die verbalen und prosodischen Informationen der Dolmetscher*innen als auch die nonverbalen Signale der Sprecher*innen, wie Mimik usw., wahr (vgl. Ahrens 2004: 72). Aus diesem Grund spricht man von einem multidirektionalen statt von einem unidirektionalen Kommunikationsdreieck, da zwischen den verschiedenen Akteur*innen der Kommunikation eine Beziehung besteht.

Was die prosodischen Elemente des Ausgangstextes betrifft, so werden diese von den Nutzer*innen nur teilweise wahrgenommen, da auch die Dolmetscher*innen akustische Informationen liefern, die die Rezeption beeinflussen (vgl. Ahrens 2004: 73). Aus diesem Grund ist es wichtig, dass die Dolmetscher*innen neben den verbalen Elementen des Originaltextes auch die prosodischen Elemente wiedergeben. Diese Elemente sind wesentlich für die intendierte Wirkung der Rede und spielen eine entscheidende Rolle in der Rezeption des Zieltextes durch die Zuhörer*innen. Die Prosodie beim Simultandolmetschen steht im Mittelpunkt der Analyse in Kapitel 4, mit besonderem Augenmerk auf der Intonation.

1.3 Kompetenzen von Simultandolmetscher*innen

Nutzer*innen haben spezifische Erwartungen an Dolmetschleistungen, die ihnen zufolge zumindest verständlich und originalgetreu sein müssen. Damit dies jedoch gewährleistet werden kann und weitere Erwartungen hinsichtlich terminologischer Genauigkeit, angenehmer

Stimme usw. erfüllt werden, müssen Dolmetscher*innen unterschiedliche Fähigkeiten entwickeln (vgl. Setton 2010: 72). Für die Ausübung des Berufs der Simultandolmetscher*innen sind spezifische Kompetenzen erforderlich. Diese werden von Dolmetscher*innen während ihrer Ausbildung erworben und verleihen dem Beruf Prestige und Anerkennung (vgl. Pöchhacker 2016: 160).

Gemäß dem Leipziger Kompetenzmodell des Dolmetschens müssen Dolmetscher*innen beim Simultandolmetschen in der Lage sein, eine Botschaft unter hohem Zeitdruck simultan aufzunehmen, zu verarbeiten und wiederzugeben. Außerdem müssen sie dazu fähig sein, in Echtzeit zwischen verschiedenen Sprachcodes zu wechseln und eine funktionale Sprachwiedergabe aufrechtzuerhalten. Sie müssen die Syntax anpassen, die Kohärenz des Textes gewährleisten, die Botschaft antizipieren, den Inhalt verstehen und die Zeit und die Zusammenarbeit in der Kabine effektiv managen (vgl. Kutz 2010: 219f.).

Ein Berufsverband, der die Qualität von Dolmetschdienstleistungen gewährleistet und maßgeblich fördert, ist die AIIC (Association Internationale des Interprètes de Conférence). Diese Organisation trägt zum Ansehen des Dolmetscher*innenberufs bei, regelt dessen Ausübung und ist eine der wichtigsten institutionellen Referenzstellen im Bereich des professionellen Dolmetschens (vgl. Pöchhacker 2016: 30). Obwohl die AIIC keine festgelegte Liste der von Dolmetscher*innen geforderten Kompetenzen erstellt, spielt sie durch die Festlegung professioneller Standards eine zentrale Rolle bei der Förderung der Qualität im Dolmetschen. Um dem Verband beizutreten, ist eine Zulassung durch das *Committee on Admissions and Language Classification* (CACL) erforderlich, das auf Grundlage professioneller Kriterien über die Aufnahme neuer Mitglieder befindet (vgl. Moser 1996: 145).

1.4 Kognitive Anforderungen und das Effort-Modell

Um besser zu verstehen, wie sich die zuvor beschriebenen Kompetenzen in der Dolmetschpraxis konkretisieren, ist es hilfreich, auf theoretische Modelle zurückzugreifen, die ihre kognitive Funktionsweise veranschaulichen. Dolmetschen als komplexes Set kognitiver Verarbeitungsprozesse zu beschreiben, ist ein theoretischer Rahmen, auf den sich die Forschung seit den 1980er Jahren konzentriert (vgl. Pöchhacker 2016: 37). Eines der einflussreichsten Instrumente zur Analyse der mentalen Dynamiken, insbesondere beim Simultandolmetschen, ist das „Effort-Modell“ von Gile (1985, 2009b). Er entwickelte dieses Modell, um die Fehler von Dolmetscher*innen bei Verdolmetschungen anhand ihres mentalen Energiemanagements zu erklären (vgl. Pöchhacker 2016: 37).

Nach Giles (1985, 2009b) „Effort-Modell“ erfordert das Simultandolmetschen verschiedene kognitive Anstrengungen: Hörverständnis und Analyse (*écoute et analyse*), Produktion (*production*) und Gedächtnis (*mémoire*). Das erste Element bezieht sich auf die Anstrengung der Wahrnehmung und des Verständnisses des Diskurses, das zweite auf die sprachliche Wiedergabe des Zieltextes und das dritte auf die Leistung der Gedächtnisoperationen (vgl. Gile 1985: 44). Außerdem gibt es die Koordination (*coordination*), d. h. das Koordinieren dieser Anstrengungen (vgl. Gile 2009b: 168). Die Energie, über die Dolmetscher*innen verfügen, wird laut Gile auf diese kognitiven Anforderungen verteilt, die einen Gleichgewichtszustand bilden. Wenn eine dieser drei Anstrengungen zunimmt, leiden eine oder beide der anderen darunter (vgl. Gile 1985: 45). Das bedeutet, dass die Dolmetschleistung Defizite aufweisen wird, die aus einer kognitiven Überlastung oder einem Ungleichgewicht zwischen den verschiedenen Anstrengungen resultieren (vgl. Pöchhacker 2016: 37). Die Dolmetscher*innen müssen daher die Gesamtbelastung durch die verschiedenen kognitiven Anstrengungen so steuern, dass die verfügbaren mentalen Ressourcen nicht überfordert werden (vgl. Ahrens 2004: 32).

1.5 Einsatzbereiche des Simultandolmetschens: Konferenz- und Mediendolmetschen

Das Simultandolmetschen kommt in einer Vielzahl unterschiedlicher Einsatzbereiche zur Anwendung. Dieses Kapitel befasst sich mit dem Konferenzdolmetschen und dem Mediendolmetschen, soweit sie den Gegenstand der vorliegenden Arbeit betreffen.

Wie bereits erwähnt, wird das Simultandolmetschen hauptsächlich im Bereich des Konferenzdolmetschens eingesetzt. Die Entstehung des Konferenzdolmetschens geht auf das Ende des Ersten Weltkriegs zurück, als multilaterale Verhandlungen und Konferenzen vom Völkerbund und der Internationalen Arbeitsorganisation (ILO) organisiert wurden. Im Rahmen dieser Zusammenkünfte wurde das Konsekutivdolmetschen eingesetzt, das bis zur Mitte des 20. Jahrhunderts die gängige Praxis im Konferenzdolmetschen darstellte (vgl. Dirikier 2015: 79). Die Präferenz für diesen Dolmetschmodus ist darauf zurückzuführen, dass dem Simultandolmetschen ursprünglich noch kein volles Vertrauen entgegengebracht wurde (vgl. Gile 2009a: 51). Nachdem die SI dann beim Nürnberger Prozess (1945-1946) eingesetzt wurde und sich internationale Organisationen und Konferenzen zu verbreiten begannen, kam es zum schnellsten und stärksten Institutionalisierungsprozess in der Geschichte des Dolmetschens (vgl. Dirikier 2015: 79). Danach erlebte das Konferenzdolmetschen einen regelrechten Aufschwung (vgl. Gile 2009a: 51). Diese Dolmetschart nahm bis in die 90er Jahre des 20. Jahrhunderts eine zentrale Rolle beim Dolmetschen ein, und erst Ende des 20. Jahrhunderts

begann man, sich mit anderen Dolmetscheinsatzbereichen zu beschäftigen, etwa dem Gerichtsdolmetschen und dem Community-Dolmetschen (vgl. Pöchhacker 2010a: 155). Im Vergleich zu den anderen Einsatzbereichen genießt das Konferenzdolmetschen nach wie vor das höchste Ansehen und die beste Vergütung (vgl. Gile 2004: 11). Dies ist einerseits auf die Schwierigkeit des Simultandolmetschens zurückzuführen, sowie auf die Tatsache, dass für die Ausübung dieses Berufs bestimmte Fähigkeiten erforderlich sind, die das Konferenzdolmetschen zu einem organisierten und autonomen Beruf machen (Setton 2010: 72).

Im Gegensatz zum Community Interpreting, das in einem intrasozialen Umfeld zwischen Dienstleister*innen und einzelnen Personen stattfindet (polizeiliche Verhöre, Gespräche zwischen Ärzt*innen und Patient*innen usw.), findet das Konferenzdolmetschen in der Regel in einem internationalen Umfeld im Rahmen von wissenschaftlichen oder politischen Konferenzen statt (Pöchhacker 2010a: 155). In dieser Arbeit wird der Begriff Konferenz nicht nur als formelle hochrangige Sitzung verstanden, sondern gemäß der Definition von Pöchhacker (1994) als „(simultan-) gedolmetschte interkulturelle Gruppenkommunikation“ (1994:51). Pöchhacker teilt Konferenzen in sieben Kategorien ein. Zur ersten Kategorie gehören satzungsgemäße Plenarversammlungen internationaler Organisationen. Eine weitere Kategorie bilden Fachkonferenzen, die sich einem spezifischen fachlichen Thema widmen. Weitere Formen sind Seminare und Schulungen, Verhandlungen im Rahmen internationaler Gremien, sowie Diskussionsforen, die sich auf internationale Veranstaltungen zu meist politischen, aktuellen Themen beziehen. Hinzu kommen Pressekonferenzen und Präsentationen, die als Informationsformate von Einzelpersonen oder Organisationen auftreten (vgl. Pöchhacker 1994: 51). Den Abschluss bildet die Kategorie der Gastvorträge, worunter Vorträge und Referate externer Redner*innen fallen, zusammengefasst unter dem Begriff „Gastvortrag“ (1994: 51). Die letztgenannte Art von Konferenz ist die Form, die für die Durchführung des Experiments für diese Masterarbeit verwendet wird, das in Kapitel 5 ausführlich erläutert wird.

Das Simultandolmetschen wird nicht nur bei mehrsprachigen Konferenzen, sondern auch im Medienbereich eingesetzt (vgl. Setton 2010: 69). Diese Form des Dolmetschens wird vor allem für Live-Übertragungen in den Massenmedien verwendet und entstand in den 1930er Jahren, als Reden von Hitler für das französische Radio simultan gedolmetscht wurden. Neben seiner prototypischen Ausprägung als Live-Übertragung findet das Mediendolmetschen auch Anwendung bei der barrierefreien Aufbereitung audiovisueller Programme für gehörlose und

schwerhörige Menschen durch den Einsatz von Gebärdensprachdolmetscher*innen (vgl. Pöchhacker 2010b: 224).

Es wird zwischen dem Dolmetschen vor Ort, das bei Kommunikationsveranstaltungen im Studio mit anwesenden Dolmetscher*innen durchgeführt wird, und dem Dolmetschen für die Übertragung von Veranstaltungen, die an entfernten Orten stattfinden, unterschieden. Im ersten Fall variieren die Modi zwischen Flüsterdolmetschen, Konsekutivdolmetschen und Simultandolmetschen. Im zweiten Fall hingegen werden live übertragene Veranstaltungen wie politische Reden oder Zeremonien simultan gedolmetscht, oft unter Verwendung der Voice-over-Technik (vgl. Pöchhacker 2010b: 224).

Die Techniken, die im Fernsehen zur Übertragung von Sendungen und Beiträgen in andere Sprachen als die Originalsprache verwendet werden, sind zahlreich, darunter die Synchronisation, die Untertitelung, das Voice-over, das Dolmetschen und die Live-Untertitelung (vgl. Elsagir 2000: 111). Im Bereich des Mediendolmetschens werden Reden überwiegend simultan unter Verwendung der Voice-over-Technik gedolmetscht (vgl. Pöchhacker 2010b: 224). Diese Technik wird in den meisten Ländern der Welt für die Neuvertonung von nicht-fiktionalen Programmen verwendet und als eine „technique in which a voice offering a translation in a given target language (TL) is heard simultaneously on top of the source language (SL) voice“ definiert (Díaz-Cintas & Orero 2006: 477). Die Lautstärke des Originals wird also verringert, damit die Verdolmetschung von den (TV-) Zuschauer*innen gehört werden kann, aber das Original dennoch im Hintergrund zu vernehmen ist (vgl. Díaz-Cintas & Orero 2006: 477).

Diese Technik wird in dem für diese Masterarbeit durchgeführten Experiment verwendet (siehe Kapitel 5). Eine Grundvoraussetzung für die Teilnahme an dem Experiment war, dass die Teilnehmenden mindestens einmal einer Dolmetschung in den Medien beigewohnt haben. Diese Form des Dolmetschens richtet sich in der Regel nämlich an ein breites, massenmediales Publikum – im Gegensatz zum Konferenzdolmetschen, das meist für einen begrenzten Kreis von Beteiligten bestimmt ist. Zur Bewertung der Qualität der für das Experiment durchgeführten Verdolmetschung war es nicht möglich, gezielt eine Personengruppe zu rekrutieren, die bereits an Konferenzen mit Dolmetscheinsatz teilgenommen hatte. Es erwies sich als deutlich naheliegender, Personen zu gewinnen, die zumindest einmal eine Medienverdolmetschung erlebt haben und somit – wenn auch nur oberflächlich – mit der Tätigkeit des Dolmetschens vertraut sind.

Das folgende Kapitel widmet sich der Qualität beim Simultandolmetschen. Es beleuchtet unterschiedliche Perspektiven, aus denen Qualität betrachtet werden kann, stellt

zentrale Qualitätskriterien vor und bietet einen Überblick über relevante Studien auf diesem Gebiet.

2 Qualität beim Simultandolmetschen

Qualität ist ein grundlegendes Konzept im Bereich des Dolmetschens. Ab den 1950er Jahren hat es dank der Organisation des Berufs auf internationaler Ebene an Bedeutung gewonnen (vgl. Zwischenberger 2010: 127). In den 1980er Jahren wurde es dann zu einem wichtigen Forschungsgegenstand der Dolmetschwissenschaft (vgl. Pöchhacker 2016: 173). Allerdings besteht noch keine Einigkeit über die Definition dieses Begriffs (vgl. Zwischenberger 2010: 128). Der internationale Verband der Konferenzdolmetscher*innen AIIC definiert Qualität als „that elusive something which everyone recognises but no one can successfully define“ (AIIC 1982: 1 zit. n. Zwischenberger 2010: 128). Qualität kann aus verschiedenen Perspektiven analysiert werden und stellt keinen absoluten Wert dar, sondern wird durch den Kontext bestimmt, in dem sie analysiert wird, wie beispielsweise das Analyse-Setting, der berufliche Hintergrund der Personen, die sie bewerten, und so weiter (vgl. Kopczynski 1994: 88). Für Kurz (2001) ist die Qualität beim Dolmetschen abhängig von der Zufriedenheit der Nutzer*innen, also derjenigen, die den Dolmetschdienst in Anspruch nehmen (vgl. 2001: 405). Das Hauptziel des Berufs des Dolmetschens besteht darin, eine effektive Kommunikation zwischen Sprecher*innen und Zuhörer*innen zu gewährleisten. Daher ist es wichtig, die Qualität der Verdolmetschung aus der Sicht der Zuhörer*innen zu untersuchen, die die Endnutzer*innen der Dienstleistung sind (vgl. Marrone 1993: 35). Nach Kurz (2001) wird die Qualität eines Dienstes an seiner Fähigkeit gemessen, die Bedürfnisse der Kund*innen zu erfüllen oder zu übertreffen. Wenn dies der Fall ist, kann die Dolmetschleistung als qualitativ hochwertig bezeichnet werden (vgl. 2001: 404). In Übereinstimmung mit diesem Prinzip wird in der vorliegenden Arbeit die Dolmetschleistung von den Nutzer*innen selbst bewertet.

Darüber hinaus wird Qualität im Rahmen dieser Arbeit auf das „Textprodukt“ bezogen, eine typische Sichtweise auf Qualität, die im Bereich des Simultandolmetschens zum Einsatz kommt. Unter diesem Begriff bezeichnet man die Verdolmetschung, wie sie von Rezipient*innen im Moment des Hörens aufgenommen und verarbeitet wird. Ihre Wahrnehmung kann dabei in Abhängigkeit von verschiedenen Kriterien variieren, die sowohl mit dem Inhalt als auch mit der Darbietung zusammenhängen (vgl. Gile 2004: 13). Laut Shlesinger (1997) sollte der Zieltext beim Konferenzdolmetschen auf drei Ebenen analysiert werden: *intertextually*, *intratextually* und *instrumentally* (1997: 128). Im ersten Fall sollte der AT mit dem ZT verglichen werden, wobei Ähnlichkeiten und Unterschiede berücksichtigt werden müssen. Im zweiten Fall sollte der ZT als eigenständiges Produkt betrachtet werden, wobei akustische, sprachliche und logische Aspekte bewertet werden. Im letzten Fall wird der

ZT als Dienstleistung für die Nutzer*innen analysiert, die sie in Anspruch nehmen (vgl. Shlesinger 1997: 128). Diese Sichtweise von Shlesinger (1997) zu den Elementen, die bei der Bewertung der Qualität im Dolmetschen zu berücksichtigen sind, ist kohärenter als andere Qualitätsmodelle. Sie vermeidet die Verwendung vager Begriffe wie „Äquivalenz“ und analysiert die Beziehung zwischen Ausgangstext und Zieltext, betrachtet dabei die intrinsischen Qualitäten des Zieltextes sowie seine tatsächliche Verwendung und Funktion. (vgl. Garzone 2002: 108).

Die Standpunkte, von denen aus man die Qualität beim Dolmetschen untersuchen kann, sind vielfältig und reichen von Dolmetscher*innen über Benutzer*innen bis hin zu den Kund*innen. Um zu überprüfen, ob die Zufriedenheit der Nutzer*innen bzw. Zuhörer*innen erreicht wurde, wird die Dolmetschleistung auf der Grundlage von Qualitätsstandards und -parametern bewertet.

2.1 Qualitätsstandards und Parameter

Professionelle Standards verleihen einer Tätigkeit ihren professionellen Charakter und spielen insbesondere im Bereich des Dolmetschens eine entscheidende Rolle. Organisationen wie die AIIC setzen sich dafür ein, solche Standards sowie die Qualität der beruflichen Praxis zu fördern und zu erhalten (vgl. Kurz 2001: 395). Laut Déjean le Féal (1990) gibt es weitgehend anerkannte Qualitätsstandards, deren Einhaltung dazu führt, dass eine Dolmetschleistung als professionell definiert wird. Beispielsweise sollten die Zuhörer*innen denselben Effekt des gedolmetschten Gesprächs erleben wie beim Original, mit gleicher Klarheit, Präzision und sprachlicher Qualität (vgl. Déjean le Féal 1990: 155).

Pöchhacker (2002) entwickelte das Modell der Qualitätsstandards, das in Abbildung 2 dargestellt ist und sich auf die Überlegungen von Viezzi (1996) bezieht. Nach diesem Modell kann das Dolmetschen als eine Tätigkeit betrachtet werden, die mit dem Textprodukt verbunden sein kann, oder als eine Dienstleistung, die in einer bestimmten Kommunikationssituation angeboten wird (vgl. Viezzi 1996: 40). Das Modell entwickelt sich zwischen zwei Polen. Auf der einen Seite wird die Qualität des Dolmetschens als Textprodukt betrachtet. Diese Sichtweise misst die Qualität des Dolmetschens anhand der Genauigkeit und Präzision des Zieltextes in Bezug auf die Wiedergabe des Ausgangstextes. Auf der anderen Seite wird das Dolmetschen als kommunikativer Akt angesehen, bei dem die Qualität anhand des Kommunikationserfolgs der Leistung gemessen wird. Zwischen diesen beiden Extremen liegen die Qualität, die sich an den Zuhörer*innen orientiert, und die Qualität, die als getreue

Wiedergabe der Aussagen der Sprecher*innen verstanden wird, wobei im Zieltext die gleichwertige Wirkung des Ausgangstextes erzielt wird (vgl. Pöchhacker 2002: 96f.).

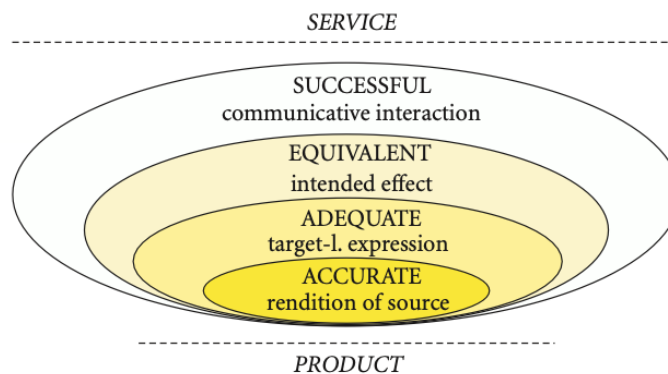


Abb. 2: Grafische Darstellung des Modells der Qualitätsstandards (Pöchhacker 2002: 97)

Um der beruflichen Praxis Orientierung zu bieten und Bezugspunkte für die Bewertung von Dolmetschleistungen zu liefern, wurden Qualitätskriterien erarbeitet (vgl. Viezzi 1996: 12). Die Festlegung solcher Kriterien erwies sich als notwendig, als man begann, sich davon zu lösen, die Qualität des Dolmetschens mit dem durch eine Universitätsausbildung erworbenen professionellen Status gleichzusetzen (vgl. Pöchhacker 2016: 173). In der Dolmetschforschung besteht Einigkeit über die Kriterien, die zur Bewertung der Qualität in den verschiedenen Interaktionsbereichen herangezogen werden. Zu den Standardkriterien, die in der Literatur zur produktorientierten Perspektive vorkommen, gehören beispielsweise Genauigkeit, Klarheit und Treue (vgl. Pöchhacker 2002: 96). Diese Kriterien stellen jedoch nur einen Teilaspekt der Gesamtqualität dar. Garzone (2002) schlägt beispielsweise einen umfassenden Ansatz für die Bewertung der Qualität beim Simultandolmetschen vor. Obwohl Genauigkeit und Treue zu den grundlegenden, unter Dolmetscher*innen anerkannten Qualitätskriterien gehören, hält Garzone (2002) es für notwendig, weitere Faktoren zu berücksichtigen, die sich auf die Qualität der Leistung auswirken. Diese betreffen sowohl die Dolmetscher*innen, d. h. ihre sprachlichen und technischen Kompetenzen, als auch das Ausgangsmaterial, wie die Aussprache der Sprecher*innen usw. (vgl. 2002: 109). Barik (1971) ist ebenfalls der Meinung, dass für eine genauere Qualitätsbewertung weitere wichtige Faktoren wie die Intonation und die Sprechpausen berücksichtigt werden sollten (vgl. 1971: 207).

2.1.1 Die Kriterien von Bühler (1986)

Bühlers Studie aus dem Jahr 1986 ist eine der ersten Arbeiten, die wesentlich dazu beigetragen haben, Kriterien für die Bewertung der Qualität von Konferenzdolmetscher*innen zu

entwickeln. Es wurden Fragebögen an 47 AIIC-Mitglieder, darunter sieben Mitglieder des CACL (Committee on Admissions and Language Classification) verteilt. Darin sollten die Mitglieder 16 Kriterien auf einer vierstufigen Skala danach bewerten, wie wichtig sie diese bei der Bewerbung von Kandidat*innen für die AIIC-Mitgliedschaft einschätzten. Die 16 Kriterien, welche in sprachliche und außersprachliche unterteilt wurden, sind zu einer Art Eckpfeiler der empirischen Qualitätsforschung im Bereich Konferenzdolmetschen geworden (vgl. Pöchhacker & Zwischenberger 2010: 2). Zu den sprachlichen Kriterien im Zusammenhang mit der semantischen Ebene zählten „native accent, fluency of delivery, logical cohesion of utterance, sense consistency with original message, completeness of interpretation, correct grammatical usage, use of correct terminology, use of appropriate style“ (Bühler 1986: 234). Zu den außersprachlichen Kriterien, die mit dem Bereich der Pragmatik zusammenhängen, gehörten „pleasant voice, thorough preparation of Conference documents, endurance, poise, pleasant appearance, reliability, ability to work in a team, positive feedback from delegates“ (1986: 234). Das Kriterium, das die höchste Bewertung erhielt, war „sense consistency with original message“ (Sinnübereinstimmung mit dem Original), das von 96% der Teilnehmer*innen als sehr wichtig eingestuft wurde (vgl. 1986: 231). Was die außersprachlichen Kriterien betrifft, wurde „pleasant voice“ (angenehme Stimme) in den meisten Fällen als wünschenswert, aber nicht als wesentlich eingestuft (vgl. 1986: 233). „Fluency of delivery“ (Flüssigkeit der Wiedergabe) wurde von 10% der CACL-Mitglieder und 49% der Gesamtgruppe als sehr wichtig bewertet (vgl. 1986: 232). Dieses Kriterium wurde den sprachlichen zugeordnet, gehört aber, wie in Kapitel 3 dargelegt, zu den prosodischen Merkmalen.

Bühler (1986) stellte in dieser Studie die Hypothese auf, dass die Erwartungen der Dolmetscher*innen in Bezug auf die Qualitätskriterien mit denen der Nutzer*innen übereinstimmen (vgl. 1986: 231). Darüber hinaus stellte sie die Annahme auf, dass die Nutzer*innen die Qualität einer Verdolmetschung anhand von Kriterien beurteilen würden, die sie als „oberflächlich“ bezeichnete, wie „native accent“ (muttersprachlicher Akzent), angenehme Stimme und Flüssigkeit der Wiedergabe, weil ihnen der Ausgangstext nicht bekannt ist (vgl. 1986: 233). Diese Hypothesen wurde jedoch durch eine Studie von Kurz aus dem Jahr 1989 in Frage gestellt (siehe Kapitel 2.2.2.1).

2.1.2 Die Bedeutung nonverbaler Parameter

Diese Masterarbeit konzentriert sich auf die Bewertung der Qualität einer Simultandolmetschleistung und betrachtet dabei die Intonation als Variable, die in zwei

Varianten auftritt. Die Intonation gehört zu den prosodischen, nonverbalen Elementen der Sprache, die mit der Vortragsweise der Dolmetschleistung zu tun haben. Die in der Forschung am häufigsten genannten Kriterien im Zusammenhang mit der Darbietungsweise umfassen eine angenehme Stimme, einen muttersprachlichen Akzent, die Intonation und die Flüssigkeit der Wiedergabe. Diese Faktoren tragen dazu bei, dass die Dolmetschung angenehmer zu verfolgen ist, und werden als Indikatoren für die Kompetenz und Zuverlässigkeit der Dolmetscher*innen wahrgenommen, die in der Lage sind, eine angemessene Wiedergabe des Ausgangstextes zu liefern (vgl. Garzone 2003: 28). Traditionell wird diesen Kriterien jedoch von Nutzer*innen und Dolmetscher*innen wenig Bedeutung beigemessen (vgl. Zwischenberger 2010: 127). Wie aus der oben beschriebenen Studie von Bühler (1986) hervorgeht, werden beim Konferenzdolmetschen inhaltsbezogene Kriterien wie die Sinnübereinstimmung mit dem Original oder „logical cohesion“ (Kohärenz) als die relevantesten für die Bewertung einer Dolmetschleistung angesehen.

Eine Ausnahme bildet das Mediendolmetschen: ein Bereich, in dem die Art und Weise der Informationsvermittlung, also die Form, teilweise eine größere Bedeutung hat als der Inhalt selbst (vgl. Mack 2002: 209). Die Qualitätsanforderungen an Dolmetscher*innen, die im Medienbereich tätig sind, sind sehr hoch und betreffen nicht nur die grammatikalische Korrektheit, sondern auch eine flüssige Präsentation und eine angenehme Stimme. Dies liegt auch daran, dass Fernsehzuschauer*innen, die daran gewöhnt sind, im Fernsehen fehlerfreie Reden zu hören, hohe Erwartungen an die Art und Weise haben, wie Informationen präsentiert werden (vgl. Kurz 1998: 311).

In den meisten Umfragestudien zu Erwartungen im Bereich des Konferenzdolmetschens wird die Tendenz bestätigt, inhaltsbezogene Kriterien höher zu bewerten als prosodische Kriterien. Allerdings stimmen die Erwartungen der Nutzer*innen hinsichtlich der Relevanz bestimmter Bewertungskriterien nicht immer mit ihren tatsächlichen Qualitätsbeurteilungen überein.

Darüber hinaus wird den darbietungsbezogenen Kriterien, also den Kriterien, die sich auf die Vortragsweise der Dolmetschleistung beziehen, im abstrakten Kontext wenig Bedeutung beigemessen. Ihre Relevanz tritt jedoch deutlich zutage, wenn die Gesamtqualität einer konkreten Leistung bewertet wird (vgl. Pöchhacker & Zwischenberger 2010: 3). Die Antworten in den Fragebögen spiegeln oft abstrakte Vorstellungen von Qualitätskriterien wider, die jedoch nicht mit einer in der Praxis umsetzbaren Definition von Qualität übereinstimmen. Wenn Fragen zu spezifischen Aspekten des Simultandolmetschens gestellt werden, drücken die Befragten Standpunkte aus, die auf Erfahrungen beruhen oder von

vorgefassten Ideen oder Einstellungen beeinflusst sind (vgl. Garzone 2003: 27). In Studien zur Qualität, die in den Unterkapiteln 2.2.1 und 2.2.2 beschrieben werden, ist dieser Faktor deutlich erkennbar. In den Studien zu den Erwartungen bewerten die Teilnehmer*innen die Qualitätskriterien auf der Grundlage von vorgefassten Meinungen und messen inhaltlichen Kriterien, die als wichtigste Qualitätsindikatoren gelten, große Bedeutung bei. Die darbietungsbezogenen Kriterien werden hingegen niedrig bewertet, wobei ihre Wichtigkeit von verschiedenen Faktoren abhängt, wie dem Setting, in dem sie bewertet werden, der Erfahrung der Teilnehmer*innen mit der Inanspruchnahme von Dolmetschdiensten und ihrem beruflichen Hintergrund und dem Kontext, in dem sie berücksichtigt werden.

In der Studie von Collados Aís aus dem Jahr 1998, die als Grundlage für diese Masterarbeit dient, wird deutlich, dass die Erwartungen der Teilnehmer*innen nicht mit der Bewertung einer konkreten Simultandolmetschleistung übereinstimmen. Die Studie analysierte die Diskrepanzen zwischen den Erwartungen von 42 spezialisierten Nutzer*innen und 15 professionellen Dolmetscher*innen und ihren tatsächlichen Bewertungen in Bezug auf eine Dolmetschleistung, die entweder mit monotoner oder lebhafter Intonation, beziehungsweise mit oder ohne inhaltliche Fehler, geliefert wurde. Die Ergebnisse zeigten, dass die Intonation, obwohl sie in dem Erwartungsfragebogen niedriger bewertet wurde als die inhaltlichen Kriterien, dennoch einen erheblichen Einfluss auf die Qualitätsbewertung hat. Die Teilnehmer*innen wurden von der monotonen Intonation beeinflusst und vergaben in dieser Version niedrigere Bewertungen für die Gesamtqualität, bemerkten jedoch nicht die inhaltlichen Fehler in der Version mit lebendiger Intonation (vgl. Collados Aís 1998). Die Studie von Collados Aís (1998) wird in Kapitel 4 ausführlich vorgestellt und zeigt, dass die darbietungsbezogenen Kriterien von Bedeutung sind und nicht „oberflächlich“, wie von Bühler (1986) festgehalten. Bei Dolmetschleistungen ist es daher wichtig, dass die Dolmetscher*innen nicht nur formale und inhaltliche Kriterien erfüllen, sondern auch die Standards, die mit der Darbietung zu tun haben, einhalten.

2.2 Qualitätsperspektive

Der Begriff der Qualität beim Dolmetschen ist vielschichtig und es gibt verschiedene Perspektiven, aus denen sie analysiert werden kann. Die beiden Perspektiven, auf die sich die Forschung am meisten konzentriert hat, ist jene der Dolmetscher*innen (INT) und jene der Nutzer*innen, d. h. der Zieltextempfänger*innen (TT-R). Zur Kategorie der Nutzer*innen gehören auch die Produzent*innen des Ausgangstextes (ST-P) als Sprecher*innen. Es gibt auch weitere Perspektiven, darunter die der Kund*innen (*client*), d. h. derjenigen, die den

Dolmetschauftrag erteilen, die von Gile (1991) in seinem „Configuration-Model“ eingeführt wurde (Gile 1991:189). Schließlich gibt es auch die Perspektive der externen Beobachter*innen, die von Kolleg*innen der Dolmetscher*innen (Coll.), von Partner*innen oder Vertreter*innen der Auftraggeber*innen oder Nutzer*innen sowie von Forscher*innen eingenommen wird, deren Ziel es ist, objektive Merkmale des Textprodukts zu messen (vgl. Pöchhacker 2001: 411f.). Diese Perspektiven sind in Abbildung 3 dargestellt.

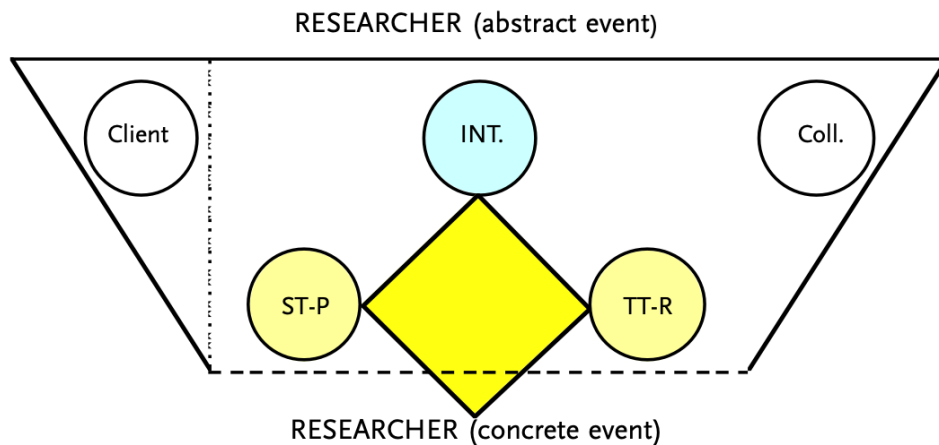


Abb. 3: Grafische Darstellung der Qualitätsperspektiven (Pöchhacker 2001: 42)

Die Forschung zur Qualität beim Dolmetschen unterscheidet außerdem zwischen einer produktorientierten und der interaktionsorientierten Perspektive. Es ist durchaus möglich, den Fokus auf das Endprodukt der Leistung zu richten oder auf den gesamten Prozess der kommunikativen Interaktion (vgl. Pöchhacker 2001: 412). Im Konferenzdolmetschen wird üblicherweise die produktorientierte Perspektive eingenommen, während in Community-basierten Bereichen, wo die Festlegung kohärenter beruflicher Standards noch im Gange ist, der Fokus auf den Fähigkeiten und Qualifikationen der Dolmetscher*innen liegt (vgl. Pöchhacker 2016: 173). Die Forscher*innen können als externe Beobachter*innen außerdem selbst entscheiden, ob sie die Fähigkeiten, Bedürfnisse und Standpunkte der Beteiligten abstrakt oder konkret erforschen möchten. Im ersten Fall können sie dies unter Berücksichtigung einer hypothetischen oder konkreten Dolmetschleistung vornehmen, im zweiten Fall können sie ein kommunikatives Ereignis in einer bestimmten kommunikativen Situation analysieren und sich so dem Geschehen selbst unmittelbarer annähern, wie in Abbildung 3 dargestellt (vgl. Pöchhacker 2001: 412).

Die methodischen Ansätze, die für die Forschung zur Qualität beim Dolmetschen zum Einsatz kommen, sind vielfältig und reichen von Umfragen bis hin zu Experimenten. Es besteht keine Einigkeit darüber, welche Perspektive für die Analyse der Qualität einer Verdolmetschung besser geeignet ist: die der Dolmetscher*innen oder die der Nutzer*innen. Im Folgenden werden die wichtigsten Studien vorgestellt, die die Qualität beim Konferenzdolmetschen aus der Perspektive der Dolmetscher*innen oder der Nutzer*innen oder aus beiden Perspektiven analysiert haben. Der Schwerpunkt liegt dabei hauptsächlich auf der Perspektive der Nutzer*innen.

2.2.1 Die Perspektive der Dolmetscher*innen: Studien zu den Erwartungen

Die meisten Studien, die zur Perspektive der Dolmetscher*innen durchgeführt wurden, sind Umfragestudien, in denen ihre Erwartungen untersucht werden. Die in diesem Kapitel ausführlich behandelten Studien sind die Studien von Chiaro und Nocella (2004) und Zwischenberger (2013).

Chiaro und Nocella (2004) führten eine Folgestudie zu Bühlers Studie von 1986 durch. Sie stellten Schwächen im Forschungsdesign der Studie von Bühler (1986) fest, in der die meisten Kriterien als „wichtig“ oder „sehr wichtig“ eingestuft wurden, die Dolmetscher*innen jedoch Schwierigkeiten hatten, zwischen diesen Kriterien Prioritäten zu setzen (vgl. Chiaro & Nocella 2004: 282). Chiaro und Nocella führten eine Online-Umfrage unter 286 Dolmetscher*innen weltweit durch (vgl. 2004: 279). Ausgehend von Bühlers Kriterien mussten die Dolmetscher*innen keine Bewertung vornehmen, sondern stufen neun sprachliche Kriterien und acht zusätzliche außersprachliche Kriterien von den wichtigsten bis zu den am wenigsten wichtigen ein (vgl. 2004: 283f.). Die Kriterien, die als besonders wichtig angesehen wurden, waren inhaltsbezogene Kriterien, und zwar die Übereinstimmung mit dem Original, Vollständigkeit und Kohärenz. Die beiden darbietungsbezogenen Kriterien angenehme Stimme und muttersprachlicher Akzent wurden hingegen als am wenigsten wichtig angesehen (vgl. Chiaro & Nocella 2004: 287).

Zwischenberger führte zwischen 2008 und 2009 zwei webbasierte Befragungen unter Mitglieder*innen des Deutschen Verbands der Konferenzdolmetscher*innen (VKD) und der AIIC durch. Die beiden Umfragen werden in Zwischenbergers Dissertation aus dem Jahr 2013 ausführlich behandelt und sind Teil eines größeren Forschungsprojekts zum Thema Qualität beim Simultandolmetschen unter der Leitung von Franz Pöchhacker (vgl. Zwischenberger 2013: 13). Der Fragebogen bestand aus drei Teilen (vgl. 2013: 149). Im ersten Teil A wurden soziodemografische Hintergrundinformationen erhoben (vgl. 2013: 150). Der zweite Teil B

stellte eine Art Replikation der Studie von Bühler aus dem Jahr 1986 dar. Die Befragten sollten auf einer vierstufigen Skala von „sehr wichtig“ bis „unwichtig“ die relative Bedeutung von elf Qualitätskriterien für eine Simultanverdolmetschung bewerten. Die Kriterien wurden aus der Studie von Bühler (1986) abgeleitet, mit Schwerpunkt auf den von ihr definierten sprachlichen Kriterien: Flüssigkeit, korrekte Terminologie, korrekte Grammatik, Sinnübereinstimmung mit dem Original, muttersprachlicher Akzent, Kohärenz, angenehme Stimme, angemessener Stil sowie Vollständigkeit, zu denen noch Synchronität und lebendige Intonation hinzukamen (vgl. 2013: 151f.). Die Teilnehmer*innen sollten außerdem bewerten, ob die den Kriterien zugewiesene Bedeutung von der Art der Konferenz abhängt, bei der eine Simultanverdolmetschung stattfindet, wie beispielsweise große Vollversammlung, Seminar, Verhandlung, Pressekonferenz. Im abschließenden Teil B des Fragebogens sollten die Teilnehmenden eine etwa einminütige Hörprobe einer professionellen Simultandolmetschung anhören und diese auf einer sechsstufigen Skala bewerten. Diese Dolmetschleistung wurde in zwei verschiedenen Versionen für jede der beiden Befragungen durchgeführt: eine mit monotoner Intonation und eine mit lebhafter Intonation für die Umfrage unter den Mitgliedern der AIIC sowie eine flüssige und eine stockende Version für die Umfrage unter den VKD-Mitgliedern (vgl. 2013: 152f.). Die Teilnehmenden jeder der beiden Befragungen wurden in zwei Gruppen aufgeteilt, von denen jede eine der beiden Versionen hörte (vgl. 2013: 155). Der dritte und letzte Teil des Fragebogens „C“ befasste sich mit der Selbstwahrnehmung der Konferenzdolmetscher*innen in Bezug auf ihre Rolle (vgl. 2013: 158). Da sich die Untersuchung unter den AIIC-Mitgliedern mit der Manipulation der Intonation befasste, werden im Folgenden ausschließlich diese Ergebnisse betrachtet.

Eine Zusammenfassung der zentralen Ergebnisse der AIIC-Befragung wurde 2010 publiziert (vgl. Pöchlhammer & Zwischenberger 2010). In Bezug auf die Ergebnisse von Teil B des Fragebogens zu den Bewertungskriterien erhielten die Kriterien Sinnübereinstimmung mit dem Original und Kohärenz die höchsten Gesamtwerte (88,3 % bzw. 74,8 % der Teilnehmenden stuften sie als „sehr wichtig“ ein), in Einklang mit den Ergebnissen der Studie von Bühler aus dem Jahr 1986 (vgl. 2010: 5). Das Kriterium der lebhaften Intonation erhielt im Vergleich niedrige Bewertungen; nur 28,2 % der Teilnehmenden stuften es als „sehr wichtig“ ein (vgl. 2010: 4). Die Bedeutung der Kriterien Sinnübereinstimmung mit dem Original und Kohärenz wurde stets als absolut vorrangig angesehen und variierte auch nicht nach Art der Konferenz. Allerdings variierte die Bedeutung der form- und darbietungsbezogenen Kriterien, wenn sie mit konkreten Dolmetschsituationen in Verbindung gebracht wurden. So wurde beispielsweise die Synchronität, die insgesamt nur eine niedrige Bewertung erhielt (15,3 % der

Teilnehmenden stuften sie als „sehr wichtig“ ein), im Kontext von Medienveranstaltungen als besonders wichtig eingestuft (vgl. 2010: 6).

Im zweiten Teil des Fragebogens B, der sich auf die Audioprobe bezog, wird deutlich, welchen Einfluss die Intonation auf die Gesamtbewertung der Qualität der Verdolmetschung hat. Obwohl dieses Kriterium niedrige Bewertungen erhielt, erzielte die Version mit monotoner Intonation niedrigere Werte als die mit lebhafter Intonation, mit einem Durchschnitt von 3,68 bzw. 3,84 (vgl. 2010: 7). Wie bereits festgestellt, wird den darbietungsbezogenen Qualitätskriterien wie der Intonation in der Regel weniger Bedeutung beigemessen, wenn sie allgemein bewertet werden, aber sie spielen eine wichtige Rolle, wenn die Gesamtqualität einer bestimmten Dolmetschleistung beurteilt werden soll (vgl. Pöchhacker & Zwischenberger 2010: 2f.). Damit wird die Bedeutung empirisch gewonnener Erkenntnisse aus realen Dolmetschsituationen gegenüber theoretisch-abstrakten Erwartungshaltungen hervorgehoben.

2.2.2 Die Perspektive der Nutzer*innen

Die Nutzer*innen sind diejenigen, die die Dolmetschleistung in Anspruch nehmen, und aus diesem Grund sollten sie auch diejenigen sein, die sie beurteilen (vgl. Seleskovitch 1986: 236). Wie die Studien aus der Perspektive der Dolmetscher*innen sind auch die Studien aus der Perspektive der Nutzer*innen größtenteils Umfrage-Studien, zu denen sich eine Reihe von Studien gesellt, die von Collados Aís (1998) begonnen wurden und die die Methodik des Experiments anwenden. Einige von ihnen konzentrieren sich auf die Erwartungen der Nutzer*innen (Kopczynski 1994, Kurz 1989, 1993, Kurz & Pöchhacker 1995, Moser 1996), während andere die Bewertungen der Nutzer*innen in konkreten Dolmetschsituationen untersuchen oder versuchen, beide Aspekte zu kombinieren (Collados Aís 1998, Garzone 2003, Gile 1990). Aus den Ergebnissen dieser Untersuchungen wird –ähnlich wie bei den Studien, die die Perspektive der Dolmetscher*innen einnehmen – deutlich, dass inhaltliche Kriterien in der Regel am höchsten bewertet werden und ihre Bedeutung meistens konstant bleibt. Die Bedeutung von darbietungsbezogenen Kriterien unterliegt hingegen Abweichungen, abhängig vom Setting, in dem sie betrachtet werden, oder des beruflichen Hintergrunds der Nutzer*innen. Darüber hinaus stimmen die Erwartungen der Nutzer*innen oft nicht mit der konkreten Bewertung der Qualität der Verdolmetschung in einer realen Situation überein.

2.2.2.1 Umfrage-Studien zu Erwartungen

Kurz (1989, 1993) führte zahlreiche Studien zur Qualität aus der Perspektive der Nutzer*innen durch. Ihr Ziel war es, die Validität der Hypothese von Bühler (1986) zu überprüfen, wonach

die Erwartungen der Dolmetscher*innen hinsichtlich der Qualitätskriterien mit denen der Nutzer*innen übereinstimmen (vgl. Kurz 1989: 143). 1988 unterbreitete sie daher 47 Teilnehmer*innen eines medizinischen Kongresses einen Fragebogen, in dem sie die ersten acht der 16 von Bühler verwendeten Kriterien nach ihrer Bedeutung für die Qualität der Dolmetschung auf einer vierstufigen Skala bewerten sollten. Die Kriterien waren: muttersprachlicher Akzent, angenehme Stimme, Flüssigkeit, Kohärenz, Sinnübereinstimmung mit dem Original, Vollständigkeit, korrekte Verwendung der Grammatik, korrekte Verwendung der Terminologie (vgl. Kurz 1989: 144). Es lässt sich eine Korrelation zwischen dieser Studie und der von Bühler (1986) feststellen, was die Bedeutung der Kriterien korrekte Verwendung der Terminologie, Kohärenz und Sinnübereinstimmung mit dem Original betrifft, die insgesamt die besten Bewertungen erhielten. Die Kriterien muttersprachlicher Akzent, angenehme Stimme und korrekte Verwendung der Grammatik wurden hingegen von den Teilnehmer*innen dieser Studie niedriger bewertet als von den Befragten in der Studie von Bühler (1986). Nach den Ergebnissen dieser Studie beurteilen die Nutzer*innen also die Qualität nicht anhand der von Bühler als „oberflächlich“ definierten Kriterien (vgl. Kurz 1989: 146). Kurz vermutete, dass die in einem bestimmten Setting erzielten Ergebnisse für andere Benutzergruppen variieren könnten, und führte daher weitere Vergleichsstudien durch (vgl. Kurz 1993: 15). Der gleiche Fragebogen, der anlässlich der internationalen Konferenz über Allgemeinmedizin vorgelegt wurde, wurde auch anlässlich einer internationalen Konferenz über Qualitätskontrolle und einer Sitzung des Europarats ausgegeben. Die Ergebnisse wurden dann mit denen der AIIC-Dolmetscher*innen in Bühlers Studie verglichen (vgl. 1993: 15). Es lässt sich feststellen, dass die Teilnehmer*innen dieser Studie im Vergleich zu den Dolmetscher*innen, die an der Studie von Bühler (1986) teilnahmen, weniger strenge Anforderungen an die Qualität der Verdolmetschung stellten (vgl. Kurz 1993: 20). Da es sich um drei verschiedene Benutzergruppen handelte (Delegierte des Europarates, Ärzt*innen und Ingenieur*innen), erwiesen sich die Bewertungsprofile als unterschiedlich. Wie in der Studie von 1989 waren die Kriterien, die die höchsten Bewertungen erhielten, Sinnübereinstimmung mit dem Original, Kohärenz und korrekte Verwendung der Terminologie, denen alle drei Gruppen etwa die gleiche Bedeutung beimaßen (vgl. 1993: 20). Die darbietungsbezogenen Kriterien erhielten dagegen erneut niedrigere Bewertungen als in der Studie von Bühler (1986). Diese Ergebnisse weichen von Bühler's Hypothese ab, wonach Nutzer*innen die Qualität der Dolmetschleistungen anhand des muttersprachlichen Akzents, der angenehmen Stimme und Flüssigkeit der Wiedergabe bewerten (vgl. Bühler 1986: 233). Die Wichtigkeit, die diesen Kriterien zugewiesen wurde, variierte zudem je nach Nutzergruppe und somit je nach deren

beruflichem Hintergrund (vgl. Kurz 1993: 20). So legten beispielsweise die Delegierten des Europarates mehr Wert auf die Flüssigkeit als Ärzt*innen und Ingenieur*innen, während Ärzt*innen der angenehmen Stimme und dem muttersprachlichen Akzent mehr Bedeutung beimaßen als die beiden anderen Gruppen (vgl. 1993: 19f). Dies unterstreicht, dass Qualität kein absoluter Wert ist, sondern von verschiedenen Faktoren bestimmt wird.

Ein weiteres anschauliches Beispiel dafür, wie sehr der berufliche Hintergrund der Nutzer*innen von Dolmetschleistungen die Wahrnehmung der Qualität beeinflusst, findet sich in der Studie von Kurz und Pöchhacker (1995). Sie verglichen die Ergebnisse der Studie von Kurz aus dem Jahr 1993 mit 19 TV-Fachleuten, die wiederum die Bedeutung der Kriterien von Bühler (1986) bewerten sollten. Die 19 Befragten maßen den darbietungsbezogenen Kriterien wie muttersprachlicher Akzent, angenehme Stimme und Flüssigkeit der Wiedergabe eine deutlich höhere Bedeutung bei als der Vollständigkeit (vgl. Kurz und Pöchhacker 1995: 351f.). Wie in Kapitel 1 dargelegt, ist im Medienbereich häufig die Art und Weise, wie Informationen übermittelt werden, wichtiger als der Inhalt selbst.

Im Jahr 1994 führte Kopczynski eine Studie durch, um die Einstellungen und Erwartungen von polnischen Nutzer*innen von Dolmetschdienstleistungen in Bezug auf Konferenzdolmetschungen anhand von selbst formulierten Kriterien zu ermitteln. Der Fragebogen richtete sich sowohl an Dolmetscherdienstnutzer*innen als auch an Redner*innen und wurde an Personen verteilt, die schon an internationalen Verhandlungen oder Konferenzen teilgenommen hatten. Insgesamt nahmen 57 Personen aus drei Berufsgruppen teil: 20 aus den Geisteswissenschaften, 23 aus Wissenschaft und Technik sowie 14 Diplomaten*innen (vgl. 1994: 91). Im ersten Teil des Fragebogens sollten die Befragten angeben, welche Funktion ihrer Meinung nach beim Konferenzdolmetschen am wichtigsten ist: „rendering the general content of T1, rendering the detailed content of T1, terminological precision, style, grammatical correctness of utterances, fluency of delivery, diction, voice qualities“ (1994: 92). Im zweiten Teil sollten sie den Faktor angeben, den sie als am stärksten störend empfanden, und eine Reihe von Störfaktoren bewerten, darunter: „faulty terminology, ungrammatical sentences, stylistic mistakes, incomplete sentences, lack of fluency poor diction, monotonous intonation, monotonous tempo, speeding up and slowing down - too general rendition of content - too detailed rendition of content“ (1994: 92). Im dritten und letzten Teil des Fragebogens wurden die Teilnehmenden zur Rolle der Dolmetscher*innen befragt und dazu, ob diese eine aktive Rolle einnehmen sollten. Die Ergebnisse hinsichtlich der Funktionen waren einheitlich, wobei der Inhalt gegenüber der Form bevorzugt wurde und „detailed content“ (inhaltliche Ausführlichkeit) und „terminological precision“ (terminologische Präzision) als oberste

Prioritäten genannt wurden. Der wichtigste Störfaktor für alle drei Gruppen war „incorrect terminology“, d.h. falsche Terminologie (vgl. 1994: 98). In dieser Studie zu Erwartungen wurde die monotone Intonation nicht als wesentlicher irritierender Faktor angesehen und erhielt bei den Sprecher*innen überwiegend die Bewertungsstufen 2 (10,5 %), und 3 (5,2 %). Bei den Empfänger*innen entfielen 5,3 % der Bewertung auf Stufe 2 und 2,6 % auf Stufe 3 (vgl. 1994: 95).

1996 veröffentlichte Moser eine Studie, die untersuchen sollte, ob unterschiedliche Nutzergruppen unterschiedliche Erwartungen an das Dolmetschen haben. Unter anderem wollte er herausfinden, ob diese Erwartungen je nach Art der Konferenz variieren oder ob es allgemeine Erwartungen gibt, die von der Art der Konferenz unabhängig sind. Während 84 verschiedenen Sitzungen wurden im Jahr 1995 insgesamt 201 standardisierte Interviews von 94 Dolmetscher*innen anhand eines Fragebogens durchgeführt. Die Befragten fungierten sowohl als Redner*innen als auch als Zuhörer*innen (vgl. Moser 1996: 149f.). Die Konferenzen ließen sich in vier Typen einteilen: große und kleine technische Konferenzen sowie große und kleine allgemeine Konferenzen (1996: 152). Die Ergebnisse zeigten, dass die Erwartungen je nach verschiedenen Faktoren wie Art der Sitzung, Geschlecht, Alter und vorherige Erfahrungen mit dem Simultandolmetschen variierten. Die Erwartung, der die größte Bedeutung beigemessen wurde, war die originalgetreue Wiedergabe, mehr noch als die Synchronität, die rhetorischen Fähigkeiten und die Qualität der Stimme (vgl. 1996: 154). Dennoch betrachteten 80% der Teilnehmer*innen eine lebhafte und animierte Stimme als wichtigen Faktor, wobei diese Einschätzung mit dem Alter tendenziell zunahm, jedoch mit zunehmender Erfahrung der Teilnehmenden abnahm. Nutzer*innen, die regelmäßig an Konferenzen mit hoher spontaner Interaktion teilnahmen, schätzten besonders lebhafte Dolmetscher*innen und empfanden eine monotone Intonation als eher störend (vgl. 1996: 173). Die Kriterien, die in den verschiedenen Konferenzsituationen weitgehend konstant blieben, waren Klarheit des Ausdrucks, Vollständigkeit und korrekte Terminologie (vgl. 1996: 174).

2.2.2.2 Umfrage-Studien zu Erwartungen und Qualitätsbewertung

Die meisten bisher vorgestellten Studien konzentrierten sich lediglich auf die Untersuchung der abstrakten Erwartungen der Nutzer*innen. Es ist jedoch offensichtlich, dass diesen einzelnen Kriterien eine bestimmte Bedeutung beigemessen wird, die isoliert betrachtet wird, während im tatsächlichen Bewertungsprozess verschiedene Faktoren ins Spiel kommen, die miteinander interagieren und zu unterschiedlichen Ergebnissen führen (vgl. Garzone 2003: 28).

Im Jahr 1990 führte Gile eine Untersuchung durch, um zu ermitteln, ob die Erwartungen der Nutzer*innen mit der Bewertung einer konkreten Dolmetschleistung übereinstimmten. Zu diesem Zweck unternahm er eine Fallstudie, in der er 18 französischsprachigen und 5 amerikanischen Teilnehmer*innen einer zweisprachigen Konferenz (Französisch-Englisch) zum Thema genetische Augenheilkunde einen Fragebogen vorlegte. Die Nutzer*innen sollten auf einer Skala von eins bis fünf die allgemeine Qualität, die terminologische Qualität, die Sinnübereinstimmung mit dem Original, die Stimme, das Tempo und die Intonation bewerten und sich zu eventuellen Schwächen der Dolmetschleistung äußern. Die Ergebnisse waren einheitlich, mit Ausnahme des Kriteriums Stimme, das von den französischsprachigen Nutzern*innen als „schlecht“ und „sehr schlecht“ bewertet wurde, wobei sie einen der beiden Dolmetscher*innen deutlich bevorzugten (vgl. Gile 1990: 67). Trotz seiner negativen Bewertung hatte dieses Kriterium keinen Einfluss auf die Gesamtbewertung der Qualität der Dolmetschleistung, die als „sehr gut“ eingestuft wurde. Die prosodischen Elemente spielten in dieser Studie also keine wesentliche Rolle. Dies steht im Gegensatz zu den Ergebnissen der Studie von Bühler aus dem Jahr 1986, in der Flüssigkeit der Wiedergabe von 100% der CACL-Mitglieder als „sehr wichtig“ eingestuft wurde, sowie zu den Ergebnissen der Studie von Kurz aus dem Jahr 1989. In dieser Studie wurden Stimme und Flüssigkeit der Wiedergabe als relevant betrachtet, obwohl sie keine besonders hohen Bewertungen erhielten, da sie von nur 17% bzw. 28% der Teilnehmer*innen als „sehr wichtig“ eingestuft wurden. Gile formuliert die Hypothese, dass Menschen, die im wissenschaftlichen Bereich tätig sind, wie die Teilnehmer*innen dieser Fallstudie, weniger empfindlich auf die Qualität der Stimme, den Rhythmus und die Intonation der Dolmetschleistung reagieren als andere Benutzergruppen (vgl. Gile 1990: 68).

Zu den Umfrage-Studien zu Erwartungen und Qualitätsbewertung gehört auch die 1998 veröffentlichte Studie von Collados Aís. Diese Studie, die in Kapitel 4 ausführlich vorgestellt wird, zeigt, dass die Erwartungen der Nutzer*innen nicht mit ihrer Bewertung einer tatsächlichen Leistung übereinstimmen. In diesem Experiment, in dem zwei Versionen derselben Dolmetschung präsentiert wurden – eine monotone, aber korrekte und eine lebhafte, aber fehlerhafte –, hatte das Kriterium der Intonation einen erheblichen Einfluss auf die Bewertung der Qualität. Die Version, die die besten Bewertungen erhielt, war die lebhafte, aber fehlerhafte. Dieses Ergebnis stand im Widerspruch zu den Antworten der Nutzer*innen, die in der Umfrage zu ihren Erwartungen die Intonation niedrig und die inhaltlichen Kriterien hoch bewertet hatten (vgl. Collados Aís 1998).

Garzone (2003) zeigte ebenfalls, wie prosodische Elemente die Bewertung der Qualität einer Dolmetschleistung beeinflussen können. Sie analysierte, wie 16 Fachleute, darunter Ärzt*innen und Ingenieur*innen, die bereits Simultandolmetschdienste in Anspruch genommen hatten, die Qualität einer Dolmetschleistung bewerteten. In der ersten Phase äußerten die Teilnehmer*innen ihre Erwartungen anhand von vier Bewertungskriterien aus Bühlers Studie (1986) und signalisierten eine Präferenz für inhaltliche Aspekte (Treue und Kohärenz) im Vergleich zu formalen Aspekten (Flüssigkeit und angenehme Stimme). In der nächsten Phase, in der zwei Verdolmetschungen desselben Textes bewertet wurden – eine flüssig und gut artikuliert, die andere durch Zögern und eine unregelmäßige Intonation gekennzeichnet –, zeigte sich jedoch, dass eine schlechte formale Wiedergabe ihre Bewertung anderer Aspekte der beiden Leistungen negativ beeinflusste (vgl. Garzone 2003: 26f.). Dies zeigt erneut eine Diskrepanz zwischen den angegebenen Erwartungen und den bewerteten Kriterien einer konkreten Leistung und verdeutlicht die Bedeutung prosodischer Faktoren beim Dolmetschen. Das nächste Kapitel konzentriert sich auf die nonverbalen Elemente der Sprache, wobei ein besonderer Schwerpunkt auf der Intonation liegt.

3 Intonation

Kommunikation ist ein grundlegender Bestandteil des menschlichen Lebens. Menschen sprechen nicht ohne Artikulation und in der Regel auch nicht ohne Stimmgebrauch. Diese Aspekte gehören zur Prosodie, die eine wesentliche Komponente der Sprache darstellt. Daraus folgt, dass es keine sprachliche Äußerung ohne Prosodie gibt (vgl. Reed 2010: 12). Die Prosodie spielt nicht nur eine entscheidende Rolle bei der Vermittlung von Bedeutung, sondern auch bei der Wahrnehmung der Sprecher*innen durch die Zuhörer*innen. Die Zuhörer*innen bewerten das Wissen, das Engagement und die Überzeugung von Redner*innen nicht nur anhand des Inhalts ihrer Äußerungen, sondern auch durch die Art und Weise, wie diese Äußerungen präsentiert werden (vgl. Brennan & Williams 1995: 383). Auch die Glaubwürdigkeit der Redner*innen wird anhand der nonverbalen Elemente der Sprache beurteilt (vgl. García Becerra 2011: 22). In der Kommunikation kommt ihnen daher eine große Bedeutung zu.

Im Bereich der Kommunikation unterscheidet Poyatos (1997) zwischen „verbal language“ (verbale Sprache), „paralanguage“ (Paralinguistik) und „kinesics“ (Kinesik) (1997: 249). Die verbale Sprache gliedert sich in zwei wesentliche Ebenen: die segmentale Ebene, die durch Phoneme und Morpheme geprägt ist und auf der Wörter sowie Sätze gebildet werden, und die suprasegmentale Ebene, zu der unter anderem auch die Intonation gehört. Diese umfasst Intonationsmuster, die an sich keine referenzielle Bedeutung besitzen, es sei denn, sie spezifizieren oder definieren das lexikalische Konstrukt (vgl. Poyatos 1993: 129). Paralinguistik, auch als paraverbale Sprache bekannt, bezieht sich auf “nonverbal voice qualities, modifiers and independent sounds and silences with which we support or contradict the simultaneous or alternating linguistic and kinesic structures” (Poyatos 1993: 130). Innerhalb der Paralinguistik unterscheidet man daher zwischen Qualitäten, die den suprasegmentalen Elementen der sprachlichen Struktur nahestehen, darunter „primary qualities“ wie Timbre, Lautstärke, Intonation, „qualifiers“ wie laryngeale und artikulatorische Kontrolle, „differentiators“ wie Lachen und Weinen, sowie „alternants“ wie Seufzer und andere Geräusche (vgl. Poyatos 1993: 130). Die Prosodie umfasst nonverbale Elemente der Sprache, nämlich tonale, dynamische und durationale Phänomene, die die mündliche Kommunikation beeinflussen. Tonale Elemente beziehen sich auf Intonation, dynamische Elemente betreffen Intensitätsvariationen wie Lautstärke und Rhythmus und durationale Elemente umfassen Dauer, Sprechgeschwindigkeit und Pausen. Es gibt auch „gemischte Phänomene“, die diese Parameter kombinieren, wie zum Beispiel Akzente und Junktoren (vgl. Ahrens 2004: 76f.).

Die Intonation gehört zu den tonalen Elementen. In der Fachliteratur finden sich zahlreiche Definitionen und Abgrenzungen des Begriffs „Intonation“ (vgl. Ahrens 2004: 78). In der linguistischen und phonetischen Literatur werden beispielsweise auch die prosodischen Merkmale von Rhythmus, Sprechtempo, Pausen, Dauer, Intensität und Stimmqualität unter dem Begriff Intonation zusammengefasst (vgl. Möbius 1993: 1). Um das Konzept der Intonation genauer zu behandeln, wird in dieser Arbeit die Intonation in ihrem engsten Sinne verstanden, nämlich als Tonhöhenverlauf einer sprachlichen Äußerung (Günther 1999: 62, vgl. Möbius 1993: 1; Peters 2021: 3). Der Tonhöhenverlauf bezieht sich auf die Schwankungen der Grundfrequenz der Stimme (F0), die das akustische Korrelat der Intonation darstellen (vgl. Möbius 1993: 1).

Beim Sprechen wird der Tonhöhenverlauf beziehungsweise die Äußerung von Sprecher*innen in mehrere Einheiten unterteilt, die sogenannten Intonationseinheiten. In der Intonationsforschung stellt diese Segmentierung eine zentrale Fragestellung dar, und die Definition dieser Einheiten variiert je nach Forschungsansatz (Ahrens 2004: 84). Sie werden beispielsweise von Möbius (1993) und Peters (2021) als „intonatorische Phrasen“ oder „Intonationsphrasen“ bezeichnet (Möbius 1993: 19; Peters 2021: 5). Laut Möbius sind diese Einheiten prosodische Segmente, die durch ihre eigene Intonation gekennzeichnet sind und von der semantischen und prosodischen Struktur abhängen, nicht jedoch von der Syntax. Lieberman (1966) hingegen argumentiert, dass diese Einheiten durch die Notwendigkeit der Sprecher*innen, Atempausen zu machen, bedingt sind (vgl. Lieberman 1966: 2).

Zur prosodischen Kategorie der durationalen Elemente gehören die Dauer, Sprechgeschwindigkeit und Pausen (vgl. Ahrens 2004: 99). Die Dauer als prosodisches Phänomen bezieht sich auf die Länge von Silben und Lauten (vgl. Reed 2010: 106). Die Lautdehnung kann als Verlangsamung des Sprachrhythmus entweder vor einer Pause oder am Ende einer Intonationseinheit auftreten. Die letztgenannte Art der Lautdehnung hat eine „delimitative“ Funktion, da sie den Abschluss einer Intonationseinheit markiert (vgl. Ahrens 2004: 100). Die Sprechgeschwindigkeit wird dann in Silben pro Sekunde gemessen (vgl. Reed 2010: 123). Pausen werden als Unterbrechungen im akustischen Signal des Sprachflusses definiert (vgl. Reed 2010: 163). Sie erfüllen verschiedene Funktionen: Unter anderem ermöglichen sie dem/der Sprecher*in eine Atempause und segmentieren den Sprechfluss (vgl. Ahrens 2004: 103f.). Wenn sie an den Grenzen von Intonationseinheiten auftreten, erfüllen sie eine delimitative Funktion, indem sie die verschiedenen Intonationseinheiten in der Sprache trennen (vgl. Goldman-Eisler 1972: 111).

Zu den dynamischen Phänomenen gehören Intensität und Rhythmus (vgl. Ahrens 2004: 97). Die Intensität wird akustisch als Lautheit wahrgenommen und gilt als schwer zu analysierender prosodischer Parameter, da ihre Wahrnehmung häufig mit der Wahrnehmung der Tonhöhe überlappt (vgl. Reed 2010: 179f.). Die Lautheit trägt zusammen mit der Dauer und der Intonation dazu bei, bestimmte Elemente in der Kommunikation hervorzuheben (vgl. Ahrens 2004: 98). Dieses Phänomen wird als Prominenz bezeichnet und verweist auf die Betonung eines Teils des Sprachsignals (vgl. Möbius 1993: 9). Der Rhythmus bezieht sich darauf, wie Sprache in regelmäßigen Abständen organisiert ist (vgl. Reed 2010: 139).

Die Interaktion unterschiedlicher tonaler, dynamischer und durationaler Faktoren führt zu gemischten Phänomenen wie Akzenten und Junktoren. Der Akzent bezieht sich auf die Hervorhebung von Silben beim Sprechen (vgl. Ahrens 2004: 107). Junktoren hingegen sind prosodische Eigenschaften, die in sprachlichen Ausdrücken eine delimitative Funktion haben; in längeren Ausdrücken fungieren sie als Begrenzer von Intonationseinheiten (vgl. Ahrens 2004: 109).

Ein Element, das zu den prosodischen Elementen zählt, aber keiner der angeführten Kategorien zugeordnet werden kann, ist der „Sprechfluss“. Einige Autor*innen betrachten ihn als die Art und Weise, in der Äußerungen artikuliert werden, die sowohl flüssig verlaufen als auch durch unnatürliche Pausen oder stockende Elemente unterbrochen sein können. Andere sehen darin das Zusammenspiel verschiedener zeitlicher Variablen der Sprache wie Pausen und beispielsweise der Klarheit der Stimme (vgl. Rennert 2010: 102). Dieses prosodische Element wird in der Qualitätsbewertung beim Simultandolmetschen als „Flüssigkeit der Wiedergabe“ bewertet, was sich auf den flüssigen Ausdruck einer Verdolmetschung bezieht (siehe Kapitel 2).

3.1 Intonatorische Gestaltung

Hinsichtlich der Struktur der Intonation unterscheidet Peters (2021) drei Formen der Intonatorischen Gestaltung: „Konturwahl“, „Phrasierung“ und „Akzentzuweisung“ (Peters 2021: 4). Diese Konzepte werden von Halliday (1967) als *tone*, *tonality* und *tonicity* bezeichnet (Halliday 1967: 19).

Die Wahl einer Intonationskontur bestimmt, wie Akzentsilben hervorgehoben werden (vgl. Peters 2021: 5). Die Intonationskonturen stellen den gesamten zeitlichen Verlauf des Parameters F0 in einer Äußerung dar (vgl. Möbius 1993: 9). Sie sind Folgen von Tönen, die mittels lokaler Tonhöhenbewegungen oder Tonhöhenstufen entstehen (vgl. Peters 2021: 3). Die Abschnitte einer Äußerung, in denen solche Töne auftreten, sind die Intonationsphrasen (vgl.

Möbius 1993: 18). Die Verwendung bestimmter Intonationskonturen, deren phonetische Realisierung, sowie die kommunikativen Funktionen, die diese Konturen erfüllen, können von Sprache zu Sprache variieren (vgl. Peters 2021: 3). Die Intonationskonturen, spielen zusammen mit dem Intensitätsverlauf und der Veränderung der Stimmqualität eine entscheidende Rolle für die Hörer*innen, da sie wesentliche Hinweise zur korrekten syntaktischen Gliederung von Äußerungen liefern (vgl. Möbius 1993: 19).

Unter Phrasierung versteht man demnach die Unterteilung von Äußerungen in zwei oder mehr Abschnitte (vgl. Möbius 1993: 19). Diese Segmente sind sowohl für die phonetische und prosodische Struktur als auch für die syntaktische Gliederung der Sätze von zentraler Bedeutung. Sie werden in Intonationsphrasen unterteilt, die den Bereich bestimmen, in dem die Intonationskonturen realisiert werden (vgl. Peters 2021: 4f.).

Die Akzentzuweisung bezieht sich letztendlich darauf, die sprachlichen Einheiten zu definieren, die in einer Intonationsphrase hervorgehoben werden. Sie ist entscheidend für die Intonation, da sie bestimmt, welche syntaktischen Elemente die jeweilige intonatorische Kontur erhalten (vgl. Peters 2021: 5).

3.2 Intonationsmodelle

In der Prosodieforschung wurden unterschiedliche theoretische Modelle zur abstrakten Beschreibung und Modellierung der Intonation entwickelt, wobei verschiedene Bezeichnungen zum Einsatz kommen, um dasselbe Konzept zu erfassen. (vgl. Ahrens 2004: 89). Zwei Theorien, auf die besonders hingewiesen wird, sind der „Tonsequenz-Ansatz“ und der „Superpositions-Ansatz“ (Möbius 1993: 51).

Der Tonsequenz-Ansatz, der von Pierrehumbert (1980) begründet und im Rahmen der autosegmentalen Phonologie weiterentwickelt wurde, ist ein lokal orientierter Ansatz (vgl. Günther 1999: 63). Gemäß dieser Theorie wird die Intonation ausschließlich auf der lokalen Ebene, d. h. auf der Wortebene, bestimmt (vgl. Ahrens 2004: 97). In diesem Modell besteht die Intonation aus einer Reihe tonaler Elemente (vgl. Ladd 1983: 40). Diese tonalen Sequenzen folgen bestimmten Realisierungsregeln, die ihre spezifische Anordnung bestimmen (vgl. Ahrens 2005: 95). Im Modell von Pierrehumbert wird eine Intonationsphrase als eine Folge von hohen (H) und tiefen (L) Tönen dargestellt (vgl. Pierrehumbert 1980: 65). Die Töne überschneiden sich nicht direkt, sondern treten im Verlauf der Äußerung nacheinander auf (vgl. Möbius 1993: 56). Laut dieser Theorie beschreiben die verschiedenen Arten von Tonsequenzen den Verlauf der Grundfrequenz einer Äußerung. Sie leiten sich aus der Funktion der Intonation ab und klassifizieren charakteristische Intonationsverläufe an Phrasengrenzen und in

akzentuierten Elementen. Man unterscheidet zwischen Grenztönen, Akzenttönen und Phrasentönen. Die Grenztöne markieren die Grenzen der Äußerungen und kennzeichnen die Modalität der Sätze. Sie sind mit der letzten Silbe in den jeweiligen Äußerungsabschnitten verbunden. Die Akzenttöne realisieren die Akzente und sind mit den betonten Silben verknüpft. Zusätzlich gibt es Phrasentöne, die den Intonationsverlauf zwischen Akzenttönen und Grenztönen gestalten (vgl. Günther 1999: 64).

Im Gegensatz zum Tonsequenz-Ansatz, der die Intonation als eine einfache Abfolge von unterscheidbaren Tönen, die aufeinander folgen, begreift, interpretiert der Superpositions-Ansatz den tonalen Tonhöhenverlauf einer Äußerung als ein komplexes Modell, das sich auf globale Tendenzen der Tonhöhe bezieht (vgl. Ahrens 2004: 96). Das deskriptive Inventar der Intonation besteht aus einer Reihe von komplexen Intonationskonturen, die unter anderem absteigend, aufsteigend-absteigend und konstant sein können. Dabei werden die Intonationskonturen ganzen Äußerungen oder deren Teilen, den sogenannten Tongruppen, zugewiesen, ohne dass festgelegt wird, auf welchen Silben die einzelnen Konturteile realisiert sind (vgl. Günther 1999: 64). Die Modelle, die auf dem Superpositions-Ansatz basieren, sind hierarchisch strukturiert. Laut diesen Modellen entsteht der Tonhöhenverlauf einer Äußerung durch die Überlagerung mehrerer Komponenten. In seinem Modell integriert Thorsen (1988) folgende Komponenten im dänischen Standard: An oberster Stelle steht der Text, der eine diskursabhängige intonatorische Gliederung hervorruft, die als „Textkontur“ bezeichnet wird. Darunter folgt die „Satzintonationskontur“, die den Einfluss des Satzes beziehungsweise der Äußerung widerspiegelt, sowie die „Phrasenkontur“, die die prosodische Phrase betrifft. Die letzte relevante Ebene wird durch Betonungsgruppenmuster repräsentiert. Diese vier Komponenten sind sprachspezifisch und können von den Sprecher*innen gezielt gesteuert werden (vgl. Thorsen 1988: 2).

Der Unterschied zwischen den beiden Ansätzen beruht hauptsächlich auf der Art und Weise, wie sie die Beziehung zwischen lokalen Bewegungen und globalen Verläufen in der Intonationskontur beschreiben (vgl. Thorsen 1985: 1207). Darüber hinaus erfordern die Superpositions-Modelle eine vorherige Planung bestimmter prosodischer Komponenten der Sprachproduktion, was sie von den Tonsequenz-Modellen unterscheidet. Es wird angenommen, dass die Sprecher*innen den globalen Tonhöhenverlauf im Voraus planen (vgl. Ahrens 2004: 1997).

Möbius (1993) argumentiert, dass die strikte Unterscheidung von Intonationsmodellen in zwei klare Kategorien – Tonsequenz-Modelle einerseits und hierarchische Modelle andererseits – sowohl theoretisch als auch methodologisch problematisch und nicht haltbar ist.

Eine genauere Analyse der verschiedenen Intonationsmodelle zeige, dass alle zumindest teilweise eine nicht-lokale Komponente enthalten. So weise beispielsweise auch das Modell von Pierrehumbert (1980), das auf sequenziellen Tönen basiert, hierarchische Elemente auf (Möbius 1993: 62). Daher sei es für die Analyse der Intonation geeigneter, sich auf Modelle zu beziehen, die sowohl lokale Elemente als auch globale Komponenten auf Satz- oder Phrasenebene umfassen (vgl. Möbius 1993: 64).

Dennoch eignen sich für Möbius (1993) die Superpositions-Modelle besonders gut für eine quantitative Analyse der Intonation. In diesen Modellen, darunter auch dem von Thorsen (1988), ist die Gesamttendenz der F0 das Ergebnis der Kombination mehrerer sich überlappender Komponenten, von denen jede mit einem bestimmten Aspekt der Sprache zusammenhängt (vgl. Möbius 1993: 65). Möbius (1993) greift auf das Modell von Fujisaki zurück (1983, 1988), das das Überlagerungsprinzip auf streng mathematische Weise anwendet, was ihm mehrere Vorteile verleiht. Da das Modell auf der Überlagerung mehrerer Komponenten beruht, ist es möglich, den Einfluss einzelner Faktoren auf bestimmte Zeitintervalle oder auf definierte sprachliche oder prosodische Einheiten, wie Akzentgruppen oder Sätze, genau zu analysieren (vgl. Möbius 1993: 65). Das bedeutet, dass an jeder Stelle einer Äußerung der Wert der Grundfrequenz (F0) berechnet werden kann, was quantitative Untersuchungen erleichtert (vgl. Möbius 1993: 66).

3.3 Akustische Korrelate der Intonation

Auf der Ebene der akustischen Struktur einer sprachlichen Äußerung lassen sich die auditiven Variablen grundsätzlich in zwei Kategorien unterteilen: Qualitätsmerkmale und dynamische Merkmale. Die akustischen Korrelate der auditiven Qualitätsmerkmale sind im Wesentlichen spektral und umfassen Aspekte wie Formantenfrequenzen und Amplituden. Im Gegensatz dazu gehört die Grundfrequenz (F0) des sprachlichen Signals, die als Korrelat der Tonhöhe dient, zu den dynamischen Merkmalen (vgl. Laver & Trudgill 1979: 5f.).

Es lassen sich drei verschiedene Arten von Schallsignalen in Bezug auf die Tonhöhe unterscheiden: Töne, Klänge und Geräusche. Ein Ton ist ein Schallereignis, das durch eine regelmäßige Schwingung erzeugt wird – er lässt sich am besten als eine reine Sinuswelle beschreiben. In diesem Zusammenhang spricht man auch von Sinustönen. Ein Klang setzt sich aus mehreren Tönen zusammen, die auch als Teiltöne oder Partialtöne bezeichnet werden. Ein Geräusch hingegen ist ein Schallereignis, das keine regelmäßigen Schwingungen aufweist. Geräusche können sich entweder über längere Zeiträume hinziehen (wie Rauschen) oder sehr plötzlich und impulsartig auftreten (wie ein Knall). In Abbildung 4 werden diese drei Arten von

Schallereignissen anhand von Oszillogrammen veranschaulicht. Ein Oszillogramm stellt die zeitliche Veränderung der elektrischen Spannung dar, die vom Mikrofon erfasst wird, und visualisiert somit die Schwankungen des Luftdrucks, die den Sprechschall repräsentieren. (vgl. Peters 2021: 13).

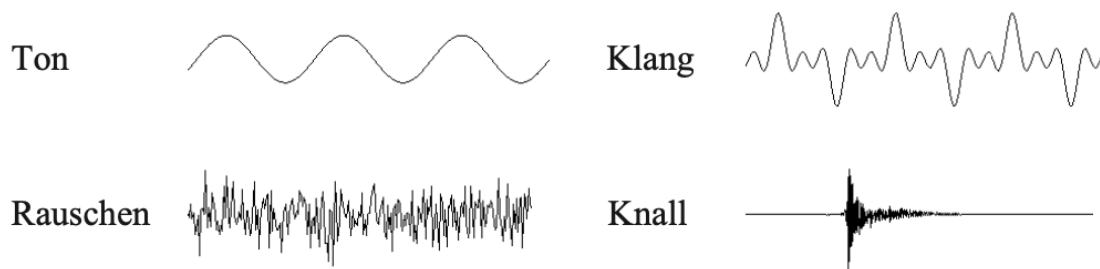


Abb. 4: Darstellung der Schallsignalen in Bezug auf die Tonhöhe, adaptiert nach Peters (2021: 13)

Der Sprechschall besteht aus harmonischen Klängen, die charakteristisch für Vokale sind, harmonischen Klängen mit Geräuschanteilen, die charakteristisch für stimmhafte Konsonanten sind, und Geräuschen, die charakteristisch für stimmlose Konsonanten sind (vgl. Peters 2021: 13). Die harmonischen Töne zeichnen sich dadurch aus, dass die Frequenzen ihrer verschiedenen Komponenten, die als Teiltöne bezeichnet werden, einem festen mathematischen Muster folgen. Im Sprechschall stehen die Frequenzen aller Teiltöne in einem einfachen, ganzzahligen Verhältnis zur Frequenz des tiefsten Teiltöns, der als Grundton bezeichnet wird. Die Frequenz dieses Grundtons ist die bereits genannte Grundfrequenz (F_0), die in Hertz (Hz) gemessen wird, was die Anzahl der Schwingungen pro Sekunde angibt. Alle anderen Teiltöne sind Vielfache der Grundfrequenz und werden als Obertöne bezeichnet. Wie in Abbildung 5 dargestellt, ist bei einem Vokal mit einer Grundfrequenz von 100 Hz zu erwarten, dass das Sprechsignal auch Frequenzen von 200 Hz, 300 Hz, 400 Hz, 500 Hz und so weiter enthält (vgl. Peters 2021: 14).

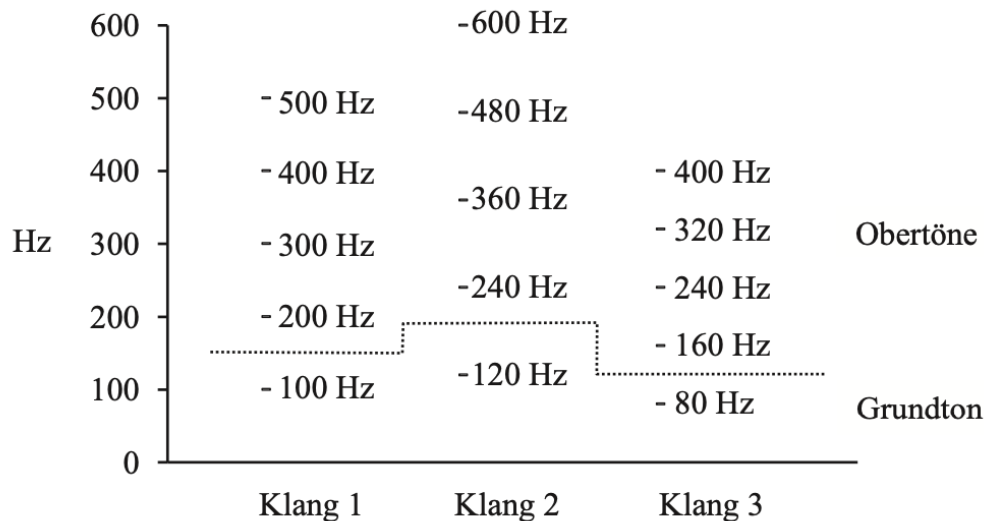


Abb. 5: Beispiel von harmonischen Klängen mit Grundfrequenzen (100 Hz, 120 Hz und 80 Hz) und Darstellung des Grundtons und der ersten vier Obertöne, adaptiert nach Peters (2021: 14)

Aus artikulatorischer Sicht entstehen die verschiedenen Melodieverläufe durch die Veränderung der Spannung der Stimmbänder und die Modifikation des subglottalen Drucks. Aus physiologischer Sicht ist die Grundfrequenz gleichzusetzen mit der Frequenz, mit der sich die Glottis bei der Phonation öffnet und schließt – also mit der Periodenfrequenz ihres Bewegungszyklus (vgl. Günther 1999: 63).

3.4 Die Wahrnehmung der Intonation

Die Grundfrequenz (F0) wird als das wichtigste akustische Korrelat für die wahrgenommene Tonhöhe betrachtet (vgl. Peters 2021: 14). Sie ist entscheidend für die Wahrnehmung und Strukturierung von Sprache. Durch die Veränderung der Spannung und Länge der Stimmbänder, das Heben oder Senken des Kehlkopfs und die Anpassung des subglottalen Drucks können Sprecher*innen die Periodizität der Stimmbandvibrationen beeinflussen und somit verschiedene Aspekte der Tonhöhe steuern. Diese Anpassungen tragen zur Wahrnehmung der syntaktischen Struktur eines Satzes bei, beeinflussen die Modalität, die Informationsstruktur, die Haltung und die Emotionen der Sprecher*innen sowie die Dialogsituation, etwa die kommunikative Absicht der Sprecher*innen oder ihre Intention, das Gespräch zu führen oder abzugeben (vgl. Vaissière 2005: 236). Laut Peters (2021) ist die Grundfrequenz zwar mit der Wahrnehmung der Tonhöhe verbunden, aber sie ist nicht unbedingt erforderlich, um die Tonhöhe zu erkennen. Im Allgemeinen wird eine Erhöhung der Grundfrequenz mit einer Wahrnehmung einer höheren Tonhöhe assoziiert, während eine

Abnahme der Grundfrequenz als eine Senkung der Tonhöhe wahrgenommen wird. Es sind jedoch zwei wesentliche Präzisierungen zu beachten: Einerseits beeinflussen mikroprosodische Schwankungen der Grundfrequenz nicht die Wahrnehmung der Tonhöhe, da sie in sehr kurzen Zeiträumen auftreten und zu gering sind, um wahrgenommen zu werden (vgl. Peters 2021: 14). Andererseits muss ein Schallereignis nicht zwingend eine explizite Grundfrequenz aufweisen, um eine bestimmte Tonhöhe wahrzunehmen. Vielmehr reicht es aus, wenn zwei nahe beieinander liegende Vielfache der Grundfrequenz vorhanden sind, die den Frequenzen der Harmonischen im Sprachsignal entsprechen (vgl. Peters 2021: 15). Auch wenn Teile des Frequenzspektrums durch andere Geräusche in der Umgebung verdeckt werden, wie es beim gleichzeitigen Sprechen der Fall ist, kann die Grundfrequenz dank des Verhältnisses der im Signal hörbaren Teiltöne noch rekonstruiert werden. Aus diesem Grund lässt sich feststellen, dass die Wahrnehmung der Intonation äußerst robust im Hinblick auf den Frequenzbereich ist (vgl. Peters 2021: 16).

3.5 Funktionen der Intonation

Die Intonation erfüllt in der mündlichen Kommunikation verschiedene Funktionen, da der Tonhöhenverlauf sowohl linguistische als auch paralinguistische Informationen transportiert (vgl. Ahrens 2004: 83). Aus linguistisch-grammatikalischer Sicht trägt die Intonation wesentlich zur Realisierung des Modus des Satzes bei, da sie eines seiner konstituierenden Elemente ist (vgl. Günther 1999: 63). Sie bestimmt, ob es sich zum Beispiel um einen Aussagesatz oder einen Fragesatz handelt (vgl. Pierrehumbert 1980: 7). Auch das Ende eines Sprachsegments oder einer Äußerung kann durch die Intonation signalisiert werden, beispielsweise durch einen abfallenden Melodieverlauf (vgl. Günther 1999: 63). Zudem unterstützt die Intonation die Zuhörer*innen beim Verstehen der gesprochenen Sprache, indem sie zur Segmentierung beziehungsweise Phrasierung der Äußerungen beiträgt. Die einzelnen Segmente der Intonationseinheiten weisen spezifische Tonmuster auf, die deren Abgrenzung und Strukturierung ermöglichen (vgl. Ahrens 2004: 83). Experimente haben gezeigt, dass die Verwendung „untypischer“ Intonationsmuster das Hörverständnis beeinträchtigen kann. (vgl. Braun et al. 2011: 366).

Aus paralinguistischer Sicht besitzt die Intonation eine expressive Funktion. (vgl. Ahrens 2004: 84). Sie wird verwendet, um Emotionen, Einstellungen und Absichten zu kommunizieren. Darüber hinaus dient die Intonation auf semantisch-pragmatischer Ebene unter anderem der Markierung der Informationsstruktur und der Lenkung der Aufmerksamkeit der Empfänger*innen (vgl. Barranco-Droege et al. 2011: 65). Schließlich hat die Intonation auch

eine soziale Funktion inne, die mit dem Beruf und der gesellschaftlichen Stellung der Sprecher*innen zusammenhängt (vgl. Ahrens 2004: 84). Wie bereits erwähnt, bewerten Zuhörer*innen das Wissen, das Engagement und die Überzeugung von Redner*innen nicht nur anhand des Inhalts ihrer Äußerungen, sondern auch anhand der Art und Weise, wie diese Äußerungen präsentiert werden (vgl. Brennan & Williams 1995: 383). Auch die Glaubwürdigkeit der Redner*innen wird anhand der nonverbalen Elemente der Sprache bewertet (vgl. García Becerra 2011: 22). Im folgenden Kapitel wird die Intonation in Bezug auf das Simultandolmetschen analysiert und ein Überblick über den aktuellen Forschungsstand in diesem Bereich vermittelt.

4 Intonation beim Simultandolmetschen

Das Dolmetschen zählt zu den Sprechberufen (vgl. Ahrens 2004: 113). Da prosodische Merkmale integraler Bestandteil der gesprochenen Sprache sind, spielt ihr Vorhandensein im Ausgangstext und ihre Wiedergabe durch die Dolmetscher*innen im Zieltext eine wesentliche Rolle. Prosodische Elemente ermöglichen ein besseres Verständnis der verbalen Elemente der Rede (vgl. Collados Aís 2008: 209). Außerdem sind sie sowohl im AT für das Verständnis der Dolmetscher*innen als auch im ZT für das Verständnis der Zuhörer*innen und für die Qualität der Verdolmetschung wichtig (vgl. Tissi 2000: 104). Aus diesem Grund müssen Dolmetscher*innen sich der suprasegmentalen Merkmale der Sprache bewusst sein (vgl. Darò 1990: 90). Es wird von Dolmetscher*innen erwartet, dass sie ein ausgeprägtes und kompetentes Prosodie-Management beherrschen, und sie sollten gemäß den AIIC-Standards keine monotone Prosodie einsetzen (vgl. Barranco-Droege et al. 2011: 76).

4.1 Merkmale der Intonation beim Simultandolmetschen

Die Intonation beim Dolmetschen zeichnet sich durch besondere Merkmale aus, die sich von denen der freien Sprechweise unterscheiden. Dies hat verschiedene Gründe. Die Prosodie kann in der Rezeptionsphase durch den Ausgangstext und damit durch die Intonation der Sprecher*innen beeinflusst werden, und zwar durch Satzabbrüche, Korrekturen und Interferenzen (vgl. Ahrens 2004: 117). Darüber hinaus kann sie durch den Dolmetschprozess selbst in der Produktionsphase und die damit zusammenhängenden kognitiven Anstrengungen beeinflusst werden (siehe Kapitel 1.1.4). Außerdem sprechen Dolmetscher*innen oft sehr schnell, was auf eine stärkere Aktivierung der rechten Gehirnhälfte hinweist, die für die Steuerung der suprasegmentalen Merkmale der Sprache verantwortlich ist. Daher könnte eine höhere Sprechgeschwindigkeit teilweise für Schwierigkeiten bei der Kontrolle von Intonation, Prosodie, Aussprache usw. verantwortlich sein (vgl. Darò 1994: 267).

Shlesinger (1994) definiert die Intonation beim Simultandolmetschen als eine Intonation *sui generis*, d. h. “marked by a set of salient features not found in any other language use” (Shlesinger 1994: 226). Die Autorin vergleicht die stimmliche Produktion beim Simultandolmetschen mit der des Vorlesens und stellt fest, dass es besondere Merkmale der Intonation von Dolmetscher*innen gibt. Sie verwendet den Begriff Intonation, obwohl sie ihn nicht nur als Tonhöhenverlauf einer sprachlichen Äußerung betrachtet, sondern Intonation in Anlehnung an die Kategorien von Halliday (1967) als die Gesamtheit von *tonality*, *tonicity* und *tone* definiert (Halliday 1967: 19). Die Studie zeigt, dass die prosodischen Merkmale von

Verdolmetschungen von den Erwartungen der Hörer*innen abweichen (vgl. Shlesinger 1994: 231). Beispielsweise neigen Dolmetscher*innen dazu, eine unverhältnismäßig große Anzahl von Pausen an „unnatürlichen“ Stellen einzufügen, und Dolmetschleistungen sind durch die Platzierung von Tonhöhenbewegungen an unnatürlichen Stellen und nicht standardmäßige Geschwindigkeitsänderungen gekennzeichnet (vgl. 1994: 230).

Auch Ahrens (2004) hat festgestellt, dass die Intonation beim Simultandolmetschen durch besondere Merkmale gekennzeichnet ist. Sie analysierte ein Korpus, das aus drei Verdolmetschungen in deutscher Sprache und einem englischen Ausgangstext bestand. Das Korpus wurde im Jahr 2000 am Fachbereich für Angewandte Sprach- und Kulturwissenschaft (FASK) der Johannes Gutenberg-Universität Mainz in Germersheim anlässlich einer Konferenz aufgezeichnet (vgl. Ahrens 2004: 137f.). Die Studie ergab, dass Dolmetscher*innen den ZT strategisch in mehr Intonationseinheiten unterteilen als im AT enthalten sind (vgl. 2004: 227). Obwohl sie den Inhalt des entsprechenden Abschnitts des Ausgangstextes noch nicht vollständig kennen, geben Dolmetscher*innen die bereits verstandenen Segmente in der Zielsprache wieder. Dies hilft, ihr Kurzzeitgedächtnis so schnell wie möglich zu entlasten (vgl. 2004: 228). Es wurde außerdem festgestellt, dass der finale Tonhöhenverlauf am Ende der Informationseinheiten nicht durch eine fallende Kontur gekennzeichnet ist, die ein klares Ende markieren würde, sondern durch weiterweisende Konturen - also steigende, gleichbleibende und fallende Verlaufsmuster. Dies liegt daran, dass die Dolmetscher*innen, die den AT nicht kennen, nicht in der Lage sind, die Abfolge der Informationseinheiten vorherzusehen und genau zu verstehen, wo eine neue Einheit beginnt und endet. Durch diese Art des finalen Tonhöhenverlaufs ist es möglich, eine Information weiterzuentwickeln, wenn dies im AT vorgegeben ist (vgl. 2004: 209).

Pelz (1999) stellte außerdem fest, dass Verdolmetschungen monotoner sind als die Originalreden (vgl. 1999: 138). Die Meinungen darüber, wie Dolmetscher*innen die Intonation der Sprecher*innen wiedergeben sollen, gehen auseinander, vor allem in Bezug auf eine monotone Intonation im Ausgangstext. Es ist umstritten, inwieweit Dolmetscher*innen als Kommunikationsexpert*innen die prosodischen Mängel des Ausgangstextes verbessern sollten (vgl. Barranco-Droege et al. 2011: 77). Dieser Aspekt wird im Abschnitt 4.4 behandelt.

4.2 Forschungsstand zur Intonation beim Simultandolmetschen

Die Bedeutung der prosodischen Elemente im Bereich der Dolmetschwissenschaft wurde in der Literatur lange Zeit vernachlässigt. Dies liegt auch daran, dass in verschiedenen Umfragen, die seit den 1980er Jahren durchgeführt wurden, Bewertungskriterien für den Inhalt einer

Verdolmetschung als wichtiger angesehen wurden als ausdrucksbezogene Kriterien (siehe Kapitel 2). Die ersten Studien, die sich mit den prosodischen Elementen befassten, konzentrierten sich hauptsächlich auf Pausen, die als Schlüssel für die simultane Aufgabe des Zuhörens und Sprechens angesehen wurden (vgl. Ahrens 2005: 54). Der bevorzugte methodische Ansatz für empirische Studien zur Bedeutung von prosodischen Merkmalen, unter anderem auch der Sprechrate, Intonation usw., war die experimentelle Forschung (vgl. Garzone 2015: 281). Die Studien, die für diese Masterarbeit und das durchgeführte Experiment relevant sind, sind diejenigen, die sich mit der Qualitätsbewertung von Simultandolmetschleistungen in Bezug auf die Intonation beschäftigen.

Die Qualität im Bereich des Dolmetschens wird aus verschiedenen Perspektiven betrachtet, die von der Sichtweise der Dolmetscher*innen über die der Kund*innen bis hin zu der der Nutzer*innen reichen. Die Nutzererwartungen und deren Wahrnehmung der Dolmetschleistung sind seit jeher ein zentraler Forschungsfokus in diesem Bereich, da ihre Zufriedenheit darüber entscheidet, ob eine Verdolmetschung als erfolgreich gilt (vgl. García Becerra 2012: 173ff.). Solche Studien zur Qualität des Dolmetschens, die sich mit den Erwartungen verschiedener Nutzergruppen beschäftigt haben, haben ergeben, dass nonverbale Aspekte im Vergleich zu verbalen Parametern eine wesentlich geringere Wirkung entfalten. In den nachfolgend beschriebenen Studien wird hingegen der Einfluss nonverbaler Elemente, insbesondere der Intonation, auf die Bewertung der Simultandolmetschleistung deutlich.

Die erste Studie, die die stimmlich-prosodischen Aspekte einer professionellen Verdolmetschung in einer realen Situation betrachtet, ist die von Shlesinger aus dem Jahr 1994 über die Auswirkungen der Intonation von Dolmetscher*innen auf das Verständnis der Zuhörer*innen. Sie stellt fest, dass abnormale Tonhöhenbewegungen und Betonungsmuster das Verstehen beeinträchtigen können (vgl. Shlesinger 1994: 231). Diese Arbeit diente als Ausgangspunkt für das von Holub (2010) durchgeführte Experiment über die Auswirkungen monotoner Intonation auf das Verständnis des Publikums und auf die Qualität der Leistung. Intonation wurde in dieser Studie als Variation der Grundfrequenz bezeichnet. Dementsprechend wurde Monotonie als das Fehlen einer solchen Variation definiert (vgl. Holub 2010: 120). Holub (2010) nutzte einen innovativen Ansatz zur Erstellung von Stimulusmaterial: Während in früheren Studien Intonationsabweichungen vorwiegend durch stimmliche Darbietung erzeugt wurden, verwendete Sie in ihrer Untersuchung das Audibearbeitungsprogramm PRAAT, um zwei Versionen derselben Verdolmetschung zu erstellen. Dank dieser Methode kann man die Intonation anpassen, während alle anderen Sprachparameter unverändert bleiben. Das Material wurde in einer simulierten

Konferenzsituation mehreren randomisierten Zuhörergruppen präsentiert (vgl. Holub 2010: 116). Es wurden zwei Versionen der gleichen Verdolmetschung erstellt: eine lebendige und eine monotone Version. Die lebendige Version wurde aus einer auf Video aufgezeichneten englischen Ausgangsrede erstellt und dann mit PRAAT manipuliert, um sie monotoner zu gestalten (vgl. Holub 2010: 121). Die Ergebnisse der Studie zeigten, dass sich Monotonie sowohl auf das Verständnis als auch auf die Leistungsbewertung der Dolmetscher*innen negativ auswirken kann (vgl. Holub 2010: 116).

4.2.1. Studien von Collados Aís zur Qualitätsbewertung

Collados Aís initiierte in den 1990er Jahren eine Reihe von experimentellen Studien, um die Intonation als Qualitätsparameter zu untersuchen (1998, 2007, 2008, 2016). Die für diese Masterarbeit maßgebliche Studie ist ein Experiment, das Collados Aís im Rahmen ihrer Dissertation durchführte, die 1998 veröffentlicht wurde. Diese Studie befasste sich mit der Wirkung monotoner Intonation bei der Bewertung einer Simultandolmetschleistung durch zwei unterschiedliche Gruppen: Nutzer*innen und professionelle Dolmetscher*innen. Das Hauptziel der Studie war es, zu untersuchen, wie monotone Intonation die Bewertung der Qualität des Simultandolmetschens beeinflusst (vgl. Collados Aís 1998: 83). Diese Studie bildet die Grundlage der vorliegenden Masterarbeit und dient zugleich als Ausgangspunkt für die partielle Replikation des Experiments.

Die Studie von 1998 zielte im experimentellen Teil unter anderem darauf ab, die Hypothese zu überprüfen, dass die monotone Intonation sowohl bei spezialisierten Nutzer*innen als auch bei professionellen Dolmetscher*innen die Bewertung der Dolmetschqualität, der Professionalität und der Zuverlässigkeit negativ beeinflusst. Zudem wurde untersucht, inwieweit eine monotone Intonation die Bewertung des/der Originalredners/Originalrednerin beeinflusst. Darüber hinaus sollte überprüft werden, ob bestimmte inhaltliche Fehler von beiden Gruppen unbemerkt bleiben und folglich keinen Einfluss auf die Bewertung der Verdolmetschung haben (vgl. Collados Aís 1998: 81ff.).

Die Teilnehmer*innen setzten sich aus zwei Hauptgruppen zusammen: 42 juristische Expert*innen, darunter Professor*innen für öffentliches Recht und verwandte Gebiete der Rechtsvergleichung an der Universität Granada, und 15 Dolmetscher*innen mit Spanisch, Französisch und/oder Englisch als A/B-Sprache. Zu den Voraussetzungen für die Nutzer*innen zählte neben ihrer fachlichen Spezialisierung die Erfahrung als Teilnehmer*innen an internationalen Kongressen, Seminaren oder Konferenzen mit Simultandolmetschangebot. Die juristischen Expert*innen erfüllten die genannte Voraussetzung aufgrund ihrer beruflichen

Tätigkeit und hätten bei der Beurteilung einer Verdolmetschung über mehr Kompetenz verfügt als Nutzer*innen ohne vorherige Erfahrung (vgl. Collados Aís 1998: 85ff.).

Was die Methodik betrifft, wurde ein deutscher Ausgangstext erstellt, der von einer Dolmetscherin ins Spanische verdolmetscht wurde. Der Text befasste sich mit der Finanzierung politischer Parteien in Deutschland und war Teil eines Seminars zur Parteienfinanzierung in Europa, das von der Abteilung für Verfassungsrecht organisiert wurde (vgl. Collados Aís 1998: 91). Es wurden drei unterschiedliche spanische Versionen der Verdolmetschung erstellt. Bei zwei Versionen wurde jeweils eine unabhängige Variable manipuliert, nämlich die Intonation und die Bedeutung des Originaltextes, die durch verschiedene inhaltliche Fehler verzerrt wurde. In der Kontrollversion wurde keine der oben genannten Variablen verzerrt (vgl. Collados Aís 1998: 89). Die drei Versionen waren wie folgt gestaltet: Version 1 wies eine monotone Intonation und keine Fehler in der Übertragung des Sinns der Originalrede auf. Version 2 war geprägt von einer nicht-monotonen Intonation, enthielt aber Fehler in der Übertragung der Bedeutung der Originalrede. Version 3 schließlich wies eine nicht-monotone Intonation und keine Fehler bei der Sinnübertragung der Originalrede auf (vgl. Collados Aís 1998: 90). Alle drei Verdolmetschungen auf Spanisch wurden der ursprünglichen AT-Rede auf Deutsch überlagert, die mit monotoner Intonation vorgetragen wurde und 10 Minuten dauerte (vgl. 1998: 91). Collados Aís entschied sich, eine monotone Ausgangstextversion zu reproduzieren, damit zukünftige Forschungen klären können, ob ein lebhaft intonierter Ausgangstext die Bewertung der Dolmetschleistung durch die Nutzer*innen, weiter negativ beeinflussen könnte (vgl. 1998: 90).

Die Dissertation umfasste vier Studien. Die erste Studie zielte darauf ab, die Grundvoraussetzungen zu ermitteln, indem überprüft wurde, ob für die Teilnehmer*innen erkennbar war, dass sich die drei Versionen in ihrer Monotonie unterschieden. Aus diesem Grund wurde Student*innen der Fakultät für Psychologie ein Fragebogen vorgelegt, mit dem Ziel, die Intonation der Dolmetscher*in und des/der Redners/Rednerin zu bestimmen (Studie 1A). Darüber hinaus wurde eine computergestützte Analyse von Tonhöhe und Intensität mithilfe des Visi-Pitch-Systems (Studie 1B) durchgeführt. Zusätzlich wurde an Studierende der Fakultät für Übersetzen und Dolmetschen ein Fragebogen verteilt, um die emotionalen Rückschlüsse, die auf der Intonation basieren, zu überprüfen (Studie 1C). Die zweite Studie hatte die Überprüfung der Erwartungen an die Qualität der Verdolmetschung seitens der Nutzer*innen (Studie 2A) und Dolmetscher*innen (Studie 2B) zum Ziel. Daher wurde jede der beiden Nutzer*innengruppen (Dolmetscher*innen und juristische Expert*innen) in drei Untergruppen aufgeteilt und ihnen wurde ein Fragebogen vorgelegt, der auf dem Fragebogen

von Bühler (1986) basierte, in dem sie verschiedene Kriterien bewerten sollten (vgl. Collados Aís 1998: 96f). Die dritte Studie stellt den zentralen Teil der Arbeit dar, in dem die Dolmetschleistung evaluiert wurde. In Studie 3A erfolgte die Bewertung der Leistung durch die Nutzer*innen, was den Ausgangspunkt dieser Masterarbeit bildet, und in Studie 3B durch die Dolmetscher*innen. In der Studie 4 wurde eine vergleichende Analyse der in den vorherigen Studien erzielten Ergebnisse durchgeführt (vgl. Collados Aís 1998: 97f.).

In Studie 3 wurde jeder Untergruppe von Nutzer*innen in ihren jeweiligen Abteilungen oder Büros eines der drei Videos auf einem Monitor gezeigt (vgl. Collados Aís 1998: 170f.). Für die Dolmetscher*innen erfolgte die Präsentation der Videos im Simultandolmetscher-Labor der Fakultät für Übersetzen und Dolmetschen (vgl. Collados Aís 1998: 194). Im Anschluss erfolgte eine Bewertung der Version durch jede Untergruppe anhand eines Fragebogens, der auf dem Fragebogen zu ihren Erwartungen basierte. Beide Fragebögen enthielten ähnliche Fragen, die auf einer fünfstelligen Skala beantwortet werden sollten (vgl. Collados Aís 1998/2002: 330).

Die Bewertung der Verdolmetschung erfolgte anhand der folgenden Kriterien: Gesamtbewertung, muttersprachlicher Akzent, angenehme Stimme, Flüssigkeit, Kohärenz, Sinnübereinstimmung mit dem Original, Vollständigkeit, korrekte Terminologie, Stil, Intonation, Diktion, Professionalität, Zuverlässigkeit, Bewertung der Originalrede (vgl. Collados Aís 1998/2002: 331-334).

Die Ergebnisse zeigten, dass die inhaltlich korrekte monotone Version schlechtere Bewertungen erhielt als die nicht-monotone Version, die fehlerhaft war. Die monotone Version wurde außerdem als wenig professionell und zuverlässig eingestuft (vgl. Collados Aís 1998: 246). Die Bewertungen der Teilnehmer*innen zeigten eine Diskrepanz zwischen den Bewertungen der Verdolmetschung und den im Fragebogen geäußerten Erwartungen, in dem der Parameter der Intonation als nicht besonders relevant angesehen wurde. Es wurde also nachgewiesen, dass eine monotone Intonation die Beurteilung der Qualität beim Simultandolmetschen beeinträchtigt (vgl. Collados Aís 1998: 249). Die detaillierten Ergebnisse der Studie 3A zu den Bewertungen der Nutzer*innen werden in Kapitel 6 ausführlich vorgestellt und mit den Ergebnissen der vorliegenden Arbeit verglichen.

4.2.2 Folgestudien zum Experiment von Collados Aís (1998)

In den folgenden Jahren führte Collados Aís mehrere Studien durch (2007, 2008, 2016), die die Qualitätsbewertung von Verdolmetschungen in Bezug auf die Intonation zum Untersuchungsgegenstand hatten.

2007 wurde die Studie von 1998 repliziert und die Ergebnisse wurden nicht bestätigt. Diese Untersuchung wurde mit Proband*innen einer anderen Herkunft im Rahmen einer breit angelegten Studie durchgeführt, in der elf Parameter manipuliert wurden (vgl. Collados Aís et al. 2007: 12). Den Teilnehmer*innen wurden dieselben Dolmetschungen aus der früheren Studie aus dem Jahr 1998 vorgespielt. Obwohl die Nutzer*innen die Monotonie der Intonation in der manipulierten Version bemerkten und negativ beurteilten – wenn auch in geringerem Ausmaß als in der vorherigen Studie – wurde die Gesamtbewertung der Verdolmetschung nicht negativ beeinflusst (vgl. Collados Aís 2007: 171). Darüber hinaus erzielte das manipulierte Video trotz der monotonen Intonation bessere Ergebnisse in Bezug auf Professionalität und Zuverlässigkeit (vgl. Collados Aís 2007: 197). Die Ursachen für diese unterschiedlichen Ergebnisse können auf verschiedene Faktoren zurückgeführt werden, zu denen die Größe der Stichprobe und die Variabilität der Proband*innen gehören. Die Fehlermöglichkeit ist bei kleinen Stichprobengrößen hoch, und nonverbale Aspekte wie die Intonation werden durch zwischenmenschliche und kulturelle Variabilität beeinflusst. Da die Proband*innen aus einer anderen Stadt stammten, ist es möglich, dass ihre Sensibilität für den Parameter Intonation geringer war. Außerdem könnten diese abweichenden Ergebnisse darauf zurückzuführen sein, dass der Grad der Erfahrung als Nutzer*innen von Simultandolmetschleistungen in den Fragebögen nicht detailliert genug angegeben wurde. Dies könnte ihre Empfindlichkeit gegenüber bestimmten Parametern beeinflusst haben (vgl. Collados Aís 2007: 172).

Nachdem die Ergebnisse von 1998 nicht bestätigt wurden, beschloss Collados Aís, mehrere Studien durchzuführen, um mögliche Beurteilungsmuster zu untersuchen, sowohl allgemein als auch in Bezug auf die Intonation. Sie reproduzierte daher das Experiment der Studie von 1998 im Jahr 2008, wobei sie Fragen zur Einstellung der Dolmetscher*innen einbezog und einen Crossover-Test einführte, um zu verstehen, wie die Reihenfolge der Verdolmetschung die Bewertung der Einstellung beeinflussen könnte (vgl. Barranco-Droege et al. 2011: 89). Die Proband*innen mussten daher beide manipulierten Videos zu unterschiedlichen Zeitpunkten ansehen und bewerten (vgl. Collados Aís 2008: 200). Die Hauptziele waren die Überprüfung der Auswirkungen monotoner Intonation auf die Qualitätsbewertung der Verdolmetschung durch die Nutzer*innen und die Analyse ihrer emotionalen Inferenzen in Bezug auf Monotonie. Der verwendete Ansatz war qualitativer als in früheren Studien, da offenere Fragebögen zur Bewertung durch die Nutzer*innen, Diskussionsgruppen und gezielte Interviews verwendet wurden (vgl. 2008: 194). Das Material bestand aus zwei spanischen Dolmetschungen eines deutschen Ausgangstextes. Eine Dolmetschung war nicht monoton und wurde als Kontrollversion bezeichnet, die andere wurde

durch Manipulation des Intonationsparameters monoton gestaltet (vgl. 2008: 201). In dieser Studie wurden die Ergebnisse aus dem Jahr 1998 bezüglich der negativen Auswirkungen einer monotonen Intonation auf die Bewertung der Qualität einer Verdolmetschung erneut bestätigt. Die Kontrollversion erzielte in allen betrachteten Parametern bessere Ergebnisse als die andere Version, mit Ausnahme der Kriterien korrekte Sinnübertragung und Kohärenz, bei denen keine besonderen Unterschiede zwischen der Bewertung der beiden Versionen festgestellt wurden (vgl. Collados Aís 2008: 202). Darüber hinaus wurde festgestellt, dass die Nutzer*innen Monotonie als negatives emotionales Signal wahrnehmen. Dies wurde deutlicher, wenn die monotone Version nach der nicht monotonen Version gehört wurde (vgl. 2008: 211).

Schließlich führte Collados Aís 2016 eine Folgestudie zu ihrer Studie aus dem Jahr 2008 durch. Hauptziel war die Analyse der Bewertungsmuster der Nutzer*innen in Bezug auf die Intonation der Dolmetscher*innen. Es sollte überprüft werden, ob die Nutzer*innen unabhängig vom Grad der Monotonie der Dolmetscher*innen konsistente Bewertungsmuster für die Bewertung der Intonation haben. Auf diese Weise war es möglich, die Auswirkungen der Intonation auf die Bewertung anderer Kriterien wie Gesamtbewertung, korrekte Sinnübertragung, Einstellung der Dolmetscher*innen, Professionalität und Zuverlässigkeit zu verstehen. Es sollte auch überprüft werden, ob die Bewertung der oben genannten Kriterien je nach Grad der Monotonie der Dolmetscher*innen variiert und ob dies auf Bewertungsmodelle der Intonation durch die Nutzer*innen zurückzuführen ist. Außerdem sollte untersucht werden, ob diese Bewertungsmodelle vom Kenntnisstand der Nutzer*innen in Bezug auf Dolmetschprozesse abhängen. Zu diesem Zweck wurden zwei Studien mit unterschiedlichen Proband*innen und Materialien durchgeführt (vgl. Collados Aís 2016: 3). In der ersten Studie bewerteten Proband*innen, die nur über geringe Kenntnisse des Simultadolmetschprozesses verfügten, vier Verdolmetschungen mit einem unterschiedlichen Grad an Monotonie. In der zweiten Studie hatten die sechs Proband*innen fortgeschrittene Kenntnisse des Simultandolmetschprozesses. Sie bewerteten ein Korpus von 30 Europäischen Parlaments-Verdolmetschungen, die vom Europe by Satellite (EbS)-Kanal aufgenommen wurden, anhand der Bewertung der Intonation der Dolmetscher*innen (vgl. 2016: 3). Die Intonation erhielt in beiden Studien niedrigere Werte als die Kriterien Gesamtbewertung, korrekte Sinnübertragung und die aus der Leistung abgeleiteten Einschätzungen zu Haltung, Professionalität und Zuverlässigkeit des/der Dolmetschers/Dolmetscherin (vgl. 2016: 17). Die Ergebnisse zeigten außerdem, dass es Gemeinsamkeiten in den Bewertungsmustern der Intonation der Dolmetscher*innen gibt, obwohl Grenzen existieren, innerhalb derer sich die meisten Proband*innen bei der Leistungsbewertung bewegen. Tatsächlich bleiben die Nutzer*innen

vorsichtig, insbesondere bei der Vergabe von hohen Bewertungen, weil sie davon überzeugt sind, dass es unmöglich ist, einige Parameter zu bewerten (vgl. 2016: 19).

4.3 Die Rolle von Dolmetscher*innen bei der Wiedergabe der Intonation

Wie bereits erwähnt, wird im Bereich des Dolmetschens darüber diskutiert, inwiefern Dolmetscher*innen eine aktive Rolle übernehmen sollten, um beispielsweise die monotone Leistung von Sprecher*innen zu verbessern (vgl. Barranco-Droege et al. 2011: 77). Einige Autor*innen sprechen sich für ein aktives Eingreifen durch die Dolmetscher*innen aus. Collados Aís (2001) argumentiert, dass Dolmetscher*innen die Leistung der Sprecher*innen verbessern sollten, die im Gegensatz zu ihnen keine professionellen Sprachmittler*innen sind (vgl. Collados Aís 2001: 108f.). Dies widerspricht jedoch der *ghost role*, wonach die Dolmetscher*innen nicht in die Rede eingreifen, sondern getreu wiedergeben sollen, was die Redner*innen sagen (vgl. Kopczynski 1994: 90). Die Kommunikationspartner*innen gehen oft davon aus, dass die Dolmetscher*innen diese Rolle übernehmen, insbesondere im Rahmen des Community Interpreting (vgl. Barranco-Droege et al. 2011: 77).

Aus den bisher beschriebenen Studien von Collados Aís lassen sich Überlegungen darüber anstellen, inwieweit die Nutzer*innen einen Eingriff der Dolmetscher*innen erwarten oder nicht. Nach ihrer Studie von 1998 verlangen die Nutzer*innen implizit ein Eingreifen der Dolmetscher*innen, welches von der ihnen typischerweise zugeschriebenen *ghost role* abweicht. Die monotone Version wurde negativer bewertet als die nicht-monotone Version, obwohl die Ausgangsrede monoton vorgetragen wurde (vgl. Collados Aís 1998: 250).

Beim Experiment von 2007 wird eine weitere Hypothese in Betracht gezogen, warum die Hörer*innen bei monotoner Intonation nachsichtiger waren: Es könnte daran liegen, dass die Nutzer*innen es gewohnt sind, Verdolmetschungen *sui generis* zu hören, die monoton klingen, sodass es nicht notwendig ist, dass die Dolmetscher*innen die Intonation des Ausgangstextes verbessern (vgl. Collados Aís 2007: 172).

In der Studie von 2008 schrieben die Nutzer*innen den Dolmetscher*innen die *ghost role* zu. Dies wird deutlich, da die Dolmetscher*innen in der nicht monotonen Version als neutral beurteilt wurden – allerdings nur, wenn diese Version vor der monotonen Version gehört wurde. In der umgekehrten Situation wurde den Dolmetscher*innen eine viel aktivere Rolle zugewiesen (vgl. Collados Aís 2008: 211). Auch dies ist auf das Modell *sui generis* der Intonation der Dolmetscher*innen zurückzuführen (vgl. 2008: 210). Dennoch wurde die monotone Version negativer bewertet als die Kontrollversion.

Die Studie von 2016 bestätigte die Präferenz der Nutzer*innen für einen Eingriff durch die Dolmetscher*innen, wie sie bereits 1998 festgestellt wurde, allerdings nur für eine bestimmte Kategorie von Nutzer*innen. Die Ergebnisse hängen sehr stark vom Kenntnisstand der Nutzer*innen bezüglich des Dolmetschprozesses ab. Ein erfahrenes Publikum neigt dazu, eine dynamischere Intonation zu schätzen, weil es weiß, dass Dolmetscher*innen aktiv sein müssen, um die Kommunikation zu erleichtern. Ein Publikum mit begrenzten Kenntnissen wird hingegen eine übermäßig lebhaft Intonation möglicherweise nicht zu schätzen wissen, da es von der Intonation *sui generis* beeinflusst wird, die die Leistungen tendenziell kennzeichnet (vgl. Collados Aís 2016: 18).

Die Meinungen der Nutzer*innen über die Rolle der Dolmetscher*innen in Bezug auf die Intonation sind daher vielfältig und nicht einheitlich. Sie können von einigen Faktoren abhängen, wie z. B. der Kenntnis des Dolmetschprozesses und dem Intonationsmodell *sui generis*, von dem viele Nutzer*innen bei der Bewertung ausgehen. Es muss betont werden, dass die verwendeten Ausgangstexte für die Studien aus den Jahren 1998 und 2007 von einer monotonen Intonation geprägt waren. Bei einem Ausgangstext mit emotionalem Charakter und dynamischer Intonation sind die Dolmetscher*innen nach Knobloch (2002) gefordert, die Emotionen der Sprecher*innen zu erfassen und entsprechend wiederzugeben (vgl. Knobloch 2002: 206). Es ist jedoch wichtig, dass die Emotionalität genau wiedergegeben wird, ohne sie zu verfälschen (vgl. Knobloch 2002: 200).

Das folgende Kapitel befasst sich mit dem Experiment, das im Rahmen dieser Masterarbeit durchgeführt wurde. Die Methodik basiert auf Collados Aís (1998), allerdings als partielle Replikation, da ein nicht monotoner Ausgangstext verwendet wird. Die Entscheidung, einen Text mit lebendiger Intonation einzusetzen, lässt sich auf die Anmerkung von Collados Aís in der Studie von 1998 zurückführen. Sie regte an, in Zukunft zu überprüfen, ob ein lebhaft intonierter Ausgangstext die Bewertung der Dolmetschleistung durch die Nutzer*innen weiter negativ beeinflussen könnte (vgl. Collados Aís 1998: 90).

5 Methodik

In den vorangegangenen Kapiteln wurde der theoretische Hintergrund dieser Arbeit dargelegt und somit die Grundlage für die Konzeption des experimentellen Teils geschaffen, der in diesem Kapitel präsentiert wird. Die Forschungsfrage und die verwendete Methodik werden detailliert vorgestellt, wobei die Analogien und Unterschiede zwischen dieser Studie und der Referenzstudie von Collados Aís (1998), von der diese Arbeit eine Teilreplikation ist, hervorgehoben werden.

5.1 Zielsetzung und Hypothesen

Diese Masterarbeit widmet sich der Qualitätsbewertung beim Simultandolmetschen aus der Perspektive der Nutzer*innen, wobei das Augenmerk auf den Parameter der Intonation gelegt wird. Untersucht wird, welchen Einfluss eine monotone im Vergleich zu einer dynamischen Dolmetschleistung auf die Wahrnehmung und Bewertung der Dolmetschqualität hat. Das Ziel besteht darin, bestimmte Hypothesen von Collados Aís aus der 1998 veröffentlichten Studie zu bestätigen. Diese besagen, dass eine monotone Intonation die Bewertung der Qualität der Verdolmetschung sowie die Wahrnehmung der Nutzer*innen hinsichtlich des subjektiven Kriteriums der Professionalität der Dolmetscher*innen negativ beeinflusst (vgl. Collados Aís 1998: 81f.). Darüber hinaus wird angenommen, dass ein lebhafterer Ausgangstext dazu führt, dass die monoton intonierte Dolmetschleistung von den Nutzer*innen noch negativer bewertet wird. Der Ausgangstext von Collados Aís war nämlich bewusst durch eine monotone Intonation gekennzeichnet. Auf diese Weise hätten zukünftige Forschungen klären können, ob eine monotone Intonation der Dolmetschleistung bei einem lebhaft intonierten Ausgangstext als noch stärker nachteilig wahrgenommen wird (vgl. 1998: 90).

Da diese Studie eine partielle Replikation der Studie von Collados Aís (1998) darstellt, basiert die verwendete Methodik auf ebendieser Studie, weist jedoch einige Unterschiede auf. Die Forschungsziele der Studie von Collados Aís (1998) umfassten insgesamt acht Punkte. Zusätzlich zu den im vorherigen Abschnitt genannten Zielsetzungen sollte ihre Studie sowohl die Qualitätserwartungen von Fachnutzer*innen als auch von professionellen Dolmetscher*innen ermitteln. Darüber hinaus sollte die Hypothese überprüft werden, dass eine monotone Intonation nicht nur einen negativen Einfluss auf die wahrgenommene Qualität und Professionalität hat, sondern auch auf das Vertrauen, das dem/der Dolmetscher*in entgegengebracht wird. Außerdem sollte analysiert werden, wie sich eine monotone Intonation

auf die Bewertung der Originalrede auswirkt, und es sollte die Hypothese überprüft werden, dass bestimmte inhaltliche Fehler von Dolmetscher*innen und Nutzer*innen unbemerkt bleiben und die Bewertung der Verdolmetschung nicht beeinflussen (vgl. Collados Aís 1998: 81ff.). In ihrer Studie wurden zwei Versionen eines Kontrollvideos erstellt, in denen jeweils eine unabhängige Variable manipuliert wurde, nämlich die Intonation und die Bedeutung des Originaltextes, die durch verschiedene inhaltliche Fehler verzerrt wurde. In der Kontrollversion wurde keine Variable verzerrt (siehe Kap. 4.2.1).

Die Zielsetzungen und Forschungshypothesen dieser Studie wurden im Vergleich zu denen der Doktorarbeit von Collados Aís (1998) reduziert und aus materiellen und zeitlichen Gründen auf den Umfang einer Masterarbeit beschränkt. Im Gegensatz zu Collados Aís (1998), die in ihrer Studie Dolmetscher*innen und spezialisierte Nutzer*innen einbezog, sind die Teilnehmer*innen dieser Studie ausschließlich nicht-spezialisierte Nutzer*innen, da keine Expert*innengruppe zur Verfügung stand. Darüber hinaus wurde kein Fragebogen zu den Erwartungen bereitgestellt, sondern lediglich ein Fragebogen zur Bewertung der Dolmetschleistung. Die manipulierte Variable beschränkte sich ausschließlich auf die Intonation; der inhaltliche Sinn der Rede blieb unverändert, und es wurde lediglich eine manipulierte Videoversion auf Grundlage des Kontrollvideos erstellt. Diese Unterschiede werden im Laufe des Kapitels ausführlich erläutert.

5.2 Versuchsdesign

Im Bereich der Übersetzungs- und Dolmetschwissenschaft sind unterschiedliche Arten von Forschung möglich, die von experimenteller, quasi-experimenteller bis hin zu beobachtungsbasierter Forschung reichen. Der Unterschied zwischen „experimentell“ und „quasi-experimentell“ ist nicht eindeutig (vgl. Mellinger & Hanson 2017: 7f.). Ein wesentlicher Unterschied besteht jedoch darin, dass die Teilnehmer*innen in der experimentellen Forschung den Gruppen zufällig zugeordnet werden und die unabhängige Variable unter kontrollierten Bedingungen manipuliert wird. In der quasi-experimentellen Forschung kann die unabhängige Variable hingegen kontrolliert oder manipuliert werden, doch die Zuordnung der Teilnehmer*innen erfolgt nicht zufällig (vgl. Mellinger & Hanson 2017: 6ff.). Gemäß dieser Definition handelt es sich bei der Studie dieser Masterarbeit um eine quasi-experimentelle Forschung. Sie basiert auf der Manipulation einer unabhängigen Variable, der Intonation, und einer nicht vollständig randomisierten Zuordnung der Proband*innen. Die Daten wurden mittels einer Umfrage erhoben und anschließend quantitativ analysiert. Dieser methodische Ansatz wurde auch von Collados Aís in ihrer Studie aus dem Jahr 1998 verwendet, allerdings

mit einer strengeren Kontrolle der Gruppen und der Umgebung als bei diesem Experiment, das online durchgeführt wurde.

Um das Experiment durchzuführen, wurde ein Video mit Ton aufgenommen, das im Rahmen einer fiktiven Konferenz zur Vorstellung eines Buches aufgenommen und mit einer mittelmäßig lebhaften Intonation vorgetragen wurde. Ausgehend von diesem Video wurden zwei Verdolmetschungen erstellt: Zum einen eine Kontrollversion, die sich durch eine lebhafte Intonation auszeichnet und die Prosodie des Originals nachahmt. Zum anderen eine Version in einer stark monotonen Ausführung. Diese wurde mit der Shadowing-Technik ausgehend von der lebhaften Version erstellt. In beiden Versionen wurde der Ausgangstext mit der Voice-Over-Technik überlagert. Die Intonation war die einzige Variable, die manipuliert wurde.

Die Proband*innen wurden in zwei Gruppen eingeteilt. Es gab bei der Auswahl der Proband*innen drei Voraussetzungen für die Teilnahme am Experiment: Italienischkenntnisse auf mindestens C1-Niveau, keine Ausbildung im Bereich Dolmetschen, und mindestens einmal eine Dolmetschleistung in den Medien verfolgt zu haben. Jede der beiden Gruppen hörte sich eine der beiden Versionen online an. Anschließend wurde ihnen ein Fragebogen zur Bewertung der Verdolmetschung anhand von fünf Qualitätskriterien vorgelegt, die auf einer Ordinalskala von 1 bis 5 zu bewerten waren. Die erhobenen Daten wurden anschließend quantitativ ausgewertet.

5.2.1 Material und Stimuli

Im Folgenden werden die für die Durchführung des Experiments verwendeten Materialien und deren Vorbereitung beschrieben.

5.2.1.1 Ausgangstext

Der Ausgangstext in deutscher Sprache wurde so konzipiert, dass er zum Kontext einer fiktiven Konferenz zur Vorstellung eines Buches passen würde. Der Text wurde in einer Word-Datei verfasst und dem Redner drei Tage vor seiner Präsentation im August 2025 zur Verfügung gestellt. Die Art der Konferenz, auf die Bezug genommen wird, ist die Kategorie Gastvorträge, worunter Vorträge und Referate von Gastredner*innen fallen (vgl. Pochacker 1994: 51).

Das Buch trägt den Titel „Al di là del fiume“ (Jenseits des Flusses) und ist ein fiktiver Roman, der zur literarischen Gattung des Bildungsromans gehört, in dem der Protagonist einen Bildungsprozess durchläuft, der zu seinem Wachstum und seiner Reifung führt (vgl. Selbmann 1994 :11). Dieses Genre entstand zwischen Ende des 18. und Anfang des 19. Jahrhunderts,

wobei Goethes Roman „Wilhelm Meisters Lehrjahre“ als klassisches Vorbild gilt (vgl. Iversen 2009: 21). Die Wahl dieser literarischen Strömung begründet sich in den charakteristischen Merkmalen des Romans, der zuweilen spannende Ereignisse beinhaltet. In der Regel durchläuft der Hauptcharakter einen Reifungsprozess, der dazu führt, dass er sich selbst wiederentdeckt, nachdem er Liebesbeziehungen und Freundschaften eingegangen ist und mit der Realität des Lebens in Konflikt geraten ist (vgl. Swales 1991: 98 zit. nach Iversen 2009: 22). Aufgrund der behandelten Themen sollte die Präsentation des Textes in der Kontrollversion mit einer lebhaften Intonation präsentiert werden. Dies war eine grundlegende Voraussetzung, um die Auswirkungen der monotonen Verdolmetschung auf die Bewertung durch die Teilnehmer*innen untersuchen zu können. Für die Durchführung dieser Masterarbeit stand kein Fachpublikum wie das von Collados Aís (1998) zur Verfügung, weshalb ein allgemeines Thema gewählt wurde, das keine terminologischen Vorkenntnisse erforderte. Der Text von Collados Aís (1998) war hingegen juristisch-technischer Natur und befasste sich mit der Finanzierung politischer Parteien in Deutschland (siehe Kapitel 4.2.1).

5.2.1.2 Textgestaltung & Präsentationsweise

Die Dauer der Rede betrug 4:28 Minuten. Diese ist für die Durchführung eines Online-Experiments mit geringem Umfang im Vergleich zur Studie von Collados Aís (1998), in der die Redezeit 10 Minuten betrug, als angemessen zu erachten. Die kurze Dauer der Rede hatte zum Ziel, die Zuhörer*innen nicht zu langweilen und ihre Aufmerksamkeit möglichst vom Anfang bis zum Ende des Videos aufrechtzuerhalten. Die Nutzer*innen waren außerdem ein allgemeines Publikum und kein Fachpublikum. Die Rede wurde von einem deutschsprachigen Schauspieler spontan und ohne beschleunigtes Sprechtempo vorgetragen, sodass sie für die Dolmetscherin keine Schwierigkeit darstellte (vgl. Ahrens 2004: 114). Der Text ist logisch aufgebaut und weist einen typisch unidirektionalen Informationsfluss auf (vgl. Pöchhacker 1994: 52). Er gehört zu der von Kopczynski (1980) definierten Kategorie des „semi-prepared oral monologue with notes“ (Kopczynski 1980: 16 zit. nach Pöchhacker 1994: 109.). Die Word-Datei, die den Text enthielt, wurde dem Redner drei Tage vor seiner Präsentation zur Verfügung gestellt. Zum Zeitpunkt der Präsentation standen ihm nur Schlüsselwörter und kurze Sätze zur Verfügung, um die Sprechgeschwindigkeit zu begrenzen. Die Sprechgeschwindigkeit hätte wiederum die Intonation der Dolmetscherin beeinflussen können, die in diesem Experiment die zentrale Variable darstellt.

Der Text umfasst eine Einführung des Autors selbst, eine Zusammenfassung der Handlung des Buches und einen kurzen Auszug, der direkt aus dem Buch vorgelesen wird. Da

es sich um eine aufgezeichnete Präsentation mit Video handelt, sind auch Gestik und Mimik sichtbar.

5.2.2 Die Vorbereitung der Dolmetscherin

Die für die Studie ausgewählte Dolmetscherin hat ihr Studium im Fachbereich Konferenzdolmetschen an der Universität Forlì in Italien abgeschlossen und verfügt über ein Jahr Berufserfahrung. In einer realen Konferenzsituation hätte sich die Dolmetscherin auf diesen Auftrag vorbereitet, indem sie das Buch gelesen und sich über die Biografie des Autors informiert hätte. Da dies nicht möglich war, wurden ihr eine Woche vor dem Dolmetschen Informationen zur Handlung des Buches in Form einer Word-Datei zur Verfügung gestellt. Die Verfügbarkeit von Vorbereitungsmaterialien wie Redemanuskripte, Abstracts oder Folienvorlagen gehört zu den Vorbereitungsstrategien für Konferenzen für Dolmetscher*innen (vgl. Kalina 1998: 116). Im vorliegenden Fall war für die Textvorbereitung keine besondere terminologische Vorarbeit erforderlich. Die Dolmetscher*in erhielt jedoch am Tag der Verdolmetschung den gesamten Abschnitt, den der Redner aus dem Buch vorlesen würde, um im Voraus die bestmögliche Entsprechung zu finden. Dies war aufgrund der literarischen Natur des Textes erforderlich, die eine spontane Verdolmetschung erschwert hätte.

5.2.3 Dolmetschversionen

Auf der Grundlage des Ausgangsvideos wurden zwei Verdolmetschungen erstellt, die sich im Grad ihrer Monotonie unterschieden. Wie bereits erwähnt, wird Intonation in dieser Masterarbeit im engeren Sinne als Tonhöhenverlauf einer sprachlichen Äußerung verstanden (siehe Kapitel 3). Collados Aís (1998) verwendet diese Begriffsbestimmung in ihrer Studie nicht und bezieht in die Definition der monotonen Intonation auch die prosodischen Parameter der Lautstärke, Geschwindigkeit und Tonhöhe mit ein (vgl. Collados Aís 1998: 75). Neuere Studien zur Qualität in Bezug auf die Intonation, wie die von Holub aus dem Jahr 2010 (siehe Kapitel 4.2), verstehen die Intonation jedoch nur als Variation der Grundfrequenz, weshalb dieser Parameter der Einzige war, der manipuliert wurde und sich von einer Version zur anderen unterschied. Damit auch die anderen prosodischen Parameter wie Pausen und Geschwindigkeit unverändert blieben, wurde die Technik des Shadowing verwendet. Die lebhaftere Kontrollversion, die in ihrer Intonation dem Ausgangstext ähnelt, wurde daher von der Dolmetscherin mittels Shadowing mit eintöniger Intonation gedolmetscht, um die monotone Version zu erstellen. Bei dieser Technik wird der Input unmittelbar und wortwörtlich in derselben Sprache wiedergegeben (vgl. Pöchhacker 2016: 121). Die Entscheidung für die

Anwendung der Shadowing-Technik ist auf die Komplexität der Verwendung von Audiotranskriptionsprogrammen zurückzuführen. Holub (2010) verwendete beispielsweise in ihrer Studie das Programm PRAAT für die Stimmmanipulation (vgl. Holub 2010: 121). Wie im folgenden Abschnitt dargelegt, wurde das Programm PRAAT in der vorliegenden Studie ausschließlich eingesetzt, um den Unterschied in der Intonation zwischen den beiden Versionen aufzuzeigen.

Die Kontrollversion beinhaltet mehrere Selbstkorrekturen, welche ebenfalls in der monotonen Version beibehalten wurden. Von beiden Dolmetschversionen wurde der erste Dolmetschversuch verwendet. Für die Kontrollverdolmetschung wurde die erste Version ausgewählt, damit die Leistung so natürlich und authentisch wie möglich war. Was die monotone Version betrifft, reichte eine Audioaufnahme aus, um das gewünschte Ergebnis der Monotonie zu erzielen, da es sich um eine professionelle Dolmetscherin handelte, die an Shadowing-Übungen zur Vorbereitung auf Konferenzen gewöhnt war.

Um grafisch darzustellen, wie sich die beiden Verdolmetschungen hinsichtlich ihrer Monotonie unterscheiden, wurde die Pitch-Analyse des PRAAT-Programms verwendet, das in der Lage ist, die Grundfrequenz zu schätzen (vgl. Holub 2010: 121). Mittels der Option „query“ von PRAAT wurde die Intonation, verstanden als Variation der Grundfrequenz (F0), quantitativ anhand der folgenden Werte der Grundfrequenz analysiert: „minimum pitch“, „maximum pitch“, „mean“ und „standard deviation“. Diese Werte wurden von Holub in ihrer Studie aus dem Jahr 2010 verwendet, um die Grundfrequenzen der monotonen Version mit denen der lebhaften Version zu vergleichen (vgl. Holub 2010: 122).

Die Audiodatei wurde zunächst in das Programm PRAAT hochgeladen. Vor der Analyse wurden ein *Pitch Floor* und ein *Pitch Ceiling* festgelegt, die die unteren und oberen Hz-Werte im Analysefenster darstellen. (vgl. Reed 2010: 26). Da Sprecher*innen in ihrem Tonumfang sehr unterschiedlich sein können, wurden die Werte für die Tonhöhenanalyse so eingestellt, dass sie gemäß der online verfügbaren PRAAT-Bedienungsanleitung für eine Sprecherin mit weiblicher Stimme geeignet sind. Der Pitch Floor wurde auf 100 Hz und der Pitch Ceiling auf 800 Hz eingestellt, was für hohe Stimmen, wie beispielsweise für eine Frauenstimme geeignet ist (vgl. PRAAT o.J.). Tabelle 1 zeigt die Werte der beiden Versionen im Vergleich. Eine geringere Standardabweichung steht für eine monotonere Intonation (vgl. Holub 2010: 122).

Tab. 1: Grundfrequenzwerte der Kontroll- und Monotonversion im Vergleich

	Kontrollversion	Monotonversion
Minimum pitch	90.28 Hz	88.42 Hz
Maximum pitch	428.45 Hz	406.86 Hz
Mean	225.81 Hz	206.16 Hz
Standard deviation	31.89 Hz	14.65 Hz

Die Abbildung 6 veranschaulicht den Unterschied in der Intonation zwischen den beiden Versionen. Mit der Option „draw“ von PRAAT wurden die Pitch-Konturen verglichen, die als Darstellung der Intonationskonturen interpretiert werden (siehe Kapitel 3.1). In der lebhaften Version auf der linken Seite sind deutliche Schwankungen und ein breiterer Tonhöhenbereich zu erkennen. In der monotonen Version rechts ist hingegen ein relativ gleichbleibender F0-Verlauf zu beobachten (siehe Abb. 6).

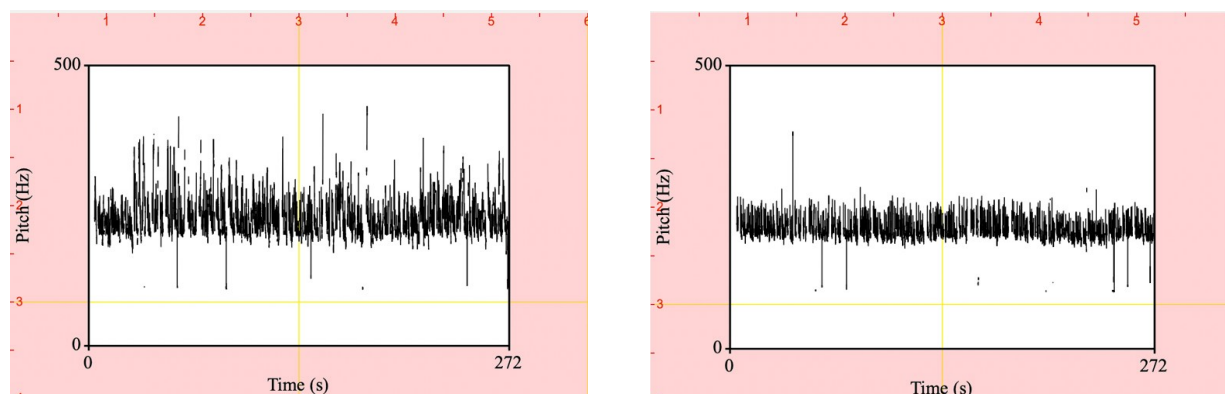


Abb. 6: Grafische Darstellung der Pitch-Konturen – links die Kontrollversion, rechts die Monotonversion

5.3 Proband*innen

Die Teilnehmer*innen wurden in Gruppe A und Gruppe B unterteilt. Gruppe A besteht aus 98 Teilnehmer*innen, darunter 87 italienischen Muttersprachler*innen und 11 Personen mit Italienischkenntnissen auf Niveau C1/C2 im Alter zwischen 20 und 70 Jahren. Gruppe B besteht aus 91 italienischen Muttersprachler*innen im Alter zwischen 19 und 79 Jahren. Neben der Beherrschung der italienischen Sprache ab Niveau C1 waren die Voraussetzungen für die Teilnahme, keine direkte Erfahrung im Bereich des Dolmetschens im Rahmen von Ausbildung

oder Beruf zu haben sowie mindestens einmal eine Medienverdolmetschung im Fernsehen gesehen zu haben. 12 Teilnehmer*innen aus Gruppe A und 2 aus Gruppe B wurden aufgrund fehlender Voraussetzungen ausgeschlossen.

Die Entscheidung, nicht nur Muttersprachler*innen, sondern auch Personen mit sehr guten Italienischkenntnissen (C1 und höher) einzubeziehen, beruht auf der Erkenntnis, dass das Sprachverständnis kein wesentlicher Faktor für die Überprüfung war, ob die Intonation die Bewertung der Dolmetschleistung beeinflusst. In jedem Fall kann ein gutes Textverständnis von Bedeutung sein. Daher wurden Personen mit einem Niveau unter C1 ausgeschlossen, da ihre Bewertungen durch andere Faktoren beeinflusst werden könnten, die mit dem Sprachverständnis zusammenhängen. Der Ausschluss von Nutzer*innen mit Erfahrung im Bereich Dolmetschen ist darauf zurückzuführen, dass erfahrene Nutzer*innen möglicherweise durch ihre eigene Erfahrung in diesem Bereich beeinflusst werden. Dies könnte dazu führen, dass sie nach anderen, subjektiven Kriterien bewerten als weniger erfahrenen Nutzer*innen. Schließlich ist die Voraussetzung, mindestens einmal einer Medienverdolmetschung beigewohnt zu haben, erforderlich, um sicherzustellen, dass die Nutzer*innen eine allgemeine Vorstellung davon haben, was Dolmetschen ist.

Die Entscheidung für ein „allgemeines“ Publikum, das im Gegensatz zu den Nutzer*innen der Studie von Collados Aís (1998) nicht spezialisiert ist, ist auf die mangelnde Verfügbarkeit von fachlich spezialisierten Nutzer*innen für Studien im Rahmen einer Masterarbeit zurückzuführen. Das allgemeine Publikum stellt potenzielle Teilnehmer*innen einer Konferenz dar, die der Vorstellung eines Buches gewidmet ist. Ihre Rekrutierung erfolgte nach nicht-probabilistischen Methoden, nämlich dem Network-Sampling und dem Schneeballverfahren. Ersteres besteht darin, ein bestehendes soziales Netzwerk zur Rekrutierung von Teilnehmenden zu nutzen, während Letzteres darin besteht, weitere Personen über Mitglieder einer bestimmten Gemeinschaft zu gewinnen (vgl. Mellinger & Hanson 2017: 13). Dank italienischsprachiger Bekannten wurde die Einladung zur Teilnahme an dem Experiment online an verschiedene Personen weitergeleitet, darunter auch an universitäre WhatsApp-Gruppen. Um eine möglichst ausgewogene Gruppeneinteilung für Gruppe A und B zu erreichen, wurde einigen WhatsApp-Gruppen die lebendige Version des Videos in der Kontrollbedingung und anderen die monotone Version des Videos zugesandt. Diese beiden Versionen werden in der vorliegenden Arbeit als Test A und Test B bezeichnet: Test A bezieht sich auf die Gruppe, die die lebendige Dolmetschleistung hörte (Kontrollversion), während Test B die Gruppe umfasst, die die monotone Version hörte. Die Empfänger*innen wurden

anschließend gebeten, den Fragebogen an weitere Personen aus ihrem Bekanntenkreis weiterzuleiten.

5.4 Versuchsablauf

Das Experiment wurde online mittels eines mit Google Forms erstellten Fragebogens durchgeführt. Dieser umfasste sowohl das Ansehen des Videos (Version A oder Version B) als auch die Bewertungsfragen. Somit fungierte der Fragebogen als zentrales Instrument zur Durchführung des Experiments und der Datenerhebung. Die Entscheidung, das Experiment online und nicht vor Ort durchzuführen, ist auf zwei Faktoren zurückzuführen: Erstens ermöglicht die Online-Durchführung eine wesentlich einfachere organisatorische Umsetzung des Experiments, und zweitens wird für diese Studie das Schneeballverfahren angewendet, wodurch es unmöglich ist, die Teilnehmer*innen planmäßig zu rekrutieren. Darüber hinaus wurde dieses Experiment im Jahr 2025 durchgeführt, also 27 Jahre nach dem von Collados Aís (1998). Heutzutage sind die meisten Menschen es gewohnt, Aktivitäten online durchzuführen, und sind eher bereit, Fragebögen in digitaler Form auszufüllen, anstatt sich vor Ort dafür einzufinden. Darüber hinaus ist es möglich, durch Online-Fragebögen in kurzer Zeit eine Vielzahl von Antworten zu erhalten (vgl. Dillman et al. 2014: 314f). Die Fragebögen wurden am 11.10.2025 online zur Verfügung gestellt und am 25.10.2025 geschlossen, nachdem eine ausreichende Anzahl für die Gruppen A (98) und B (91) für die Datenanalyse erreicht worden war.

Wie in Kapitel 2 dargelegt, ist der Fragebogen als Methode zur Datenerhebung in der Dolmetschforschung weit verbreitet. Fragebögen können auf verschiedene Weise durchgeführt werden, etwa per Post, Telefon, persönlich oder online (vgl. Zwischenberger 2015: 407).

Der für diese Masterarbeit verwendete Fragebogen wurde online mit Google Forms erstellt und wies folgende Struktur auf: eine kurze Einführung zum Rahmen der Konferenz und zum Experiment selbst, ein Eingangsfragebogen zur Auswahl der Proband*innen, um zu ermitteln, ob sie die Teilnahmebedingungen erfüllten, die Ansicht einer der beiden Versionen der Verdolmetschung, ein Fragebogen zur Bewertung der Dolmetschleistung und zwei abschließende Fragen zu Alter und Geschlecht der Teilnehmer*innen.

Bei der Erstellung von Fragebögen ist es von entscheidender Bedeutung, dass die Fragen so formuliert sind, dass sie für die Nutzer*innen verständlich sind. Sie müssen klare Anweisungen zur Beantwortung enthalten, z. B. „Hast du schon einmal an einer Konferenz mit Simultandolmetschen vor Ort teilgenommen? Wenn ja, wie häufig? (Bitte Zahl eingeben, z. B. 3)“ (vgl. Anhang IV). Die Fragen des Fragebogens wurden klar, einfach und direkt formuliert,

wobei zu technische Formulierungen vermieden wurden. Sie können außerdem zusätzliche Anweisungen oder Beispiele enthalten, die auf das Verständnis der Nutzer*innen ausgerichtet sind (vgl. Dillman et al. 2014: 141). Die Frage „Haben Sie jemals eine Fernsehsendung gesehen, in der simultan gedolmetscht wurde? Zum Beispiel in „Sanremo“ oder „Che tempo che fa“, wenn die Gäste eine Fremdsprache sprechen und der Ton ins Italienische übersetzt wird, damit das Publikum ihn verstehen kann“ wurde durch ein praktisches Beispiel mit Verweis auf zwei in Italien bekannte Sendungen ergänzt, in denen häufig ausländische Gäste eingeladen werden, die eine Simultanverdolmetschung benötigen.

Um die Privatsphäre der Teilnehmenden zu schützen, wurden ihre Antworten anonym erfasst, wie in der Einleitung zum Fragebogen angemerkt. Die Anonymität ist nämlich eine der Grundvoraussetzungen für die informierte Einwilligung (vgl. Mellinger 2017: 19). Dies ist ein grundlegendes Element in der Forschung, bei der Menschen beteiligt sind (vgl. 2017: 15).

Es wurde eine inklusive und respektvolle Sprache gegenüber allen Teilnehmer*innen verwendet. In der Einleitung wurde die Formulierung „Benvenuto“ (Willkommen) mit Sternchen verwendet, das im Italienischen verwendet wird, um alle Geschlechter einzuschließen. In der abschließenden Frage zum Geschlecht wurden folgende Antwortmöglichkeiten angeboten: „männlich“, „weiblich“, „divers“ und „keine Angabe“, um diskriminierende Dichotomien zu vermeiden (vgl. Anhang IV). Diese Entscheidung steht im Einklang mit der Aufmerksamkeit, die Gender-Themen in der historischen Periode, in der diese Arbeit veröffentlicht wurde, gewidmet wird.

5.5 Erhebungsinstrument: Fragebogen

5.5.1 Gestaltung und Layout

Das online durchgeführte Experiment erfolgte nach den Kriterien für webbasierte Experimentumfragen. Nachdem die Teilnehmer*innen auf den in der WhatsApp-Gruppe bereitgestellten Link geklickt hatten, gelangten sie zunächst auf eine einführende Seite, auf der sie begrüßt wurden. Dort war der Titel der Umfrage angeführt und die Forschende bedankte sich bei den Teilnehmer*innen für ihre Zeit und informierte sie über den Inhalt des Experiments. Diese Einführung stellt in Umfragen den ersten direkten Kontakt der Proband*innen mit dem Experiment dar und muss daher einladend sein und die Teilnehmer*innen dazu motivieren, den Fragebogen auszufüllen (vgl. Dillman et al. 2014: 326). Die letzte Seite der online Umfrage enthielt eine Danksagung für die Teilnahme (vgl. 2014: 328).

In Bezug auf die Layoutgestaltung wurden die Fragen des Eingangsfragebogens einzeln

nacheinander auf dem Bildschirm angezeigt. Dies ist typisch für Web-Umfragen, bei denen die Teilnehmer*innen jeweils auf eine Schaltfläche klicken müssen, um mit dem Ausfüllen fortzufahren. Dies hat den Vorteil, dass die Befragenden mehr Kontrolle darüber haben, wie jede Frage den Befragten angezeigt wird, und dass die Antworten auf diese Weise nach jeder Frage an den Server übermittelt und in der Datenbank gespeichert werden (vgl. 2014: 324).

Eine separate Seite für jede Frage erwies sich außerdem als praktisch, weil Teilnehmer*innen, die die Teilnahmebedingungen nicht erfüllten, aus dem Experiment ausgeschlossen werden konnten. Beispielsweise wurden Nutzer*innen, die die Frage „Hast du aktive Erfahrungen im Bereich Dolmetschen? (z.B. Ausbildung oder Beruf als Dolmetscher/in)?“ (vgl. Anhang IV) mit „Ja“ beantworteten, umgehend aus dem Fragebogen ausgeschlossen. Sie wurden auf die Seite mit der Danksagung weitergeleitet, auf der ihnen mitgeteilt wurde, dass sie den Fragebogen nicht weiter ausfüllen konnten (vgl. Mariel et al. 8: 2021).

Die Fragen in Bezug auf die Bewertungskriterien wurden auf einer einzigen Seite aufgeführt. Der Fragebogen bestand also aus zwei Teilen: dem Eingangsfragebogen, der vor dem Ansehen des Videos ausgefüllt wurde, und dem Fragebogen zur Qualitätsbewertung, der nach dem Ansehen des Videos ausgefüllt wurde. Im Folgenden werden der Fragebogen vor dem Ansehen des Videos und der Fragebogen nach dem Ansehen des Videos beschrieben.

5.5.2 Eingangsfragebogen

Wichtige Elemente, die beim Versuchsdesign zu berücksichtigen sind, sind die Kriterien für die Eignung der Teilnehmer*innen. Hierbei wird zwischen Einschlusskriterien und Ausschlusskriterien unterschieden. Erstere umfassen Kriterien, die erfüllt werden müssen, um die Teilnahme der Proband*in am Versuch zu ermöglichen. Zweitere umfassen Kriterien, die die Teilnahme verhindern, da sie die Forschungsziele beeinträchtigen (vgl. Sue & Ritter 2016: 4).

In diesem Zusammenhang wurden den Teilnehmer*innen, nachdem sie über das Thema des Experiments informiert worden waren, Fragen gestellt, um die Voraussetzungen für die Teilnahme zu überprüfen. Sie wurden beispielsweise zu ihren Italienischkenntnissen befragt, die in drei Kategorien unterteilt wurden: keine Kenntnisse, Grundkenntnisse (A1-A2), fortgeschrittene Kenntnisse (B1-B2) und sehr gute Kenntnisse (C1-C2), wobei nur Personen mit C1 bis C2-Niveau die Befragung fortsetzen durften. Die Entscheidung, auch Nicht-Muttersprachler*innen einzubeziehen, beruht auf der Tatsache, dass sich das Experiment nicht ausschließlich an eine Gruppe von Muttersprachler*innen als Vertreter*innen der italienischen

Kultur richtet, sondern an Personen, die die Sprache an sich beherrschen. Personen mit einem Niveau unter C1 wurden hingegen ausgeschlossen, um mögliche Störungen aufgrund sprachlicher Unterschiede zwischen den Teilnehmer*innen zu vermeiden. Sie hätten den Zieltext zur Bewertung zwar verstehen können, doch die Bewertung hätte aufgrund des geringeren Sprachlevels womöglich auf Basis anderer Kriterien als der Qualität der Verdolmetschung selbst erfolgen können.

Ein weiteres Kriterium zur Überprüfung der Eignung der Teilnehmer*innen war die mindestens einmalige Beobachtung einer Medienverdolmetschung, um sicherzustellen, dass die Teilnehmer*innen über grundlegende Kenntnisse im Bereich Dolmetschen verfügten. Die letzte grundlegende Voraussetzung betraf das Fehlen direkter Erfahrungen mit Dolmetschen im Rahmen der Ausbildung oder der beruflichen Laufbahn. Dolmetscher*innen und Dolmetschstudierende könnten durch eine Intonation *sui generis* beeinflusst sein, die die Bewertung der monotonen Version beeinträchtigen würde (siehe Kap. 4.3). Alle Teilnehmer*innen, deren Antworten diese Anforderungen nicht erfüllten, wurden ausgeschlossen.

Der Eingangsfragebogen enthielt zwei weitere Fragen, die als Störvariablen behandelt wurden. Diese hatten keinen Einfluss auf die Einbeziehung oder den Ausschluss aus dem Test, sondern dienten der Differenzierung der Teilnehmer*innen sowie der Kontrolle einer potenziellen Veränderung der Beziehung zwischen den abhängigen und unabhängigen Variablen (vgl. Mellinger 2017: 4). Eine Frage betraf die Deutschkenntnisse. Teilnehmer*innen mit fließenden Deutschkenntnissen hätten die Leistung möglicherweise anders bewerten können, da sie durch das Verständnis beider Texte beeinflusst waren. Eine weitere Frage betraf die Häufigkeit der vorherigen Teilnahme an Konferenzen, bei denen Simultandolmetschen zum Einsatz kam. Ziel war es, eine mögliche Veränderung der Qualitätswahrnehmung zu untersuchen, die auf eine Gewöhnung an eine Intonation *sui generis* bei simultan gedolmetschten Konferenzen zurückzuführen sein könnte.

5.5.3 Fragebogen zur Qualitätsbewertung

Nach dem Ansehen des Videos füllten die Teilnehmenden den zweiten Teil des Fragebogens aus, in dem sie die Dolmetschleistung anhand verschiedener Kriterien bewerteten. Die Bedeutung dieser Kriterien wurde vor dem Ansehen des Videos in einfachen Worten erläutert, um sie so verständlich wie möglich zu machen. Die zu vergebenden Punkte wurden auf einer Skala von 1 bis 5 verteilt.

Die für den Test ausgewählten Kriterien basieren auf denen, die Collados Aís (1998) in

ihrer Studie „3A“ zur Bewertung der Qualität des Simultandolmetschens durch die Nutzer*innen verwendete, die Teil ihrer Doktorarbeit aus dem Jahr 1998 ist (vgl. Collados Aís 1998: 165). Um die Kohärenz und Signifikanz der Daten zu gewährleisten, wurden einige Kriterien ausgeschlossen. Einige Kriterien wurden weggelassen, da sie nur relevant gewesen wären, wenn diese Masterarbeit auch den Fragebogen zu den Erwartungen an die Qualität einer Dolmetschleistung enthalten hätte. Andere wurden weggelassen, weil sie womöglich vom Gegenstand der betreffenden Forschung ablenken könnten.

Im Folgenden werden die Gründe für den Ausschluss beziehungsweise die Aufnahme bestimmter Kriterien von Collados Aís (1998) erläutert. In ihrer Studie wurden folgende Kriterien für die Befragung herangezogen: Gesamtbewertung, muttersprachlicher Akzent, angenehme Stimme, Flüssigkeit, Kohärenz, Sinnübereinstimmung mit dem Original, Vollständigkeit, korrekte Terminologie, Stil, Intonation, Diktion, Professionalität, Zuverlässigkeit, Bewertung der Originalrede. Davon wurden folgende Kriterien in die vorliegende Arbeit aufgenommen: Gesamtbewertung, Flüssigkeit, Kohärenz, Intonation und Professionalität. Das Kriterium der Gesamtbewertung wurde beibehalten, da es sich um einen globalen Indikator handelt, der die allgemeine Wahrnehmung der Leistung zusammenfasst und von grundlegender Bedeutung ist, um zu überprüfen, wie sich eine monotone Intonation negativ auf die Bewertung der Qualität der Dolmetschung auswirkt. Das Kriterium des muttersprachlichen Akzents wurde ausgeschlossen, da es nur relevant gewesen wäre, wenn den Teilnehmer*innen ein Erwartungsfragebogen vorgelegt worden wäre. Die Dolmetscherin war italienische Muttersprachlerin, sodass eine Bewertung ihres muttersprachlichen Akzents irrelevant gewesen wäre. Das Kriterium der angenehmen Stimme wurde ausgeschlossen, da es die Nutzer*innen bei der Bewertung der Intonation hätte ablenken können und redundant gewesen wäre. Das Kriterium der Flüssigkeit wurde hingegen aufgenommen, da es von den Nutzer*innen leicht zu bewerten und zu erkennen war. Darüber hinaus wurde auch das Kriterium Kohärenz einbezogen. Obwohl im Gegensatz zur Studie von Collados Aís (1998) der Sinn des Originaltextes nicht manipuliert wurde, konnten die Nutzer*innen dennoch eine Bewertung der Kohärenz des Textes vornehmen. Die Kriterien Sinnübereinstimmung mit dem Original und Vollständigkeit wurden hingegen ausgeschlossen, da die meisten Teilnehmer*innen den Ausgangstext nicht kannten und diese Parameter daher nur für einen Erwartungsfragebogen relevant gewesen wären. Außerdem wurde korrekte Terminologie ausgeschlossen, da der Text nur Basisvokabular und keine Fachtermini enthielt. Die Kriterien Stil und Diktion wurden nicht einbezogen, da sie den Fokus von der Intonation hätten ablenken können. Die Intonation wurde einbezogen, da sie die unabhängige Variable darstellte. Das

Kriterium der Professionalität wurde als entscheidend für die Beantwortung der Forschungsfrage angesehen, inwiefern Monotonie die wahrgenommene Professionalität von Dolmetscher*innen beeinflusst. Die Bewertung der Originalrede wurde ausgeschlossen, da sie für die Forschungsziele nicht relevant war. Letztlich wurde das Kriterium der Zuverlässigkeit weggelassen, da eine Verwechslung mit Professionalität nicht ausgeschlossen werden konnte, was zu einer Verwässerung der Daten geführt hätte. Die Teilnehmer*innen der Studie hatten keine Erfahrung im Bereich des Dolmetschens und hätten die Bedeutung von Zuverlässigkeit in diesem Kontext nicht nachvollziehen können.

Im Folgenden werden die Antwortskalen des Fragebogens erläutert, die sowohl im Eingangsfragebogen als auch im Bewertungsfragebogen verwendet wurden.

5.5.4 Verwendete Skalen

Der Fragebogen bestand überwiegend aus geschlossenen Fragen. Bei diesem Fragetyp können die Proband*innen eine oder mehrere Optionen aus einer Liste von Antwortkategorien auswählen (vgl. Dillman et al. 2014: 163). Diese Kategorien können nominal oder ordinal sein (vgl. Dillman et al. 2014: 143). In der vorliegenden Studie wurden sowohl ordinale als auch nominale Skalen verwendet. Die Ordinalskala bietet eine Reihe von Antwortmöglichkeiten, die auf einem Kontinuum angeordnet sind (vgl. Dillman et al. 2014: 143).

Alle Skalen des Fragebogens zur Leistungsbewertung waren ordinale Skalen, beispielsweise: „Gesamtbewertung: 1 = überhaupt nicht gut, 2, 3, 4, 5 = sehr gut“. Die Ordinalskalen waren unipolar, d. h. sie maßen ein Kriterium mit zunehmender Intensität entlang einer Skala: z. B. 1 = überhaupt nicht gut, 5 = sehr gut (vgl. Döring 2023: 247). Das galt auch für die beiden Fragen zu den Italienisch- und Deutschkenntnissen: „keine, Grundkenntnisse (A1-A2), fortgeschrittene Kenntnisse (B1-B2), sehr gute Kenntnisse (C1-C2)“ (vgl. Anhang IV). Obwohl der Abstand zwischen den Werten mathematisch nicht präzise ist, werden Ordinalskalen häufig verwendet, da sie den Aufwand für die Proband*innen reduzieren, die keine subjektiven Meinungen oder Zustände quantifizieren müssen (vgl. Dillman et al. 2014: 178). Sie umfassten maximal fünf Antwortkategorien, da Skalen dieses Umfangs valider und reliabler sind (vgl. Dillman et al. 2014: 195). Jedes zu bewertende Kriterium wurde mit dem Wert 1 als Mindestwert (z. B. für „nicht gut“) und 5 als Höchstwert (für „sehr gut“) definiert. Die Werte 2 bis 4 wurden hingegen nicht definiert, um eine Überforderung der Proband*innen durch zu viele verbale Kategorien zu vermeiden. Diese Art von Skala wird als *polar point labeled scale* bezeichnet (Dillman et al. 2014: 185).

Die Nominalskala umfasst eine Reihe von Kategorien, die keine natürliche Reihenfolge

untereinander aufweisen (vgl. Dillman et al. 2014: 143). Die meisten Fragen des Eingangsfragebogens enthielten diese Art von Skala, z. B. „Ist Italienisch deine Muttersprache?“ Mögliche Antworten: Ja oder Nein (vgl. Anhang IV).

Für die Fragen zur Häufigkeit der Teilnahme an Konferenzen als Publikum sowie zur Altersangabe wurde keine geschlossene Frage verwendet. Stattdessen kam eine offene Frage mit Frageansatz und verbaler numerischer Anleitung zum Einsatz (vgl. Dillman et al. 2014: 141). Die Fragen lauteten: Hast du schon einmal an einer Konferenz mit Simultandolmetschen vor Ort teilgenommen? Wenn ja, wie häufig? (Bitte Zahl eingeben, z. B. 3) Antwort: „ca ... Mal“ und: „Dein Alter (bitte als Zahl eingeben, z. B. 25) Antwort: ... Jahre“ (vgl. Anhang IV).

5.6 Datenanalyse

Die Datenanalyse diente der Überprüfung der Forschungshypothesen. Ziel war es, festzustellen, ob eine monotone Intonation die Bewertung der Qualität der Dolmetschleistung und die Wahrnehmung der Professionalität der Dolmetscherin durch die Nutzer*innen negativ beeinflusst. Zudem sollte überprüft werden, ob ein lebhafterer Ausgangstext zu noch niedrigeren Bewertungen in der monotonen Dolmetschversion führen könnte.

Zunächst wurden die Ergebnisse von Test A mit denen von Test B der vorliegenden Studie verglichen. Danach wurden die Mittelwertdifferenzen zwischen den lebhaften und monotonen Versionen der vorliegenden Studie (Test A und Test B) mit denen der Studie von Collados Aís (1998) (Video 3 und Video 1) verglichen. Anschließend wurde ein direkter Vergleich zwischen den monotonen Versionen (Test B und Video 1) der beiden Studien durchgeführt. In Anlehnung an den methodischen Ansatz von Collados Aís (1998) wurden für die Datenanalyse sowohl deskriptive als auch inferenzstatistische Methoden angewendet, unter Einsatz nichtparametrischer Tests und begleitet von einer grafischen Darstellung der Daten (vgl. Collados Aís 1998: 98). Alle nichtparametrischen Tests wurden mit dem Programm JASP durchgeführt, mit Unterstützung des Handbuchs von Andy Field (2025): „Discovering Statistics Using JASP“ (vgl. Field et al. 2025).

5.6.1 Datenüberprüfung für nichtparametrische Tests

Um die Bewertungen der Tests A und B für jedes Bewertungskriterium zu vergleichen, wurde der nichtparametrische Mann-Whitney-U-Test angewendet, der üblicherweise zum Vergleich zweier unabhängiger Gruppen verwendet wird (vgl. Mellinger 2017: 88). Für die vergleichende Analyse zwischen den Gruppen verwendete Collados Aís (1998) hingegen den Kruskal-Wallis-Test, was durch das Vorhandensein von drei zu vergleichenden Gruppen erforderlich war (vgl.

Collados Aís 1998: 98). Der Mann-Whitney-U-Test weist in Fällen, in denen die Daten nicht normalverteilt sind, eine höhere statistische Aussagekraft auf als parametrische Methoden. Obwohl der Test häufig bei relativ kleinen Stichproben verwendet wird, beispielsweise bei weniger als 30 Beobachtungen pro Gruppe, wurde er für diese Arbeit verwendet, da die Stichprobengröße nicht der einzige entscheidende Faktor für die Wahl des Tests ist. Vielmehr ist die Normalität (oder Nicht-Normalität) der zugrunde liegenden Population das entscheidende Kriterium (vgl. Mellinger 2017: 100f.). Zur Überprüfung der Verteilung der Daten dieser Studie wurde eine grafische Darstellung herangezogen, da grafische Methoden eine höhere Zuverlässigkeit aufweisen als Voruntersuchungen zur Identifizierung von Nicht-Normalität (vgl. Mellinger 2017: 93).

Abbildung 7 zeigt ein Q-Q-Plot-Diagramm. Dieses dient dazu, die Form der Verteilung einer Variablen zu beobachten. Dabei werden die Quantile der beobachteten Daten mit den erwarteten Quantilen verglichen, die auftreten würden, wenn die Daten einer bestimmten Verteilung, zum Beispiel der Normalverteilung, folgen (vgl. Field et al. 2025: 247). Der Begriff Quantile bezeichnet Werte, die einen gegebenen Datensatz in gleiche Teile aufteilen (vgl. Field et al. 2025: 725). Der mit JASP erstellte Q-Q-Plot links in Abbildung 7 illustriert die Verteilung der Bewertungen des Schlüsselkriteriums Intonation der Tests A und B. Unter der Prämisse der Normalverteilung der Daten wäre eine diagonale Linie zu erwarten, wie sie in Abbildung 7, rechts, im Diagramm dargestellt ist. Da die Daten im Diagramm links dieser diagonalen Linie nicht folgen, wird die Annahme einer normalen Verteilung der Daten widerlegt (vgl. Field et al. 2025: 252). Aus dieser Feststellung ergibt sich die Entscheidung, die vergleichende Analyse der Ergebnisse der Tests A und B mit nichtparametrischen Tests durchzuführen.

Q-Q Plot

Intonation

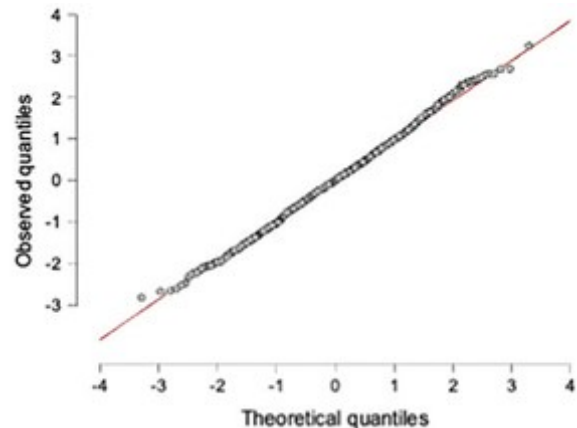
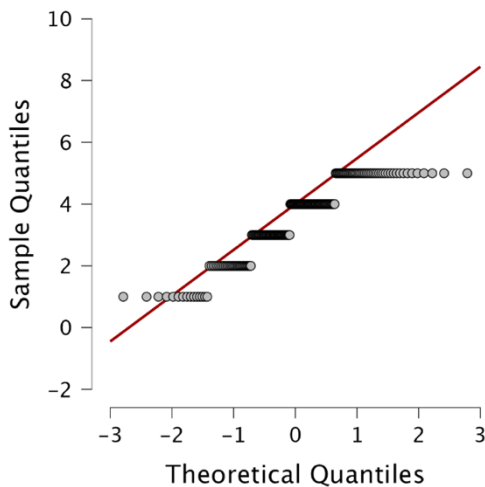


Abb. 7: Links: Verteilungsdarstellung der Bewertungen des Kriteriums Intonation für Test A und B; Rechts: Beispiel einer Normalverteilung der Daten (Field et al. 2025: 251)

5.6.2 Nichtparametrische Tests zur Untersuchung von Bewertungskriterien und Störvariablen

Für jedes Bewertungskriterium wurden zunächst die Häufigkeiten der einzelnen Bewertungen für die Gruppen A und B in Tabellen erfasst. Jede Häufigkeitstabelle wurde zur besseren Übersichtlichkeit mit einem Diagramm versehen und mit dem Mann-Whitney-U-Test auf eine statistisch signifikante Differenz zwischen den Bewertungen in Test A und B überprüft. Die Signifikanz wurde anhand des p-Wertes gemessen, der, wenn er kleiner als 0,05 ist, die Nullhypothese widerlegt (vgl. Mellinger 2017: 75). Diese besagt, dass kein Unterschied oder Zusammenhang zwischen den Variablen besteht (vgl. Mellinger 2017: 6).

Anschließend wurden die beiden Störvariablen „Teilnahme an Konferenzen“ und „Deutschkenntnisse“ analysiert. Zunächst wurde der Mann-Whitney-U-Test separat für Gruppe A und B durchgeführt, um die Teilnehmer*innen, die bereits an Konferenzen mit Dolmetscher*innen teilgenommen hatten, mit denen zu vergleichen, die noch nie an solchen Konferenzen beteiligt waren. Die beiden Gruppen waren für beide Tests ungleich groß. Im Test A hatten 16 Teilnehmende bereits an Konferenzen teilgenommen, 82 nicht. Im Test B hatten 15 Teilnehmende bereits an Konferenzen teilgenommen, 76 nicht. Der Mann-Whitney-U-Test ist jedoch dafür geeignet, da er auch bei ungleichen Gruppengrößen angewendet werden kann (vgl. Conover 1999: 285).

Um zu überprüfen, ob die Häufigkeit der Teilnahme an Konferenzen wiederum einen

Einfluss auf die Leistungsbewertung hatte, wurde der nichtparametrische Kendall-Tau-Korrelationstest durchgeführt. Eine häufigere Teilnahme an Konferenzen könnte nämlich zu einer abweichenden Wahrnehmung der Leistungsqualität führen, da Teilnehmer*innen mit mehr Erfahrung in der Konferenzteilnahme möglicherweise eher an eine Intonation *sui generis* gewöhnt sind (siehe Kapitel 4.3). Der Kendall-Test eignet sich optimal zur Schätzung der Korrelation in der Population und wird verwendet, wenn ein kleiner Datensatz vorliegt (vgl. Field et al. 2025: 716). Der Stichprobenumfang von Gruppe A umfasste 16 Nutzer*innen, die bereits an Konferenzen teilgenommen hatten, während es in Gruppe B 15 waren.

In Bezug auf die Störvariable der Deutschkenntnisse wurde der Mann-Whitney-U-Test durchgeführt, obwohl die akustische Wahrnehmung des Ausgangstextes im Vergleich zur akustischen Wahrnehmung der Verdolmetschung eingeschränkt war. Ziel war die Überprüfung, ob sich die Bewertungen zwischen den Teilnehmer*innen mit und ohne Deutschkenntnissen wesentlich unterschieden. Es wurde nicht überprüft, ob die Leistungsbewertung je nach Sprachniveaus variieren könnte. Dies war nämlich nicht das Ziel dieser Analyse. Vielmehr ging es um die Identifizierung möglicher grober Unterschiede zwischen den beiden Gruppen.

5.6.3 Deskriptive Analyse: Intra- und Interstudien-Vergleich mit der Studie von Collados Aís (1998)

Abschließend wurde überprüft, wie stark sich die monotone und die lebhaft Version innerhalb jeder Studie in Bezug auf die Mittelwerte unterscheiden. Das Ziel war es, zu überprüfen, ob die monotone Version der Dolmetschleistung bei einem lebhaften Ausgangstext noch negativere Bewertungen erzeugt, indem die Größe des Unterschieds zwischen den Mittelwerten der Bewertungen der lebhaften und der monotonen Version innerhalb jeder Studie gemessen wurde. Die Videos aus der Studie von Collados Aís (1998), die für die Analyse herangezogen wurden, sind Folgende: Video 3 der Kontrollversion und Video 1 der monotonen Version, wobei Video 2, in dem der Sinn der Rede manipuliert wurde, ausgelassen wurde (siehe Kapitel 4.2.1). Zu diesem Zweck wurde für jedes Bewertungskriterium in beiden Studien die Differenz zwischen dem Mittelwert der lebhaften Version und dem der monotonen Version berechnet. Dieser Indikator ermöglicht es, die Effektstärke zwischen den beiden Bedingungen innerhalb jeder Studie zu beobachten (vgl. Field et al. 2025: 383). Auf diese Weise kann bewertet werden, ob die Auswirkung der Monotonie auf die Bewertung durch die Nutzer*innen in der vorliegenden Arbeit stärker ausgeprägt ist als in der von Collados Aís (1998).

Darüber hinaus wurde für eine umfassendere Analyse ein deskriptiver Vergleich zwischen den Bewertungen der monotonen Versionen beider Studien (Test B und Video 1)

durchgeführt. Dies ermöglicht es, die Ergebnisse der vorliegenden Studie in einen vergleichenden Kontext zu stellen. Eine vergleichende Analyse der Kontrollversionen der beiden Studien (Test A und Video 3) wurde nicht vorgenommen, da dies nicht zu den Zielen der Untersuchung gehörte.

5.6.3.1 Vorgehen der deskriptiven Analysen

Der deskriptive Ansatz bietet ein klares Verständnis und eine visuelle Darstellung der analysierten Daten, sowohl grafisch als auch numerisch. Im Gegensatz zum inferentiellen Ansatz sieht er keine Überprüfung der Hypothesen der beobachteten Daten vor, und jede Schlussfolgerung, die über eine reine Interpretation der Daten hinausgeht, ist spekulativ (vgl. Bortz & Schuster 2010: 3). Deshalb können die Schlussfolgerungen, die in Kapitel 7 gezogen werden, nicht statistisch auf die Gesamtbevölkerung angewandt werden, sie spiegeln nur die in den Daten sichtbaren Tendenzen wider.

Die Publikation von Collados Aís (1998), die als Grundlage für die Erstellung dieser Masterarbeit herangezogen wurde, gibt für jedes bewertete Kriterium ausschließlich die Häufigkeitsverteilungen der Bewertungen in Tabellen und die Mittelwerte an (vgl. Collados Aís 1998: 171). Diese Mittelwerte wurden genutzt, um die Effektstärke auf die Bewertungsmittelwerte für die Versionen von Collados Aís (1998) zu berechnen, während in der vorliegenden Studie die Mittelwerte mit dem Programm JASP ermittelt wurden. Die Berechnung der Effektstärken wurde in beiden Studien manuell durchgeführt. Der Vergleich beschränkt sich auf die deskriptive Analyse der Unterschiede zwischen den Mittelwerten, ohne die statistische Signifikanz zwischen den beiden Stichproben zu überprüfen. Dies ist auf das Fehlen der Originaldaten von Collados Aís (1998) zurückzuführen. Um beispielsweise einen standardisierten Effektwert wie den von Cohen's d zu berechnen, wären die für diese Berechnung erforderlichen Standardabweichungen erforderlich (vgl. Field et al. 2025: 129).

Auch die vergleichende Analyse der beiden monotonen Versionen der beiden Studien erfolgte auf deskriptive Weise. Die Entscheidung für eine deskriptive Analyse ergab sich aus der Tatsache, dass die zu vergleichenden Daten aus zwei unterschiedlichen Studien stammten. Mellinger (2017) beschreibt in seinem Werk den Einsatz nichtparametrischer Tests zum Vergleich von Gruppen innerhalb eines einzigen Studienkontexts (vgl. 2017: 87). Da hier jedoch unabhängige Studien vorliegen, wurde eine deskriptive Vorgehensweise gewählt, die den Gegebenheiten der Daten entspricht. Darüber hinaus enthielt die Publikation von Collados Aís (1998), wie bereits erwähnt, nicht die Originaldaten, die für die Durchführung nichtparametrischer Tests erforderlich sind.

Die Analyse wurde sowohl anhand von Häufigkeitstabellen als auch mit dem Programm JASP durchgeführt. Die Häufigkeitsmaße können als absolute oder relative Häufigkeiten dargestellt werden. Erstere geben die exakte Anzahl der Beobachtungen pro Kategorie an und sind bei kleinen Stichproben häufiger anzutreffen. Letztere geben den Prozentsatz der Beobachtungen an, die zu jeder Kategorie im Verhältnis zur Gesamtstichprobe gehören, und werden für größere Stichproben verwendet (vgl. Mellinger 2017: 41). In ihrer Studie hat Collados Aís (1998) die Tabellen der absoluten Häufigkeiten für jede Kategorie bereitgestellt (vgl. Collados Aís 1998: 171). Diese wurden in einzelne Bewertungen umgewandelt und in einer Excel-Tabelle organisiert, um die Berechnung der deskriptiven Statistiken mit JASP durchzuführen. Für jedes Kriterium wurden sowohl die absoluten als auch die relativen Häufigkeiten der beiden Studien verglichen, da die Stichprobengrößen unterschiedlich waren. Die Studie von Collados Aís (1998) mit Video 1 umfasste eine kleine Stichprobe von 14 Teilnehmer*innen während an der Studie mit dem Tests B 91 Teilnehmer*innen teilnahmen. Die relativen Häufigkeiten wurden für beide Datensätze auf Grundlage der absoluten Werte berechnet.

Mittels der Option „descriptive statistics“ (deskriptive Statistik) des Programmes JASP, wurden für jedes Kriterium die wichtigsten Merkmale der Stichprobendaten angegeben, darunter die Lagemaße wie Median und Mittelwert sowie Streuungsmaße wie Standardabweichung (vgl. Mellinger 2017: 40).

Die Lagemaße sind Werte, die eine Datenreihe am besten repräsentieren (vgl. Mellinger 2017: 41). In der vorliegenden Untersuchung wurden der Median sowie der arithmetische Mittelwert berechnet (vgl. 2017: 42f.). Was die Variabilitätsmaße betrifft, also Maße, die die Verteilung oder Variabilität der Daten angeben, wurde die Standardabweichung berechnet. (vgl. 2017: 47). Die Standardabweichung gibt Aufschluss darüber, inwieweit die Werte vom Mittelwert abweichen. Je größer der Wert, desto stärker sind die Daten gestreut, d. h., desto stärker weichen sie vom Mittelwert ab; je kleiner der Wert, desto geringer ist die Variabilität der Daten und desto näher liegen sie am Mittelwert (vgl. Mellinger 2017: 49).

Um einen vollständigen Überblick über die Daten zu bieten, wurden alle inferentiellen Analysen durch die Tabelle „Group Descriptives“ ergänzt, in der die Anzahl der Teilnehmer*innen, der Mittelwert und die Standardabweichung angegeben wurden.

6 Ergebnisse

Die für Test A gesammelten Antworten wurden von 96 Proband*innen im Alter zwischen 18 und 79 Jahren bereitgestellt, von denen 60,2 % weiblich, 1 % ohne Angabe und 38,8 % männlich sind. Test B erhielt insgesamt 91 Antworten. Die Altersspanne der Teilnehmenden an Test B liegt zwischen 20 und 83 Jahren, wobei 74,4 % der Proband*innen weiblich und 25,6 % männlich sind. Zwölf Antworten für Test A und zwei Antworten für Test B wurden von Google Forms automatisch ausgeschlossen, da sie die Teilnahmekriterien nicht erfüllten. Die Datenerhebung erfolgte vom 11.10.2025 bis zum 25.10.2025, dem Tag, an dem die Option zur Teilnahme am Test auf der Online-Plattform deaktiviert wurde, da eine zufriedenstellende Anzahl von Antworten erreicht wurde.

Im Folgenden werden zunächst die Ergebnisse der Tests A und B für jedes Kriterium aufgeführt, die mittels des Mann-Whitney-U-Tests verglichen und durch Häufigkeitstabellen und Säulendiagramme ergänzt werden. Anschließend wird für jeden Test separat analysiert, ob die Störvariablen die Bewertungen beeinflussen. Abschließend erfolgt ein deskriptiver Vergleich zwischen der vorliegenden Studie und der Studie von Collados aus dem Jahr 1998, wobei Häufigkeitstabellen und Lagemaße angegeben werden.

Wie bereits erwähnt, wurden die statistischen Tests mit dem Programm JASP durchgeführt. Es wurden ausschließlich die für die Interpretation der Ergebnisse relevanten Statistiken angegeben. In den folgenden Abschnitten bezeichnet „N“ die Stichprobengröße, „SD“ die Standardabweichung und „p“ das Signifikanzniveau.

6.1 Vergleich der Ergebnisse von Test A und Test B

Im Folgenden werden die Ergebnisse der Bewertungen für Test A und B für jedes Bewertungskriterium der Dolmetschleistung verglichen, wobei die Reihenfolge des Fragebogens, der den Nutzer*innen vorgelegt wurde, beibehalten wird.

Wie aus Abbildung 8 hervorgeht, liegen die Antworten der Proband*innen von Test B beim Kriterium der Gesamtbewertung am häufigsten auf den mittleren Werten (3 und 4), mit 36 bzw. 37 Antworten. Im Gegensatz dazu vergibt die Mehrheit der Proband*innen (53) in Test A die höchste Bewertung (5), während 38 Personen den Wert 4 auswählten (Tab. 2). Die Unterschiede zwischen den beiden Gruppen sind gemäß dem Mann-Whitney-U-Test statistisch signifikant, wobei $p < 0,001$ beträgt, was darauf hindeutet, dass die Bewertungen von Test A signifikant höher sind als die von Test B (Tab. 3). Die deskriptiven Statistiken zeigen, dass die

Gesamtbewertung in Test A mit einem Mittelwert von 4,47 und einer Standardabweichung von 0,63 höher und homogener ist als in Test B, der einen Mittelwert von 3,67 und eine Standardabweichung von 0,79 aufweist (Tab. 4). Dies bedeutet, dass die Teilnehmer*innen Test A insgesamt positiver und einheitlicher bewertet haben, während die Bewertungen in Test B niedriger ausfielen und eine größere Streuung aufweisen.

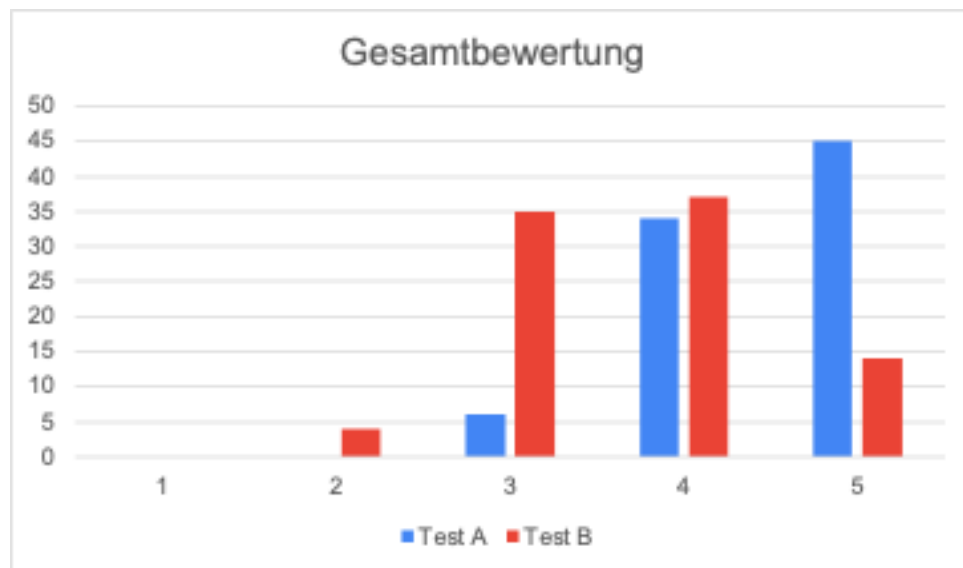


Abb. 8: Verteilung der Bewertungen für das Kriterium Gesamtbewertung

Tab. 2: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung

Bewertung	Test A	Test B
1	0	0
2	0	4
3	7	36
4	38	37
5	53	14

Tab. 3: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung

	U	p
Gesamtbewertung	6829	< .001

Tab. 4: Deskriptive Statistiken für das Kriterium Gesamtbewertung

Gesamtbewertung	Gruppe	N	Mittelwert	SD
	Test A	98	4.469	0.629
	Test B	91	3.670	0.790

Für das Kriterium Flüssigkeit vergaben die meisten Nutzer*innen von Test A die höheren Werte 4 und 5, mit jeweils 42 und 41 Bewertungen. In Test B liegen die meisten Bewertungen um den Wert 3. Die Werte von Test A sind daher tendenziell höher als die von Test B (Abb. 9 und Tab. 5). Gemäß dem Mann-Whitney-U-Test sind die Unterschiede zwischen den beiden Gruppen mit $p < 0,001$ statistisch signifikant (Tab. 6). Der Mittelwert von Test A beträgt 4,27 mit einer Standardabweichung von 0,71, was darauf hindeutet, dass die Bewertungen insgesamt hoch und relativ konzentriert um den Mittelwert liegen. Der Mittelwert von Test B beträgt 3,33 mit einer SD = 0,98, was auf niedrigere Bewertungen und eine geringere Homogenität der Werte im Vergleich zu Test A hindeutet (Tab. 7).

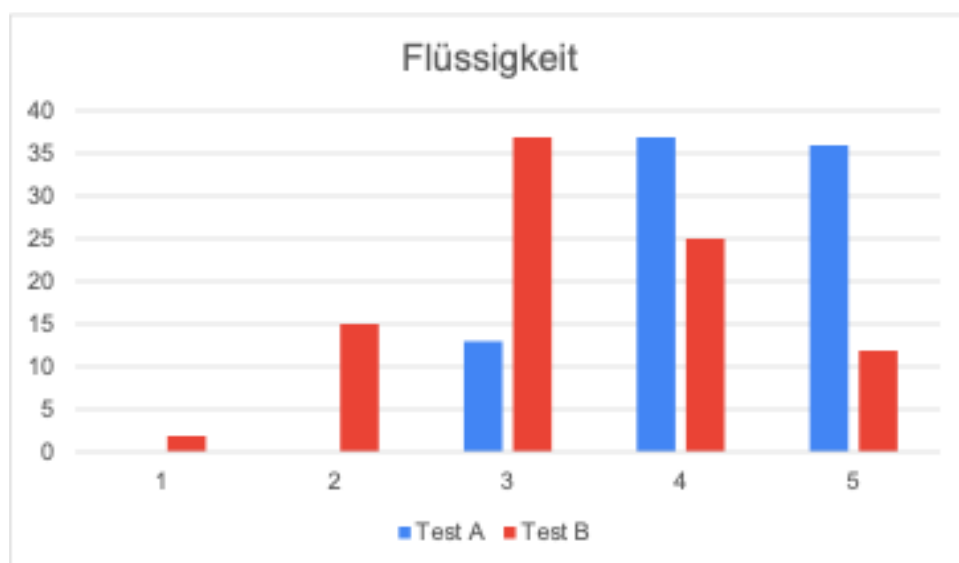


Abb. 9: Verteilung der Bewertungen für das Kriterium Flüssigkeit

Tab. 5: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit

Bewertung	Test A	Test B
1	0	2
2	0	15
3	15	37
4	42	25
5	41	12

Tab. 6: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit

	U	p
Flüssigkeit	6811	< .001

Tab. 7: Deskriptive Statistiken für das Kriterium Flüssigkeit

Flüssigkeit	Gruppe	N	Mittelwert	SD
	Test A	98	4.265	0.711
	Test B	91	3.330	0.978

Beim Kriterium Intonation vergaben die meisten Teilnehmer*innen des Tests A hohe Bewertungen: 35 Personen wählten den Wert 4, 45 Personen den Wert 5. Im Gegensatz dazu liegt die Verteilung in Test B eher bei niedrigen und mittleren Werten, wobei die meisten Teilnehmer*innen die Werte 2 und 3 vergaben (29 bzw. 26 Personen), während nur 4 Personen den höchsten Wert 5 vergaben (Abb. 10 und Tab. 8). Diese Daten zeigen deutlich eine Tendenz der Teilnehmer*innen, Test A positiver zu bewerten als Test B. Der Unterschied in den Bewertungen zwischen den beiden Tests für die Intonation ist statistisch signifikant (Tab. 9). Auf der deskriptiven Ebene weist Test A einen Mittelwert von 4,27 auf, der höher ist als der Mittelwert von Test B mit 2,63. In Test B, wo die SD bei 1,10 liegt, sind die Bewertungen heterogener als in Test A, der eine SD von 0,78 aufweist (Tab. 10).

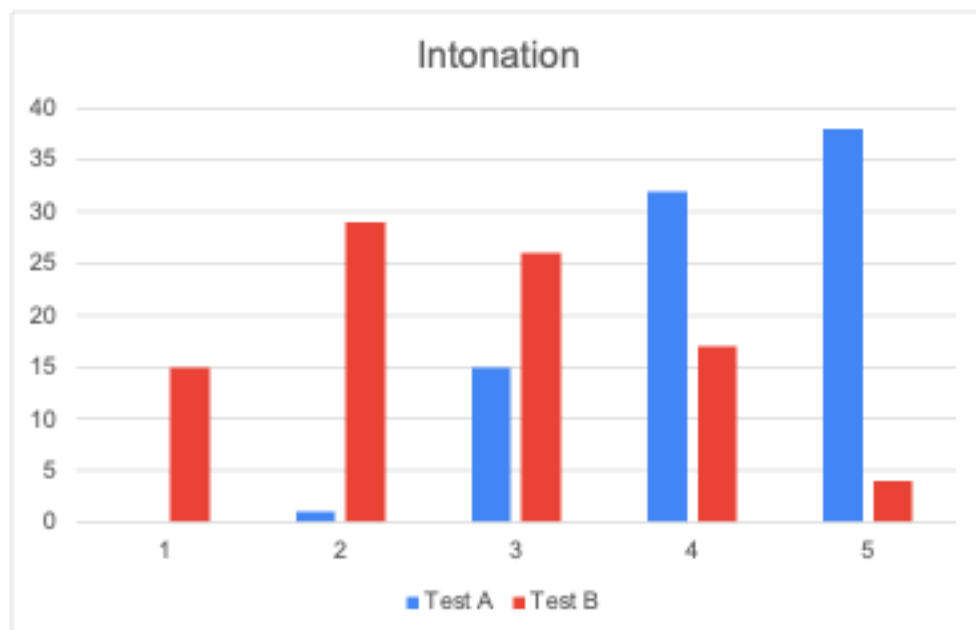


Abb. 10: Verteilung der Bewertungen für das Kriterium Intonation

Tab. 8: Häufigkeiten der Bewertungen für das Kriterium Intonation

Bewertung	Test A	Test B
1	0	15
2	1	29
3	17	26
4	35	17
5	45	4

Tab. 9: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation

	U	p
Intonation	7751	< .001

Tab. 10: Deskriptive Statistiken für das Kriterium Intonation

Intonation	Gruppe	N	Mittelwert	SD
	Test A	98	4.265	0.781
	Test B	91	2.626	1.102

Beim Kriterium Kohärenz konzentrieren sich die Bewertungen für beide Tests überwiegend auf hohe Werte. Im Test A vergaben die meisten Teilnehmer*innen die Werte 4 und 5, die jeweils von 33 und 58 Personen ausgewählt wurden. Auch im Test B vergaben die meisten Nutzer*innen eine hohe Bewertung, wobei 44 Personen dieses Kriterium mit 4 bewerteten (Abb. 11 und Tab. 11). Obwohl sich die Bewertungen beider Tests auf hohe Werte konzentrieren, ist die Verteilung der Bewertungen in den beiden Tests unterschiedlich. Die Bewertungen in Test A sind homogener (SD: 0,63), während die in Test B heterogener sind (SD: 0,98). Der Mittelwert von Test A liegt mit 4,52 höher als der Mittelwert von Test B mit 3,70 (Tab. 13). Der Unterschied zwischen Test A und B für dieses Kriterium ist signifikant (Tab. 12)

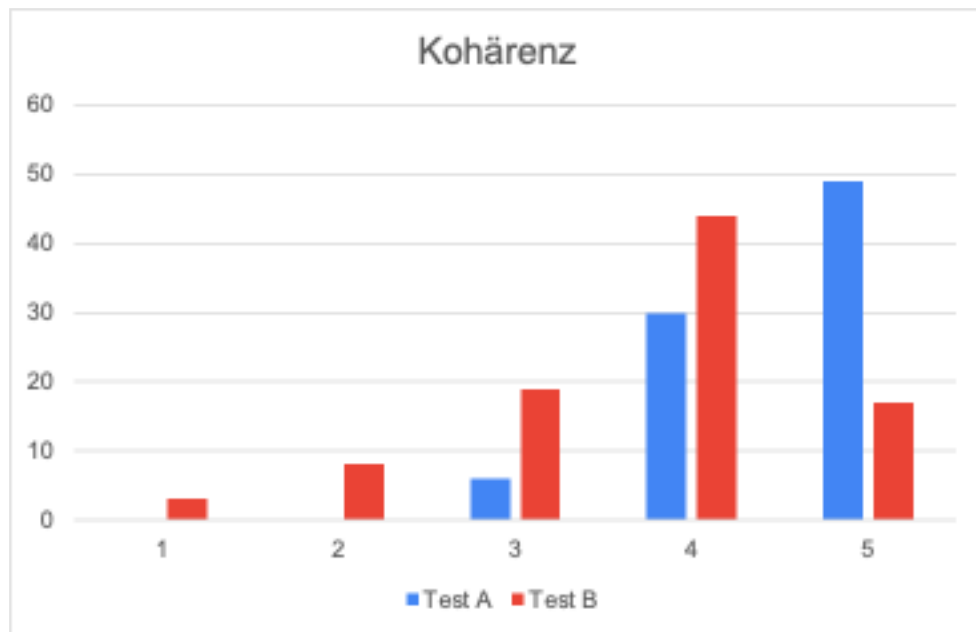


Abb. 11: Verteilung der Bewertungen für das Kriterium Kohärenz

Tab. 11: Häufigkeiten der Bewertungen für das Kriterium Kohärenz

Bewertung	Test A	Test B
1	0	3
2	0	8
3	7	19
4	33	44
5	58	17

Tab. 12: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz

	U	p
Kohärenz	6645	< .001

Tab. 13: Deskriptive Statistiken für das Kriterium Kohärenz

Kohärenz	Gruppe	N	Mittelwert	SD
	Test A	98	4.520	0.630
	Test B	91	3.703	0.983

Wie aus Abbildung 12 hervorgeht, konzentrieren sich die Antworten hinsichtlich des Kriteriums Professionalität im Test A größtenteils auf den höchsten Wert, mit 64 Antworten für den Wert 5. Die Mehrheit der Nutzer*innen bewertete Test B mit 3 und 4 – jeweils 26 und

28 Personen (Abb.12 und Tab. 14). Der Unterschied zwischen den beiden Tests ist statistisch signifikant (Tab. 15). Der Mittelwert von Test A liegt mit 4,52 höher als der Mittelwert von Test B mit 3,52. Die Werte von Test A konzentrieren sich stärker um den Mittelwert mit einer SD von 0,736, während sie in Test B mit einer SD von 1,06 heterogener sind (Tab. 16).

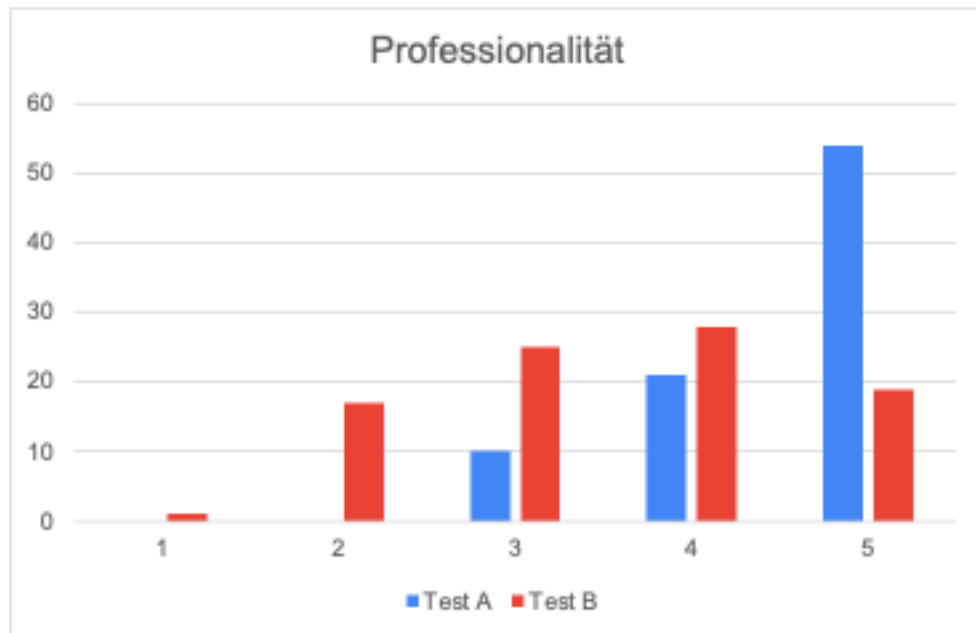


Abb. 12: Verteilung der Bewertungen für das Kriterium Professionalität

Tab. 14: Häufigkeiten der Bewertungen für das Kriterium Professionalität

Bewertung	Test A	Test B
1	0	1
2	1	17
3	11	26
4	22	28
5	64	19

Tab. 15: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität

	U	p
Professionalität	6843	< .001

Tab. 16: Deskriptive Statistiken für das Kriterium Professionalität

Professionalität	Gruppe	N	Mittelwert	SD
	Test A	98	4.520	0.735
	Test B	91	3.516	1.058

Die Tabellen 17 und 18 fassen die oben beschriebenen Ergebnisse zusammen. Tabelle 17 fasst die Ergebnisse der Mann-Whitney-U-Tests für alle untersuchten Kriterien zusammen, und Tabelle 18 enthält die deskriptiven Statistiken (Tab. 17 und 18).

Tab. 17: Zusammenfassung der Mann-Whitney-U-Testergebnisse für Test A und Test B

	U	p
Gesamtbewertung	6829	< .001
Flüssigkeit	6811	< .001
Intonation	7751	< .001
Kohärenz	6645	< .001
Professionalität	6843	< .001

Tab. 18: Zusammenfassung deskriptiver Statistiken für Test A und Test B

	Group	N	Mittelwert	SD
Gesamtbewertung	Test A	98	4.469	0.629
	Test B	91	3.670	0.790
Flüssigkeit	Test A	98	4.265	0.711
	Test B	91	3.330	0.978
Intonation	Test A	98	4.265	0.781
	Test B	91	2.626	1.102
Kohärenz	Test A	98	4.520	0.630
	Test B	91	3.703	0.983
Professionalität	Test A	98	4.520	0.735
	Test B	91	3.516	1.058

6.2 Analyse der Störvariable „Teilnahme an Konferenzen“ für Test A und Test B

Im vorliegenden Abschnitt erfolgt eine Gegenüberstellung der Bewertungen von Nutzer*innen, die an Konferenzen teilgenommen haben, bei denen Dolmetscher*innen eingesetzt wurden, und derjenigen, die noch nie an einer solchen Konferenz teilgenommen haben. Der Vergleich wird separat für Test A und B unter Verwendung des Mann-Whitney-U-Tests durchgeführt. Des Weiteren wird mittels des Kendall-Tau-Tests evaluiert, ob bei den Konferenzteilnehmer*innen eine Korrelation zwischen der Häufigkeit der Teilnahme und der abgegebenen Bewertung besteht.

Für das Kriterium Gesamtbewertung vergaben beide Gruppen hohe Bewertungen, wobei die Bewertungen 4 und 5 am häufigsten vorkommen: bei jeweils 6 und 9 der Teilnehmenden, die bereits an Konferenzen teilgenommen haben, und bei 32 bzw. 44 derjenigen, die noch nie teilgenommen haben (Tab. 19). Mit $p = 0,841$ zeigt sich kein statistisch signifikanter Unterschied zwischen den beiden Untergruppen (Tab. 20). Die Mittelwerte sind beide hoch und ähnlich, und die Streuung der Ergebnisse ist für beide Gruppen gleich, was auf eine vergleichbare Variabilität der Bewertungen innerhalb der beiden Gruppen hindeutet (Tab. 21). Gemäß dem Kendall-Tau-Test besteht auch kein Zusammenhang zwischen der Häufigkeit der Teilnahme an Konferenzen mit Dolmetschdienst und der abgegebenen Bewertung (Tab. 22).

Tab. 19: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme

Bewertung	Test A - Teilnehmende mit Konferenzterfahrung	Test A - Teilnehmende ohne Konferenzterfahrung
1	0	0
2	0	0
3	1	6
4	6	32
5	9	44

Tab. 20: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme

	U	p
Gesamtbewertung	637.0	.841

Tab. 21: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme

Gesamtbewertung	Gruppe	N	Mittelwert	SD
	Ohne Konferenzteilnahme	82	4.463	0.632
	Mit Konferenzteilnahme	16	4.500	0.632

Tab. 22: Ergebnisse des Kendall's Tau-Tests für das Kriterium Gesamtbewertung für Test A nach Konferenzteilnahme

Gesamtbewertung	Kendall's tau B	p
Konferenzteilnahme	-0.153	.494

Auch im Test B zeigte der Vergleich zwischen den beiden Untergruppen für das Kriterium Gesamtbewertung keine statistisch signifikanten Unterschiede, mit $p = 0,35$ (Tab. 24). Der Großteil der Nutzer*innen vergab Bewertungen zwischen 3 und 4 – jeweils 9 und 4 Nutzer*innen mit Konferenzterfahrung und 27 und 33 Nutzer*innen ohne Konferenzterfahrung (Tab. 23). Der Mittelwert der Bewertungen von Teilnehmer*innen ohne Konferenzterfahrung ist etwas höher als der Mittelwert von Teilnehmer*innen mit Konferenzterfahrung. In beiden Fällen konzentrieren sich die Bewertungen moderat um den Mittelwert, wie die Standardabweichungen von 0,800 bzw. 0,743 zeigen (Tab. 25). Es besteht kein signifikanter Zusammenhang zwischen der Häufigkeit der Konferenzteilnahme in der Vergangenheit und der abgegebenen Bewertung (Tab. 26).

Tab. 23: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme

Bewertung	Test B- Teilnehmende mit Konferenzterfahrung	Test B- Teilnehmende ohne Konferenzterfahrung
1	0	0
2	0	4
3	9	27
4	4	33
5	2	12

Tab. 24: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme

	U	p
Gesamtbewertung	652.5	.346

Tab. 25: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme

Gesamtbewertung	Gruppe	N	Mittelwert	SD
	Ohne Konferenzteilnahme	76	3.697	0.800
	Mit Konferenzteilnahme	15	3.533	0.743

Tab. 26: Ergebnisse des Kendall's Tau-Tests für das Kriterium Gesamtbewertung für Test B nach Konferenzteilnahme

Gesamtbewertung	Kendall's tau B	p
Konferenzteilnahme	0.130	.563

Im Test A vergaben die meisten Teilnehmer*innen unter beiden Bedingungen hohe Bewertungen für das Kriterium Flüssigkeit, und mit $p = 0,448$ besteht kein statistisch signifikanter Unterschied (Tab. 27 und 28). Der Mittelwert der von den Teilnehmer*innen ohne Konferenzerfahrung vergebenen Bewertungen ist etwas höher als der der Gruppe mit Erfahrung, während die Bewertungen der letzteren etwas heterogener ausfallen, wie die Standardabweichung von 0,806 gegenüber 0,694 in der Gruppe ohne Erfahrung zeigt (siehe Tab. 29). Aus dem Kendall-Tau-Test geht keine Korrelation zwischen der Häufigkeit der Teilnahme an Konferenzen und den für die Leistung vergebenen Bewertungen hervor (siehe Tab. 30).

Tab. 27: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme

Bewertung	Test A- Teilnehmende mit Konferenzerfahrung	Test A- Teilnehmende ohne Konferenzerfahrung
1	0	0
2	0	0
3	4	11
4	6	36
5	6	35

Tab. 28: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme

	U	p
Flüssigkeit	729.0	.448

Tab. 29: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme

Flüssigkeit	Gruppe	N	Mittelwert	SD
	Ohne Konferenzteilnahme	82	4.293	0.694
	Mit Konferenzteilnahme	16	4.125	0.806

Tab. 30: Ergebnisse des Kendall's Tau-Tests für das Kriterium Flüssigkeit für Test A nach Konferenzteilnahme

Flüssigkeit	Kendall's tau B	p
Konferenzteilnahme	-0.138	.552

Auch in Test B zeigt der Vergleich zwischen den beiden Untergruppen hinsichtlich des Kriteriums Flüssigkeit keinen statistisch signifikanten Unterschied, mit $p = 0,682$ (Tab. 32). Die Bewertungen konzentrieren sich in beiden Untergruppen überwiegend auf die Werte 3 und 4, was insgesamt auf eine mittelmäßig positive Wahrnehmung hinweist (Tab. 31). Die Mittelwerte der beiden Untergruppen sind sehr ähnlich (3,34 ohne Erfahrung; 3,20 mit Erfahrung). Die Streuung ist in beiden Fällen moderat ($SD = 0,989$ und $SD = 0,941$), was auf insgesamt vergleichbare Bewertungen hindeutet (Tab. 33). Nach dem Korrelationstest von Kendall's Tau ergibt sich kein statistisch signifikanter Zusammenhang zwischen der Häufigkeit der Teilnahme an Konferenzen und der Bewertung der Flüssigkeit (Tab. 34).

Tab. 31: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme

Bewertung	Test B- Teilnehmende mit Konferenzserfahrung	Test B- Teilnehmende ohne Konferenzserfahrung
1	1	1
2	1	14
3	8	29
4	4	21
5	1	11

Tab. 32: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme

	U	p
Flüssigkeit	607.0	.682

Tab. 33: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme

	Gruppe	N	Mittelwert	SD
Flüssigkeit	Ohne Konferenzteilnahme	76	3.335	0.989
	Mit Konferenzteilnahme	15	3.200	0.941

Tab. 34: Ergebnisse des Kendall's Tau-Tests für das Kriterium Flüssigkeit für Test B nach Konferenzteilnahme

Flüssigkeit	Kendall's tau B	p
Konferenzteilnahme	0.218	.319

Im Test A vergaben die meisten Teilnehmer*innen beider Gruppen für das Kriterium Intonation hohe Bewertungen. Nutzer*innen mit Konferenzerfahrung neigen dazu, die Bewertungen gleichmäßiger zwischen 3, 4 und 5 zu verteilen, wobei mehr Nutzer*innen hohe Werte vergeben (4 und 5; 6 und 6). Nutzer*innen ohne Konferenzerfahrung zeigen eine ähnliche Verteilung, jedoch mit mehr Teilnehmer*innen bei den Werten 4 und 5, nämlich 29 bzw. 39 (Tab. 35). Der Unterschied zwischen den beiden Untergruppen ist mit $p = 0,408$ statistisch nicht signifikant (Tab. 36). Die deskriptiven Statistiken zeigen, dass Teilnehmer*innen ohne Konferenzerfahrung in Test A der Intonation eine etwas höhere Durchschnittsbewertung geben (Mittelwert = 4,29; SD = 0,78) als diejenigen mit Konferenzerfahrung (Mittelwert = 4,13; SD = 0,81). Darüber hinaus weist die etwas höhere Standardabweichung in der Gruppe mit Erfahrung auf eine größere Variabilität der Bewertungen im Vergleich zur Gruppe ohne Erfahrung hin (Tab. 37). Nach dem Korrelationstest von Kendall's Tau ergibt sich kein statistisch signifikanter Zusammenhang zwischen der Häufigkeit der Teilnahme an Konferenzen und der Bewertung der Intonation (Tab. 38).

Tab. 35: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test A nach Konferenzteilnahme

Bewertung	Test A- Teilnehmende mit Konferenzerfahrung	Test A- Teilnehmende ohne Konferenzerfahrung
1	0	0
2	0	1
3	4	13
4	6	29
5	6	39

Tab. 36: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test A nach Konferenzteilnahme

	U	p
Intonation	736.0	.408

Tab. 37: Deskriptive Statistiken für das Kriterium Intonation für Test A nach Konferenzteilnahme

Intonation	Gruppe	N	Mittelwert	SD
	Ohne Konferenzteilnahme	82	4.293	0.778
	Mit Konferenzteilnahme	16	4.125	0.806

Tab. 38: Ergebnisse des Kendall's Tau-Tests für das Kriterium Intonation für Test A nach Konferenzteilnahme

Intonation	Kendall's tau B	p
Konferenzteilnahme	0.085	.694

Auch in Test B zeigt sich kein statistisch signifikanter Unterschied zwischen den beiden Untergruppen hinsichtlich des Kriteriums Intonation, mit $p = 0,912$ (Tab. 40). Die Bewertungen der beiden Untergruppen sind ähnlich, und beide Gruppen neigen dazu, niedrige bis moderate Bewertungen zu vergeben, die sich auf die Werte 2 und 3 konzentrieren, jeweils 25 und 22 Nutzer*innen ohne Konferenz Erfahrung und 4 Nutzer*innen mit Erfahrung (Tab. 39). Die Mittelwerte sind nahezu identisch (2,62 für die Gruppe ohne und 2,68 für die Gruppe mit Erfahrung) und die Standardabweichung zeigt, dass die Bewertungen in der Gruppe, die an den Konferenzen teilgenommen hat, mit einer SD von 1,234 gegenüber 1,083 etwas variabler sind (Tab. 41). Gemäß dem Korrelationstest von Kendall's Tau ergibt sich kein statistisch signifikanter Zusammenhang zwischen der Häufigkeit der Teilnahme an Konferenzen und der Bewertung der Intonation (Tab. 42).

Tab. 39: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test B nach Konferenzteilnahme

Bewertung	Test B- Teilnehmende mit Konferenz Erfahrung	Test B- Teilnehmende ohne Konferenz Erfahrung
1	3	12
2	4	25
3	4	22
4	3	14
5	1	3

Tab. 40: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test B nach Konferenzteilnahme

	U	p
Intonation	559.5	.912

Tab. 41: Deskriptive Statistiken für das Kriterium Intonation für Test B nach Konferenzteilnahme

	Gruppe	N	Mittelwert	SD
Intonation	Ohne Konferenzteilnahme	76	2.618	1.083
	Mit Konferenzteilnahme	15	2.667	1.234

Tab. 42: Ergebnisse des Kendall's Tau-Tests für das Kriterium Intonation für Test B nach Konferenzteilnahme

Intonation	Kendall's tau B	p
Konferenzteilnahme	0.142	.501

In Test A bewerteten beide Gruppen das Kriterium Kohärenz insgesamt positiv, jedoch neigten die Teilnehmer*innen ohne Konferenzterfahrung dazu, etwas höhere Bewertungen abzugeben. Die Mehrheit vergab eine Bewertung von 5 (52 Personen), während in der Untergruppe ohne Konferenzteilnahme die Mehrheit (9) eine Bewertung von 4 vergab (Tab. 43). Der Unterschied zwischen den beiden Untergruppen ist mit $p = 0,087$ nicht signifikant (Tab. 44). Die Mittelwerte zeigen einen geringen Unterschied: 4,56 für Nutzer*innen ohne Konferenzterfahrung gegenüber 4,31 für Nutzer*innen mit Konferenzterfahrung, und die Standardabweichung ist ähnlich, was darauf hindeutet, dass die Streuung der Bewertungen in beiden Gruppen vergleichbar ist (Tab. 45). Der Zusammenhang zwischen der Häufigkeit der Teilnahme an Konferenzen und der Bewertung dieses Kriteriums ist nicht signifikant (Tab. 46).

Tab. 43: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test A nach Konferenzteilnahme

Bewertung	Test A- Teilnehmende mit Konferenzterfahrung	Test A- Teilnehmende ohne Konferenzterfahrung
1	0	0
2	0	0
3	1	6
4	9	24
5	6	52

Tab. 44: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test A nach Konferenzteilnahme

	U	p
Kohärenz	811.0	.087

Tab. 45: Deskriptive Statistiken für das Kriterium Kohärenz für Test A nach Konferenzteilnahme

	Gruppe	N	Mittelwert	SD
Kohärenz	Ohne Konferenzteilnahme	82	4.561	0.630
	Mit Konferenzteilnahme	16	4.313	0.602

Tab. 46: Ergebnisse des Kendall's Tau-Tests für das Kriterium Kohärenz für Test A nach Konferenzteilnahme

Kohärenz	Kendall's tau B	p
Konferenzteilnahme	-0.324	.146

Im Test B konzentrieren sich die Bewertungen für das Kriterium Kohärenz für beide Untergruppen auf die mittleren Werte 3 und 4 (Tab. 47). Der Mann-Whitney-U-Test ergab keine Signifikanz zwischen den beiden Gruppen, mit $p = 0,779$ (Tab. 48). Der Mittelwert der Bewertungen der Teilnehmer*innen ohne Konferenzerfahrung liegt bei 3,71 und entspricht damit etwa dem Mittelwert der Teilnehmerinnen mit Konferenzerfahrung, der bei 3,67 liegt (Tab. 49). Die Korrelation zwischen der Häufigkeit der Teilnahme an Konferenzen und den für dieses Kriterium vergebenen Bewertungen ist nicht signifikant (Tab. 50).

Tab. 47: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test B nach Konferenzteilnahme

Bewertung	Test B- Teilnehmende mit Konferenzerfahrung	Test B- Teilnehmende ohne Konferenzerfahrung
1	0	3
2	2	6
3	2	17
4	10	34
5	1	16

Tab. 48: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test B nach Konferenzteilnahme

	U	p
Kohärenz	595.0	.779

Tab 49: Deskriptive Statistiken für das Kriterium Kohärenz für Test B nach Konferenzteilnahme

	Gruppe	N	Mittelwert	SD
Kohärenz	Ohne Konferenzteilnahme	76	3.711	1.017
	Mit Konferenzteilnahme	15	3.667	0.816

Tab. 50: Ergebnisse des Kendall's Tau-Tests für das Kriterium Kohärenz für Test B nach Konferenzteilnahme

Kohärenz	Kendall's tau B	p
Konferenzteilnahme	0.067	.763

In Bezug auf das Kriterium Professionalität gaben beide Gruppen ähnliche Bewertungen ab, die sich auf die höchsten Werten konzentrieren. Unter denjenigen, die noch nie an Konferenzen teilgenommen haben, gaben 56 von 82 Personen die Bewertung 5 ab, und unter denjenigen, die bereits teilgenommen haben, gaben 8 von 16 die Bewertung 5 ab (Tab. 51). Der Mann-Whitney-Test ist mit $p = 0,167$ nicht signifikant (Tab. 52). Die Mittelwerte der beiden Untergruppen sind sehr ähnlich, mit einer leichten Tendenz zu höheren Bewertungen für die Nutzer*innen ohne Konferenzerfahrung (4,56), während der Durchschnittswert für die Nutzer*innen mit Konferenzerfahrung bei 4,31 liegt (Tab. 53). Nach dem Kendall's Tau-Test besteht keine signifikante Korrelation zwischen der Anzahl der Konferenzen, an denen die Nutzer*innen teilgenommen haben, und den Bewertungen, die sie für das Kriterium Professionalität abgegeben haben (Tab. 54).

Tab. 51: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test A nach Konferenzteilnahme

Bewertung	Test A- Teilnehmende mit Konferenzzerfahrung	Test A- Teilnehmende ohne Konferenzzerfahrung
1	0	0
2	0	1
3	3	8
4	5	17
5	8	56

Tab. 52: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test A nach Konferenzteilnahme

	U	p
Professionalität	777.5	.167

Tab 53: Deskriptive Statistiken für das Kriterium Professionalität für Test A nach Konferenzteilnahme

	Gruppe	N	Mittelwert	SD
Kohärenz	Ohne Konferenzteilnahme	82	4.561	0.722
	Mit Konferenzteilnahme	16	4.313	0.793

Tab. 54: Ergebnisse des Kendall's Tau-Tests für das Kriterium Professionalität für Test A nach Konferenzteilnahme

Professionalität	Kendall's tau B	p
Konferenzteilnahme	-0.110	.614

In Test B neigt die Gruppe mit Konferenzzerfahrung dazu, etwas höhere Bewertungen abzugeben: 7 von 15 Personen vergaben den Wert 4, während in der Gruppe ohne Erfahrung 23 von 76 den Wert 3 vergaben (Tab. 55). Der Unterschied zwischen den beiden Gruppen ist jedoch mit $p = 0,934$ statistisch nicht signifikant (Tab. 56). Die Mittelwerte der beiden Gruppen sind nahezu identisch, mit 3,51 für Teilnehmer*innen ohne Erfahrung und 3,53 für Teilnehmer*innen mit Erfahrung (Tab. 57). Nach dem Kendall's Tau-Test besteht kein Zusammenhang zwischen der Anzahl der Konferenzteilnahmen und der Bewertung des Kriteriums Professionalität (Tab. 58).

Tab. 55: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test B nach Konferenzteilnahme

Bewertung	Test B- Teilnehmende mit Konferenzzerfahrung	Test B- Teilnehmende ohne Konferenzzerfahrung
1	0	1
2	3	14
3	3	23
4	7	21
5	2	17

Tab. 56: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test B nach Konferenzteilnahme

	U	p
Professionalität	562.0	.934

Tab. 57: Deskriptive Statistiken für das Kriterium Professionalität für Test B nach Konferenzteilnahme

	Gruppe	N	Mittelwert	SD
Professionalität	Ohne Konferenzteilnahme	76	3.513	1.077
	Mit Konferenzteilnahme	15	3.533	0.990

Tab. 58: Ergebnisse des Kendall's Tau-Tests für das Kriterium Professionalität für Test B nach Konferenzteilnahme

Professionalität	Kendall's tau B	p
Konferenzteilnahme	-0.070	.747

6.3 Analyse der Störvariable „Deutschkenntnisse“ für Test A und Test B

Im folgenden Abschnitt werden die Bewertungen der Nutzer*innen mit Deutschkenntnissen und die Bewertungen derjenigen ohne Deutschkenntnisse für Test A und B verglichen. Obwohl die Teilnehmer*innen im Fragebogen gebeten wurden, ihr Deutschniveau (A1–C2) anzugeben, wurden die Daten für diese Analyse in zwei Gruppen eingeteilt: „JA“ für diejenigen mit Deutschkenntnissen und „NEIN“ für diejenigen ohne Deutschkenntnisse. Um die Bewertungen dieser beiden Hauptgruppen zu vergleichen, wurde der Mann-Whitney-U-Test durchgeführt, wobei die verschiedenen Sprachkenntnisstufen nicht berücksichtigt wurden. Die Entscheidung, keine Korrelationsanalyse auf der Grundlage der unterschiedlichen Sprachkenntnisse durchzuführen, beruht darauf, dass die Audioaufnahme des Ausgangstextes in weiten Teilen

von der Verdolmetschung übertönt wird, was ein optimales Zuhören verhindert. Unter diesem Umstand wird die Hypothese aufgestellt, dass es keine signifikanten Unterschiede in der Bewertung aufgrund der unterschiedlichen Deutschkenntnisse geben kann. Es wird nur allgemein analysiert, ob das Verständnis der deutschen Sprache die Bewertung beeinflusst hat.

Für das Kriterium Gesamtbewertung vergab die Mehrheit der Nutzer*innen in Test A hohe Bewertungen (4-5), während keine der beiden Gruppen niedrige Bewertungen (1-2) vergab. Die Nutzer*innen mit Deutschkenntnissen zeigen eine leichte Häufung bei 5 (12 von 19), während in der Gruppe ohne Kenntnisse die 4er-Bewertungen überwiegen (Tab. 59). Die Gruppe mit Deutschkenntnissen weist einen etwas höheren Durchschnitt auf, der Unterschied zur Gruppe ohne Deutschkenntnisse ist jedoch nicht signifikant (Tab. 60 und 61).

Tab. 59: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test A nach Deutschkenntnissen

Bewertung	Test A-Teilnehmende mit Deutschkenntnissen	Test A-Teilnehmende ohne Deutschkenntnissen
1	0	0
2	0	0
3	0	7
4	7	31
5	12	41

Tab. 60: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test A nach Deutschkenntnissen

	U	p
Gesamtbewertung	859.5	.271

Tab. 61: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test A nach Deutschkenntnissen

	Gruppe	N	Mittelwert	SD
Gesamtbewertung	Ja	19	4.632	0.496
	Nein	79	4.430	0.654

Die Analyse der Ergebnisse von Test B für die Gesamtbewertung zeigt, dass die meisten Teilnehmer*innen mittlere bis hohe Bewertungen vergeben (Tab. 62). Die Gruppe ohne Deutschkenntnisse hat einen etwas höheren Mittelwert (3,72 gegenüber 3,40), aber der Mann-Whitney-U-Test zeigt, dass der Unterschied statistisch nicht signifikant ist (Tab. 63 und 64).

Tab. 62: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Test B nach Deutschkenntnissen

Bewertung	Test B-Teilnehmende mit Deutschkenntnissen	Test B-Teilnehmende ohne Deutschkenntnissen
1	0	0
2	0	4
3	10	26
4	4	33
5	1	13

Tab. 63: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Gesamtbewertung für Test B nach Deutschkenntnissen

	U	p
Gesamtbewertung	425.5	.098

Tab. 64: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Test B nach Deutschkenntnissen

Gesamtbewertung	Gruppe	N	Mittelwert	SD
	Ja	15	3.400	0.632
	Nein	76	3.724	0.810

Die Analyse des Kriteriums Flüssigkeit zeigt, dass die Mehrheit der Teilnehmer*innen beider Untergruppen im Test A hohe Bewertungen von 4 bis 5 vergibt (Tab. 65). Die Gruppe mit Deutschkenntnissen weist einen leicht höheren Durchschnitt auf (4,37 gegenüber 4,24), aber der Mann-Whitney-U-Test zeigt, dass der Unterschied statistisch nicht signifikant ist (Tab. 66 und 67).

Tab. 65: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test A nach Deutschkenntnissen

Bewertung	Test A-Teilnehmende mit Deutschkenntnissen	Test A-Teilnehmende ohne Deutschkenntnissen
1	0	0
2	0	0
3	4	11
4	4	38
5	11	30

Tab. 66: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test A nach Deutschkenntnissen

	U	p
Flüssigkeit	846.0	.353

Tab. 67: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test A nach Deutschkenntnissen

Flüssigkeit	Gruppe	N	Mittelwert	SD
	Ja	19	4.368	0.831
	Nein	79	4.241	0.683

Für das Kriterium Flüssigkeit in Test B vergaben die Teilnehmer*innen beider Gruppen überwiegend mittlere Bewertungen, wobei in beiden Gruppen den Wert 3 überwiegt (Tab. 68). Die Mittelwerte sind nahezu identisch, und der Mann-Whitney-U-Test zeigt keine signifikanten Unterschiede zwischen den Bewertungen der beiden Untergruppen (Tab. 69 und 70).

Tab. 68: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Test B nach Deutschkenntnissen

Bewertung	Test B-Teilnehmende mit Deutschkenntnissen	Test B-Teilnehmende ohne Deutschkenntnissen
1	0	2
2	3	12
3	7	30
4	2	23
5	3	9

Tab. 69: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Flüssigkeit für Test B nach Deutschkenntnissen

	U	p
Flüssigkeit	552.5	.848

Tab. 70: Deskriptive Statistiken für das Kriterium Flüssigkeit für Test B nach Deutschkenntnissen

Flüssigkeit	Gruppe	N	Mittelwert	SD
	Ja	15	3.333	1.047
	Nein	76	3.329	0.971

Für das Kriterium Intonation vergaben beide Untergruppen im Test A hohe Bewertungen. In der Gruppe mit Deutschkenntnissen konzentrieren sich die Bewertungen auf die höchsten Werte, insbesondere 5 (12 Fälle) und 4 (5 Fälle). In der Gruppe ohne Deutschkenntnisse sind die Bewertungen breiter verteilt, wobei die Werte 5 (33) und 4 (30) überwiegen, jedoch mit einer größeren Anzahl von mittleren Bewertungen, da 15 Nutzer*innen die Bewertung 3 vergaben (Tab. 71). Die Mittelwerte zeigen eine etwas höhere Bewertung in der Gruppe mit Deutschkenntnissen mit einem Durchschnitt von 4,53 im Vergleich zur Gruppe ohne Kenntnisse, die einen Durchschnitt von 4,20 aufweist (Tab. 73). Der Mann-Whitney-U-Test zeigt jedoch keinen statistisch signifikanten Unterschied (Tab. 72).

Tab. 71: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test A nach Deutschkenntnissen

Bewertung	Test A-Teilnehmende mit Deutschkenntnissen	Test A-Teilnehmende ohne Deutschkenntnissen
1	0	0
2	0	1
3	2	15
4	5	30
5	12	33

Tab. 72: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test A nach Deutschkenntnissen

	U	p
Intonation	922.0	.096

Tab. 73: Deskriptive Statistiken für das Kriterium Intonation für Test A nach Deutschkenntnissen

Intonation	Gruppe	N	Mittelwert	SD
	Ja	19	4.526	0.697
	Nein	79	4.203	0.791

Im Test B fielen die Bewertungen der Intonation in beiden Gruppen insgesamt mittel bis niedrig aus. Die Teilnehmer*innen mit Deutschkenntnissen vergaben vor allem die Bewertung 2 (8 Fälle), während diejenigen ohne Deutschkenntnisse ihre Bewertungen überwiegend zwischen 2 und 3 verteilten, nämlich 21 bzw. 24 Personen (Tab. 74). Der Mittelwert ist in der Gruppe

mit Deutschkenntnissen (2,40) etwas niedriger als in der Gruppe ohne Deutschkenntnisse (2,67), aber der Unterschied ist statistisch nicht signifikant (Tab. 75 und 76).

Tab. 74: Häufigkeiten der Bewertungen für das Kriterium Intonation für Test B nach Deutschkenntnissen

Bewertung	Test B-Teilnehmende mit Deutschkenntnissen	Test B-Teilnehmende ohne Deutschkenntnissen
1	2	13
2	8	21
3	2	24
4	3	14
5	0	4

Tab. 75: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Intonation für Test B nach Deutschkenntnissen

	U	p
Intonation	488.0	.367

Tab. 76: Deskriptive Statistiken für das Kriterium Intonation für Test B nach Deutschkenntnissen

Intonation	Gruppe	N	Mittelwert	SD
	Ja	15	2.400	0.986
	Nein	76	2.671	1.124

In Test A wird die Kohärenz in beiden Gruppen sehr positiv bewertet. Die Teilnehmer*innen mit Deutschkenntnissen vergaben fast ausschließlich die Werte 5 (13) und 4 (6), während diejenigen ohne Deutschkenntnisse ihre Bewertungen zwischen 5 (45), 4 (27) und 3 (7) verteilten (Tab. 77). Der Mittelwert ist in der Gruppe mit Deutschkenntnissen (4,68) etwas höher als in der Gruppe ohne Deutschkenntnisse (4,48), jedoch ist der Unterschied statistisch nicht signifikant (Tab. 78 und 79).

Tab. 77: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test A nach Deutschkenntnissen

Bewertung	Test A-Teilnehmende mit Deutschkenntnissen	Test A-Teilnehmende ohne Deutschkenntnissen
1	0	0
2	0	0
3	0	7
4	6	27
5	13	45

Tab. 78: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test A nach Deutschkenntnissen

	U	p
Kohärenz	857.5	.270

Tab. 79: Deskriptive Statistiken für das Kriterium Kohärenz für Test A nach Deutschkenntnissen

Kohärenz	Gruppe	N	Mittelwert	SD
	Ja	19	4.684	0.478
	Nein	79	4.481	0.658

In Test B wurde die Kohärenz in beiden Gruppen auf einem mittleren bis hohen Niveau bewertet. Die Teilnehmer*innen mit Deutschkenntnissen vergaben vor allem die Bewertung 4 (7), gefolgt von 3 und 5 (jeweils 3), während diejenigen ohne Deutschkenntnisse ihre Bewertungen auf 4 (37) und 3 (16) konzentrierten, mit einigen hohen und wenigen niedrigen Bewertungen (Tab. 80). Die Mittelwerte sind nahezu identisch, mit 3,73 für diejenigen mit Deutschkenntnissen, und 3,70 für diejenigen ohne Deutschkenntnisse, und der Mann-Whitney-U-Test zeigt keine signifikanten Unterschiede zwischen den Gruppen (Tab. 81 und 82).

Tab. 80: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Test B nach Deutschkenntnissen

Bewertung	Test B-Teilnehmende mit Deutschkenntnissen	Test B-Teilnehmende ohne Deutschkenntnissen
1	0	3
2	2	6
3	3	16
4	7	37
5	3	14

Tab. 81: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Kohärenz für Test B nach Deutschkenntnissen

	U	p
Kohärenz	574.5	.963

Tab. 82: Deskriptive Statistiken für das Kriterium Kohärenz für Test B nach Deutschkenntnissen

Kohärenz	Gruppe	N	Mittelwert	SD
	Ja	15	3.733	0.961
	Nein	76	3.697	0.994

In Test A wird die Professionalität in beiden Gruppen sehr positiv bewertet. Die Teilnehmer*innen mit Deutschkenntnissen vergaben vor allem die Bewertung 5 (14), gefolgt von 3 und 4 (jeweils 3), während diejenigen ohne Deutschkenntnisse ihre Bewertungen auf 5 (51) und 4 (19) konzentrierten, mit wenigen niedrigeren Bewertungen (Tab. 83). Die Mittelwerte sind mit 4,53 für diejenigen mit Deutschkenntnissen und 4,52 für diejenigen ohne Deutschkenntnisse nahezu identisch, und der Mann-Whitney-U-Test zeigt keine signifikanten Unterschiede zwischen den Gruppen (Tab. 84 und 85).

Tab. 83: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test A nach Deutschkenntnissen

Bewertung	Test A-Teilnehmende mit Deutschkenntnissen	Test A-Teilnehmende ohne Deutschkenntnissen
1	0	0
2	0	1
3	3	7
4	3	19
5	14	51

Tab. 84: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test A nach Deutschkenntnissen

	U	p
Professionalität	766.0	.873

Tab. 85: Deskriptive Statistiken für das Kriterium Professionalität für Test A nach Deutschkenntnissen

Professionalität	Gruppe	N	Mittelwert	SD
	Ja	19	4.526	0.772
	Nein	79	4.519	0.731

Im Test B vergaben die Teilnehmer*innen mit Deutschkenntnissen vor allem die Bewertung 2 (6) und in geringerem Maße 4 (4) und 3 (3), während diejenigen ohne Deutschkenntnisse ihre Bewertungen hauptsächlich zwischen 4 (24) und 3 (23) verteilten, wobei auch hohe Bewertungen vorkommen (Tab. 86). Der Mittelwert liegt in der Gruppe mit Deutschkenntnissen (3,13) etwas unter dem der Gruppe ohne Kenntnisse (3,59), jedoch ist der Unterschied statistisch nicht signifikant (Tab. 87 und 88).

Tab. 86: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Test B nach Deutschkenntnissen

Bewertung	Test B-Teilnehmende mit Deutschkenntnissen	Test B-Teilnehmende ohne Deutschkenntnissen
1	0	1
2	6	11
3	3	23
4	4	24
5	2	17

Tab. 87: Ergebnisse des Mann-Whitney-U-Tests für das Kriterium Professionalität für Test B nach Deutschkenntnissen

	U	p
Professionalität	432.5	.129

Tab. 88: Deskriptive Statistiken für das Kriterium Professionalität für Test B nach Deutschkenntnissen

Professionalität	Gruppe	N	Mittelwert	SD
	Ja	19	15	3.133
	Nein	79	76	3.592

6.4 Ergebnisse des Intra- und Interstudien-Vergleichs mit der Studie von Collados Aís (1998)

Der folgende Abschnitt widmet sich dem deskriptiven Vergleich der Ergebnisse der Studie von Collados Aís (1998) mit denen der vorliegenden Studie. Die Analyse gliedert sich in zwei Ebenen: Einerseits wird innerhalb jeder Studie der Unterschied zwischen der lebhaften und der monotonen Version durch die Berechnung der Differenz zwischen den jeweiligen Mittelwerten verglichen; andererseits wird ein Vergleich zwischen den monotonen Versionen der beiden Studien durchgeführt.

Für den Intra-Studien-Vergleich wurden in der Studie von Collados Aís (1998) die monotone Version 1 und die lebhafte Version 3 der Kontrollgruppe der Studie 3A über die

Nutzer*innen herangezogen, während in der vorliegenden Studie die monotone Version (Test B) und die Kontrollversion (Test A) miteinander verglichen wurden. Die Mittelwerte der Studie von Collados Aís (1998) wurden aus Kapitel 8.3 entnommen (vgl. Collados Aís 1998: 223). Die Berechnung der Mittelwertdifferenz zur Überprüfung des Umfangs der Differenz zwischen den Mittelwerten der beiden Versionen in allen beiden Studien wurde manuell durchgeführt.

Für den Inter-Studien-Vergleich wurden die beiden monotonen Versionen beider Studien miteinander verglichen, d. h. Video 1 von Collados Aís (1998) und Test B der vorliegenden Studie. Da Collados Aís (1998) ausschließlich absolute Häufigkeitstabellen angibt, wurden diese Daten verwendet, um die Verteilung der Bewertungen zu rekonstruieren, um mit Hilfe des Programms JASP deskriptive Statistiken berechnen zu können. In den Tabellen werden für jedes Kriterium die absoluten und relativen Häufigkeiten beider Studien angegeben. Die relativen Häufigkeiten wurden auf der Grundlage der jeweiligen absoluten Werte berechnet. Darüber hinaus werden für jedes Kriterium die wichtigsten deskriptiven Indikatoren der Stichprobe angegeben, darunter die Lagemaße (Mittelwert und Median) und die Streuungsmaße (Standardabweichung). Der Vergleich zwischen den Studien erfolgt ausschließlich in deskriptiver Form, ohne dass auf inferenzstatistische Verfahren zurückgegriffen wird.

Was das Kriterium der Gesamtbewertung betrifft, so geht aus Tabelle 89 hervor, dass in beiden Studien die lebhafte Version höhere Mittelwerte aufweist als die monotone Version. Die Mittelwertdifferenz zwischen den beiden Versionen beträgt in der vorliegenden Studie 0,80 und in der Studie von Collados Aís (1998) 1,00 (Tab. 89).

Tab. 89: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Gesamtbewertung

Kriterium	Studie	Lebhafte Version (Test A /Video 3, Mittelwert)	Monotone Version (Test B /Video 1, Mittelwert)	Mittelwertdifferenz
Gesamtbewertung	Vorliegende Studie	4.47	3.67	0,80
Gesamtbewertung	Collados Aís	4.35	3.35	1,00

Was den Vergleich zwischen der vorliegenden Studie und der Studie von Collados Aís (1998) bei den monotonen Versionen betrifft, so zeigt sich in beiden Stichproben eine Konzentration der Antworten auf die mittleren bis hohen Werte, insbesondere in den Kategorien 3 und 4. In der Stichprobe von Collados Aís (1998) vergab die Mehrheit der Teilnehmer*innen den Wert 3 (42,9 %), gefolgt von Wert 4 (28,6 %). In Test B ist die Verteilung ähnlich, jedoch leicht in Richtung höherer Bewertungen verschoben: Der Wert 4 kommt am häufigsten vor (40,7 %), gefolgt von Wert 3 (39,6 %) (Tab. 90). Der Mittelwert der Gesamtbewertung in Test B von 3,67 ist, wie bereits erwähnt, etwas höher als der von Collados Aís (1998) für Video 1 angegebene Mittelwert von 3,36 (Tab. 91). Es sei darauf hingewiesen, dass die in dieser Studie und in der Originalstudie von Collados Aís (1998) angegebenen Mittelwerte für dieses Kriterium leicht unterschiedlich gerundet sind: hier wurde der Mittelwert 3,36 angegeben, während Collados Aís für Video 1 den Wert 3,35 beibehalten hat (Tab. 89).

Tab. 90: Häufigkeiten der Bewertungen für das Kriterium Gesamtbewertung für Video 1 (Collados Aís) und Test B

Bewertung	Collados Aís Video 1 (N=14)	Collados Aís Video 1 %	Test B (N=91)	Test B_%
1	1	7,1	0	0,0
2	1	7,1	4	4,4
3	6	42,9	36	39,6
4	4	28,6	37	40,7
5	2	14,3	14	15,4

Tab. 91: Deskriptive Statistiken für das Kriterium Gesamtbewertung für Video 1 (Collados Aís) und Test B

Gesamtbewertung	Gruppe	Median	Mittelwert	SD
	Collados Aís Video 1	3.000	3.357	1.082
	Test B	4.000	3.670	0.790

Aus Tabelle 92 geht hervor, dass in beiden Untersuchungen die lebhaftere Version höhere Mittelwerte aufweist als die monotone Version für das Kriterium Flüssigkeit. Die Mittelwertdifferenz zwischen den beiden Versionen ist in der vorliegenden Studie mit 0,94 größer als in der Studie von Collados Aís (1998) mit einem Unterschied von 0,50 (Tab. 92).

Tab. 92: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Flüssigkeit

Kriterium	Studie	Lebhaftes Version (Test A /Video 3, Mittelwert)	Monotone Version (Test B /Video 1, Mittelwert)	Mittelwertdifferenz
Flüssigkeit	Vorliegende Studie	4.27	3.33	0,94
Flüssigkeit	Collados Aís	4.57	4.07	0,50

Aus den Daten geht ein recht deutlicher Unterschied in der Bewertung des Kriteriums Flüssigkeit zwischen Video 1 von Collados Aís (1998) und Test B hervor. Im ersten Fall konzentrieren sich die Bewertungen auf die höheren Werte: Die Hälfte der Bewertungen entspricht dem Wert 4 und fast ein Drittel dem Wert 5, während niedrige Werte nicht vorkommen. In Test B ist die Verteilung hingegen breiter und verschiebt sich in Richtung der Mittelwerte. Der häufigste Wert ist 3 (40,7 %), gefolgt von 4 (27,5 %), und es gibt auch niedrige Bewertungen, die in der ersten Gruppe nicht vorkommen (Tab. 93). Die deskriptiven Statistiken bestätigen diesen Trend: Der Mittelwert ist für Video 1 höher (4,07 gegenüber 3,33), was auf eine insgesamt bessere Bewertung hindeutet, während die höhere Standardabweichung in Test B auf eine größere Variabilität in den Bewertungen hinweist (Tab. 94).

Tab. 93: Häufigkeiten der Bewertungen für das Kriterium Flüssigkeit für Video 1 (Collados Aís) und Test B

Bewertung	Collados Aís Video 1 (N=14)	Collados Aís Video 1 _ %	Test B (N=91)	Test B _ %
1	0	0,00	2	2,2
2	0	0,00	15	16,5
3	3	21,4	37	40,7
4	7	50,0	25	27,5
5	4	28,6	12	13,2

Tab. 94: Deskriptive Statistiken für das Kriterium Flüssigkeit für Video 1 (Collados Aís) und Test B

Flüssigkeit	Gruppe	Median	Mittelwert	SD
	Collados Aís Video 1	4.000	4.071	0.730
	Test B	3.000	3.330	0.978

In der vorliegenden Studie beträgt der Mittelwert für das Kriterium Intonation 4,27 in der lebhaften Version und 2,63 in der monotonen Version, mit einer Mittelwertdifferenz von 1,64. In der Studie von Collados Aís (1998) liegt der Mittelwert der lebhaften Version bei 3,28 und der der monotonen Version bei 2,07, mit einer Mittelwertdifferenz von 1,21, die geringer ist als in Test B (Tab. 95).

Tab. 95: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Intonation

Kriterium	Studie	Lebhaftes Version (Test A /Video 3, Mittelwert)	Monotone Version (Test B /Video 1, Mittelwert)	Mittelwertdifferenz
Intonation	Vorliegende Studie	4.27	2.63	1,64
Intonation	Collados Aís	3.28	2.07	1,21

Es zeigt sich eine deutliche Diskrepanz zwischen den Einzelbewertungen der beiden Studien: Bei Video 1 konzentrieren sich die Bewertungen überwiegend auf die niedrigeren Antwortmöglichkeiten: 35,7 % der Teilnehmer*innen vergaben den Wert 1 und weitere 35,7 % den Wert 2. Die höchsten Werte (4) machen nur 14,3 % aus, und der Höchstwert 5 wurde nie vergeben. In Test B ist die Verteilung ausgewogener und leicht zu den Mittelwerten verschoben. Der häufigste Wert ist 2 (31,9 %), aber ein erheblicher Anteil der Bewertungen liegt auch auf den Werten 3 (28,6 %) und 4 (18,7 %), während der Wert 5 zwar selten, aber dennoch vorhanden ist (Tab. 96). Der Mittelwert von Video 1 (2,07) liegt unter dem Mittelwert von Test B (2,63), bei einer ähnlichen Streuung in beiden Stichproben (Tab. 97).

Tab. 96: Häufigkeiten der Bewertungen für das Kriterium Intonation für Video 1 (Collados Aís) und Test B

Bewertung	Collados Aís. Video 1 (N=14)	Collados Aís Video 1 %	Test B (N=91)	Test B %
1	5	35,7	15	16,5
2	5	35,7	29	31,9
3	2	14,3	26	28,6
4	2	14,3	17	18,7
5	0	0,00	4	4,4

Tab. 97: Deskriptive Statistiken für das Kriterium Intonation für Video 1 (Collados Aís) und Test B

Intonation	Gruppe	Median	Mittelwert	SD
	Collados Aís Video 1	2.000	2.071	1.072
	Test B	3.000	2.626	1.102

Was das Kriterium der Kohärenz betrifft, so beträgt der Mittelwert der lebhaften Version in der vorliegenden Studie 4,52, während der der monotonen Version bei 3,70 liegt, und eine Mittelwertdifferenz von 0,82 aufweist. In der Studie von Collados Aís (1998) beträgt der Mittelwert der lebhaften Version 4,42 und der der monotonen Version 4,14, mit einer geringeren Mittelwertdifferenz als bei Video 1, nämlich 0,28 (Tab. 98).

Tab. 98: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Kohärenz

Kriterium	Studie	Lebhaftes Version (Test A /Video 3, Mittelwert)	Monotone Version (Test B /Video 1, Mittelwert)	Mittelwertdifferenz
Kohärenz	Vorliegende Studie	4.52	3.70	0,82
Kohärenz	Collados Aís	4.42	4.14	0,28

Die monotonen Versionen beider Studien weisen eine insgesamt positive Bewertung auf, allerdings mit einem leichten Unterschied zugunsten von Video 1. In der Stichprobe von

Collados Aís (1998) konzentrieren sich die Antworten deutlich auf die hohen Werte: 50 % vergaben den Wert 4 und 35,7 % den Wert 5, während niedrige Bewertungen fast nicht vorkommen. Auch in Test B überwiegt der Wert 4 (48,4 %), jedoch ist eine breitere Verteilung in Richtung der Mittelwerte und ein geringerer Anteil an Höchstbewertungen (18,7 %) zu beobachten (Tab. 99). Der Mittelwert von Video 1 (4,14) ist höher als der von Test B (3,70), und die Streuung ist in beiden Gruppen ähnlich (Tab. 100).

Tab. 99: Häufigkeiten der Bewertungen für das Kriterium Kohärenz für Video 1 (Collados Aís) und Test B

Bewertung	Collados Aís Video 1 (N=14)	Collados Aís Video 1 %	Test B (N=91)	Test B_%
1	0	0,0	3	3,3
2	1	7,1	8	8,8
3	1	7,1	19	20,9
4	7	50,0	44	48,4
5	5	35,7	17	18,7

Tab. 100: Deskriptive Statistiken für das Kriterium Kohärenz für Video 1 (Collados Aís) und Test B

Kohärenz	Gruppe	Median	Mittelwert	SD
	Collados Aís Video 1	4.000	4.143	0.864
	Test B	4.000	3.703	0.983

In der vorliegenden Studie beträgt der Mittelwert der lebhaften Version für das Kriterium Professionalität 4,52, während der Mittelwert der monotonen Version bei 3,51 liegt, und eine Mittelwertdifferenz von 1,01 aufweist. Bei Video 1 beträgt der Mittelwert der lebhaften Version 4,42 und der der monotonen Version 3,64, mit einer mittleren Differenz von 0,78, die geringer ist als die der vorliegenden Studie (Tab. 101).

Tab. 101: Mittelwertdifferenz zwischen der lebhaften und der monotonen Version nach Studie für das Kriterium Professionalität

Kriterium	Studie	Lebhafte Version (Test A / Video 3, Mittelwert)	Monotone Version (Test B / Video 1, Mittelwert)	Mittelwertdifferenz
Professionalität	Vorliegende Studie	4.52	3.51	1,01
Professionalität	Collados Aís	4.42	3.64	0,78

In den monotonen Versionen der beiden Studien äußern sich beide Gruppen tendenziell positiv. Bei Video 1 vergab die Hälfte der Teilnehmer*innen den Wert 4, gefolgt von mittleren Bewertungen (3: 21,4 %), während die Höchstbewertung 5 seltener vergeben wurde (14,3 %). Es wurden keine Mindestbewertungen vergeben. Auch in Test B ist der Wert 4 der häufigste (30,8 %), aber es gibt auch einen erheblichen Anteil an Bewertungen mit 3 (28,6 %) und 5 (20,9 %) sowie eine geringe Anzahl an niedrigen Bewertungen (Tab. 102). Der Mittelwert von Video 1 (3,64) ist etwas höher als der von Test B (3,52), und die höhere Standardabweichung in Test B (1,058 gegenüber 0,929) deutet auf eine größere Streuung der Antworten hin (Tab.103).

Tab. 102: Häufigkeiten der Bewertungen für das Kriterium Professionalität für Video 1 (Collados Aís) und Test B

Bewertung	Collados Aís Video 1 (N=14)	Collados Aís Video 1 %	Test B (N=91)	Test B_%
1	0	0,0	1	1,1
2	2	14,3	17	18,7
3	3	21,4	26	28,6
4	7	50,0	28	30,8
5	2	14,3	19	20,9

Tab. 103: Deskriptive Statistiken für das Kriterium Professionalität für Video 1 (Collados Aís) und Test B

Professionalität	Gruppe	Median	Mittelwert	SD
	Collados Aís Video 1	4.000	3.643	0.929
	Test B	4.000	3.516	1.058

Abbildung 13 stellt die Mittelwertdifferenzen jeder Studie zwischen den lebhaften Kontrollversionen (Video 3 und Test A) und den monotonen Versionen (Video 1 und Test B) visuell dar.

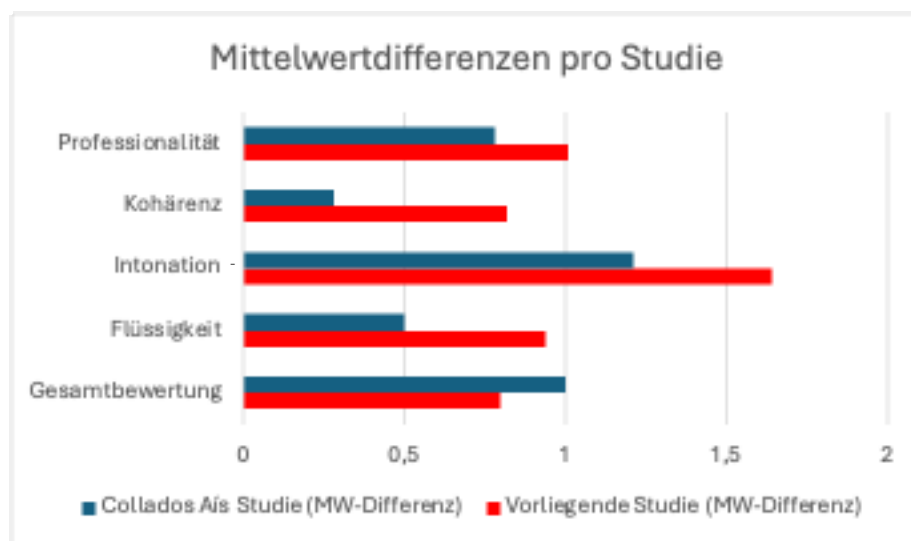


Abb. 13: Vergleich der Mittelwertdifferenzen zwischen der lebhaften und der monotonen Version nach Studie über alle Kriterien

Abbildung 14 stellt die Bewertungsunterschiede zwischen Test B und Video 1 für jedes Kriterium visuell zusammenfassend dar.

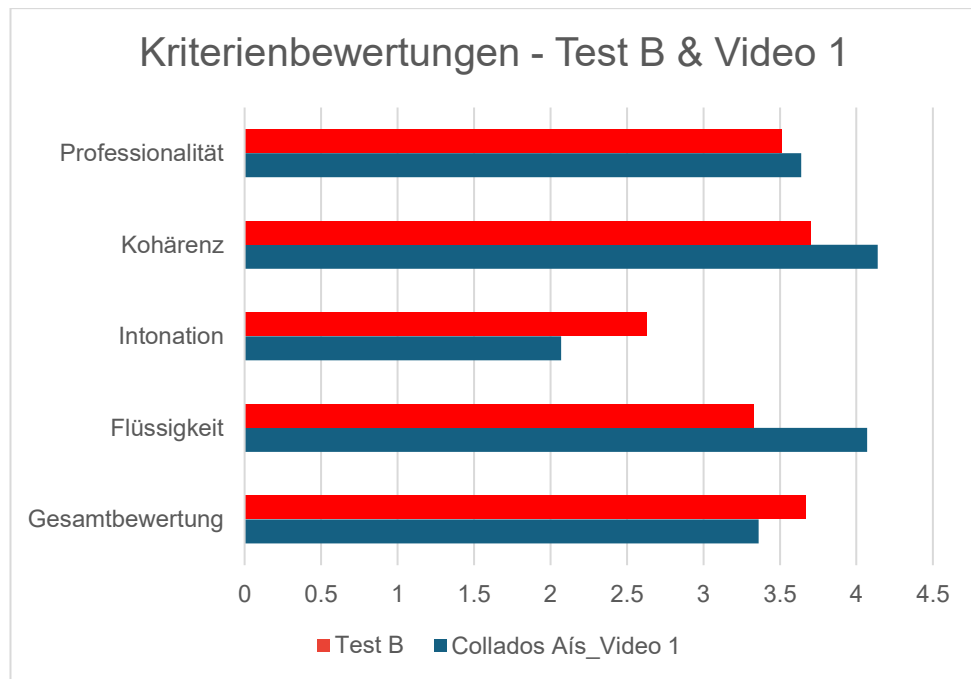


Abb. 14: Vergleich der Bewertungen über alle Kriterien – Video 1 (Collados Aís) und Test B

7 Diskussion und Schlussfolgerungen

7.1 Ergebnisse des Vergleichs zwischen Test A und Test B sowie der Störvariablen

Aus den im vorigen Kapitel dargestellten Ergebnissen geht hervor, dass die Hypothesen der Studie von Collados Aís (1998), die in der vorliegenden Studie erneut überprüft werden sollten, bestätigt werden konnten: Es hat sich gezeigt, dass sich Monotonie negativ auf die Wahrnehmung der Qualität der Dolmetschleistung durch die Nutzer*innen und auf ihre Wahrnehmung der Professionalität der Dolmetscherin auswirkt.

Der durch monotone Intonation gekennzeichnete Test B wurde von den Nutzer*innen in Bezug auf alle fünf Kriterien Gesamtbewertung, Flüssigkeit, Intonation, Kohärenz und Professionalität negativer bewertet als Test A. Das Kriterium mit dem größten Unterschied in der Bewertung zwischen Test A und B ist die Intonation, deren Mittelwert in Gruppe A bei 4,27 und in Test B bei 2,63 liegt, wobei der Unterschied statistisch signifikant ist (Tab. 10). Die Verteilung der Bewertungen zwischen Test A und Test B unterscheidet sich für alle fünf betrachteten Kriterien signifikant, wie die statistischen Mann-Whitney-U-Tests zeigen. Das zweite Kriterium mit dem deutlichsten Unterschied zwischen den beiden Gruppen ist die Professionalität, mit einem Mittelwert von 4,52 für Test A und 3,52 für Test B, gefolgt von Flüssigkeit (Mittelwert von 4,26 für Test A und 3,33 für Test B) und Kohärenz (Mittelwert Test A: 4,52 und Test B: 3,70) (Tab. 16, 7 und 13). Das Kriterium mit dem geringsten Unterschied zwischen den beiden Gruppen mit einem Mittelwert von 4,47 für Test A und 3,67 für Test B ist die Gesamtbewertung (Tab. 4). Die sehr niedrigen Bewertungen im Kriterium Intonation im Test B lassen sich auf den monotonen Charakter der Dolmetschleistung zurückführen, der bereits im Kapitel 5.2.3.1 visuell veranschaulicht wurde, sowie auf die entsprechende Wahrnehmung dieser Monotonie seitens der Nutzer*innen. Die Manipulation der Intonation führte auch bei den anderen Kriterien zu niedrigeren Bewertungen als beim Test A, wodurch die Dolmetscherin als weniger professionell und die Leistung als weniger flüssig und weniger kohärent empfunden wurde. Dies bestätigt die Ergebnisse der Studien von Holub (2010) und Collados Aís (1998, 2008, 2016), wonach eine monotone Intonation zu einer schlechteren Bewertung der Dolmetschleistung führt (siehe Kap. 4.2.). Wie in Kapitel 2.2.2.1 dargelegt, neigen Nutzer*innen in Studien zu Erwartungen (Kurz 1989, 1993, Kopczynski 1994, Moser 1996) dazu, Kriterien, die sich auf den Inhalt beziehen, höhere Bewertungen zuzuweisen. Wenn hingegen die Intonation manipuliert wird und eine konkrete Leistung bewertet werden muss, wie im Fall der vorliegenden Studie, kommt die Bedeutung der darbietungsbezogenen Kriterien zum Tragen. Trotz der negativeren Bewertungen für die anderen Kriterien weist die allgemeine

Wahrnehmung des Tests B für die Gesamtbewertung einen Mittelwert von 3,67 auf – das zweitbeste Kriterium nach dem Kriterium Kohärenz. Außerdem weist dieses Kriterium den geringsten – wenn auch statistisch signifikanten – Unterschied zwischen Test A und Test B auf. Die Gesamtbewertung stellte in diesem Fall nicht den arithmetischen Mittelwert aller Kriterien dar, der 3,29 hätte betragen müssen, sondern bezog sich auf die generelle Wahrnehmung der Qualität der Leistung. Insgesamt bewerteten die Nutzer*innen Test B mit einem hohen Wert. Dies kann als Hinweis auf den ganzheitlichen Charakter der Gesamtbewertung interpretiert werden. Die Nutzer*innen scheinen, obwohl sie von der Monotonie in Test B beeinflusst sind, der allgemeinen Verständlichkeit und dem Kommunikationserfolg mehr Bedeutung beizumessen als einzelnen formalen Aspekten. Die einzelnen Kriterien, wie Intonation oder Professionalität der Dolmetscherin, werden hingegen stärker durch die Monotonie beeinträchtigt. Dies deutet darauf hin, dass die Monotonie der Intonation zwar alle Kriterien negativ beeinflusst, aber für sich genommen nicht ausreicht, um die Gesamtwahrnehmung der Qualität der Dolmetschleistung vollständig zu beeinträchtigen.

Was die Störvariable der Teilnahme an Konferenzen betrifft, wurde für beide Tests separat mittels des Mann-Whitney-U-Tests überprüft, ob dieses Element die Leistungsbewertung beeinflussen könnte. Es sollte also überprüft werden, ob Nutzer*innen, die über mehr Erfahrung im Bereich Dolmetschen verfügen, nach anderen Bewertungskriterien beurteilen als diejenigen, die noch nie an Konferenzen teilgenommen haben, bei denen Dolmetscher*innen eingesetzt wurden. Wie beispielsweise in der Studie von Collados Aís aus dem Jahr 2007 zu sehen ist, die eine Wiederholung der Studie von 1998 darstellt, waren die Nutzer*innen bei der Bewertung der monotonen Version nachsichtiger, was auf ihre Gewöhnung an die Intonation *sui generis* der Verdolmetschungen zurückzuführen sein könnte (vgl. Collados Aís 2007: 172). Die Mann-Whitney-U-Tests haben keine statistisch signifikanten Unterschiede für die einzelnen Kriterien ergeben. Die Mittelwerte aller Bewertungskriterien sind in beiden Tests für Teilnehmer*innen mit und ohne Konferenzerfahrung sehr ähnlich. Bei beiden Tests vergaben die Nutzer*innen, die noch nie an Konferenzen teilgenommen haben, beim Kriterium Flüssigkeit höhere Bewertungen als die Nutzer*innen mit Erfahrung (Test A Flüssigkeit: 4,29 gegenüber 4,12, Test B Flüssigkeit: 3,33 gegenüber 3,20). In Test B vergaben die Nutzer*innen mit Konferenzerfahrung höhere Bewertungen mit einem Mittelwert von 2,67 gegenüber 2,62. Die Unterschiede sind in jedem Fall minimal, sodass keine Rückschlüsse auf den Zusammenhang zwischen der Erfahrung als Konferenzteilnehmer*in und der Wahrnehmung der Intonation gezogen werden können. Ebenso ist die Häufigkeit der Teilnahme in Bezug auf die Bewertungen für keinen der beiden

Tests statistisch signifikant, wie durch die Kendall-Tau-Tests bestätigt wurde (siehe Kapitel 6.2).

Auch hinsichtlich der Störvariable Deutschkenntnisse zeigen die Ergebnisse des Mann-Whitney-U-Tests, dass dieses Element weder im Test A noch im Test B einen signifikanten Einfluss auf die Bewertungen der analysierten Kriterien hat. In Test A sind die Mittelwerte für die Kriterien Gesamtbewertung, Flüssigkeit und Intonation bei den Teilnehmer*innen mit Deutschkenntnissen etwas höher. Dennoch sind diese Unterschiede statistisch nicht signifikant. In Test B vergaben diejenigen ohne Deutschkenntnisse etwas höhere Bewertungen für die Kriterien Gesamtbewertung, Intonation und Professionalität, jedoch sind die Unterschiede zwischen den beiden Gruppen minimal (siehe Kapitel 6.3). Diese Ergebnisse lassen vermuten, dass die Kenntnis der Originalsprache weder die Gesamtbewertung noch die einzelnen formalen Aspekte der Dolmetschleistung zu beeinflussen scheint. Dies könnte auch auf die schlechte akustische Qualität des Ausgangstextes zurückzuführen sein, der von der Dolmetschleistung überlagert wurde.

7.2 Ergebnisse des Intra- und Interstudienvergleichs

In ihrer Studie aus dem Jahr 1998 stellte Collados Aís fest, dass zukünftige Forschungen überprüfen könnten, ob eine monotone Intonation der Dolmetschleistung bei einem lebhaft intonierten Ausgangstext als noch stärker nachteilig wahrgenommen wird (vgl. 1998: 90). Ihr Text wies eine monotone Intonation auf – auch aufgrund des behandelten Themas, das keine besondere Ausdruckskraft erforderte. Die vorliegende Studie hat sich zum Ziel gesetzt, anhand eines lebhaften Texts zu überprüfen, ob diese Hypothese bestätigt werden kann. Gemäß den Ergebnissen der deskriptiven Analyse stützen die Daten diese Hypothese: Bei allen untersuchten Bewertungskriterien, mit Ausnahme des Kriteriums der Gesamtbewertung, fällt der Unterschied der Mittelwerte zwischen der lebhaften und der monotonen Version in der vorliegenden Studie stärker aus als in der Studie von Collados Aís (1998). Das bedeutet, dass in dieser Studie zwischen der lebhaften Version A und der monotonen Version B deutlicher zu erkennen ist, dass die monotone Version schlechter bewertet wurde als die lebhafte Version. Das Kriterium, bei dem der Unterschied zwischen den beiden Versionen am größten ist, ist die Intonation, mit einem Mittelwert von 4,27 im Test A und 2,63 im Test B und einer Mittelwertdifferenz von 1,64, verglichen mit einer Differenz von 1,21 in der Studie von Collados Aís (1998), mit einem Mittelwert von 3,28 bei Video 3 und 2,07 bei Video 1 (Tab. 95). Auf dieses Kriterium folgt das Kriterium Professionalität: Im Test A erreicht es einen Mittelwert von 4,52, während dieser im Test B bei 3,51 liegt, was einer Differenz von 1,01

entspricht. Diese Differenz ist höher als jene von 0,78, die in der Studie von Collados Aís (1998) festgestellt wurde (Tab. 101). Auf das Kriterium der Professionalität folgt das Kriterium der Flüssigkeit mit einem Mittelwert von 4,27 im Test A gegenüber 3,33 im Test B und einer Differenz von 0,94 gegenüber einer Differenz von 0,50 in der Studie von Collados Aís (1998) (Tab. 92). Das Kriterium Kohärenz weist einen geringeren Unterschied als die vorherigen Kriterien auf, mit einem Mittelwert von 4,52 im Test A und 3,70 im Test B und einer Mittelwertdifferenz von 0,82, die jedoch immer noch größer ist als in der Studie von Collados Aís (1998), die eine Mittelwertdifferenz von 0,28 aufweist (Tab. 98). Das einzige Kriterium, dessen Mittelwertdifferenz in der vorliegenden Studie geringer ist als in der Studie von Collados Aís (1998), ist die Gesamtbewertung. In der Studie von Collados Aís (1998) weist sie eine Abweichung von 1,00 auf, in der vorliegenden Studie beträgt die Abweichung 0,80 (Tab. 89).

In Bezug auf die vergleichende Analyse, die für die beiden monotonen Versionen der zwei Studien durchgeführt wurde, lässt sich in der vorliegenden Studie eine Tendenz zu einer negativeren Bewertung der Dolmetschleistung hinsichtlich der Kriterien Flüssigkeit (Mittelwerte: Video 1: 4,07 vs. 3,33 Test B), Kohärenz (Mittelwerte: Video 1: 4,14 Test B: 3,70) und Professionalität (Mittelwerte: Video 1: 3,64 vs. Test B: 3,52) beobachten. Die Kriterien Gesamtbewertung (Mittelwerte: Video 1: 3,36 vs. Test B: 3,67) und Intonation (Mittelwerte: Video 1: 2,07 vs. Test B: 2,63) erhielten in der vorliegenden Studie hingegen höhere Bewertungen. Diese Ergebnisse zeichnen somit ein differenziertes Bild: Einerseits ist in der vorliegenden Studie die Mittelwertdifferenz zwischen der lebhaften und der monotonen Version größer, andererseits fällt die Bewertung der monotonen Version für sich genommen nicht systematisch negativer aus als in der Studie von Collados Aís (1998). Insbesondere beim Kriterium Intonation und Gesamtbewertung erhält die monotone Version von Collados Aís (Video 1) in absoluten Zahlen niedrigere Bewertungen, während die Bewertung dieser Kriterien in der vorliegenden Studie (Test B) weniger streng ausfällt.

Es ist außerdem zu beachten, dass in beiden Studien die beiden Bedingungen von unabhängigen Gruppen bewertet wurden. Die beobachteten Unterschiede spiegeln daher Abweichungen zwischen den Gruppen wider und nicht einen direkten Vergleich durch die Teilnehmer*innen selbst.

Solche Ergebnisse können auf verschiedene Faktoren zurückzuführen sein. Daher ist es sinnvoll, zwischen der relativen Distanz zwischen den Bedingungen innerhalb der jeweiligen Studie und den absoluten Bewertungsniveaus bzw. den vergebenen Punktwerten in der vorliegenden Studie und in der Studie von Collados Aís (1998) zu unterscheiden.

Einer der Gründe, warum in der vorliegenden Studie die Mittelwertdifferenz zwischen der lebhaften und der monotonen Version deutlicher ausfällt, lässt sich auf die von Collados Aís (1998) aufgestellte Hypothese zurückführen. Diese besagt, dass die Nutzer*innen im Test B die Monotonie der Dolmetschleistung möglicherweise aufgrund der nicht-monotonen Natur des Ausgangstextes stärker wahrgenommen haben. Dies verstärkt den wahrgenommenen Unterschied zwischen der lebhaften Intonation des Ausgangstextes und der Monotonie des Zieltextes. Eine weitere Erklärung lässt sich auf das im Ausgangstext behandelte Thema zurückführen, das sich im Gegensatz zum Thema von Collados Aís (1998) besser für eine lebhafte Intonation eignet. Eine weitere mögliche Ursache lässt sich auf die unterschiedliche Intonation der lebhaften und monotonen Versionen jeder Studie zurückführen. Es ist möglich, dass in der vorliegenden Studie die lebhafte Version im Vergleich zur Studie von Collados Aís (1998) lebhafter war, was zu positiveren Bewertungen der lebhaften Version geführt haben könnte, was wiederum zu einer deutlicheren Mittelwertdifferenz zwischen den beiden Versionen in der vorliegenden Studie geführt hätte.

Was den Vergleich zwischen den beiden monotonen Versionen der Studien betrifft, so wurde unter Berücksichtigung der absoluten Werte und nicht der Mittelwertdifferenzen festgestellt, dass die Kriterien Flüssigkeit, Kohärenz und Professionalität in dieser Studie niedrige Mittelwerte aufweisen, während Intonation und Gesamtbewertung in der Studie von Collados Aís (1998) negativer bewertet werden (siehe Kapitel 6.4). Diese Ergebnisse können im Lichte einiger Limitationen der vorliegenden Studie interpretiert werden. Im Folgenden werden diese sowie weitere methodische Einschränkungen näher erläutert.

7.3 Limitationen

Eine Limitation betrifft die Verschiedenartigkeit der Stichprobe. Die Nutzer*innen der vorliegenden Studie stellen eine ganz andere Stichprobe dar als die Nutzer*innen der Studie von Collados Aís (1998), die sich aus juristischen Expert*innen, also spezialisierten Nutzer*innen, zusammensetzte. In der vorliegenden Studie handelt es sich bei den Nutzer*innen um ein allgemeines Publikum, dessen Voraussetzungen lediglich darin bestanden, bereits Simultanverdolmetschungen in den Medien gesehen zu haben, Italienisch auf dem Niveau C1 oder höher zu sprechen und keine Erfahrung im Bereich Dolmetschen auf beruflicher oder Ausbildungsstufe zu haben. Sie stellen eine Stichprobe potenzieller Teilnehmer*innen einer Konferenz zur Vorstellung eines Buches dar. Es ist möglich, dass die Teilnehmenden dieser Studie der allgemeinen Verständlichkeit ein größeres Gewicht beimessen als spezifischeren technischen Parametern. Dies könnte darauf hindeuten, dass sie

nicht genau wissen, wie eine professionelle Simultanverdolmetschung klingen sollte, und ihre Bewertungen daher tendenziell milder ausfallen als jene erfahrenerer Nutzer*innen.

Eine weitere Limitation betrifft den Kontext der Testdurchführung. Die Studie von Collados Aís (1998) wurde in einem kontrollierten Umfeld durchgeführt, in dem die Teilnehmer*innen vermutlich konzentriert und vollständig auf das Anhören der Verdolmetschung fokussiert waren, während die vorliegende Studie online durchgeführt wurde. Ohne den Haupteffekt der Monotonie auf die Bewertungen in Frage zu stellen, könnte dies sich auf die Aufmerksamkeit der Teilnehmer*innen und damit auf die abgegebenen Bewertungen ausgewirkt haben. Die Nutzer*innen könnten dabei schneller ein Urteil gefällt haben, das auf ihren ersten Eindrücken beruht.

Eine weitere Einschränkung betrifft die begrenzte Vergleichbarkeit der vorliegenden Studie mit der von Collados Aís (1998). Es lässt sich nämlich nicht feststellen, inwiefern sich die unterschiedlichen Intonationen in den beiden Studien unterscheiden, da die von Collados Aís (1998) verwendeten Videos nicht zur Verfügung stehen. Dieser Unterschied in der Intonation könnte zu Abweichungen bei den Bewertungen geführt haben. Es ist möglich, dass die Monotonie in Video 1 von Collados Aís (1998) ausgeprägter war als in der vorliegenden Studie und daher negativer bewertet wurde. Andererseits könnte die größere Mittelwertdifferenz beim Kriterium Intonation in der vorliegenden Studie darauf zurückzuführen sein, dass die lebhaftere Version lebhafter war als die von Collados Aís (1998).

Neben den bereits erörterten Aspekten weist die vorliegende Studie weitere methodische Limitationen auf, die für die Interpretation der Ergebnisse von Bedeutung sind. Eine davon betrifft die durchgeführten inferentiellen Tests. Es ist bei der Analyse der Unterschiede zwischen Test A und Test B zu beachten, dass zwar alle Ergebnisse statistisch signifikant waren, jedoch die Möglichkeit eines Fehlers vom Typ I in Betracht gezogen werden muss (vgl. Field et al. 2025: 730). Dennoch ist deutlich zu erkennen, dass Test B im Vergleich zu Test A schlechter bewertet wurde. Darüber hinaus erfolgte der Vergleich zwischen den Unterschieden im Mittelwert der Bewertungen für jeden Test und der Vergleich zwischen der monotonen Version der vorliegenden Studie und der von Collados Aís (1998) mittels eines deskriptiven Ansatzes. Im Gegensatz zum inferentiellen Ansatz kann dieser Ansatz keine Hypothesen überprüfen, sondern liefert lediglich vorläufige Hinweise und Beobachtungen, ohne statistische Schlussfolgerungen zu ermöglichen, die sich verallgemeinern lassen. Aus diesem Grund sind die Interpretationen der Ergebnisse als spekulativ zu betrachten (vgl. Bortz & Schuster 2010: 3).

Eine weitere Einschränkung betrifft die Ungleichverteilung zwischen den

Untergruppen. Was beispielsweise die Störvariablen der Teilnahme an Konferenzen und der Deutschkenntnisse betrifft, so hatten in Gruppe A nur 16 von 98 Teilnehmer*innen bereits an Konferenzen teilgenommen, während 19 angaben, über Deutschkenntnisse zu verfügen. In Gruppe B hatten 15 von 91 Teilnehmer*innen bereits an Konferenzen teilgenommen und weitere 15 verfügten über Deutschkenntnisse. Daher kann das Ausbleiben signifikanter Effekte für beide Störvariablen einen möglichen Einfluss dieser beiden Faktoren nicht vollständig ausschließen. Auch beim Vergleich zwischen der Version B der vorliegenden Studie und dem Video 1 der Studie von Collados Aís (1998) ist die Stichprobengröße stark unausgewogen ($n = 14$ vs. $n = 90$), was die direkte Vergleichbarkeit der beobachteten Ergebnisse einschränken kann.

Eine weitere Einschränkung des Vergleichs zwischen den monotonen Versionen besteht in der zeitlichen Distanz zwischen den beiden Untersuchungen. In den letzten Jahrzehnten haben sich nicht nur die technischen Bedingungen, sondern auch die Erwartungen der Nutzer*innen und ihre Auseinandersetzung mit multimedialen Produkten erheblich verändert. Diese kontextuellen Unterschiede könnten die Bewertungskriterien der Nutzer*innen beeinflusst und die direkte Vergleichbarkeit der Ergebnisse eingeschränkt haben.

7.4 Schlussfolgerungen

Ziel dieser Studie war es, zwei der Hypothesen von Collados Aís (1998) zu überprüfen, wonach Monotonie einen negativen Einfluss auf die Bewertung der Qualität der Dolmetschleistung sowie auf die Wahrnehmung der Professionalität der Dolmetsche*innen durch die Nutzer*innen hat. Wie die Ergebnisse zeigen, wurden diese Hypothesen bestätigt. Die Teilnehmer*innen dieser Studie bewerteten den monotonen Test B bei allen fünf Bewertungskriterien statistisch signifikant schlechter. Dies bestätigt somit die Bedeutung prosodischer Parameter bei der Bewertung einer Dolmetschleistung, wie bereits ausführlich erörtert und durch die in dieser Arbeit analysierten Forschungen von Holub (2010) und Collados Aís (1998, 2008, 2016) belegt wurde. Dennoch erzielte das Kriterium der Gesamtbewertung in der monotonen Version dieser Arbeit eine recht positive Bewertung und belegte mit einem Mittelwert von 3,67 den zweiten Platz. Dies bedeutet, dass für die nicht spezialisierten Nutzer*innen, die an der vorliegenden Studie teilgenommen haben, die monotone Intonation die ganzheitliche Wahrnehmung der Qualität der Dolmetschleistung nicht vollständig beeinträchtigt. Wahrscheinlich wird dabei die kommunikative Wirksamkeit stärker gewichtet als einzelne technische Parameter.

Ein weiteres Ziel dieser Studie war es, zu überprüfen, ob eine monotone Intonation der

Dolmetschleistung bei einem lebhaft intonierten Ausgangstext als noch stärker nachteilig wahrgenommen wird. Diese Hypothese scheint sich zu bestätigen, da in der vorliegenden Studie die Mittelwertdifferenz zwischen der monotonen und der lebhaften Version bei allen Kriterien – mit Ausnahme der Gesamtbewertung – größer ist als in der Studie von Collados Aís (1998). Dies könnte darauf zurückzuführen sein, dass die Wahrnehmung von Monotonie deutlicher zutage tritt, wenn der Ausgangstext eine lebhafte Intonation aufweist. Eine weitere mögliche Erklärung könnte darin liegen, dass die lebhafte Version der vorliegenden Studie eine lebhaftere Intonation aufweist als die Version, die in der Studie von Collados Aís (1998) verwendet wurde. Ein direkter Vergleich zwischen den verschiedenen prosodischen Realisierungen ist jedoch nicht möglich, da die Original-Videoaufnahmen der Studie von Collados Aís (1998) nicht vorliegen.

Was den Vergleich der monotonen Versionen der beiden Studien betrifft, so werden die Kriterien Flüssigkeit, Kohärenz und Professionalität in der vorliegenden Studie schlechter bewertet als in der Studie von Collados Aís (1998), doch weisen Intonation und Gesamtbewertung höhere Durchschnittswerte auf. Dies könnte an der Tatsache liegen, dass es sich um ein nicht spezialisiertes Publikum handelt, das der allgemeinen Verständlichkeit möglicherweise mehr Bedeutung beimisst als anderen, spezifischeren Parametern. Darüber hinaus sind die abgegebenen Bewertungen weitgehend subjektiv und können von verschiedenen Faktoren abhängen, darunter auch vom Setting, in dem die Studie durchgeführt wurde. Da der Fragebogen online ausgefüllt wurde, lässt sich nicht ausschließen, dass Faktoren wie die Aufmerksamkeitsspanne der Teilnehmer*innen oder das vollständige Ansehen des Videos die Ergebnisse beeinflusst haben könnten.

Auf jeden Fall lässt sich feststellen, dass die Ergebnisse dieser Studie zeigen, dass die Intonation im Bereich Dolmetschen sehr wichtig ist, da sie die Bewertung der meisten Kriterien in der monotonen Version beeinflusst und dazu führt, dass die Nutzer*innen die Dolmetscherin als weniger professionell wahrnehmen.

Wie in Kapitel 2.2.2 dargelegt, sind es die Nutzer*innen, die eine Dolmetschleistung bewerten sollten, da sie diejenigen sind, die diese Leistung in Anspruch nehmen (vgl. Seleskovitch 1986: 236).

Die Tatsache, dass die Intonation einen so erheblichen Einfluss auf die Bewertung der Leistung durch die Nutzer*innen hatte, bestätigt, wie wichtig es ist, auch im Bildungsbereich den prosodischen Kriterien mehr Aufmerksamkeit zu schenken, die zwar nicht direkt den Inhalt betreffen, aber die Wahrnehmung der Qualität der Verdolmetschung maßgeblich beeinflussen. Es muss jedoch betont werden, dass, obwohl sich eine monotone Intonation negativ auf einzelne

Kriterien der Dolmetschleistung auswirkt, die Teilnehmer*innen der vorliegenden Studie in ihrer Gesamtbewertung tendenziell der allgemeinen Verständlichkeit und dem Kommunikationserfolg den Vorrang geben. Das bedeutet, dass die monotone Intonation zwar als negativer Aspekt erkannt wird, ihre Auswirkung auf die Gesamtwahrnehmung der Qualität jedoch dadurch gemildert wird, dass die Botschaft klar und kohärent bleibt.

Es wäre interessant, wenn zukünftige Forschungen die Studie von Collados Aís (1998) mit einem lebhaften Ausgangstext replizieren könnten. Dabei sollten spezialisierte Nutzer*innen und nicht ein allgemeines Publikum einbezogen werden, um zu prüfen, ob die Ergebnisse ähnlich ausfallen und in der monotonen Version niedrigere Bewertungen für das Kriterium der Gesamtbewertung vergeben werden. Darüber hinaus könnten weitere Studien durchgeführt werden, die neben Spanisch-Deutsch und Italienisch-Deutsch auch andere Sprachpaare berücksichtigen.

Literaturverzeichnis

- Ahrens, Barbara (2004). *Prosodie beim Simultandolmetschen*. Frankfurt am Main/Wien: Lang.
- AIIC (1982). *Practical Guide for Professional Interpreters*. Genève: AIIC.
- AIIC (2022). *AIIC Code of Professional Ethics*. Genève: AIIC.
- Alexieva, Brista (1988). Analysis of the simultaneous interpreter's output. In: Nekeman, Paul (ed) *Translation, our future. Proceedings of the XIth World Congress of FIT*. Maastricht: Euroterm, 484–488.
- Barik, Henri C. (1971). A Description of Various Types of Omissions, Additions and Errors of Translation Encountered in Simultaneous Interpretation. *Meta* 16 (4), 199–210.
- Barik, Henri C. (1975). Simultaneous interpretation: Qualitative and linguistic data. *Language and Speech* 18 (3), 272–297.
- Barranco-Droege, Rafael; Collados Aís, Ángela & Breñaña, José Manuel Pazos (2011). Intonation. In: Collados Aís, Ángela; Fernández Iglesias, Emilia; Pradas Macías, Esperanza Macarena & Stévaux, Elisabeth (eds) *Qualitätsparameter beim Simultandolmetschen. Interdisziplinäre Perspektiven*. Tübingen: Narr, 61–92.
- Bortz, Jürgen, & Schuster, Christof (2010). *Statistik für Human- und Sozialwissenschaftler*. 7. Aufl. Berlin, Heidelberg: Springer.
- Braun, Bettina; Audra, Dainora & Ernestus, Miriam (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes* 26 (3), 350–375.
- Brennan, Susan E. & Williams, Maurice (1995). The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language* 34, 383–398.
- Bühler, Hildegund (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua* 5 (4), 231–235.
- Chiaro, Delia & Nocella, Giuseppe (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the world wide web. *Meta* 49 (2), 278–293.
- Collados Aís, Ángela (1998). *La evaluación de la calidad en interpretación simultánea: La importancia de la comunicación no verbal*. Granada: Editorial Comares.
- Collados Aís, Ángela (1998/2002). Quality Assessment in Simultaneous Interpreting: The importance of nonverbal communication. In: Pöchhacker, Franz & Shlesinger, Miriam

- (eds) *The Interpreting Studies Reader*. London/New York: Routledge, 326–336.
- Collados Aís, Ángela (2007). La incidencia del parámetro entonación. In: Collados Aís; Pradas Macías; Stévaux & García Becerra (eds), 159–174.
- Collados Aís, Ángela (2008). Evaluacion de la calidad en interpretacion simultanea contrastes de exposicion e inferencias emocionales. Evaluacion de la evaluacion. In: Hansen, Gyde; Chesterman, Andrew & Gerzymisch-Arbogast, Heidrun (eds) *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*. Amsterdam: John Benjamins, 193–214.
- Collados Aís, Ángela (2010). La evaluación de la calidad en interpretación simultánea: pautas evaluadoras según usuarios. In: Rabadán, Rosa; Guzmán, Trinidad & Fernández López, Marisa (eds) *Lengua, traducción, recepción. En honor de Julio César Santoyo*. León: Universidad de León, 67–90.
- Collados Aís, Ángela (2016). Quality Assessment and Intonation in Simultaneous Interpreting: Evaluation Patterns. In: Calvo Rigual, Cesáreo & Spinolo, Nicoletta (eds) *Translating Orality. MonTI Special Issue 3*trans, 1–24.
- Collados Aís, Ángela; Pradas Macías, Esperanza Macarena; Stévaux, Elisabeth & García Becerra, Olalla (eds) (2007). *La evaluación de la calidad en interpretación simultánea :parámetros de incidencia*. Granada: Comares.
- Collados Aís, Ángela; Fernández Iglesias, Emilia; Pradas Macías, Esperanza Macarena & Stévaux, Elisabeth (eds) (2011). *Qualitätsparameter beim Simultandolmetschen. Interdisziplinäre Perspektiven*. Tübingen: Narr.
- Conover, William Jay (1999). *Practical Nonparametric Statistics*. 3d ed. New York, NY [u.a.]: Wiley
- Darò, Valeria (1990). Voice frequency and simultaneous interpretation. *The Interpreters' Newsletter* 3, 88–92.
- Darò, Valeria (1994). Non-linguistic factors influencing simultaneous interpretation. In: Lambert, Sylvie & Moser-Mercer, Barbara (eds) *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins , 249–272.
- Déjean le Féal, Karla (1990). Some Thoughts on the Evaluation of Simultaneous Interpretation. In: Bowen, David & Bowen, Margareta (eds) *Interpreting- yesterday, today, and tomorrow*. Binghamton, NY: SUNY, 154–60.
- Díaz-Cintas, Jorge & Orero Clavero, Pilar (2006). Voice-Over. In: Brown, Keith (ed) *Encyclopedia of Language & Linguistics* 13 (2). Oxford: Elsevier, 477–479.
- Dirikier, Ebru (2015). Conference interpreting. In: Pöchhacker, Franz (ed) *Routledge*

- encyclopedia of interpreting studies*. Routledge, 78–82.
- Dillman, Don A; Leah Melani, Christian & Smyth, Jolene D (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, New Jersey: Wiley.
- Döring, Nicola (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 6. Aufl. Berlin Heidelberg: Springer.
- Elsagir, Ina M. (2000). Anforderungen an Dolmetschleistungen im Fernsehen aus Zuschauersicht: Eine Fallstudie. In: Kalina, Sylvia, Buhl, Silke & Gerzymisch-Arbogast, Heidrun (eds) *Dolmetschen: Theorie, Praxis, Didaktik; mit ausgewählten Beiträgen der Saarbrücker Symposien*. St. Ingbert: Röhrig Universitätsverlag, 107–123.
- Field, Andy; van Doorn, Johnny & Wagenmakers, Eric-Jan (2025). *Discovering Statistics Using JASP*. 1st ed. United Kingdom of Great Britain & Northern Ireland: Sage Publications Ltd (UK).
- Fujisaki, Hiroya (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In: MacNeilage, Peter (ed) *The production of speech*. New York: Springer, 39–55.
- Fujisaki, Hiroya (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In: Fujimura, Osamu (ed) *Vocal physiology: voice production, mechanisms and functions*. New York: Raven, 347–355.
- García Becerra, Olalla (2011). Eindrucksbildung. In: Aís; Fernández; Pradas Macías & Stévaux (eds), 15–32.
- García Becerra, Olalla (2012). First impressions in interpreting quality assessment, the incidence of nonverbal communication. In: Jiménez Ivars, María Amparo (ed) *Interpreting Brian Harris: Recent Developments in Translatology*. Bern Wien [u.a.]: Lang, 173–192.
- Garzone, Giuliana Elena & Viezzi, Mauro (eds) (2002). *Interpreting in the 21st Century: Challenges and Opportunities*. Amsterdam: John Benjamins.
- Garzone, Giuliana & Viezzi, Mauro (2002). Quality and norms in interpretation. In: Garzone & Viezzi (eds), 107–119.
- Garzone, Giuliana Elena (2003). Reliability of quality criteria evaluation in survey research. In: Collados Aís, Ángela (ed) *La evaluación de la calidad en interpretación: docencia*

- y profesión. *Actas del 1. Congreso Internacional sobre Evaluación de la Calidad en Interpretación de Conferencias, Almuñécar, 2001*. Granada: Comares, 23–30.
- Garzone, Giuliana Elena (2015). Norms. In: Pöchhacker, Franz (ed) *Routledge encyclopedia of interpreting studies*. Routledge, 281–283.
- Gile, Daniel (1985). Le modèle d'efforts et l'équilibre d'interprétation en interprétation simultanée. *Meta* (30) 1, 44–48.
- Gile, Daniel (1990). L'évaluation de la qualité de l'interprétation par les délégués: une étude de cas. *The Interpreters' Newsletter* 3, 66–71.
- Gile, Daniel (1991). A communication-oriented analysis of quality in non-literary translation and interpretation. In: Larson, Mary-Lynn (ed) *Translation: Theory and Practice, Tension and Interdependence*. Binghamton (NY): SUNY, 188–200.
- Gile, Daniel (2004). Translation research versus interpreting research: Kinship, differences and prospects for partnership. In: Schäffner, Christina (ed) *Translation Research and Interpreting Research: Traditions, Gaps and Synergies*. Clevedon: Multilingual Matters, 10–34.
- Gile, Daniel (2009a). Conference Interpreting, Historical and Cognitive Perspectives. In: Baker, Mona & Saldanha, Gabriela (eds) *Routledge Encyclopedia of Translation Studies*. London/New York: Routledge, 71–75.
- Gile, Daniel (2009b). *Basic Concepts and Models for Interpreter and Translator Training*. Revised Edition. Amsterdam/Philadelphia: John Benjamins.
- Goldman-Eisler, Frieda (1972). Pauses, clauses, sentences. *Language and Speech* 15 (2), 103–113.
- Günther, Carsten (1999). *Prosodie und Sprachproduktion*. Berlin/Boston: De Gruyter.
- Halliday, Michael A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.
- Hardin, James N (1991). *Reflection and Action: Essays on the Bildungsroman*. Columbia, S.C: University of South Carolina Press.
- Herbert, Jean (1952). *Manuel de l'interprète*. Genf: Librairie de l'Université Georg.
- Herbert, Jean (1978). How conference interpretation grew. In: Gerver, David & Sinaiko, H. Wallace (eds) *Language Interpretation and Communication*. Boston (MA): Springer, 5–10.
- Holub, Elisabeth (2010). Does intonation matter? The impact of monotony on listener comprehension. *The Interpreters' Newsletter* 15, 117–126.
- Harris, Brian (1990). Norms in Interpretation. *Target* 2, 115–119.

- Iversen, Anniken Telnes (2009). *Change and continuity: the Bildungsroman in English*. Dissertation, The University of Toronto.
- Kade, Otto (1968). *Zufall und Gesetzmäßigkeit in der Übersetzung*. Leipzig: Verlag Enzyklopädie.
- Kalina, Sylvia (1998). *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Narr.
- Kapp, Volker (1974). *Übersetzer und Dolmetscher: Theoretische Grundlagen, Ausbildung, Berufspraxis*. 3. Aufl. Tübingen: Francke.
- Knobloch, Klaus (2002). Sprecherziehung und rhetorische Kommunikation-wichtige Bausteine im Dolmetschstudium. In: Best, Joanna & Kalina, Sylvia (eds) *Übersetzen und Dolmetschen. Eine Orientierungshilfe*. Tübingen: Francke, 196–208.
- Kopczynski, Andrzej (1980). Conference interpreting. Some linguistic and communicative problems. *Uniwersytet Imienia Adama Mickiewicza w Poznaniu. Seria Filologia Angielska Poznan* 13, 1–140.
- Kopczynsky, Andrew (1994). Quality in conference interpreting: some pragmatic problems. In: Lambert, Silvie & Moser-Mercer, Barbara (eds) *Bridging the Gap. Empirical Research in Simultaneous Interpretation*. Amsterdam: John Benjamins, 87–99.
- Kurz, Ingrid (1989). Conference Interpreting—User Expectations. In: Hammond, Deanna L. (ed) *Coming of Age. Proceedings of the 30th Annual Conference of the American Translators Association*. Medford (NJ): Learned Information, 143–148.
- Kurz, Ingrid (1993). Conference Interpretation: Expectations of Different User Groups. *The Interpreters' Newsletter* 5, 13–21.
- Kurz, Ingrid (1998). Mediendolmetschen. In: Snell-Hornby, Mary; Hönig, Hans G.; Kußmaul, Paul & Schmitt, Peter A. (eds) *Handbuch Translation*. Tübingen: Stauffenburg, 311–312.
- Kurz, Ingrid (2001). Conference interpreting: quality in the ears of the user. *Meta* 46 (2), 394–409.
- Kurz, Ingrid (2003). Quality from the user perspective. In: Collados Aís, Ángela; Fernández Sánchez, María Manuela & Gile, Daniel (eds) *La evaluación de la Calidad en interpretación: investigación*. Granada: Comares, 3–22.
- Kutz, Wladimir (2010). *Was muss der Dolmetscher wissen und können*. München: Europäischer Verlag.
- Ladd, David Robert (1983). Peak features and overall slope. In: Cutler, Anne & Ladd, D Robert (eds) *Prosody: models and measurements*. Berlin: Springer 39–52.

- Laver, John & Trudgill, Peter (1979). Phonetic and linguistic markers in speech. In: Scherer, Klaus Rainer & Giles, Howard (eds) *Social markers in speech*. Cambridge: Cambridge University Press, 1–32.
- Lieberman, Philip (1966). *Intonation, perception, and language*. Dissertation, Massachusetts Institute of Technology.
- Mack, Gabriele (2002). New Perspectives and Challenges for Interpretation: The Example of Television. In: Garzone & Viezzi (eds), 203–213.
- Mariel, Petr; Hoyos, David; Meyerhoff, Jürgen; Czajkowski, Mikolaj; Dekker, Thijs; Glenk, Klaus; Jacobsen, Jette Bredahl; Liebe, Ulf; Olsen, Søren Bøye; Sagebiel, Julian & Thiene, Mara (2021). *Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis*. Berlin: Springer Nature.
- Marrone, Stefano (1993). Quality: a shared objective. *The Interpreters' Newsletter* 5, 35–41.
- Mellinger, Christopher D & Hanson, Thomas A (2017). *Quantitative Research Methods in Translation and Interpreting Studies*. London: Routledge.
- Moser, Peter (1996). Expectations of users of conference interpretation *Interpreting* 1 (2), 145–178.
- Möbius, Bernd (1993). *Ein quantitatives Modell der deutschen Intonation: Analyse und Synthese von Grundfrequenzverläufen*. Tübingen: Niemeyer.
- Peters, Jörg (2021). *Intonation*. 3. Aufl. Heidelberg: Universitätsverlag Winter.
- Phelan, Mary (2019). A Brief History of Interpreting. In: *The Interpreter's Resource*. Bristol, Blue Ridge Summit: Multilingual Matters, 1–5.
- Pierrehumbert, Janet (1980). *The phonology and phonetics of English intonation*. Dissertation, Indiana University Linguistics Club.
- Pöchhacker, Franz (1994). *Simultandolmetschen als komplexes Handeln*. Tübingen: Gunter Narr.
- Pöchhacker, Franz (2001). Quality Assessment in Conference and Community Interpreting. *Meta* 46 (2), 410–425.
- Pöchhacker, Franz (2002). Researching interpreting quality. In: Garzone & Viezzi (eds), 95–106.
- Pöchhacker, Franz (2010a). Interpreting. In: Van Doorslaer & Gambier (eds), 153–157.
- Pöchhacker, Franz (2010b). Media interpreting. In: Van Doorslaer & Gambier (eds), 224–226.
- Pöchhacker, Franz (2016). *Introducing interpreting studies*. 2nd ed. London/New York: Routledge.
- Pöchhacker, Franz & Zwischenberger, Cornelia (2010). *Survey on quality and role:*

- conference interpreters' expectations and self-perceptions*. AIIC.
<https://aiic.org/document/9646/Survey%20on%20quality%20and%20role%20conference%20interpreters%E2%80%99%20expectations%20and%20self-perceptions.pdf>
 (Stand: 09.07.2025).
- Poyatos, Fernando (1993). *Paralanguage: a linguistic and interdisciplinary approach to interactive speech and sound*. Amsterdam: John Benjamins.
- Poyatos, Fernando (1997). The reality of multichannel verbal-nonverbal communication in simultaneous and consecutive interpretation. *Benjamins Translation Library* 17, 249–282.
- PRAAT (o.J.). Intro 4.2. Configuring the pitch contour (Stand: 10.01.2026).
- Reed, Beatrice Szczepek (2010). *Analysing Conversation: An Introduction to Prosody*. London: Bloomsbury Academic.
- Rennert, Sylvi (2010). The impact of fluency on the subjective assessment of interpreting quality. *The Interpreters' Newsletter* 15, 101–115.
- Setton, Robin (2010). Conference interpreting. In: Van Doorslaer & Gambier (eds), 66–74.
- Selbmann, Rolf (1994). *Der deutsche Bildungsroman*. 2. Aufl. Stuttgart [u.a.]: Metzler.
- Seleskovitch, Danica (1986). Comment: Who should assess an interpreter's performance? *Multilingua* 5 (4), 236.
- Shlesinger, Miriam (1989). Extending the Theory of Translation to Interpretation: Norm as a Case in Point. *Target* 1, 111–115.
- Shlesinger, Miriam (1994). Intonation in the production and perception of simultaneous interpretation. In: Lambert, Sylvie & Moser-Mercer, Barbara (eds) *Bridging the gap. Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 225–236.
- Shlesinger, Miriam (1997). Quality in Simultaneous Interpreting. In: Gambier, Yves; Gile, Daniel & Taylor, Christopher (eds) *Conference Interpreting: Current Trends in Research*. Amsterdam/Philadelphia: John Benjamins, 123–131.
- Sue, Valerie M & Ritter, Lois A (2016). *Conducting Online Surveys*. 2nd ed. Los Angeles: Sage.
- Swales, Martin (1991). Trony and the Novel. In: Hardin, James N (ed) *Reflection and Action: Essays on the Bildungsroman*. Columbia, S.C.: University of South Carolina Press, 46–68.
- Thorsen, Nina Grønnum (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America* 77 (3), 1205–1216.

- Thorsen, Nina Grønnum (1988). Standard Danish intonation. Annual Report of the Institute of Phonetics. University of Copenhagen, *ARIPUC* 22, 1–23.
- Tissi, Benedetta (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreters' Newsletter* 10, 103–127.
- Van Doorslaer, Luc & Gambier, Yves (eds) (2010). *Handbook of Translation Studies*. Amsterdam/Philadelphia: John Benjamins.
- Viaggio, Sergio (1992). Translators and interpreters: professionals or shoemakers? In: Dollerup, Cay & Loddegaard, Anne (eds) *Teaching translation and interpreting: training, talent and experience. Papers from the first language international conference, Elsinore, Denmark, 31 May – 2 June 1991*. Amsterdam: Benjamins, 307–314.
- Vaissière, Jacqueline (2005). Perception of intonation. In: Pisoni, David Blair & Remez, Robert Ellis (eds) *The handbook of speech perception*. Malden, Mass: Blackwell, 236–263.
- Viezzi, Mauro (1996). *Aspetti della qualità in interpretazione*. Trieste: Scuola Superiore di Lingue Moderne per Interpreti e Traduttori.
- Williams, Sarah (1995). Observation of anomalous stress in interpreting. *The Translator* 1, 47–64.
- Zwischenberger, Cornelia (2010). Quality criteria in simultaneous interpreting: An international vs. a national view. *The Interpreters' Newsletter* 15, 127–142.
- Zwischenberger, Cornelia (2013). *Qualität und Rollenbilder beim simultanen Konferenzdolmetschen*. Berlin: Frank & Timme.
- Zwischenberger, Cornelia (2015). Survey Research. In: Pöchhacker, Franz (ed) *Routledge encyclopedia of interpreting studies*. London: Routledge, 407–408.

Anhang

Anhang I: Ausgangsrede

Meine sehr geehrten Damen und Herren,

es ist mir eine große Freude, heute bei Ihnen sein zu dürfen. Dem Österreichische Kulturforum Rom möchte ich für die Einladung herzlich danken. Eure Arbeit zur Förderung der österreichischen Kultur in Italien ist auch mir persönlich sehr wichtig. Sie trägt zum Austausch bei, und zur Vertiefung der Beziehung zwischen unseren zwei Ländern, die so untrennbar miteinander verbunden sind. Jedes Mal, wenn ich nach Italien komme, fühle ich mich ein Stück weit wie zuhause. Und ich verspreche Ihnen: Eines Tages werde ich ganz sicher Ihre so schöne Sprache lernen. Heute jedoch muss ich noch die Unterstützung einer Dolmetscherin in Anspruch nehmen, und dafür möchte ich mich nochmal ganz herzlich bedanken. *Grazie*. Für die, die mich nicht kennen: Mein Name ist Roland Pohl und ich bin Schriftsteller. Heute möchte ich Ihnen mein neuestens Buch vorstellen. Es erscheint bald im „Prima“ Verlag, und ist wahrscheinlich das Werk, mit dem ich persönlich meine zweite Heimat Italien verbinde. Dieses besondere Verhältnis zu Italien zeigt sich zuallererst mal am Ort des Geschehens: Die Geschichte spielt in Rasiglia, in Umbrien, das „Das Dorf der Bäche“ genannt wird. Vor knapp 10 Jahren war ich mit meiner Familie das erste Mal dort, und habe diesen Ort sofort ins Herz geschlossen und wusste, dass ich ihn irgendwann zu einem Schauplatz eines späteren Buches machen muss. Oft werde ich gefragt, ob es sich hierbei um einen Jugendroman handelt. Ich persönlich bezeichne ihn lieber als „Bildungsroman“. Denn auch wenn Marco, der Bursch um den es sich in der Geschichte geht, selbst ein Jugendlicher ist, richtet sich das Buch keineswegs nur an jugendliche Leser. Ganz im Gegenteil: Ich empfehle es gerade Eltern, die vielleicht die Welt aus Sicht ihrer Kinder besser verstehen wollen, deren Sorgen und die Hoffnungen in einer moderneren Welt, als sie es selbst noch von früher kennen, zu verstehen. Und auch wenn ich selbst noch selbst keine Kinder habe, ist meine Jugend noch nicht lange her. Ich hoffe doch, dass es mir deshalb ganz gut gelungen ist, einen ehrlichen, einen direkten Einblick in diese Welt zu geben, sozusagen aus erster Hand. Der Titel meines Romans lautet „Jenseits des Flusses“. An dieser Stelle möchte ich Ihnen einen kurzen Einblick geben, und eine Stelle, ganz vom Anfang vorlesen, die auch tatsächlich der erste Satz war, den ich von „Jenseits des Flusses“, vor 10 Jahren selbst in Rasiglia geschrieben habe:

„Manchmal spürte Marco, wie das Rauschen des Flusses lauter war als die Stimmen seiner Familie. In diesen Momenten wusste er, dass irgendwo dahinter ein anderes Leben auf ihn wartete – und dass er eines Tages den Mut finden musste, es selbst zu suchen.“

Marco gerät in einem Streit mit seinem Vater über seine Zukunft, der Vater möchte, dass Marco den Hof seines Vaters übernimmt, und ein sicheres Leben hat. Marco, für dein diesen Schritt der absolute Albtraum wäre, bricht von zu Hause aus, und flüchtet in die Großstadt, nicht nur, um einen neuen Ort zu entdecken, sondern auch, um sich selbst kennenzulernen. Seine Reise ist damit sowohl eine äußere als auch eine innere. Er begegnet Einsamkeit, das erste Mal in seinem Leben Unsicherheit, Rückschläge, aber er knüpft auch neue Freundschaften, die ihm eine andere Sicht auf die Welt eröffnen. Besonders prägend ist die Begegnung mit Sofia, einer jungen Malerin, die entschlossen ihren Weg geht, trotz aller Hindernisse. Durch sie stellt Marco sich die Frage: Was bedeutet es eigentlich, seiner Leidenschaft zu folgen, Und wie viel Mut es braucht, den eigenen Träumen selbst zu folgen? Über Umwege, Irrtümer und neue Einsichten wächst Marco an seinen Erfahrungen. Schließlich kehrt er in sein Dorf zurück – als ein junger Mann, der gelernt hat, seine eigenen Wünsche vor die Erwartungen anderer zu stellen. Wie die Begegnung mit dem Vater abläuft, und ob doch noch mehr passiert, dafür müsst ihr wohl das Buch lesen.

So, in diesen sinne möchte ich mich für die Aufmerksamkeit bedanken und freu mich auf ihre fragen.

Mille grazie.

Anhang II: Zieltext

Signore e signori,

sono molto lieto poter, di poter essere qui da voi, al Forum Austriaco di Cultura di Roma, che ringrazio per l'invito. Il vostro lavoro per la promozione della cultura austriaca in Italia mi sta molto a cuore perché contribuisce allo scambio e all'approfondimento delle relazioni fra i nostri due paesi, che sono così indivisibili. Ogni volta che vengo in Italia mi sento proprio come a casa e vi prometto, un giorno imparerò la vostra lingua così bella. Purtroppo però oggi devo utilizzare un'interprete, che ringrazio. Grazie. Per coloro che non mi conoscono, il mio nome è Roland Pohl, sono uno scrittore e oggi vorrei presentarvi il mio il mio ultimo libro che uscirà nelle nella casa editrice Prima, ed è l'opera con cui io, che io collego al mio alla mia seconda

casa, l'Italia. La storia avviene a Rasiglia, in Umbria, ovvero il borgo dei ruscelli. Quasi 10 anni fa sono stato lì per la prima volta e questo posto mi è subito entrato nel cuore e sapevo che prima o poi sarebbe diventato il teatro della mia opera.

Molto spesso mi chiedono se si tratta di un romanzo per ragazzi. Personalmente io lo descriverei come un romanzo di formazione, perché anche se Marco, il ragazzo protagonista della storia, è un adolescente, il libro non, non è non sì, non è dedicato solo ai ragazzi, proprio il contrario. Io lo consiglio proprio ai genitori che vogliono capire meglio il mondo dei loro figli e le speranze, le preoccupazioni del mondo moderno, di un mondo moderno, diverso dal loro. E anche se io non ho figli, la mia gioventù non è passata da molto tempo e quindi spero che mi sia riuscito bene darvi una visione sincera di questo mondo in prima persona. Il titolo del romanzo è “Al di là del fiume” e quindi vorrei leggersi un piccolo passaggio che ho scritto 10 anni fa, proprio a Rasiglia.

“A volte Marco sentiva che il rumore del fiume era più forte delle voci della sua famiglia. In quei momenti, sapeva che da qualche parte là fuori lo aspettava un'altra vita, e che un giorno avrebbe dovuto trovare il coraggio di rincorrerla”

Marco inceppa in un litigio con suo padre sul suo futuro. Il padre vorrebbe che lui prenda in mano la fattoria di famiglia e Marco, per cui questo passo sarebbe un incubo, scappa in una grande città e non solo per trovare un nuovo luogo, ma per trovare proprio sé stesso. Il suo viaggio è un viaggio esterno e interno. Incontra la solitudine, per la prima volta nella sua vita, insicurezza, le disavventure, ma conosce anche nuovi amici, che gli danno nuovi, nuove visioni sulla vita. Un'importante incontro è quello con Sofia, una pittrice giovane, che segue la sua strada nonostante gli ostacoli. Attraverso lei, Marco si chiede che cosa significhi seguire la propria passione, e quanto coraggio ci voglia per seguire i propri sogni. Per vie traverse, errori e nuovi punti di vista, Marco riesce dalle proprie, dalle sue esperienze, torna al suo paese come un giovane uomo che ha imparato mettere di mettere i propri sogni davanti alle aspettative. Come sarà l'incontro con il padre e se capiterà altro, ecco, questo lo scoprirete nel libro.

In questo senso vorrei ringraziarvi per l'attenzione e sono lieto di ascoltare le vostre domande.

Grazie mille.

Anhang III: Fragebogen

Sezione 1 di 16:

Benvenuto!

Grazie per dedicare un po' del tuo tempo al mio studio. La compilazione del questionario richiederà circa 10 minuti.

Mi chiamo Lorenza e la mia tesi di laurea magistrale presso l'Università di Vienna ha come oggetto di studio la qualità dell'interpretazione simultanea.

Quest'ultima consiste nel trasferire oralmente un discorso da una lingua a un'altra in tempo reale.

Ti chiederò di valutare una performance di interpretariato dal tedesco all'italiano. Il contesto è fittizio: si tratta di una conferenza sulla letteratura austriaca, dedicata alla presentazione di un libro inventato e attribuito a un autore immaginario.

Le tue risposte resteranno anonime :)

Sezione 2 di 16:

L'italiano è la tua lingua madre?

☐ Sì

☐ No

Qualora sia stata selezionata l'opzione "No":

Sezione 3 di 16:

Qual è il tuo livello di conoscenza dell'italiano?

☐ Base (A1–A2)

☐ Intermedio (B1–B2)

☐ Avanzato (C1–C2)

Qualora siano state selezionate le opzioni "base" o "intermedio":

Sezione 4 di 16:

Grazie per il tuo interesse! Purtroppo, secondo le risposte fornite, non soddisfi i criteri necessari per partecipare a questo studio. Ti ringrazio comunque per il tempo dedicato.

Sezione 5 di 16:

Qual è il tuo livello di conoscenza del tedesco?

- ☐ Nessuno
- ☐ Base (A1–A2)
- ☐ Intermedio (B1–B2)
- ☐ Avanzato (C1–C2)

Sezione 6 di 16:

Hai esperienza diretta nel campo dell'interpretariato? (ad esempio formazione o lavoro come interprete)

- ☐ Sì
- ☐ No

Qualora sia stata selezionata l'opzione "Sì":

Sezione 7 di 16:

Grazie per il tuo interesse! Purtroppo, secondo le risposte fornite, non soddisfi i criteri necessari per partecipare a questo studio. Ti ringrazio comunque per il tempo dedicato.

Sezione 8 di 16:

Hai mai visto un programma in TV in cui è stata effettuata un'interpretazione simultanea? Ad esempio, come a "Sanremo" o a "Che tempo che fa", quando gli ospiti parlano una lingua straniera e l'audio viene interpretato in italiano per renderlo comprensibile al pubblico.

- ☐ Sì
- ☐ No

Qualora sia stata selezionata l'opzione "No":

Sezione 9 di 16:

Grazie per il tuo interesse! Purtroppo, secondo le risposte fornite, non soddisfi i criteri necessari per partecipare a questo studio. Ti ringrazio comunque per il tempo dedicato.

Sezione 10 di 16:

Hai mai partecipato come pubblico a una conferenza con interpretariato simultaneo?

- ☐ Sì
- ☐ No

Qualora sia stata selezionata l'opzione "Sì":

Sezione 11 di 16:

Se sì, quanto spesso? (inserisci il numero in cifre, es. 3)

Circa...volte

Sezione 12 di 16:

Ora stai per visionare un video in tedesco interpretato in italiano. Al termine, ti verrà chiesto di valutare l'interpretazione su diversi aspetti:

- Valutazione complessiva: quanto ritieni buona l'interpretazione nel complesso
- Scorrevolezza: quanto scorrevole e naturale ti è sembrata
- Intonazione: quanto vivace ed espressiva ti è sembrata la voce dell'interprete
- Professionalità: quanto professionale ti è sembrata la performance interpretativa
- Coerenza: quanto chiaro e coerente ti è sembrato il discorso

Sezione 13 di 16:

VISIONE DEL VIDEO (A o B)

Sezione 14 di 16:

Valutazione dell'interpretazione

Valuta l'interpretazione appena ascoltata usando una scala da 1 (valore più basso) a 5 (valore più alto) per ciascun criterio

Valutazione complessiva

- ☐ 1 – per niente buona
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – molto buona

Scorrevolezza

- ☐ 1 – per niente scorrevole
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – molto scorrevole

Intonazione

- ☐ 1 – monotona
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – vivace

Coerenza

- ☐ 1 – per niente coerente
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – molto coerente

Professionalità

- ☐ 1 – per niente professionale
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – Molto professionale

Sezione 15 di 16:

Ti chiedo solo ancora due informazioni anagrafiche:

la tua età (inserisci il numero in cifre, es. 25)

...anni

Il tuo sesso:

- ☐ maschile
- ☐ femminile
- ☐ altro
- ☐ preferisco non specificare

Sezione 16 di 16:

Grazie per aver partecipato! Le tue risposte sono preziose per il mio studio.

Anhang IV: Übersetzung des Fragebogens

Abschnitt 1 von 16:

Willkommen!

Vielen Dank, dass du dir die Zeit nimmst, an meiner Studie teilzunehmen. Das Ausfüllen des Fragebogens dauert ungefähr 10 Minuten.

Ich heiße Lorenza und untersuche in meiner Masterarbeit an der Universität Wien die Qualität des Simultandolmetschens. Simultandolmetschen bedeutet, einen Vortrag in Echtzeit mündlich von einer Sprache in eine andere zu übertragen.

Ich werde dich bitten, eine Dolmetschleistung aus dem Deutschen ins Italienische zu bewerten. Der Kontext ist fiktiv – eine Konferenz über österreichische Literatur, bei der ein erfundenes Buch eines imaginären Autors vorgestellt wird.

Deine Antworten bleiben selbstverständlich anonym :)

Abschnitt 2 von 16:

Ist Italienisch deine Muttersprache?

- ☐ Ja
- ☐ Nein

Falls die Antwort „Nein“ lautet:

Abschnitt 3 von 16:

Wie gut sind deine Italienischkenntnisse?

- ☐ Grundkenntnisse (A1–A2)
- ☐ fortgeschrittene Kenntnisse (B1–B2)
- ☐ sehr gute Kenntnisse (C1–C2)

Falls „Grundkenntnisse“ oder „fortgeschrittene Kenntnisse“ ausgewählt wurde:

Abschnitt 4 von 16:

Vielen Dank für dein Interesse! Leider erfüllst du aufgrund deiner Antworten nicht die Voraussetzungen für die Teilnahme an dieser Studie. Trotzdem vielen Dank für deine Zeit.

Abschnitt 5 von 16:

Wie gut sind deine Deutschkenntnisse?

- ☐ Keine
- ☐ Grundkenntnisse (A1–A2)
- ☐ fortgeschrittene Kenntnisse (B1–B2)
- ☐ sehr gute Kenntnisse (C1–C2)

Abschnitt 6 von 16:

Hast du aktive Erfahrungen im Bereich Dolmetschen? (z. B. Ausbildung oder Beruf als Dolmetscher/in)

- ☐ Ja
- ☐ Nein

Falls „Ja“ ausgewählt wurde:

Abschnitt 7 von 16:

Vielen Dank für dein Interesse! Leider erfüllst du aufgrund deiner Antworten nicht die Voraussetzungen für die Teilnahme an dieser Studie. Trotzdem vielen Dank für deine Zeit.

Abschnitt 8 von 16:

Hast du schon einmal im Fernsehen eine Sendung gesehen, in der simultan gedolmetscht wurde? Beispielsweise bei Sendungen wie „Sanremo“ oder „Che tempo che fa“, wenn Gäste in einer Fremdsprache sprechen und dies für das Publikum ins Italienische gedolmetscht wird.

- ☐ Ja
- ☐ Nein

Falls „Nein“ ausgewählt wurde:

Abschnitt 9 von 16:

Vielen Dank für dein Interesse! Leider erfüllst du aufgrund deiner Antworten nicht die Voraussetzungen für die Teilnahme an dieser Studie. Trotzdem vielen Dank für deine Zeit.

Abschnitt 10 von 16:

Hast du schon einmal an einer Konferenz mit Simultandolmetschen vor Ort teilgenommen?

- ☐ Ja
- ☐ Nein

Falls „Ja“ ausgewählt wurde:

Abschnitt 11 von 16:

Wenn ja, wie häufig? (Bitte Zahl eingeben, z. B. 3)

ca. ... Mal

Abschnitt 12 von 16:

Nun wirst du ein Video auf Deutsch sehen, das ins Italienische gedolmetscht wurde.

Am Ende wirst du gebeten, die Dolmetschleistung anhand verschiedener Kriterien zu bewerten:

- Gesamtbewertung: wie gut findest du die Verdolmetschung insgesamt?
- Flüssigkeit: wie flüssig erschien dir die Verdolmetschung?
- Intonation: wie lebendig und ausdrucksstark erschien dir die Stimme der Dolmetscherin?
- Professionalität: wie professionell wirkte die Dolmetschleistung?
- Kohärenz: wie klar und kohärent erschien dir der Vortrag?

Abschnitt 13 von 16:

VIDEO ANSEHEN (A oder B)

Abschnitt 14 von 16:

Bewertung der Dolmetschleistung

Bewerte die gerade gehörte Verdolmetschung auf einer Skala von 1 (geringster Wert) bis 5 (höchster Wert) für jedes Kriterium:

Gesamtbewertung

- ☐ 1 – gar nicht gut
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – sehr gut

Flüssigkeit

- ☐ 1 – gar nicht flüssig
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – sehr flüssig

Intonation

- ☐ 1 – monoton
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – lebendig

Kohärenz

- ☐ 1 – gar nicht kohärent
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – sehr kohärent

Professionalität

- ☐ 1 – gar nicht professionell
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – sehr professionell

Abschnitt 15 von 16:

Nun bitte noch zwei demographische Angaben:

Dein Alter (bitte als Zahl eingeben, z. B. 25)

... Jahre

Dein Geschlecht:

- ☐ männlich
- ☐ weiblich
- ☐ divers
- ☐ keine Angabe

Abschnitt 16 von 16:

Vielen Dank für deine Teilnahme! Deine Antworten sind für meine Studie von großem Wert!

Abstract

Die vorliegende Arbeit widmet sich der Qualitätsbewertung beim Simultandolmetschen aus der Perspektive der Nutzer*innen, wobei das Augenmerk hauptsächlich auf den Parameter der Intonation gelegt wird. Die Studie versteht sich als Teilreplikation der Studie von Collados Aís (1998), die gezeigt hat, dass die Intonation eine zentrale Rolle bei der Bewertung der Dolmetschleistung spielt. In ihrer Studie wurden drei verschiedene Dolmetschversionen auf der Grundlage eines Ausgangstextes erstellt, der durch eine monotone Intonation gekennzeichnet war. Dabei wurden Intonation und Inhalt gezielt manipuliert: Eine Version wies eine monotone Intonation auf und enthielt keine inhaltlichen Abweichungen vom Ausgangstext. Eine weitere wurde nicht monoton vorgetragen, wies jedoch inhaltliche Fehler auf. Die dritte war eine nicht monotone Kontrollversion ohne inhaltliche Fehler. Jede Version wurde von einer unterschiedlichen Gruppe von Nutzer*innen bewertet. Die Ergebnisse zeigten, dass die monotone Version, die inhaltlich jedoch korrekt war, schlechter bewertet wurde als die nicht monotone Version mit inhaltlichen Fehlern, was die Bedeutung der Intonation im Bereich des Dolmetschens unterstreicht.

Die vorliegende Studie zielt darauf ab, die Hypothese von Collados Aís (1998) erneut zu bestätigen, wonach eine monotone Intonation die Bewertung der Dolmetschleistung sowie die Wahrnehmung der Professionalität der Dolmetscherin negativ beeinflusst. Zudem wird, ausgehend von einer Forschungsanregung von Collados Aís (1998), ein weiteres Ziel formuliert, nämlich zu untersuchen, ob eine monotone Intonation bei einem lebhaft intonierten Ausgangstext als noch stärker nachteilig wahrgenommen wird als in der Studie von 1998.

In Anlehnung an die Methodik von Collados Aís (1998) wurden zwei Dolmetschversionen (eine monotone und eine nicht monotone) auf der Grundlage eines Ausgangstextes mit lebhafter Intonation erstellt und von zwei Gruppen von Nutzer*innen mittels eines Online-Fragebogens bewertet. Die Ergebnisse bestätigen erneut die Hypothese von Collados Aís (1998): Die Teilnehmenden bewerteten die monotone Version in allen Kriterien signifikant schlechter als die lebhafte Version. Darüber hinaus scheint sich auch die Hypothese zu bestätigen, dass die monotone Version bei einem lebhaften Ausgangstext noch stärker nachteilig wahrgenommen wird als in der Studie von Collados Aís (1998). Die Mittelwertdifferenz zwischen der monotonen und der lebhaften Version ist in der vorliegenden Studie in der Mehrheit der Kriterien größer als in der Studie von Collados Aís (1998). Dies könnte darauf zurückzuführen sein, dass die Wahrnehmung von Monotonie deutlicher hervortritt, wenn der Ausgangstext eine lebhafte Intonation aufweist. Eine weitere mögliche

Erklärung ist, dass die lebhafte Dolmetschversion der vorliegenden Studie eine stärker ausgeprägte Intonation aufweist als die lebhafte Version in der Studie von Collados Aís (1998), was in der vorliegenden Studie zu einer größeren Mittelwertdifferenz zwischen der monotonen und der lebhaften Version führen kann. Dieser Aspekt lässt sich jedoch mangels der Originaldaten der Studie von Collados Aís (1998) nicht überprüfen.