



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Menschliche und LLM-gestützte literarische Übersetzung im
Vergleich am Beispiel ausgewählter Erzählungen von Anton
Tschechow aus dem Russischen ins Deutsche

verfasst von | submitted by
Ekaterina Djordjevic

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 070 360 331

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

Masterstudium Translation Russisch Deutsch

Betreut von | Supervisor:

Ass.-Prof. Mag. Dr. Waltraud Kolb

Inhaltsverzeichnis

Abbildungsverzeichnis	5
Tabellenverzeichnis	6
1 Einleitung	8
2 Theoretische Grundlagen der Übersetzungsqualität und ihrer Evaluation.....	11
2.1 Der Qualitätsbegriff in der Translationswissenschaft	11
2.2 Skopostheorie.....	14
2.3 Ansätze und Modelle zur Analyse der Qualität maschineller Übersetzungen	16
2.3.1 Automatische Evaluierungsmethoden.....	17
2.3.1.1 Bilingual Evaluation Understudy (BLEU)	17
2.3.1.2 Metric for Evaluation of Translation with Explicit Ordering (METEOR)	17
2.3.1.3 Translation Edit Rate (TER)	18
2.3.2 Menschliche Evaluierungsmethoden	19
2.3.2.1 Multidimensional Quality Metrics (MQM)	19
2.3.2.2 Smart Computer-aided Translation Environment (SCATE)	20
3 Maschinelle Übersetzung	22
3.1 Historische Entwicklung maschineller Übersetzung	22
3.2 Arten der maschinellen Übersetzung	24
3.2.1 Regelbasierte maschinelle Übersetzung	24
3.2.2 Statistische maschinelle Übersetzung	26
3.2.3 Neuronale maschinelle Übersetzung	28
3.2.3.1 LLM-gestützte Übersetzungssysteme	30
3.2.3.1.1 DeepL.....	31
3.2.3.1.2 ChatGPT.....	33
4 Literarische maschinelle Übersetzung	34
4.1 Besonderheiten literarischer maschineller Übersetzung	34
4.2 Herausforderungen der Evaluation literarischer maschineller Übersetzung	37
5 Aktueller Stand der Forschung	39
6 Untersuchungskorpus.....	43
6.1 Ausgewählte Erzählungen von Anton Tschechow	43
6.2 Kurze Erzählung als Untersuchungsgegenstand	44
6.3 Über den Autor	45

6.4 Inhaltliche Charakterisierung der Erzählungen	45
6.4.1 Der Tod eines Beamten	45
6.4.2 Der Dicke und der Dünne	46
6.4.3 Ein Chamäleon	47
7 Methode	48
7.1 Analysegrundlage: erweiterte SCATE-Taxonomie nach Tezcan et al. (2019) und Key- Feature-Analyse nach Webster et al. (2020)	48
7.2 Methodisches Vorgehen	50
7.3 Quantitative Analyse: lexikalische Vielfalt	51
7.4 Qualitative Analyse: lokale Kohäsion, Kohärenz, Stil & Register	53
7.4.1 Lokale Kohäsion	53
7.4.2 Kohärenz	54
7.4.3 Stil & Register	55
8 Übersetzungsanalyse	56
8.1 Quantitative Analyse: Type-Token-Ratio	56
8.2 Lokale Kohäsion	59
8.3 Kohärenz	64
8.4 Stil & Register	69
9 Ergebnisse	75
10 Fazit	78
11 Bibliographie	81
Primärliteratur	81
Sekundärliteratur	82
Anhang	91
Anhang A: Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register zu <i>Der Tod eines Beamten</i> (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a)	91
Anhang 2: Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register zu <i>Der Dicke und der Dünne</i> (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b)	96
Anhang 3: Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register zu <i>Ein Chamäleon</i> (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)	102
Humanübersetzung: <i>Der Tod eines Beamten</i> (Tschechow 2003a)	116
DeepL Pro-Übersetzung: <i>Der Tod eines Beamten</i> (DeepL Pro 2026a)	119
ChatGPT-5.3-Übersetzung: <i>Der Tod eines Beamten</i> (ChatGPT-5.3 2026a)	122
Humanübersetzung: <i>Der Dicke und der Dünne</i> (Tschechow 2023b)	125
DeepL-Pro-Übersetzung: <i>Der Dicke und der Dünne</i> (DeepL Pro 2026b)	127
ChatGPT-5.3-Übersetzung: <i>Der Dicke und der Dünne</i> (ChatGPT-5.3 2026b)	129

Humanübersetzung: <i>Ein Chamäleon</i> (Tschekow 2023c).....	131
DeepL-Pro-Übersetzung: <i>Ein Chamäleon</i> (DeepL Pro 2026c).....	135
ChatGPT-5.3-Übersetzung: <i>Ein Chamäleon</i> (ChatGPT-5.3 2026c)	139
Abstract.....	142

Abbildungsverzeichnis

Abb. 1: Die erweiterte SCATE-Fehlerannotation (Tezcan et al. 2019).....	49
---	----

Tabellenverzeichnis

Tab. 1: Type-Token-Ratio der untersuchten Übersetzungen von Der Tod eines Beamten	57
Tab. 2: Type-Token-Ratio der untersuchten Übersetzungen von Der Dicke und der Dünne ..	57
Tab. 3: Type-Token-Ratio der untersuchten Übersetzungen von Ein Chamäleon	58
Tab. 4: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	59
Tab. 5: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	61
Tab. 6: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus <i>Ein Chamäleon</i> (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	61
Tab. 7: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Der Tod eines Beamten (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a).....	64
Tab. 8: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Der Dicke und der Dünne (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b).....	65
Tab. 9: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	65
Tab. 10: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	66
Tab. 11: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	67
Tab. 12: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	67
Tab. 13: Übersicht der in die finale Auswertung aufgenommenen Kohärenzfehler nach Unterkategorie und Übersetzungsvariante	68
Tab. 14: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Der Tod eines Beamten (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a).....	70

Tab. 15: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Der Tod eines Beamten (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a).....	70
Tab. 16: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Der Dicke und der Dünne (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b).....	71
Tab. 17: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Der Dicke und der Dünne (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b).....	71
Tab. 18: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	72
Tab. 19: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	73
Tab. 20: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c).....	73
Tab. 21: Übersicht der in die finale Auswertung aufgenommenen Fehler im Bereich Stil & Register nach Unterkategorie und Übersetzungsvariante	74

1 Einleitung

Die Entwicklung von Large Language Models (LLMs) hat den Bereich der maschinellen Übersetzung (MÜ) in den letzten Jahren wesentlich geprägt und neue Perspektiven für die Übersetzung eröffnet (Gao et al. 2024: 2). Auch im Bereich der literarischen Übersetzung rücken LLMs zunehmend in den Fokus der Forschung (Zhang et al. 2025b: 29099). Obwohl sich ihre künftige Rolle in diesem Bereich noch nicht eindeutig prognostizieren lässt, gewinnen LLM-gestützte Übersetzungssysteme in der literarischen Übersetzung weiter an Bedeutung (Walter 2026: 32). Zugleich macht die aktuelle Forschung deutlich, dass LLM-gestützte literarische Übersetzung kritisch bewertet werden muss. Zhang et al. (2025a: 10961) verweisen darauf, dass eine Überschätzung der Qualität LLM-generierter Übersetzungen sowohl berufliche Folgen für professionelle Übersetzer*innen als auch Auswirkungen auf ästhetische Aspekte literarischer Übersetzungen haben kann.

In diesem Zusammenhang untersucht die vorliegende Arbeit literarische maschinelle Übersetzung im Kontext LLM-gestützter Übersetzungstechnologien. Die Forschungsfrage lautet: In welcher Hinsicht unterscheiden sich die LLM-gestützten Übersetzungen ausgewählter literarischer Texte von menschlichen Übersetzungen aus dem Russischen ins Deutsche? Aus dieser Forschungsfrage werden vier Annahmen abgeleitet, die im Folgenden dargestellt werden.

Die erste Annahme betrifft die lexikalische Vielfalt und geht davon aus, dass LLM-gestützte Übersetzungen eine geringere lexikalische Vielfalt als Humanübersetzungen aufweisen. Ausgangspunkt dafür bilden Ergebnisse aus der Forschung zur literarischen maschinellen Übersetzung. Webster et al. (2020: 12) stellen anhand eines Korpus aus vier Romanen und durchgeführter Analyse fest, dass die Humanübersetzungen eine größere Zahl unterschiedlicher Wörter als die neuronalen maschinellen Übersetzungen enthalten.

Die zweite Annahme bezieht sich auf die lokale Kohäsion. Ausgehend von dieser Annahme zeigen LLM-gestützte Übersetzungen eine geringere lokale Kohäsion im Vergleich zu menschlichen Übersetzungen. Diese Annahme stützt sich auf die Studie von Webster et al. (2020: 14f.), die im Bereich literarischer neuronaler maschineller Übersetzung für maschinelle Übersetzungen niedrigere Kohäsionswerte als für Humanübersetzungen feststellen.

Die dritte Annahme betrifft die Kohärenz und geht davon aus, dass LLM-gestützte Übersetzungen im Vergleich zu menschlichen Übersetzungen weniger kohärent sind. Tezcan et al. (2019:45) betonen die Relevanz von Kohärenz für die Untersuchung literarischer maschineller Übersetzung und stellen im Rahmen ihrer Studie fest, dass kohärenzbezogene Probleme einen

der zentralen Fehlerbereiche darstellen. Auch Fonteyne et al. (2020: 3795) kommen zu einem vergleichbaren Ergebnis. In ihrer Fehleranalyse einer neuronalen maschinellen Übersetzung eines Romans gehört Kohärenz zu dem am stärksten vertretenen Fehlerbereichen.

Die vierte Annahme bezieht sich auf Stil & Register. Demnach weisen LLM-gestützte Übersetzungen im Vergleich zu Humanübersetzungen Schwächen bei stilistischen Nuancen und Registerwahl auf. Auch diese Annahme stützt sich auf die Ergebnisse aus der Forschung zur literarischen maschinellen Übersetzung. Tezcan et al. (2019: 45) zeigen anhand ihrer Analyseergebnisse, dass Stil & Register nach Mistranslation und Kohärenz den dritthäufigsten Fehlerbereich bildet. Stil- und registerbezogene Probleme betreffen dabei vor allem Formulierungen, die in der Zielsprache nicht flüssig wirken. Dies weist darauf hin, dass Stil & Register für neuronale maschinelle Übersetzung einen signifikanten Problembereich darstellen können. Auch die Ergebnisse von Fonteyne et al. (2020: 3796) sprechen dafür, Stil & Register als wichtigen Analysebereich für literarische maschinelle Übersetzung zu berücksichtigen. In ihrer Analyse im Bereich der literarischen maschinellen Übersetzung zählt Stil & Register zu den häufigsten Fehlerbereichen.

Die genannten Studien untersuchen keine LLM-gestützten Übersetzungen, sondern verschiedene Formen maschineller Übersetzung. Ihre Ergebnisse bilden dennoch einen wichtigen Ausgangspunkt für die vorliegende Arbeit, da sie zeigen, dass lexikalische Vielfalt, lokale Kohäsion, Kohärenz sowie Stil & Register zentrale Problembereiche bei der maschinellen Übersetzung literarischer Texte darstellen können. In diesem Zusammenhang wird überprüft, ob sich vergleichbare Tendenzen auch bei LLM-gestützten literarischen Übersetzungen im untersuchten Korpus zeigen.

Das Ziel der vorliegenden Arbeit besteht darin, die aus der Forschungsfrage abgeleiteten Annahmen anhand eines literarischen Korpus zu überprüfen. Gegenstand der Arbeit ist die literarische maschinelle Übersetzung im Kontext LLM-gestützter Übersetzungstechnologien. Als Untersuchungsmaterial dienen drei kurze Erzählungen von Anton Tschechow: *Смерть чиновника* (Tschechow 1950a), *Толстый и тонкий* (Tschechow 1950b) und *Хамелеон* (Tschechow 1950c). Das Korpus umfasst neben den Erzählungen auf Russisch jeweils eine Humanübersetzung pro Erzählung: *Der Tod eines Beamten* (Tschechow 2003a), *Der Dicke und der Dünne* (Tschechow 2003b) und *Ein Chamäleon* (Tschechow 2003c). Darüber hinaus werden für jede Erzählung je eine mit DeepL Pro und eine mit ChatGPT-5.3 Instant erstellte Übersetzungsvariante herangezogen. Für ChatGPT-5.3 Instant wird im weiteren Verlauf der Arbeit die

Kurzform ChatGPT-5.3 verwendet. Aufgrund des Erscheinungszeitpunkts unterliegen die herangezogenen Erzählungen keinem urheberrechtlichen Schutz mehr. Ihre Verwendung als Ausgangstexte für die im Rahmen dieser Arbeit erstellten LLM-gestützten Übersetzungen ist daher urheberrechtlich unproblematisch (UrhG 1936: § 60 Abs. 1).

Zur Überprüfung der Annahmen wird ein kombinierter Analyseansatz angewendet, der quantitative und qualitative Methoden verbindet. Der Ansatz orientiert sich an der bei Webster et al. (2020) vorgenommenen Kombination aus Fehlerannotation auf Grundlage der erweiterten SCATE-Taxonomie nach Tezcan et al. (2019) und Key-Feature-Analyse nach Webster et al. (2020) und wird an die Fragestellung der vorliegenden Arbeit angepasst. Die lexikalische Vielfalt wird anhand der Type-Token-Ration quantitativ untersucht, während die lokale Kohäsion qualitativ analysiert wird. Für die Bereiche Kohärenz sowie Stil & Register erfolgt eine manuelle Fehlerannotation auf Grundlage der erweiterten SCATE-Taxonomie.

Die Arbeit ist wie folgt aufgebaut: Zunächst werden im Kapitel 2 die theoretischen Grundlagen der Übersetzungsqualität und ihrer Evaluation dargestellt. Dabei wird der Qualitätsbegriff im Rahmen der Translationswissenschaft diskutiert. Anschließend werden die Skopostheorie sowie ausgewählte Ansätze und Modelle zur Bewertung maschineller Übersetzung behandelt, wobei sowohl automatische als auch menschliche Evaluationsverfahren berücksichtigt werden. Das dritte Kapitel behandelt das Thema der maschinellen Übersetzung. Neben der historischen Entwicklung der MÜ werden unterschiedliche Formen maschineller Übersetzung vorgestellt. Darüber hinaus werden DeepL und ChatGPT als für die vorliegende Arbeit zu untersuchende Systeme beschrieben. Dabei werden auch die im Analyseteil verwendeten Versionen DeepL Pro und ChatGPT-5.3 berücksichtigt. Darauf aufbauend widmet sich das Kapitel 4 der literarischen maschinellen Übersetzung. Dabei werden die Besonderheiten literarischer Texte sowie zentrale Herausforderungen bei der Evaluation literarischer maschineller Übersetzung dargestellt. Kapitel 5 schafft einen Überblick über den aktuellen Forschungsstand. Im Mittelpunkt stehen Studien, die sich mit lexikalischer Vielfalt, lokaler Kohäsion, Kohärenz sowie Stil & Register in maschinell erstellten literarischen Übersetzungen befassen. Im Kapitel 6 wird das Untersuchungskorpus vorgestellt. Dazu gehören die ausgewählten Erzählungen von Anton Tschechow, ihre inhaltliche Charakterisierung sowie biographische Angaben zum Autor. Außerdem wird die kurze Erzählung als Untersuchungsgegenstand dargestellt. Anschließend wird im Kapitel 7 das methodische Vorgehen der Untersuchung beschrieben. Dabei werden der kombinierte Analyseansatz sowie für die Analyse relevanten Kategorien lexikalische Vielfalt, lokale Kohäsion, Kohärenz sowie Stil & Register erläutert. Im Kapitel 8

wird die Übersetzungsanalyse präsentiert. Die Analyse der lexikalischen Vielfalt und der lokalen Kohäsion wird vollständig in den entsprechenden Unterkapiteln 8.1 und 8.2 dargestellt. Die Unterkapitel 8.3 und 8.4 behandeln die Kategorien Kohärenz sowie Stil & Register anhand ausgewählter Beispiele aus der Fehlerannotation. Die vollständige Fehlerannotation ist im Anhang dokumentiert. Abschließend werden im Kapitel 9 die Analyseergebnisse zusammengefasst und im Hinblick auf die formulierten Annahmen diskutiert. Darüber hinaus wird ein Ausblick auf mögliche weiterführende Untersuchungen gegeben.

2 Theoretische Grundlagen der Übersetzungsqualität und ihrer Evaluation

2.1 Der Qualitätsbegriff in der Translationswissenschaft

Im Folgenden werden zentrale Auffassungen von Übersetzungsqualität dargestellt, um die Vielfalt der in der Translationswissenschaft vertretenen Qualitätskonzepte beschreiben zu können. Eine ausführliche kritische Gegenüberstellung dieser Ansätze ist im Rahmen der vorliegenden Arbeit nicht beabsichtigt. Vielmehr dient ihre Darstellung der theoretischen Einordnung des Qualitätsbegriffs und der Herleitung des funktionalen Bezugsrahmens der Untersuchung.

Die Bewertung von Übersetzungsqualität stellt in der Translationswissenschaft eine zentrale, zugleich jedoch komplexe Fragestellung dar. Übersetzungsqualität ist nicht eindeutig zu bestimmen, da Translation als Prozess zugleich kognitive, sprachliche, soziale, kulturelle sowie technologische Dimensionen umfasst. Gerade diese Vielschichtigkeit erschwert eine einheitliche Bestimmung und Bewertung von Übersetzungsqualität (Castilho et al. 2018: 10). Wie Colina (2020: 458) hervorhebt, besteht weder hinsichtlich des Begriffs der Übersetzungsqualität noch hinsichtlich dessen, was als Übersetzung gilt, Einigkeit. Qualität ist zudem als relativer Begriff zu verstehen, dessen Bewertung von historischen, sozialen und kulturell geprägten Wertvorstellungen abhängt. Die translationswissenschaftliche Diskussion ist durch eine Vielfalt theoretischer Zugänge gekennzeichnet, die unterschiedliche Aspekte von Übersetzung in den Vordergrund rücken, etwa den Übersetzungszweck, die Rolle der Übersetzenden oder Fragen der Äquivalenz. Erschwert wird die Bestimmung von Übersetzungsqualität zudem dadurch, dass nicht alle Bewertungsansätze auf explizit formulierten Theorien beruhen, sondern teilweise unausgesprochene, im professionellen Kontext verankerte Vorstellungen widerspiegeln.

Auch House (2024: 69) weist darauf hin, dass Aussagen über die Qualität einer Übersetzung stets ein bestimmtes Verständnis von Übersetzen voraussetzen und damit an eine Übersetzungstheorie gebunden sind. Unterschiedliche theoretische Positionen führen folglich auch zu unterschiedlichen Auffassungen von Übersetzungsqualität sowie zu verschiedenen Verfahren ihrer Bewertung. Auch Schäffner (1998: 1) greift die Frage auf, anhand welcher Kriterien ein Zieltext als „gute“ Übersetzung beurteilt werden kann. Sie betont, dass solche Kriterien nicht allgemeingültig feststehen, sondern je nach Zweck der Beurteilung und dem zugrunde gelegten theoretischen Rahmen variieren.

Der Äquivalenzbegriff gehört zu den zentralen Konzepten der translationswissenschaftlichen Qualitätsdiskussion und spielt insbesondere in frühen Übersetzungstheorien eine grundlegende Rolle (House 2015: 5). In äquivalenzorientierten Ansätzen wird Translation vor allem als möglichst genaue Reproduktion des Ausgangstextes im Zieltext konzipiert. Da eine vollständige Äquivalenz zwischen Ausgangs- und Zieltext auf allen Ebenen nicht erreichbar ist, wurde das Konzept im translationswissenschaftlichen Diskurs weiter ausdifferenziert (Lau-scher 2000: 151). Der amerikanische Linguist Eugene Nida (1964: 159) differenziert zwischen formaler und dynamischer Äquivalenz. Formale Äquivalenz zielt auf eine möglichst enge Orientierung an Form und Inhalt des Ausgangstextes. Dynamische Äquivalenz definieren Nida und Taber (1969: 23) demgegenüber als den Grad, in dem die Rezipient*innen des Zieltexts auf die Botschaft in wesentlich ähnlicher Weise reagieren wie die Rezipierenden des Ausgangstextes. Zugleich betonen sie, dass eine solche Reaktion aufgrund unterschiedlicher kultureller und historischer Kontexte nicht identisch sein kann, dass aber ein hohes Maß an Äquivalenz der Reaktion erreicht werden sollte. Andernfalls verfehlt die Übersetzung ihren Zweck. Darüber hinaus halten Nida und Taber fest, dass dynamische Äquivalenz Vorrang vor formaler Entsprechung hat (1969: 13).

Koller und Henjum (2020) unterscheiden dabei denotative, konnotative, textnormative, pragmatische sowie formal-ästhetische Äquivalenztypen. Die Typologie geht auf Werner Koller zurück und wurde bereits in früheren Auflagen der *Einführung in die Übersetzungswissenschaft* entwickelt (vgl. Prunč 2013: 60). Zugleich betonen Koller und Henjum (2020: 311f.), dass für jeden konkreten Übersetzungsfall eine „Hierarchie der in der Übersetzung zu erhaltenden Werte“ festgelegt werden muss. Daraus ergibt sich, welche Äquivalenzforderungen im konkreten Fall Vorrang haben. Welche Form von Äquivalenz besonders stark zu gewichten ist, hängt somit vom jeweiligen Text und von den Ergebnissen einer übersetzungsrelevanten Textanalyse ab (Koller & Henjum 2020: 312). Für Newmark (1991: 111) ist eine gute Übersetzung vor allem

dadurch gekennzeichnet, dass sie „as accurate as possible, as economical as possible, in denotation and in connotation, referentially and pragmatically“ ist. Außerdem hebt er hervor, dass Genauigkeit sich dabei auf den Ausgangstext bezieht, insbesondere auf die von der Autor*in intendierte Bedeutung beziehungsweise auf den im Text enthaltenen Sachverhalt. Zugleich kann sie an die intellektuellen und emotionalen Voraussetzungen der vorgesehenen Leserschaft angepasst werden.

Einen frühen systematischen Ansatz zur Übersetzungsbewertung entwickelt Katharina Reiß (1971) im Rahmen ihrer Übersetzungskritik. Der Ansatz basiert auf einem äquivalenzorientierten Verständnis von Übersetzen, bei dem Äquivalenz nicht nur zwischen Ausgangs- und Zieltext als Ganzem, sondern auch auf der Ebene einzelner Texteinheiten angestrebt wird. Reiß zufolge liegt optimale Äquivalenz dann vor, wenn die zielsprachliche Entsprechung dem sprachlichen und situativen Kontext, der Stil- und Sprachebene sowie der Intention der Autor*in entspricht und im Zieltext „die gleiche Wertigkeit“ wie die ausgangssprachliche Einheit aufweist (Reiß 1971: 11f.). Für Reiß ist die Bestimmung des Texttyps für die Beurteilung von Übersetzungen von zentraler Bedeutung. Wie sie ebenfalls hervorhebt, ohne diesen Schritt werde eine Übersetzung nach „unzutreffenden Maßstäben“ bewertet (1971: 24). Entsprechend entwickelt sie eine Texttypologie, die inhaltsbetonte, formbetonte und appellbetonte Texte unterscheidet. Diese Typologie wird in der späteren Rezeption häufig vereinfachend als Unterscheidung zwischen informativen, expressiven und operativen Texttypen wiedergegeben (vgl. Prunč 2013: 82). Übersetzungskritik ist nach Reiß (1971: 115) dann sachgerecht, wenn sich ihre Kriterien am jeweiligen Texttyp sowie an den textimmanenten innersprachlichen Instruktionen und den außersprachlichen Determinanten orientieren.

Einen Gegensatz zu den präskriptiven äquivalenzorientierten Ansätzen bildet der deskriptiv-zielkulturorientierte Ansatz von Gideon Toury, der nicht von normativen Äquivalenzforderungen ausgeht, sondern den Zieltext und seine Stellung in der Zielkultur zum Ausgangspunkt der Analyse macht. Damit verschiebt sich der Fokus von der Erfüllung vorgegebener Äquivalenzforderungen auf den Zieltext und seine Funktion innerhalb der Zielkultur (Toury 1995: 24).

Insgesamt zeigt sich, dass frühe translationswissenschaftliche Qualitätskonzepte vor allem äquivalenzorientiert waren und Übersetzungsqualität primär über ihre Relation zum Ausgangstext bestimmten. Im weiteren Verlauf der translationswissenschaftlichen Diskussion kam es jedoch zu einem Paradigmenwechsel, in dessen Zuge der Zieltext, seine Funktion und seine Einbettung in die Zielkultur stärker in den Vordergrund rückten. Funktionalistische Ansätze der

Übersetzungsbewertung beruhen auf der Annahme, dass Translation nicht vorrangig durch den Ausgangstext, sondern vor allem durch zielkulturelle Faktoren bestimmt ist. Damit verschiebt sich auch der Maßstab der Qualitätsbeurteilung: Entscheidend ist nun nicht mehr ausschließlich die Äquivalenz zum Ausgangstext, sondern in zunehmendem Maß die Frage, ob der Zieltext seine vorgesehene Funktion in der Zielkultur erfüllt (vgl. Prunč 2013: 151f.; Lauscher 2000: 156). Für die vorliegende Arbeit ist in diesem Zusammenhang insbesondere die funktionalistische Perspektive relevant, wie sie in der Skopostheorie ausgearbeitet wird.

2.2 Skopostheorie

In den 1980er Jahren kam es in der Translationswissenschaft zu einer grundlegenden Neuorientierung des wissenschaftlichen Diskurses. Bis zu diesem Zeitpunkt waren translationswissenschaftliche Ansätze überwiegend sprachwissenschaftlich geprägt und stark an der Äquivalenzorientierung ausgerichtet. Im Zuge eines Paradigmenwechsels löste sich die Translationswissenschaft von der Sprachwissenschaft und etablierte sich als eigenständige Disziplin. Einen wesentlichen Impuls für diese Entwicklung stellte die Veröffentlichung zentraler Werke dar: Katharina Reiß und Hans J. Vermeer legten 1984 ihre *Grundlegung einer allgemeinen Translationstheorie* vor, während Justa Holz-Mänttari im selben Jahr das Werk *Translatorisches Handeln* veröffentlichte (Prunč 2013: 151).

Mit der von Reiß und Vermeer veröffentlichten *Grundlegung einer allgemeinen Translationstheorie* wurde ein funktionalistischer Ansatz formuliert, der den Zweck der Translation in den Mittelpunkt stellt. Zentraler Begriff dieses Ansatzes ist der Skopos, ein aus dem Griechischen stammender Begriff für „Ziel“ oder „Zweck“, der in der Skopostheorie auf den Zweck der Translation bezogen wird (Reiß & Vermeer 1984: 96).

Reiß und Vermeer (1984: 119) konkretisieren ihre allgemeine Translationstheorie durch ein Regelgefüge, das mehrere miteinander verbundene Prinzipien umfasst. Diese sind nicht gleichrangig, sondern hierarchisch geordnet, wobei die oberste Position die Skoposregel einnimmt. Reiß und Vermeer formulieren die Regeln wie folgt:

- (1) Ein Translat ist skoposbedingt. [...]
- (2) Ein Translat ist ein Informationsangebot in einer Zielkultur und -sprache über ein Informationsangebot in einer Ausgangskultur und -sprache. [...]
- (3) Ein Translat bildet ein Informationsangebot nicht umkehrbar eindeutig ab. [...]
- (4) Ein Translat muss in sich kohärent sein. [...]

- (5) Ein Translat muss mit dem Ausgangstext kohärent sein. [...] (1984: 119)

Die hierarchische Anordnung der von Katharina Reiß und Hans J. Vermeer formulierten Regeln verdeutlicht die Überordnung der Skoposregel gegenüber den übrigen Prinzipien. Der Zweck der Translation bildet demnach das übergeordnete Kriterium, an dem sich alle translatorischen Entscheidungen orientieren und von dem sowohl die Auswahl der zu übertragenden Inhalte als auch die konkrete Translationsstrategie abhängen (Reiß & Vermeer 1984: 95f.). Der Skopos einer Translation ist dabei nicht abstrakt vorgegeben, sondern ergibt sich aus der konkreten Situation, in der der Zieltext verwendet werden soll (Reiß & Vermeer 1984: 101).

Die Kohärenzregeln spielen innerhalb der Skopostheorie ebenfalls eine zentrale Rolle. Die Forderung nach intratextueller Kohärenz gewährleistet, dass das Translat für die Rezipient*innen in der jeweiligen Situation verständlich und sinnvoll ist (Reiß & Vermeer 1984: 109). Die intertextuelle Kohärenz bezieht sich hingegen auf die Beziehung zwischen Translat und Ausgangstext und stellt sicher, dass zwischen beiden Texten ein nachvollziehbarer Zusammenhang besteht (1984: 114).

Innerhalb der Skopostheorie wird zwischen Adäquatheit und Äquivalenz klar unterschieden. Adäquatheit bezeichnet bei Reiß und Vermeer die Relation zwischen Ziel- und Ausgangstext, wobei der Skopos systematisch berücksichtigt wird. Eine Translation ist demnach adäquat, wenn die Auswahl der sprachlichen Mittel in der Zielsprache systematisch dem Zweck der Translation untergeordnet wird. Äquivalenz bestimmen Reiß und Vermeer als einen besonderen Fall adäquater Translation, bei dem Funktionskonstanz zwischen dem Ausgangs- und Zieltext besteht (Reiß & Vermeer 1984: 139). Daraus ergibt sich, dass das Äquivalenzkriterium nicht länger das primäre Kriterium für die Beurteilung von Translation bildet. Äquivalenz wird vielmehr nur dann relevant, wenn der Skopos der Translation die Beibehaltung der Funktionskonstanz zwischen dem Ausgangs- und Zieltext vorsieht (Prunč 2013: 158).

Reiß und Vermeer (1984: 103) betonen, dass der Skopos eines Translats nicht notwendigerweise mit dem Skopos des Ausgangstextes übereinstimmen muss. Translation wird als eigenständige Produktionshandlung verstanden, die prinzipiell anderen Zwecken dienen kann als die Herstellung des Ausgangstextes. Die Auffassung, dass eine Translation den Zweck des Ausgangstextes beizubehalten habe, ist nach Vermeer keine allgemeingültige Forderung einer Translationstheorie, sondern eine kulturspezifische Regel. Gleichzeitig impliziert diese Position jedoch kein Verbot der Zweckbeibehaltung. Vielmehr wird diese als eine mögliche, situations-

und kulturspezifisch begründete Entscheidung verstanden, die insbesondere im Bereich der literarischen Übersetzung häufig eine zentrale Rolle spielt (vgl. Reiß & Vermeer 1984: 102f.).

Für die vorliegende Untersuchung wird der Skopos der Übersetzung als ausgangstextbezogene und zugleich adressatenorientierte Zieltextgestaltung verstanden. Die Übersetzung soll zentrale Merkmale des literarischen Ausgangstextes, insbesondere in den Bereichen Stil & Register, Kohärenz und lokale Kohäsion, im Deutschen zielsprachlich angemessen realisieren. Die Funktion der Übersetzung besteht dabei darin, für ein deutschsprachiges Zielpublikum als literarischer Text verständlich, stilistisch angemessen und registerbezogen plausibel zu wirken. Der Skopos wird daher als Funktionskonstanz verstanden: Die genannten Merkmale sollen im Deutschen so wiedergegeben werden, dass sie für das deutschsprachige Zielpublikum nachvollziehbar und zielsprachlich angemessen sind.

Auf dieser Grundlage lässt sich der funktionalistische Theorierahmen mit dem in Kapitel 7.1 beschriebenen Analyseansatz verbinden, der die erweiterte Smart Computer-Aided Translation Environment (SCATE)-Fehlertaxonomie nach Tezcan et al. (2019) mit einer für die vorliegende Arbeit adaptierten Key-Feature-Analyse nach Webster et al. (2020) kombiniert. Während der Skopos die übergeordnete Zielsetzung der Übersetzung bestimmt, ermöglicht die erweiterte SCATE-Taxonomie eine systematische Einordnung kohärenz-, stil- und registerbezogener Auffälligkeiten in den untersuchten Zieltexten. Die adaptierte Key-Feature-Analyse erweitert den Analyseansatz um Untersuchung der lexikalischen Vielfalt und der lokalen Kohäsion.

2.3 Ansätze und Modelle zur Analyse der Qualität maschineller Übersetzungen

In der Forschung zur maschinellen Übersetzung lassen sich grundsätzlich zwei zentrale Ansätze zur Bewertung von Übersetzungsqualität unterscheiden: automatische Evaluierungsmethoden und menschliche Evaluierungsmethoden. Für die Evaluation von Übersetzungssystemen wird dabei häufig auf die Urteile menschlicher Expert*innen zurückgegriffen, da diese sowohl eine satzweise Beurteilung als auch eine zusammenfassende Einschätzung der Übersetzungsleistung ermöglichen. Gleichzeitig ist diese Form der Evaluation mit erheblichem zeitlichem, organisatorischem und finanziellem Aufwand verbunden, was ihren Einsatz für häufige und wiederholte Evaluationen deutlich erschwert. In diesem Zusammenhang wurden automatische Evaluierungsmethoden entwickelt, die eine schnelle und kostengünstige Beurteilung maschinell er-

zeugter Übersetzungen erlauben sollen. Diese Verfahren beruhen in der Regel auf dem Vergleich des maschinellen Outputs mit einer Referenzübersetzung und liefern quantifizierbare Bewertungsergebnisse (Koehn 2020: 52f.).

2.3.1 Automatische Evaluierungsmethoden

2.3.1.1 Bilingual Evaluation Understudy (BLEU)

Eine der bekanntesten automatischen Evaluierungsmetriken ist Bilingual Evaluation Understudy (BLEU), die von Papineni et al. (2002) vorgestellt wurde. Die Metrik wurde mit der Absicht entwickelt, eine kostengünstige und effiziente Alternative zur zeit- und ressourcenintensiven manuellen Bewertung maschineller Übersetzungen bereitzustellen (Papineni et al. 2002: 311). BLEU basiert auf dem Vergleich maschineller Übersetzungen mit einer oder mehreren menschlichen Referenzübersetzungen und misst deren Übereinstimmung auf der Ebene von Wortfolgen (n-Grammen) (2002: 312). Dabei fließen neben einzelnen Wörtern auch längere Wortfolgen in die Bewertung ein, sodass die Wortreihenfolge explizit berücksichtigt wird (Koehn 2020: 54). Die zugrunde liegende Annahme ist, dass Übersetzungen umso besser sind, je stärker sie den menschlichen Referenzen ähneln (Papineni et al. 2002: 311).

Der BLEU-Score nimmt Zahlenwerte zwischen 0 und 1 an, wobei höhere Zahlenwerte eine stärkere Übereinstimmung mit den Referenzübersetzungen suggerieren. Wie Papineni et al. (2002: 315) hervorheben, erreichen nur wenige Übersetzungen den Maximalwert, da dieser eine weitgehende Identität mit einer Referenzübersetzung voraussetzt und selbst menschliche Übersetzungen diesen Wert nicht zwangsläufig erzielen. Zudem hängt die Höhe des BLEU-Scores von der Anzahl der verfügbaren Referenzübersetzungen ab, da mit zunehmender Anzahl an Referenzen die Wahrscheinlichkeit von n-Gramm-Übereinstimmungen steigt, was insbesondere bei Texten mit mehreren möglichen Übersetzungsvarianten die Vergleichbarkeit der Ergebnisse einschränken kann. In diesem Zusammenhang ist BLEU vor allem für Vergleiche auf der Ebene größerer Textmengen geeignet, während Aussagen zur Qualität einzelner Übersetzungen daraus nur eingeschränkt abgeleitet werden können (2002: 318).

2.3.1.2 Metric for Evaluation of Translation with Explicit Ordering (METEOR)

Eine weitere verbreitete automatische Evaluierungsmetrik ist Metric for Evaluation of Translation with Explicit Ordering (METEOR), entwickelt von Alon Lavie und Michael J. Denkowski.

Die Metrik wurde erstmals 2004 vorgestellt und mit dem Ziel konzipiert, eine höhere Übereinstimmung mit menschlichen Bewertungen über Übersetzungsqualität zu erreichen, insbesondere auf Satzebene (Lavie & Denkowski 2009: 105). Zur Bewertung einer Übersetzung berechnet METEOR einen Score, der auf einer expliziten Zuordnung auf Wortebene zwischen der maschinellen Übersetzung (Hypothese) und einer oder mehreren Referenzübersetzungen beruht. Im Falle, wenn mehrere Referenzübersetzungen vorhanden sind, wird die Hypothese jeweils einzeln mit jeder Referenz verglichen. Für die weitere Bewertung wird anschließend das Hypothese-Referenz-Paar mit dem höchsten Score herangezogen. Der Abgleich erfolgt dabei nicht in einem einzigen Schritt, sondern durch den schrittweisen Aufbau eines Alignments zwischen Hypothese und Referenz (2009: 107). Dabei ist der Vergleich nicht auf identische Wortformen beschränkt, sondern umfasst auch morphologische Varianten sowie Synonyme (2009: 106).

Die Entwicklung von METEOR erfolgte unter Berücksichtigung der in der Forschung beschriebenen Einschränkungen von BLEU, insbesondere hinsichtlich der Berechnung verlässlicher Bewertungen auf Satzebene (2009: 106). Gleichzeitig weist METEOR im Vergleich zu BLEU eine deutlich höhere methodische Komplexität auf, da für die Berechnung zusätzliche linguistische Ressourcen wie Synonymdatenbanken benötigt werden und der Wortabgleich mit einem rechnerisch aufwendigen Alignment-Verfahren verbunden ist (Koehn 2020: 55).

2.3.1.3 Translation Edit Rate (TER)

Eine weitere automatische Bewertungsmetrik ist Translation Edit Rate (TER), die von Snover et al. (2006) entwickelt wurde. TER basiert auf der Anzahl von Editieroperationen, die erforderlich sind, um den maschinellen Output so zu verändern, dass er mit einer Referenzübersetzung übereinstimmt (Snover et al. 2006: 223). Berücksichtigt werden dabei Einfügungen, Löschungen, Ersetzungen sowie Verschiebungen von Wortsequenzen, wobei alle Editieroperationen gleich gewichtet sind (2006: 225). Wie Snover et al. (2006: 226) betonen, bildet TER nicht ab, inwieweit Hypothese und Referenz semantisch äquivalent sind, da die Metrik keine semantische Äquivalenz berücksichtigt. Auch Übersetzungen, die flüssig sind und dieselbe Bedeutung wie die Referenz ausdrücken, können daher einen hohen Editieraufwand aufweisen. TER eignet sich insbesondere für die Abschätzung des post-editorischen Aufwands, da der Score direkt als Maß für notwendige Korrekturen interpretierbar ist. Gleichzeitig heben die Autor*innen selbst hervor, dass TER ohne zusätzliche menschliche Annotation keine zuverlässigen Aussagen über die semantische Angemessenheit einer Übersetzung erlaubt.

In der Forschung wurden jedoch wiederholt kritische Anmerkungen zur Verwendung automatischer, referenzbasierter Evaluierungsmetriken für die Bewertung neuronaler maschineller Übersetzung geäußert, insbesondere aufgrund ihrer überwiegend satzbasierten Ausrichtung (Fonteyne et al. 2020: 3790f.). Für literarische Texte, bei denen solche Aspekte wie Kohärenz, Kohäsion und stilistische Gestaltung eine zentrale Rolle spielen, erweist sich diese Evaluierungsperspektive als limitiert. Gerade im Bereich literarischer Übersetzung wird daher die Notwendigkeit einer stärkeren Berücksichtigung dokumentenbasierter Evaluierungsansätze verlangt, die über die Satzebene hinausgehende Aspekte erfassen können (vgl. Fonteyne et al. 2020: 3791).

2.3.2 Menschliche Evaluierungsmethoden

In diesem Abschnitt werden ausgewählte menschliche Evaluierungsansätze vorgestellt, die für die Analyse der Qualität maschineller Übersetzungen von besonderer Relevanz sind. Im Mittelpunkt stehen dabei manuelle Fehlerannotationsmodelle, die eine differenzierte Untersuchung konkreter Übersetzungsphänomene ermöglichen. Ergänzend dazu werden Ansätze berücksichtigt, die deskriptive, teils quantifizierbare textuelle Merkmale heranziehen, ohne als automatische Evaluierungsmetriken im engeren Sinne zu gelten. Häufig dienen sie dazu, manuelle Fehlerannotationen um zusätzliche, beschreibende Informationen auf Dokumentebene zu ergänzen.

2.3.2.1 Multidimensional Quality Metrics (MQM)

Multidimensional Quality Metrics ist ein von Lommel, Uszkoreit und Burchardt entwickeltes Framework zur Bewertung von Übersetzungsqualität, das eine strukturierte Sammlung von Fehlerkategorien (*issue types*) bereitstellt (Lommel et al. 2014: 455). Das MQM-Framework ist sowohl für die Bewertung menschlicher als auch maschineller Übersetzungen konzipiert und wird unter anderem auch auf Übersetzungen angewendet, die mithilfe künstlicher Intelligenz (KI) erstellt wurden (Lommel et al. 2024: 79).

MQM verbindet eine hierarchisch organisierte Typologie von Übersetzungsfehlern mit einem Scoring Model, das die identifizierten Fehler nach festgelegten Kriterien gewichtet und zu einem Gesamturteil zusammenführt. Die Gesamtheit aller im Framework definierten Fehlertypen wird als *MQM-Full* bezeichnet. In der praktischen Anwendung findet diese vollständige Menge an Kriterien jedoch nur selten Verwendung. Stattdessen erfolgt kontextabhängig eine Auswahl spezifischer Kategorien, die eine den jeweiligen Anforderungen entsprechende Granularität der Analyse ermöglichen (Lommel et al. 2024: 79). Zu diesem Zweck bietet MQM

mit *MQM-Core* eine vordefinierte Auswahl von Fehlertypen, die sich auf zentrale übergeordnete Fehlerdimensionen und häufig verwendete Unterkategorien beschränkt (2024: 79f.). Die einzelnen Fehlerkategorien sind dabei durch klar definierte Bezeichnungen und präzise Beschreibungen charakterisiert, die eine eindeutige Interpretation im Rahmen der Annotation sicherstellen sollen (2024: 80).

Neben der Fehlertypologie bildet das Scoring-Modell die zweite zentrale Komponente des MQM-Frameworks. Dieses legt fest, wie auf Grundlage der annotierten Fehler ein Qualitätswert (*Quality Score*) ermittelt wird. Die konkrete Ausgestaltung des Scoring-Modells ist anwendungsspezifisch: Abhängig vom jeweiligen Evaluationsziel werden bestimmte Fehlertypen ausgewählt, unterschiedlich gewichtet und mit definierten Schweregraden versehen. Die annotierten Fehler werden dabei den jeweiligen Fehlertypen und Schweregraden zugeordnet und bilden die Grundlage für die anschließende Berechnung des Qualitätswerts (2024: 80). Während MQM ein allgemein einsetzbares Framework für die manuelle Fehlerannotation darstellt, greift die vorliegende Arbeit für die Analyse maschineller Übersetzungen auf eine andere Fehlertaxonomie zurück. Konkret wird eine adaptierte Version der Smart Computer-aided Translation Environment (SCATE)-Fehlertaxonomie verwendet, die in ihrer Struktur MQM ähnelt und in früheren Studien für die Annotation maschineller Übersetzungen eingesetzt wurde (Tezcan et al. 2019; Fonteyne et al. 2020; Webster et al. 2020).

2.3.2.2 Smart Computer-aided Translation Environment (SCATE)

SCATE (Smart Computer-aided Translation Environment) ist eine spezialisierte Fehlertaxonomie, die ursprünglich für die Analyse maschineller Übersetzungen entwickelt wurde und in späteren Arbeiten gezielt für die Untersuchung literarischer Übersetzungen erweitert wurde (Tezcan et al. 2017: 219). In ihrer Struktur weist die SCATE-Fehlertaxonomie Parallelen zu MQM auf, erlaubt jedoch die Annotation mehrerer Fehler auf demselben Textsegment und sieht bei *Accuracy*-Fehlern eine explizite Verknüpfung von Ausgangs- und Zieltexten im Rahmen der Annotation vor (Fonteyne et al. 2020: 3791).

Tezcan et al. (2017: 219) beschreiben SCATE (Smart Computer-aided Translation Environment) als eine hierarchisch aufgebaute Fehlertaxonomie für maschinelle Übersetzungen, die auf der grundlegenden Unterscheidung zwischen *Accuracy* und *Fluency* beruht. Ziel der SCATE-Fehlertaxonomie ist es, eine systematische Annotation von Übersetzungsfehlern in maschinellen Übersetzungen zu ermöglichen. Die SCATE-Fehlertaxonomie ist hierarchisch aufgebaut und gliedert sich in mehrere Ebenen. Auf der obersten Ebene wird zwischen *Accuracy*-

und *Fluency*-Fehlern unterschieden, die jeweils in weitere Unterkategorien unterteilt sind (2017: 225). Im Rahmen der SCATE-Annotation werden *Fluency*-Fehler ausschließlich im Zieltext annotiert. *Accuracy*-Fehler hingegen werden durch die gleichzeitige Betrachtung von Ausgangs- und Zieltext erfasst und die entsprechenden Annotationen miteinander verbunden (2017: 219). Darüber hinaus sieht SCATE eine bilinguale Annotation vor. Übersetzungsfehler werden nicht ausschließlich im Zieltext annotiert, sondern bei *Accuracy*-Fehlern zusätzlich mit den entsprechenden Stellen im Ausgangstext verknüpft. Auf diese Weise werden Fehlerquellen transparenter, und es wird möglich, bestimmte Fehlerarten, wie z.B. Auslassungen, systematisch zu erfassen. Ebenso geht SCATE davon aus, dass Übersetzungsfehler nicht isoliert auftreten müssen. Mehrere unterschiedliche Fehlerarten können sich auf dieselbe Textstelle beziehen, da Fehler häufig miteinander zusammenhängen oder sich gegenseitig bedingen. Die Taxonomie erlaubt daher die Annotation mehrerer Fehler an einer gemeinsamen Textstelle, um die Komplexität realer MÜ-Fehler angemessen abzubilden (2017: 224).

Bei der Ausarbeitung der SCATE-Fehlertaxonomie wurden mehrere zentrale Aspekte berücksichtigt. Ein Schwerpunkt liegt auf MÜ-spezifischen Fehlern. Da maschinelle Übersetzungssysteme Sprache nicht auf dieselbe Weise verarbeiten wie menschliche Übersetzer*innen, treten im MÜ-Output charakteristische Fehlerarten auf, die in Humanübersetzungen seltener vorkommen. Die Taxonomie ist daher darauf ausgerichtet, typische Fehler maschineller Übersetzungen abzubilden, etwa das unveränderte Übernehmen unbekannter Ausgangstextsegmente oder grammatische Abweichungen im Zieltext. Ein weiterer zentraler Aspekt ist die Trennung zwischen *Accuracy*- und *Fluency*-Fehlern. Diese Unterscheidung ermöglicht es, unterschiedliche Dimensionen der Übersetzungsqualität getrennt zu analysieren und trägt zugleich dazu bei, die Komplexität des Annotationsprozesses zu reduzieren (2017: 224).

Aufbauend auf den beschriebenen Eigenschaften der SCATE-Fehlertaxonomie wird im Folgenden die Fehlerklassifikation innerhalb des Modells näher erläutert. Dabei wird zunächst auf die Kategorie der *Fluency*-Fehler eingegangen, bevor anschließend die *Accuracy*-Fehler beschrieben werden. Innerhalb der Kategorie der *Fluency*-Fehler unterscheidet SCATE mehrere Hauptkategorien, darunter grammatische, lexikalische und orthographische Fehler sowie eine Sammelkategorie für Fälle, in denen mehrere *Fluency*-Fehler gleichzeitig auftreten und nicht eindeutig voneinander abgegrenzt werden können (Tezcan et al. 2017: 226). Innerhalb der Kategorie der *Accuracy*-Fehler differenziert SCATE mehrere Hauptkategorien, darunter u. a. Hinzufügungen (*Addition*), Auslassungen (*Omission*), nicht übersetzte Elemente (*Untranslated*), *Do-not-translate*-Fehler (DNT), bei denen Ausgangstextelemente übersetzt wurden, obwohl sie

im ZIELtext unverändert hätten übernommen werden sollen, Fehlübersetzungen (*Mistranslation*), mechanische Fehler, terminologische Probleme, Quelltextfehler (*Source errors*) sowie eine Kategorie für sonstige *Accuracy*-Fehler (2017: 227). Das SCATE-Modell bietet damit eine Grundlage zur systematischen Annotation von Übersetzungsfehlern in maschinellen Übersetzungen (2017: 219).

Für die Analyse literarischer Texte, bei denen über die Satzebene hinausgehende stilistische und textuelle Phänomene relevant sind, wurde das SCATE-Modell in späteren Studien gezielt erweitert. Insbesondere Tezcan et al. (2019) sowie Fonteyne et al. (2020) weisen darauf hin, dass bestimmte für literarische Texte zentrale Phänomene etwa Kohärenz, Kohäsion oder kulturelle Referenzen mit der ursprünglichen SCATE-Taxonomie nur eingeschränkt erfasst werden können und schlagen entsprechende Erweiterungen vor.

3 Maschinelle Übersetzung

3.1 Historische Entwicklung maschineller Übersetzung

Im Frühjahr 1947 lassen sich zwei Ereignisse als Ausgangspunkte für das entstehende Forschungsinteresse an maschineller Übersetzung identifizieren. Zum einen formulierte Warren Weaver in einem Brief an den Kybernetiker Norbert Wiener die Überlegung, ob sich ein Rechner entwickeln lässt, der Übersetzungen automatisiert durchführen kann. Dabei zog er Parallelen zur Kryptographie und verstand Übersetzen als eine Form des „Dekodierens“ von Information. Weaver betonte zudem, dass bereits ein System hilfreich wäre, das sich auf wissenschaftliche Texte beschränkt und zunächst vor allem verständliche Übersetzungen liefert, auch wenn diese nicht in jeder Hinsicht stilistisch ausgearbeitet sind (Hutchins 1997: 195). Ein zweites einschlägiges Ereignis fand wenige Tage später statt: Weaver traf Andrew Booth, einen britischen Kristallographen, der sich für die Finanzierung eines britischen Rechners interessierte. In diesem Gespräch sprach Weaver Übersetzen als mögliches Anwendungsfeld nicht-numerischer Computernutzung an und wies darauf hin, dass Fördermittel eher für solche Anwendungen zu gewinnen waren als für rein numerische Aufgaben (1997: 196).

Ergänzend ist darauf hinzuweisen, dass im 20. Jahrhundert bereits technische Vorläufer entstanden, die zunächst vor allem auf die Mechanisierung mehrsprachiger Wörterbücher zielten, bevor später auch Konzepte einer umfassenderen Automatisierung des Übersetzungsprozesses formuliert wurden. Hutchins verweist dabei auf das Jahr 1933, in dem unabhängig voneinander Patente angemeldet wurden: in Frankreich durch Georges Artsrouni und in Russland

durch Petr Troyanskii. Troyanskii führte diese Ansätze weiter und entwickelte seine Überlegungen bis in die 1940er Jahre. Um 1947 ging er laut Hutchins über die reine Wörterbuchmechanisierung hinaus und schlug eine weitergehende Automatisierung des Übersetzungsprozesses vor (Hutchins 1997: 197). Im Juli 1949 präsidierte Weaver seine Überlegungen zur maschinellen Übersetzung in einem Memorandum mit dem Titel *Translation* (1997: 204). Ziel des Memorandums war es, den damaligen Stand der Überlegungen zusammenzufassen und mögliche Richtungen weiterer Forschung zu skizzieren (1997: 205). Weaver hob hervor, dass maschinelle Verfahren für literarische Übersetzung kaum geeignet waren. Für technische Texte hielt er jedoch selbst eine wortwörtliche Übersetzung für potenziell nützlich, da Mehrdeutigkeit innerhalb eines Fachkontexts typischerweise reduziert ist (1997: 206). Zugleich hielt Weaver fest, dass eine vollständig fehlerfreie Übersetzung kaum erreichbar ist. Stattdessen rückte er die Idee in den Vordergrund, dass Übersetzungen auch dann praktisch brauchbar sein können, wenn sie nur einen bestimmten Grad an Genauigkeit erreichen und eine begrenzte Fehlerrate aufweisen (1997: 207). Darüber hinaus skizzierte Weaver zentrale Themenfelder der MÜ-Forschung, darunter statistische Ansätze, die Berücksichtigung von Kontext sowie das Problem der Mehrdeutigkeit (1997: 208).

Bereits Anfang der 1950er Jahre verlagerten sich die zuvor eher konzeptionellen Überlegungen in institutionelle Kontexte. Hutchins (1997: 215) zufolge wurde die maschinelle Übersetzung am Massachusetts Institute of Technology (MIT) als prüfenswertes Forschungsfeld eingeschätzt. Einen wichtigen Meilenstein dieser frühen Phase bildete die erste MÜ-Konferenz am MIT im Juni 1952 (1997: 224). Die Entwicklung der maschinellen Übersetzung zu einem praktisch orientierten Forschungsfeld zeigte sich in den folgenden Jahren besonders deutlich im Georgetown-IBM-Experiment (1997: 239).

Die Zusammenarbeit zwischen der Georgetown University und dem Unternehmen International Business Machines (IBM) stellte ein zentrales Ereignis in der frühen Geschichte der maschinellen Übersetzung dar (Hutchins 2004: 102). Die öffentliche Demonstration eines maschinellen Übersetzungssystems für das Sprachpaar Russisch-Englisch im Jahr 1954 erregte großes Interesse und sollte zeigen, dass die automatische Übersetzung zwischen natürlichen Sprachen grundsätzlich möglich ist (2004: 102f.). Allerdings basierte das Experiment auf einer kontrollierten Auswahl von Beispielsätzen und thematisch eingeschränktem Wortschatz, wodurch seine Aussagekraft deutlich begrenzt war (2004: 103). Das Georgetown-IBM-Experiment wurde von den beteiligten Forscher*innen unterschiedlich bewertet. Während einzelne

Beteiligte darin einen Ausgangspunkt für weiterführende Entwicklungen sahen, betonten andere Forscher*innen die erheblichen Einschränkungen des Experiments (2004: 111). Ungeachtet seiner starken Einschränkungen trug das Experiment maßgeblich dazu bei, maschinelle Übersetzung als eigenständiges Forschungsfeld zu etablieren und institutionelle sowie finanzielle Unterstützung für weitere Arbeiten zu legitimieren (2004: 114).

In der weiteren Entwicklung markierte der Bericht des Automatic Language Processing Advisory Committee (ALPAC) von 1966 einen deutlichen Wendepunkt in der Geschichte der maschinellen Übersetzung. Im Bericht wurden die bis dahin erzielten Ergebnisse kritisch beurteilt. Außerdem wurde festgestellt, dass maschinelle Übersetzung im Vergleich zur menschlichen Übersetzung weder hinsichtlich Qualität noch hinsichtlich Geschwindigkeit und Kosten überlegen sei. In der Folge wurde die staatliche Förderung der MÜ-Forschung in den USA weitgehend eingestellt, was die weitere Entwicklung des Feldes für längere Zeit erheblich verlangsamte, auch wenn die Forschung in anderen Ländern fortgeführt wurde (Hutchins 2023: 130). Auf dieser historischen Entwicklung aufbauend werden im Folgenden die zentralen Ansätze der maschinellen Übersetzung systematisch dargestellt und erläutert.

3.2 Arten der maschinellen Übersetzung

In diesem Kapitel werden die wichtigsten Arten der maschinellen Übersetzung skizziert, die sich historisch herausgebildet und die Entwicklung der MÜ maßgeblich geprägt haben. Im Mittelpunkt stehen dabei regelbasierte, statistische und neuronale Verfahren, bevor anschließend aktuelle, LLM-gestützte Systeme eingeordnet werden.

3.2.1 Regelbasierte maschinelle Übersetzung

Regelbasierte maschinelle Übersetzung (RBMÜ, engl. RBMT) war bis in die frühen 2000er Jahre hinein in vielen Systemen verbreitet und galt lange als Stand der Technik (Kenny 2019: 434). In der Literatur wird sie üblicherweise in drei Haupttypen unterteilt: direktes Modell, Transfer- und Interlingua-Modell (Hutchins & Somers 1992: 71; Stein 2009: 8; Kenny 2019: 434; Bai & Yu 2023: 195f.).

Beim direkten Modell handelt es sich primär um lexikalischen Transfer zwischen Ausgangs- und Zielsprache. Ausgangssätze werden dabei als Wortfolgen verarbeitet. Für diese werden über ein bilinguales Wörterbuch zielsprachige Äquivalente bestimmt, die anschließend in begrenztem Umfang morphologisch angepasst und durch Umstellungsmechanismen in eine

zielsprachliche Reihenfolge gebracht werden. Das hierfür benötigte linguistische Wissen ist entweder in den Wörterbucheinträgen hinterlegt oder wird in Form von algorithmischen Prozeduren beschrieben und im Programmcode umgesetzt. Der direkte Ansatz eignet sich damit vor allem für eine begrenzte Zahl von Satzmustern und wird insbesondere für Sprachpaare als geeignet beschrieben, deren Syntax in zentralen Punkten vergleichsweise ähnlich ist. Zugleich bleibt der Übersetzungsprozess in diesem Modell stark vereinfacht (Bai & Yu 2023: 195). Die Schwächen des direkten Ansatzes werden besonders deutlich darin, dass die Ausgabe oft einer Wort-für-Wort-Übertragung mit nur minimaler Umstellung entspricht. Dadurch entstehen häufig Fehler in der Wortwahl und Satzstrukturen, die sich zu eng an der Syntax des Ausgangstextes orientieren (Hutchins & Somers 1992: 72).

Das Transfermodell lässt sich als Weiterentwicklung des direkten Ansatzes verstehen und geht davon aus, dass die Übersetzungsqualität steigt, wenn der Ausgangstext zunächst linguistisch analysiert wird (Kenny 2019: 435). In der Analysephase wird der Ausgangstext syntaktisch und gegebenenfalls auch semantisch verarbeitet. Ergebnis ist eine Zwischenrepräsentation des Ausgangssatzes, die zwischensprachliche Unterschiede teilweise neutralisiert (Schwartz 2018: 173; Kenny 2019: 435). In der Transferphase wird diese ausgangssprachliche Zwischenrepräsentation in eine zielsprachliche Repräsentation übertragen. Auf dieser Grundlage wird in der abschließenden Generierungsphase schließlich ein Zielsatz in seiner endgültigen Form erzeugt (Hutchins & Somers 1992: 75; Schwartz 2018: 173; Kenny 2019: 435).

Das Interlingua-Modell folgt grundsätzlich derselben Analyselogik wie das Transfermodell. Die Analyse erfolgt zunächst vergleichbar, zielt aber idealerweise auf eine Bedeutungsrepräsentation, die keine ausgangssprachenspezifischen Merkmale mehr enthält. Aus dieser nicht an eine konkrete natürliche Sprache gebundenen Repräsentation kann anschließend ein Satz in unterschiedlichen Zielsprachen generiert werden (Hutchins & Somers 1992: 72; Schwartz 2018: 174; Kenny 2019: 435). Damit ist dieses Modell auf eine universelle Notation angewiesen, mit deren Hilfe Bedeutungen so ausgedrückt werden können, dass sie prinzipiell für alle im System berücksichtigten Sprachen generierbar sind (Kenny 2019: 435). Schwartz (2018: 174) betont, dass Interlingua-Systeme zwar eine strikt sprachunabhängige Zwischenebene anstreben, die realisierten Repräsentationen dieses Ziel in der Praxis aber nur unvollständig erfüllen.

Trotz seiner langen Dominanz ist regelbasierte maschinelle Übersetzung mit Grenzen konfrontiert, die unmittelbar aus der Komplexität natürlicher Sprachen und des Übersetzungsprozesses resultieren. In der Forschung werden insbesondere die unzureichende Regel- und

Wörterbuchabdeckung, der hohe Aufwand und die Kosten für deren Erstellung sowie Schwierigkeiten im Umgang mit Mehrdeutigkeit und idiomatischen Ausdrücken hervorgehoben (Bai & Yu 2023: 199). Damit ist RBMÜ insgesamt ressourcenintensiv, da umfangreiche Regeln je Sprachpaar von qualifizierten Linguist*innen entwickelt werden müssen (Kenny 2022: 35).

3.2.2 Statistische maschinelle Übersetzung

Statistische maschinelle Übersetzung (SMÜ, engl. SMT) war bis in die Mitte der 2010er Jahre hinein über längere Zeit der prägende Ansatz in der MÜ-Forschung und -Praxis (Kenny 2022: 37). Die Grundidee statistischer maschineller Übersetzung lässt sich bis zu Warren Weavers Memorandum von 1949 zurückverfolgen. Eine praktische Umsetzung statistischer Ansätze erfolgte jedoch erst deutlich später. Mit zunehmender Computerleistung und der wachsenden Verfügbarkeit umfangreicher maschinenlesbarer Texte ergab sich Ende der 1980er und Anfang der 1990er Jahre eine neue Ausgangslage. Mit dem Projekt *Candide* griff eine Forschungsgruppe am IBM TJ Watson Research Center statistische Verfahren für die maschinelle Übersetzung erneut auf und entwickelte sie gezielt weiter (Liu & Zhang 2023: 208). *Candide* markiert in diesem Zusammenhang einen Wendepunkt gegenüber regelbasierten Verfahren, weil Analyse und Generierung als statistische Prozesse umgesetzt wurden, ohne dass explizite linguistische Regeln zur Anwendung kamen (Hutchins 2023: 136).

Im Rahmen der SMÜ wählt das System für einen gegebenen Ausgangssatz die wahrscheinlichste Zielübersetzung aus. Grundlage dafür sind ein oder mehrere probabilistische Übersetzungsmodelle, die aus Parallelkorpora gelernt werden sowie ein Wahrscheinlichkeitsmodell der Zielsprache, das auf einem monolingualen Korpus der Zielsprache basiert (Kenny 2019: 435). Die Modellbildung erfolgt in mehreren Phasen, die *Training*, *Tuning* und *Decoding* umfassen. Das Erlernen der Modelle erfolgt in einer ersten Phase, die als *Training* bezeichnet wird. In einer zweiten Phase, dem *Tuning*, werden die Gewichtungen der einzelnen Modelle so kalibriert, dass das Gesamtergebnis der Übersetzung insgesamt optimiert wird (Kenny 2019: 436). In der abschließenden Phase *Decoding* generiert das System eine große Zahl möglicher Zieltextvarianten und wählt jene Hypothese aus, die unter Berücksichtigung der gelernten Modelle und ihrer Gewichtung als am wahrscheinlichsten gilt (Kenny 2019: 436; Liu & Zhang 2023: 209).

Historisch lassen sich innerhalb der SMÜ unterschiedliche Modellvarianten unterscheiden, wobei frühe Ansätze wortbasiert waren und spätere Systeme häufig phrasenbasiert arbeiten. Wortbasierte SMÜ-Ansätze gehören zu den früheren statistischen Modellen, in denen das

Wort als grundlegende Übersetzungseinheit dient (Stein 2009: 11; Liu & Zhang 2023: 209). Wortalignments spielen dabei eine wichtige Rolle, da sie Zuordnungen zwischen Wörtern eines Ausgangs- und Zielsatzes in einem Parallelkorpus herstellen (Liu & Zhang 2023: 209). Die Beschränkung auf Einzelwörter bringt mehrere Probleme mit sich. Mehrwortausdrücke und zusammengehörige Wortgruppen werden nicht als Einheit erfasst. Besonders deutlich wird dies bei Einheiten, deren Bestandteile im Satz getrennt auftreten können, z. B. bei trennbaren Verben, da die isolierte Betrachtung einzelner Wortteile zu Bedeutungsverschiebungen führen kann (Stein 2009: 12). Hinzu kommt ein morphologisches Problem: Morphologisch verwandte Wörter werden als getrennte Symbole behandelt (Manning & Schütze 1999: 491). Stein (2009: 11) weist zudem darauf hin, dass wortbasierte Modelle insbesondere dort an Grenzen stoßen, wo eine Bedeutung in einer Sprache durch mehrere Wörter ausgedrückt wird, in der anderen jedoch durch ein einzelnes Wort realisiert ist. Darüber hinaus können sich die Einschränkungen bei Sprachpaaren mit deutlich unterschiedlicher Syntax bzw. Wortstellung besonders stark auswirken, da die Übersetzung nicht nur lexikalische Entsprechungen, sondern auch Wortstellungsanpassungen erfordert (2009: 12).

Als Reaktion auf diese Einschränkungen etablierten sich phrasenbasierte SMÜ-Modelle. Die Grundidee besteht darin, nicht einzelne Wörter, sondern Wortsequenzen (*phrases*) als Übersetzungseinheiten zu behandeln und diese Einheiten auch als Ganzes zu übersetzen und falls erforderlich in ihrer Reihenfolge im Satz umzuordnen (Liu & Zhang 2023: 211). Die dabei identifizierten Zuordnungen werden typischerweise in einer sogenannten *phrase table* erfasst (Kenny 2022: 36). Der Begriff *phrase* ist dabei nicht im streng linguistischen Sinne zu verstehen, sondern bezeichnet zusammenhängende Wortsequenzen unterschiedlicher Länge (n-Gramme), die im Trainingsmaterial vorkommen (2022 :37). Dadurch wird es möglich, sowohl mehrere Wörter mit einem zu übersetzen als auch vice versa, was wortbasierte Modelle nur begrenzt leisten. Darüber hinaus liefert die Betrachtung von Wortsequenzen einen größeren Kontext, der bei Mehrdeutigkeit die Wahl der passenden Übersetzung unterstützen kann. Je nach System und Sequenzlänge kann dieser Ansatz außerdem Unterschiede in der Wortstellung bzw. Satzstruktur besser ausgleichen (Stein 2009: 12).

Die statistische maschinelle Übersetzung prägte die maschinelle Übersetzung insbesondere zwischen den 2000er-Jahren und der Mitte der 2010er-Jahre. Mit dem Aufkommen neuronaler Modelle setzte jedoch ein Paradigmenwechsel ein, in dessen Verlauf SMÜ zunehmend von neuronalen maschinellen Ansätzen abgelöst wurde (Kenny 2022: 39).

3.2.3 Neuronale maschinelle Übersetzung

Im Unterschied zu regelbasierten und statistischen Ansätzen beruht die neuronale maschinelle Übersetzung (NMÜ, engl. NMT) auf Deep-Learning-Verfahren. NMÜ-Modelle werden auf Parallelkorpora trainiert, in denen Ausgangstexte und ihre Übersetzungen satzpaarweise einander zugeordnet sind. Kenny zufolge setzte sich NMÜ um 2016 als *State of the Art* in der automatischen Übersetzung durch (Kenny 2025: 92). Im Folgenden werden zunächst die grundlegenden Bausteine neuronaler Netze sowie zentrale Konzepte wie *Embeddings* und *Attention* skizziert, bevor die Encoder–Decoder- und Transformer-Architektur erläutert wird.

Künstliche neuronale Netze bestehen aus vielen, teils Tausenden oder Millionen miteinander verbundenen Neuronen, die die zentralen Bausteine dieser Netze bilden. Die Aktivierung der Neuronen, also wie stark sie angeregt oder gehemmt sind, hängt von den eingehenden Signalen anderer Neuronen sowie von der Stärke der Verbindungen (den Gewichten), über die diese Signale weitergegeben werden (Pérez-Ortiz et al. 2022: 143f.). In einem neuronalen Netz übernehmen unterschiedliche Neuronengruppen unterschiedliche Rollen. Einige Neuronen nehmen externe Signale als Eingaben (*inputs*) auf. Eine weitere Gruppe von Neuronen bilden sogenannte versteckte Neuronen (*hidden neurons*), die Signale ausschließlich von anderen Neuronen erhalten. Ausgabeneuronen (*output neurons*) repräsentieren schließlich die Ausgabe bzw. Lösung der Aufgabe (2022: 146).

Beim Aufbau eines neuronalen Netzes wird zunächst seine Architektur festgelegt, also die grundlegende Struktur, mit der ein bestimmtes Problem gelöst werden soll. Dazu gehören Entscheidungen darüber, wie viele Neuronen das Netz umfasst, wie sie miteinander verbunden sind und welche Neuronen externe Eingaben aufnehmen bzw. welche als Ausgabeneuronen fungieren. Welche konkreten Berechnungen das Netz tatsächlich ausführt, hängt von den Gewichten der Verbindungen ab. Diese Gewichte können durch Training anhand von Beispielen bestimmt werden (2022: 146).

Damit ein neuronales Netz Sprache verarbeiten kann, werden Wörter in der NMÜ als Vektoren repräsentiert, also als Zahlenfolgen fester Länge (Kenny 2022: 41). Die vektorbasierten Wortrepräsentationen, die das Modell im Lernprozess erwirbt, werden als *Word Embeddings* bezeichnet. Dabei gilt: Wörter, die in den Trainingsdaten wiederholt in gleichen oder ähnlichen Kontexten vorkommen, erhalten tendenziell ähnliche *Embeddings*. *Word Embeddings* entstehen nicht in einem einzigen Schritt, sondern werden über aufeinanderfolgende Schichten (*layers*)

des Netzes schrittweise aufgebaut. Ein künstliches neuronales Netz, das zwischen seinen äußeren Schichten, also Eingabe- und Ausgabeschicht, mehrere dazwischenliegende Schichten aufweist, wird als tiefes neuronales Netz (*deep neural network*) bezeichnet (Kenny 2022: 42).

Darüber hinaus muss im Kontext der NMÜ der *Attention*-Mechanismus erwähnt werden, da er eine wesentliche Rolle spielt und dem neuronalen Netzwerk ermöglicht, kontextualisierte Wortrepräsentationen (*contextual word embeddings*) zu berechnen. Die Einbettung eines Wortes wird dabei in Abhängigkeit vom jeweiligen Satzzusammenhang berechnet, sodass die Repräsentation die kontextspezifische Bedeutung des Wortes widerspiegelt. *Attention* beruht auf mathematischen Operationen, die im Rahmen des Trainings vom Lernalgorithmus erlernt werden (Pérez-Ortiz et al. 2022: 153).

Hervorzuheben ist, dass kontextualisierte *Word Embeddings* nicht nur dazu führen, dass unterschiedliche Wortbedeutungen unterschiedliche Repräsentationen erhalten. Auch bei gleicher Bedeutung können sich die Repräsentationen unterscheiden, wenn sich der Kontext eines Wortes oder die Wortstellung verändert. Entsprechend können identische Wortformen innerhalb eines Satzes unterschiedliche kontextualisierte Vektorrepräsentationen erhalten, weil ihr jeweiliger Kontext verschieden ist (Pérez-Ortiz et al. 2022: 154).

Zur Berechnung kontextualisierter *Word Embeddings* wird zunächst von nicht-kontextualisierten *Word Embeddings* ausgegangen: Für einen Satz mit n -Wörtern liegen entsprechend n -Vektorrepräsentationen vor, die als Ausgangsbasis dienen. Anschließend erzeugt das neuronale Netz mittels mathematischer Operationen einen *Attention*-Vektor, der für jedes Wort einen *Attention*-Grad angibt. Diese Werte drücken aus, in welchem Ausmaß die einzelnen Wörter bei der Berechnung der kontextualisierten Repräsentation eines bestimmten Wortes berücksichtigt werden (Pérez-Ortiz et al. 2022: 153). Diese kontextualisierten Repräsentationen entstehen im Rahmen des Kodierungsprozesses, der in vielen NMÜ-Systemen über eine Encoder–Decoder-Architektur realisiert wird (vgl. Pérez-Ortiz et al. 2022: 143). Der Übersetzungsprozess in NMÜ wird als die Abfolge einer Kodierungs- und einer Dekodierungsphase beschrieben. In der Kodierung werden zunächst für die Wörter des Ausgangssatzes neuronale Repräsentationen gebildet, die anschließend unter Berücksichtigung des gesamten Satzkontexts zu kontextualisierten Repräsentationen transformiert werden. Auf dieser Grundlage erzeugt die Dekodierung die Zielsequenz schrittweise, indem das Modell das jeweils nächste Zielwort aus dem bereits generierten Kontext vorhersagt. Encoder und Decoder sind dabei als zwei neuronale Teilmodelle in einem gemeinsamen Gesamtsystem gekoppelt (Pérez-Ortiz et al. 2022: 143).

Viele NMÜ-Systeme basieren auf der Transformer-Architektur, die als Encoder–Decoder-Modell auf Basis von *Attention* beschrieben wird. Das Modell besteht aus zwei Modulen: Das erste Modul (Encoder) erzeugt kontextualisierte Wortrepräsentationen für die Wörter des Ausgangssatzes, während das zweite Modul (Decoder) den Zielsatz schrittweise generiert, indem er nacheinander Zielwörter vorhersagt (Pérez-Ortiz et al. 2022: 156). Bei jeder Vorhersage berücksichtigt der Decoder mittels *Attention* sowohl die Repräsentationen des gesamten Ausgangssatzes als auch die Repräsentationen der bereits generierten Zielwörter (Pérez-Ortiz et al. 2022: 156f.).

Eine wesentliche Voraussetzung für das Training von NMÜ-Systemen ist die Verfügbarkeit umfangreicher Trainingsdaten in Form von Parallelkorpora. In vielen Sprachkombinationen sowie in bestimmten Textsorten fehlt jedoch eine ausreichend große und geeignete Datenbasis, sodass der praktische Einsatz von NMÜ nur eingeschränkt möglich ist (Pérez-Ortiz et al. 2022: 162).

Die neuronale maschinelle Übersetzung hat sich seit der Mitte der 2010er Jahre als dominierendes Paradigma der maschinellen Übersetzung etabliert (Kenny 2025: 92). Neuere Entwicklungen zeigen jedoch, dass sich aktuelle Übersetzungssysteme nicht mehr ausschließlich im Rahmen klassischer NMÜ beschreiben lassen. Seit Ende 2022 werden große Sprachmodelle wie ChatGPT von OpenAI zunehmend auch für Übersetzungsaufgaben eingesetzt. Mit DeepL wurde 2024 bereits ein auf LLM-Infrastruktur basierendes, für Übersetzungen spezialisiertes Next-Generation-Modell eingeführt (Kenny 2025: 92; DeepL 2024). Vor diesem Hintergrund werden im Folgenden LLM-gestützte Übersetzungssysteme näher eingeordnet.

3.2.3.1 LLM-gestützte Übersetzungssysteme

Die Abkürzung LLM steht für Large Language Model. Nach Moorkens et al. (2025: 188) handelt es sich dabei um Modelle der künstlichen Intelligenz, die mithilfe von Deep-Learning-Verfahren auf großen Textmengen trainiert werden und in der Lage sind, menschenähnliche Texte zu generieren.

Nach Kenny (2025: 92) rückten große Sprachmodelle insbesondere Ende 2022 stärker in den öffentlichen Fokus, als OpenAI mit ChatGPT einen breiteren Zugang zu GPT-3 ermöglichte. Ein wesentlicher Unterschied zwischen neuronaler maschineller Übersetzung und großen Sprachmodellen besteht vor allem darin, dass NMÜ-Modelle auf eine klar umrissene Aufgabe, nämlich die automatische Übersetzung, spezialisiert sind, während LLMs nicht auf eine

einzelne Aufgabe beschränkt sind und für eine Vielzahl unterschiedlicher Aufgaben eingesetzt werden können (Kenny 2025: 92f.). Ein weiterer wichtiger Unterschied betrifft die Architektur sowie die Trainingslogik der Systeme. Große Sprachmodelle unterscheiden sich von NMÜ-Systemen in diesen beiden Aspekten grundlegend. Während neuronale Übersetzungssysteme in der Regel auf einer Encoder-Decoder-Struktur beruhen, sind LLMs typischerweise Decoder-basiert. Darüber hinaus werden LLMs nicht mit multilingualen Parallelkorpora, sondern überwiegend mit monolingualen Daten trainiert (Walter 2026: 31).

Large Language Models gelten als eine Form generativer KI, da sie auf der Grundlage von Eingaben neue Inhalte erzeugen können (Kenny 2025: 93). Moorkens et al. (2025: 189) betonen jedoch, dass LLMs ihre Ausgaben auf der Grundlage statistisch gelernter Muster erzeugen, indem sie wahrscheinliche Fortsetzungen von Wortsequenzen vorhersagen. Ihre Funktionsweise beruht damit nicht auf Verstehen im menschlichen Sinn, sondern auf Mustererkennung in großen Datenbeständen. Obwohl große Sprachmodelle wie ChatGPT nicht speziell für Übersetzungszwecke konzipiert wurden, zeigte sich, wie Walter (2026: 31) anmerkt, rasch, dass sie für Übersetzungsaufgaben eingesetzt werden können und dabei eine Qualität erreichen, die mit jener von NMÜ-Systemen vergleichbar ist.

Die wachsende Entwicklung LLM-gestützter Übersetzungssysteme spiegelt sich auch in einer zunehmenden Zahl einschlägiger Studien wider, die ihre Eignung für Übersetzungsaufgaben untersuchen. Auf diese Forschung wird im Kapitel zum aktuellen Stand der Forschung näher eingegangen. Im Folgenden werden zwei Systeme näher vorgestellt: DeepL einschließlich der kostenpflichtigen Variante DeepL Pro sowie ChatGPT auf Basis von GPT-5.3, da beide Systeme (DeepL Pro und GPT-5.3) bei der Analyse der in der vorliegenden Arbeit untersuchten Texte zum Einsatz kommen.

3.2.3.1.1 DeepL

DeepL (DeepL SE, Köln) ist Anbieter eines kommerziellen Systems zur maschinellen Übersetzung und stellt mit dem DeepL Translator sowie ergänzenden Produkten und Integrationslösungen ein breites Portfolio für professionelle Sprachverarbeitung bereit (DeepL o.J.a; DeepL o.J.e). In einem von DeepL (2021) veröffentlichten technischen Blogbeitrag beschreibt das Unternehmen die Übersetzungserzeugung als den Einsatz künstlicher neuronaler Netze, die auf vielen Millionen übersetzter Texte trainiert werden. DeepL führt Qualitätsfortschritte dabei auf

vier Optimierungsfelder zurück, nämlich Verbesserungen in den Bereichen der Netzwerkarchitektur, der Trainingsdaten, der Trainingsmethodik sowie der Netzwerkgröße. Diese Einordnung spiegelt die Perspektive des Unternehmens auf die Systementwicklung wider (DeepL 2021).

Im Juli 2024 kündigte DeepL (2024) eine neue Modellgeneration an, die das Unternehmen als LLM-basiert und auf Übersetzung sowie Editing ausgerichtet positioniert. DeepL beschreibt diese Modellgeneration als Zusammenspiel dreier Komponenten. Erstens beschreibt DeepL (2024) ein spezialisiertes, auf Sprache feinabgestimmtes LLM, das nach Unternehmensangaben „menschlicher“ wirkende Formulierungen unterstützen und zugleich das Risiko von Halluzinationen reduzieren soll. Zweitens verweist DeepL (2024) auf eine proprietäre Datenbasis, die über mehrere Jahre für die Übersetzungserstellung aufgebaut worden ist. Drittens hebt DeepL (2024) das *Human Model Tutoring* hervor, also die gezielte Anleitung des Modells durch tausende ausgewählte Sprachexpert*innen.

Nach Angaben des Unternehmens kann im DeepL Translator je nach Sprachpaar zwischen dem *Classic Language Model* und dem *Next-Generation Language Model* gewählt werden. Das *Classic Language Model* beschreibt DeepL als klassische KI-Neuralnetzarchitektur für Übersetzung und ordnet es weiterhin als sehr präzise und nuanciert ein. Das *Next-Generation Model* verortet DeepL in einer LLM-basierten Infrastruktur und betont, dass hierfür unternehmenseigene, für Übersetzung spezialisierte LLMs eingesetzt werden. Damit verbindet DeepL vor allem eine Qualitätssteigerung bei der Übersetzung längerer Eingabetexte (DeepL o.J.b).

DeepL bietet mehrere kostenpflichtige Abonnements, die sich im Funktionsumfang und in den Nutzungsmöglichkeiten unterscheiden, beispielsweise hinsichtlich Team-/Nutzerverwaltung, Glossarfunktionen, die Integration in Computer-Assisted-Translation-Tools (CAT-Tools) sowie weiterer administrativer Einstellungen (DeepL o.J.c). DeepL Pro ist laut DeepL (o.J.c) insbesondere für Einzelpersonen und Unternehmen vorgesehen, die schnelle und präzise Übersetzungen mit maximaler Datensicherheit benötigen, sowie für professionelle Übersetzer*innen, die DeepL in ihren Workflow integrieren möchten. Für die vorliegende Arbeit ist insbesondere DeepL Pro (Individual, Next-Generation Language Model) relevant, da diese DeepL-Version im Analyseteil zur Übersetzung literarischer Texte eingesetzt wurde.

Ein zentraler Unterschied zwischen der kostenfreien Nutzung und DeepL Pro liegt im Umgang mit Datenschutz und Vertraulichkeit (DeepL o.J.b). Für die kostenfreien Dienste weist

DeepL darauf hin, dass übermittelte Inhalte sowie resultierende Übersetzungen bzw. Verbesserungen in die Optimierung und das Training der verwendeten Modelle einbezogen werden können. Für DeepL Translator Pro betont DeepL (o.J.d) die Ausrichtung auf professionelle Nutzung und erhöhte Datensicherheit. Übermittelte Inhalte werden demnach nur soweit verarbeitet, wie es für die Bereitstellung des Dienstes nötig ist, anschließend gelöscht und nicht zur Verbesserung bzw. Weiterentwicklung des Dienstes verwendet. Darüber hinaus muss erwähnt werden, dass DeepL Pro Nutzer*innen im Translator für bestimmte Sprachpaare zwischen dem *Classic Language Model* und dem *Next-Generation Model* wählen können. Die Umschaltung erfolgt über ein Dropdown-Menü in der Toolbar (DeepL o.J.b).

3.2.3.1.2 ChatGPT

ChatGPT ist ein dialogbasierter KI-Assistent, der laut OpenAI bei einer Vielzahl sprachbezogener Aufgaben eingesetzt werden kann, darunter das Beantworten von Fragen, das Verfassen und Überarbeiten von Texten sowie das Übersetzen zwischen Sprachen (OpenAI o.J.b). Die Abkürzung GPT steht bei OpenAI für *Generative Pre-trained Transformer* (OpenAI Developers o.J.) ChatGPT wurde von OpenAI entwickelt, einem Unternehmen im Bereich der KI-Forschung und -Anwendung (OpenAI o.J.a). OpenAI wurde 2015 als Non-Profit gegründet und etablierte 2019 eine gewinnorientierte Tochtergesellschaft, um Forschung und Deployment auszuweiten (OpenAI o.J.c).

Am 3. März 2026 stellte OpenAI das Modell GPT-5.3 Instant vor. Für die vorliegende Arbeit ist diese GPT-Version relevant, da die untersuchten Übersetzungen mit dieser erstellt wurden. OpenAI (2026) beschreibt GPT-5.3 Instant als Weiterentwicklung, die durch präzisere Antworten, bessere Kontextualisierung sowie einen flüssigeren Gesprächsverlauf gekennzeichnet ist. Außerdem werden Verbesserungen im Bereich der Texterstellung hervorgehoben, insbesondere bei der Erstellung von Prosa mit hoher Ausdrucksstärke, Kreativität und immersiver Wirkung. Darüber hinaus muss angemerkt werden, dass OpenAI in der herangezogenen Beschreibung keine spezifischen Angaben von ChatGPT-5.3 Instant bei Übersetzungsaufgaben macht.

Auch in der translationswissenschaftlichen Forschung wird ChatGPT zunehmend im Hinblick auf seine Eignung für Übersetzungsaufgaben untersucht (Gao et al. 2024; Du et al. 2025; Kong & Macken 2025). Die Ergebnisse deuten darauf hin, dass ChatGPT-Modelle trotz fortbestehender Schwierigkeiten bei der Interpretation und Übersetzung bestimmter Formulierungen ein gewisses Potenzial für literarische Übersetzungen aufweisen (Gao et al. 2024). Die

Ergebnisse von Du et al. (2025) deuten jedoch darauf hin, dass das in der Studie eingesetzte ChatGPT-Modell GPT-4o in seiner Gesamtleistung weiterhin nicht das Niveau menschlicher Übersetzungen erreicht.

Im Unterschied zu spezialisierten Übersetzungssystemen ist ChatGPT nicht ausschließlich für Übersetzungsaufgaben konzipiert, sondern als allgemein einsetzbarer KI-Assistent. Für die vorliegende Arbeit ist dabei besonders relevant, dass das System zwar grundsätzlich eine genauere Steuerung des Outputs durch zusätzliche Prompt-Vorgaben erlaubt, diese Möglichkeit im Rahmen der Analyse jedoch bewusst nicht genutzt wurde. Stattdessen erfolgte die Eingabe mit einer knappen Standardanweisung zur Übersetzung ins Deutsche, um eine möglichst hohe Vergleichbarkeit mit den durch DeepL Pro erzeugten Übersetzungen zu gewährleisten.

4 Literarische maschinelle Übersetzung

4.1 Besonderheiten literarischer maschineller Übersetzung

Die literarische maschinelle Übersetzung hat sich als eigener und besonders anspruchsvoller Untersuchungsbereich innerhalb der MÜ-Forschung etabliert (Karakanta et al. 2025: 278). Trotz der bemerkenswerten Fortschritte großer Sprachmodelle in zahlreichen sprachbezogenen Aufgaben ist ihre Eignung für kreative und stark nuancierte Bereiche wie die literarische Übersetzung weiterhin nicht abschließend geklärt (Zhang et al. 2025b: 29099). Dieser Untersuchungsbereich ist mit mehreren spezifischen Besonderheiten und Herausforderungen verbunden.

Eine zentrale Besonderheit literarischer Übersetzung besteht darin, dass nicht nur semantische Inhalte übertragen, sondern auch die spezifische Leseerfahrung des Textes bewahrt werden soll. Dementsprechend erfordert literarisches Übersetzen eine besonders bewusste Auswahl aus verschiedenen Übersetzungsmöglichkeiten (Toral & Way 2015: 240). Gerade bei literarischen Texten beschränkt sich die Übersetzung daher nicht auf die möglichst genaue Übertragung von Bedeutungen. Vielmehr müssen auch Stil, Ton und kreative Nuancen des Ausgangstextes berücksichtigt werden, da literarische Texte auf eine Wirkung zielen, die nicht allein informationsbezogen ist (Du et al. 2025: 578). Damit wird deutlich, dass literarische Übersetzung Anforderungen stellt, die über eine rein informationsbezogene Übertragung hinausgehen.

Vor diesem Hintergrund weist Hadley (2023: 99) darauf hin, dass MÜ- und CAT-Technologien in ihrer bisherigen Entwicklung nicht primär auf die Anforderungen literarischer

Übersetzung ausgerichtet waren. Zugleich betont er, dass die Übersetzung literarischer Texte mit besonderen Herausforderungen verbunden ist, die sich unter anderem in den Bereichen Ironie, Stil, Wortspiele, Intertextualität, kulturspezifische Elemente, Register und Dialekt zeigen. Auch Wang et al. (2023: 55) betonen die Relevanz kultureller Aspekte und Redewendungen für die Anpassung von Übersetzungssystemen an literarische Texte.

Auch Brusasco (2022: 142) hebt hervor, dass literarische Übersetzung im Hinblick auf den Gesamttext betrachtet werden muss und beschreibt sie als einen dynamischen Prozess, in dem einzelne Übersetzungsentscheidungen fortlaufend im Hinblick auf den Gesamttext überprüft und gegebenenfalls angepasst werden müssen. Maßgeblich sind dabei Aspekte wie Kohäsion, Kohärenz, Stil und beabsichtigte Wirkung. Wie Wang et al. (2023: 55) hervorheben, setzt die Erstellung kohäsiver und kohärenter Übersetzungen durch maschinelle Übersetzungssysteme ein Verständnis der intendierten Bedeutung sowie der Textstruktur auf Diskursebene voraus.

Ein weiterer zentraler Aspekt betrifft die Stimme der Autor*innen. Wie Taivalkoski-Shilov (2019: 695) hervorhebt, sind literarische Texte häufig durch eine Mehrschichtigkeit von Stimmen und bewusste Ambiguität geprägt, was ihre Übersetzung besonders anspruchsvoll macht. Gerade die Komplexität der Autor*innenstimme und ihre teilweise implizite Konstruktion zeigen die Grenzen automatischer literarischer Übersetzung (2019: 696). Zugleich weist Taivalkoski-Shilov (2019: 697) darauf hin, dass bei maschinell gestützten Verfahren das Risiko besteht, dass der Stil einer einzelnen Autor*in uneinheitlicher wird oder die Stile verschiedener Autor*innen einander angeglichen werden.

Die Übersetzung literarischer Texte ist mit spezifischen Herausforderungen verbunden, da kreative Werke eine hohe sprachliche und stilistische Komplexität aufweisen und zudem die angemessene Wiedergabe der Autor*innenstimme sowie kontextueller Feinheiten erfordern (Karpinska & Iyyer 2023: 419). Die literarische maschinelle Übersetzung ist auch deshalb besonders anspruchsvoll, weil kreative Texte nicht nur sprachlich komplex sind, sondern auch durch hohe interne Variabilität, unterschiedliche kulturelle, zeitliche und individuelle stilistische Prägungen sowie komplexe referenzielle Bezüge über Satzgrenzen hinweg gekennzeichnet sind (Hadley et al. 2022: 6). Hinzu kommt, dass sich literarische Übersetzung nicht auf die interlinguale Übertragung von Inhalten beschränkt, sondern zugleich die Rekonstruktion der ästhetischen Wirkung und stilistischen Eigenart des Ausgangstextes umfasst (Kong & Macken 2025: 52). Gerade beim Einsatz maschineller Übersetzung im Bereich kreativer Texte stellt sich

daher die Frage, inwieweit kulturelle und kreative Ausdrucksformen in der Zielübersetzung angemessen erfasst werden können (2025: 53).

Darüber hinaus stellt auch die Wiedergabe metaphorischer Sprache eine besondere Herausforderung dar. Wie Dorst (2023: 174) hervorhebt, gelten Metaphern in der Übersetzungsforschung als besonders anspruchsvoll, da sie sprachlich und kulturell eingebettet sind. Dorst (2023: 182) weist zugleich darauf hin, dass Schwierigkeiten bei der neuronalen maschinellen Übersetzung von Metaphern nicht pauschal vorausgesetzt werden sollten, sondern auch mit ihrer sprachlichen Realisierung, der Satzstruktur und der Art der Metapher zusammenhängen können.

Ein weiterer zentraler Faktor ist die begrenzte Verfügbarkeit geeigneter Trainingsdaten im literarischen Bereich. Wang et al. (2023: 55) weisen darauf hin, dass urheberrechtlich geschützte, diskursbasierte Parallelkorpora für literarische Texte nur eingeschränkt verfügbar sind. Dadurch wird die Entwicklung maschineller Übersetzungssysteme erschwert, die mit den sprachlichen, stilistischen und strukturellen Besonderheiten literarischer Texte angemessen umgehen können (2023: 55f.). Daraus lässt sich ableiten, dass die Grenzen literarischer maschineller Übersetzung nicht ausschließlich auf Modellarchitekturen zurückzuführen sind, sondern auch mit der eingeschränkten Verfügbarkeit geeigneter Trainingsdaten zusammenhängen. Wang et al. (2023: 55) betonen zudem, dass literarische Texte aufgrund ihres größeren Kontextumfangs besondere Anforderungen an maschinelle Übersetzungssysteme stellen. Damit Übersetzungskonsistenz und angemessene lexikalische Entscheidungen über längere Textabschnitte hinweg gewährleistet werden können, müssen solche Systeme in der Lage sein, Langkontext zu modellieren.

Für die vorliegende Arbeit folgt daraus, dass literarische Übersetzungen nicht auf eine reine inhaltliche Äquivalenz reduziert werden können. Zusammenfassend lässt sich festhalten, dass beim maschinellen Übersetzen literarischer Texte zahlreiche Aspekte eine wesentliche Rolle spielen. Da diese jedoch nicht alle im Rahmen der vorliegenden Untersuchung gleichermaßen berücksichtigt werden können, konzentriert sich die Analyse auf eine begrenzte Zahl zentraler Kriterien, nämlich lexikalische Vielfalt, lokale Kohäsion, Kohärenz sowie Stil und Register.

4.2 Herausforderungen der Evaluation literarischer maschineller Übersetzung

Die Evaluation literarischer maschineller Übersetzung stellt ein besonders anspruchsvolles Feld dar. Dies liegt nicht nur an der Komplexität literarischer Texte selbst, sondern auch daran, dass angemessene Bewertungsverfahren mit erheblichem Aufwand verbunden sind. Die menschliche Evaluation, die zur Bewertung literarischer Qualität besonders geeignet ist, ist bei größeren Textmengen aus zeitlichen und vor allem ökonomischen Gründen nur begrenzt umsetzbar (Zhang et al. 2025b: 29099).

In diesem Zusammenhang heben Zhang et al. (2025a: 10962) hervor, dass die Evaluation literarischer maschineller Übersetzung bislang überwiegend auf unterschiedlichen Formen menschlicher Bewertung beruht, wobei zugleich offen bleibt, inwieweit etablierte Verfahren wie MQM, die sich für nicht-literarische Texte als wirksam erwiesen haben, auch für literarische Übersetzungen in gleicher Weise geeignet sind. Zhang et al. (2025a: 10969f.) zeigen zudem, dass fehlerannotierende Verfahren wie MQM bei der Evaluation literarischer Übersetzungen an Grenzen stoßen. Als Gründe nennen sie erstens die hohe Komplexität des Verfahrens, zweitens das Problem, dass MQM bestimmte in literarischen Übersetzungen bewusst gewählte zielsprachliche Äquivalenzen als Fehler erfassen kann, und drittens die Tatsache, dass aktuelle LLM-Systeme insgesamt weniger klar identifizierbare Fehler produzieren als frühere MÜ-Systeme. Damit wird deutlich, dass etablierte Fehlertaxonomien für literarische Texte nur eingeschränkt geeignet sind.

Ein weiterer wesentlicher Aspekt ist die Analyseebene. Wie Karpinska & Iyyer (2023: 421) zeigen, kann literarische maschinelle Übersetzung nicht sinnvoll auf eine rein satzweise Betrachtung reduziert werden. Da die intendierte Bedeutung oder der Stil häufig aus einem einzelnen Satz nicht eindeutig hervorgehen und literarische Übersetzungen zudem Umstellungen über Satz- und Absatzgrenzen hinweg erfordern können, behandeln Karpinska und Iyyer den Absatz als minimale diskursive Einheit. In ähnlicher Weise betonen auch Fonteyne et al. (2020: 3790), dass für die Beurteilung literarischer Übersetzungen eine dokumentbasierte Analyse besonders wichtig ist, da Aspekte wie Kohärenz und Kohäsion auf Satzebene nur eingeschränkt erfasst werden können. Literarische maschinelle Übersetzung muss daher im größeren textuellen Zusammenhang bewertet werden.

Die Bewertung literarischer maschineller Übersetzungen stellt zudem eine besondere methodische Herausforderung dar. Wie Wang et al. (2023: 55) hervorheben, lässt sich die Qualität literarischer Texte nicht angemessen durch eine automatische Metrik auf Grundlage nur

einer einzelnen Referenzübersetzung erfassen. Bei der Beurteilung literarischer Übersetzungen müssen neben der inhaltlichen Wiedergabe auch die Textstruktur sowie sprachliche Feinheiten und Komplexitäten des Ausgangstextes berücksichtigt werden. Verfahren, die sich lediglich auf eine einzige Referenzübersetzung stützen, erweisen sich daher als nur eingeschränkt aussagekräftig. Sinnvoll erscheint stattdessen eine Kombination aus menschlicher Evaluation, klar definierten Fehlertypologien und ergänzenden automatischen Verfahren. In ähnlicher Weise weist auch Matusov (2019: 17) darauf hin, dass automatische Evaluationen auf Basis einer einzelnen, möglicherweise fehlerhaften oder schlecht alignierten Referenzübersetzung unzuverlässig sein können.

Ein Beispiel für die literaturspezifische Erweiterung bestehender Fehlertaxonomien findet sich bei Tezcan et al. (2019). Die erweiterte SCATE-Taxonomie berücksichtigt als menschliche Evaluierungsmethode gezielt Kategorien wie Kohärenz, Stil und Register, die sich auch in den Studien von Fonteyne et al. (2020) und Webster et al. (2020) als besonders relevant für die Bewertung literarischer maschineller Übersetzung erweisen. Aus diesem Grund erscheint die erweiterte SCATE-Taxonomie für die Evaluation literarischer Übersetzung besonders geeignet und bildet zugleich die methodische Grundlage der vorliegenden Analyse.

Neben den Grenzen einzelner Bewertungsverfahren wird die Evaluation literarischer maschineller Übersetzung auch durch methodische Probleme der bisherigen Forschung erschwert. Bislang fehlt eine systematische Erprobung neuerer automatischer Metriken wie XCOMET, GEMBA-MQM oder Prometheus für literarische Texte. Darüber hinaus fehlt es an einer umfassenden vergleichenden Untersuchung menschlicher Bewertungsverfahren wie MQM (Multidimensional Quality Metrics), SQM (Scalar Quality Metrics) und BWS (Best-Worst Scaling) im Bereich der literarischen Übersetzung, sodass bisherige Ergebnisse häufig nur eingeschränkt miteinander vergleichbar sind (Zhang et al. 2025a: 10962). Dies zeigt, dass für die Evaluation literarischer Texte weiterhin methodischer Entwicklungsbedarf besteht.

Neuere Arbeiten reagieren auf diese Probleme mit der Entwicklung literaturspezifischer automatischer Evaluationsverfahren. Wie Zhang et al. (2025b: 29099) argumentieren, sind bestehende Evaluationsmetriken im literarischen Bereich eher auf formale Genauigkeit als auf ästhetische und künstlerische Dimensionen ausgerichtet. Dadurch kann es zu einer systematischen Überschätzung maschineller Übersetzungen gegenüber professionellen menschlichen Übersetzungen kommen, was auch die Bewertung von Übersetzungsqualität und kultureller Authentizität beeinflussen kann. Ein grundlegendes Problem besteht darin, dass literarische

Übersetzung kreative Neuinterpretationen über sprachliche und kulturelle Grenzen hinweg erfordern kann. Damit stoßen insbesondere referenzbasierte Verfahren an Grenzen, da für literarische Texte mehrere legitime Übersetzungslösungen zugleich existieren können. Mit LiTransProQA reagieren Zhang et al. (2025b) auf die aus ihrer Sicht unzureichende Eignung bestehender Bewertungsverfahren für literarische Übersetzung.

Zusammenfassend lässt sich festhalten, dass die Evaluation literarischer maschineller Übersetzung in mehrfacher Hinsicht problematisch ist. Die besonderen Anforderungen literarischer Texte betreffen nicht nur Stil, Autor*innenstimme, kulturelle Nuancen und textübergreifende Kohärenz, sondern stellen auch bestehende Bewertungsverfahren vor erhebliche Herausforderungen (vgl. Hadley et al. 2022: 6). Sowohl fehlerannotierende Verfahren als auch automatische Metriken stoßen bei der Erfassung dieser Dimensionen an Grenzen. Die Bewertung literarischer maschineller Übersetzung muss daher als methodisch komplexes Feld verstanden werden, in dem sowohl textästhetische als auch dokumentenbezogene und kulturgebundene Aspekte berücksichtigt werden müssen (Zhang et al. 2025a; Zhang et al. 2025b).

5 Aktueller Stand der Forschung

Im Folgenden werden ausgewählte Studien zur literarischen maschinellen Übersetzung vorgestellt, die für die vorliegende Arbeit besonders relevant sind. Im Mittelpunkt stehen dabei Untersuchungen zur Qualität maschineller Übersetzungen literarischer Texte, einschließlich LLM-gestützter Ansätze, sowie zu deren Evaluation.

Eine frühe einschlägige Studie ist die Arbeit von Matusov (2019), in der untersucht wird, ob neuronale maschinelle Übersetzung für den literarischen Bereich geeignet ist. Hierfür werden *State of the Art* NMÜ-Modelle für die Sprachrichtungen Englisch–Russisch und Deutsch–Englisch an literarische Texte angepasst. Matusov vergleicht Googles Online-MÜ-System mit einem allgemeinen AppTek-NMÜ-System sowie einer an literarische Texte adaptierten Variante desselben Systems (2019: 10). Für die automatische Qualitätsmessung wurden die Metriken BLEU und TER eingesetzt (2019: 12). Ergänzend entwickelt Matusov (2019: 13) eine auf literarische NMÜ-Ausgaben zugeschnittene Fehlerklassifikation und kombiniert die automatische Evaluation mit einer manuellen Fehleranalyse ausgewählter Textausschnitte durch eine bilinguale Evaluator*in. Die Ergebnisse der Studie deuten darauf hin, dass die aus dem Deutschen ins Englische erstellten NMÜ-Ausgaben in der menschlichen Bewertung insgesamt besser abschneiden als die Übersetzungen aus dem Englischen ins Russische. Zugleich betont Matusov (2019: 17), dass die Qualität der maschinellen Übersetzungen in vielen Fällen

bereits ausreicht, um den Inhalt bzw. den Handlungsverlauf zu verstehen. Darüber hinaus zeigt die Analyse, dass längere Sätze hinsichtlich ihrer syntaktischen Struktur häufig gelungen wiedergegeben werden, sodass sich die erforderliche Nachbearbeitung vielfach auf lokale und geringfügige Korrekturen beschränkt. Matusov zufolge wurden in der manuellen Analyse bis zu 30% der untersuchten Segmente als akzeptabel eingestuft. Abschließend verweist Matusov (2019: 17f.) darauf, dass weiteres Verbesserungspotenzial vor allem in der Berücksichtigung satzübergreifenden Kontexts liegt. Dies betrifft insbesondere die konsistente Übersetzung von Figuren- und Ortsnamen, die korrekte Zuordnung von Pronomen sowie stilistische Entscheidungen, etwa im Hinblick auf formelle und informelle Ausdrucksweisen.

Fonteyne et al. (2020: 3790) untersuchen am Beispiel von Agatha Christies Roman, *The Mysterious Affair at Styles*, inwieweit ein general-domain-NMÜ-System literarische Texte in hoher Qualität und unter Berücksichtigung von Aspekten auf Dokumentenebene wie Kohärenz und Kohäsion übersetzen kann. Zu diesem Zweck wird der gesamte Roman mit Googles NMÜ-System aus dem Englischen ins Niederländische übersetzt und auf Dokumentenebene mithilfe einer von Tezcan et al. (2019) für literarische NMÜ adaptierten SCATE-Taxonomie annotiert (Fonteyne et al. 2020: 3790f.). Die Ergebnisse zeigen, dass 44% der annotierten Sätze keine Fehler enthalten (2020: 3796). Unter den festgestellten Fehlertypen ist Mistranslation am häufigsten, gefolgt von Kohärenz sowie Stil und Register (2020: 3795). Darüber hinaus zeigt die Studie, dass die überwiegende Mehrheit der gemeinsam auftretenden Fehler auf die Kombination von Kohärenz- und Mistranslation-Fehlern zurückgeht (2020: 3796). Besonders relevant ist dabei, dass Kohärenz sowie Stil und Register als Kategorien gezielt in die Taxonomie aufgenommen wurden, um literarische NMÜ auf Dokumentenebene angemessen erfassen zu können, da gerade diese Kategorien für die Bewertung literarischer NMÜ von zentraler Bedeutung sind (2020: 3792). Die feingranulare Fehleranalyse macht zudem sichtbar, dass die Annotator*innen einzelne Fehlerkategorien nicht immer eindeutig voneinander abgrenzen. Besonders Überschneidungen zwischen Kohärenz und Stil & Register, Lexik sowie Fehlübersetzungen und nicht übersetzte Elemente deuten darauf hin, dass einzelne Kategorien in literarischen NMÜ-Ausgaben eng miteinander verbunden sein können (2020: 3796).

Eine weitere einschlägige Studie ist die Arbeit von Webster et al. (2020), in der die Qualität neuronaler maschineller Übersetzung im literarischen Bereich untersucht wird. Dafür werden vier Romane aus dem Englischen ins Niederländische mit zwei Systemen, nämlich Google Translate und DeepL, übersetzt (2020: 1). Die Analyse gliedert sich in zwei Teile: Zum

einen wird eine Fehleranalyse auf der Grundlage einer erweiterten SCATE-Taxonomie durchgeführt, wobei jeweils die ersten 3000 Wörter der Romane annotiert werden. Zum anderen wird eine Key-Feature-Analyse an den Romanen als Ganzen vorgenommen. Die Key-Feature-Analyse dient der Untersuchung von Unterschieden zwischen Humanübersetzungen und maschinell erzeugten Übersetzungen in Bezug auf lexikalische Vielfalt, lokale Kohäsion, syntaktische Divergenz und Stil (2020: 2). Die Ergebnisse der Fehleranalyse zeigen zunächst, dass in den maschinellen Übersetzungen mehr *Fluency*- als *Accuracy*-Fehler auftreten. Für Google Translate fällt der Anteil an *Fluency*-Fehlern etwas höher aus als für DeepL, während DeepL leicht mehr *Accuracy*-Fehler aufweist. Unter den festgestellten Fehlertypen ist Mistranslation die häufigste Kategorie, gefolgt von Kohärenz sowie Stil und Register (2020: 10). Darüber hinaus zeigt die Key-Feature-Analyse, dass Humanübersetzungen lexikalisch vielfältiger sind als die maschinellen Übersetzungen. Webster et al. (2020: 12) weisen darauf hin, dass die Zahl der unterschiedlichen Wörter in den Humanübersetzungen höher ist als in den mit Google Translate und DeepL erzeugten Versionen, was auf einen geringeren lexikalischen Reichtum der NMÜ-Ausgaben hindeutet. Auch im Bereich der Kohäsion erreichen die maschinellen Übersetzungen nicht das Niveau der Humanübersetzungen. Obwohl die maschinellen Übersetzungen das Niveau der lexikalischen Kohäsion des Ausgangstextes aufrechterhalten, erreichen sie weder das lexikalische noch das semantische Kohäsionsniveau der Humanübersetzungen (2020: 14f.). Darüber hinaus zeigt die Studie, dass sich die Humanübersetzungen stilistisch deutlich von den maschinellen Übersetzungen unterscheiden, während zwischen den mit Google Translate und DeepL erzeugten Versionen eine bemerkenswerte stilistische Ähnlichkeit besteht (2020: 17). Insgesamt legt die Studie nahe, dass die untersuchten NMÜ-Systeme im Vergleich zu Humanübersetzungen insbesondere hinsichtlich lexikalischer Vielfalt, Kohäsion und syntaktischer Diversität geringere Werte aufweisen (2020: 18).

Ebenfalls relevant für die vorliegende Arbeit ist die Studie von Zhang et al. (2025a), die die Evaluation literarischer Übersetzungen und die Leistungsfähigkeit verschiedener Bewertungsverfahren systematisch untersucht. Zhang et al. (2025a: 10962f.) stellen ein absatzbasiertes Parallelkorpus vor, das geprüfte menschliche Übersetzungen und Ausgaben von neun maschinellen Übersetzungssystemen für vier Sprachpaare (Deutsch–Englisch, Englisch–Deutsch, Deutsch–Chinesisch und Englisch–Chinesisch) umfasst. Dabei werden kommerzielle MÜ-Systeme wie Google Translate und DeepL berücksichtigt, zwei frühere leistungsstarke Transformermodelle wie NLLB-3.3b und M2M_100-1.3B sowie GPT-4o und mehrere Open-Source-LLMs. Auf dieser Grundlage vergleichen die Autor*innen drei menschliche Bewertungsschemata, nämlich MQM (Multidimensional Quality Metrics), SQM (Scalar Quality Metrics) und

BWS (Best-Worst Scaling) und prüfen zusätzlich mehrere automatische LLM-basierte Metriken, darunter Prometheus 2, XCOMET-XL/XCOMET-XXL und GEMBA-MQM (2025a: 10967). Die Studie zielt darauf ab, die Angemessenheit und Konsistenz menschlicher Evaluationsschemata für literarische Übersetzung, Unterschiede zwischen studentischen und professionellen Bewertungen sowie die Leistungsfähigkeit automatischer Metriken zu untersuchen (2025a: 10961). Die Ergebnisse zeigen, dass die untersuchten menschlichen Übersetzungen die maschinellen Systeme übertreffen. Unter den untersuchten Systemen erzielt GPT-4o die besten Resultate, gefolgt von Google Translate und DeepL. Zugleich macht die Studie deutlich, dass die Wahl des Bewertungsverfahrens erheblichen Einfluss auf die Beurteilung literarischer Übersetzungen hat. Während MQM für den literarischen Bereich zumindest teilweise ungeeignet erscheint, erweist sich BWS als das geeignetste Verfahren für den Vergleich qualitativ hochwertiger Übersetzungen. Die Aussagekraft von SQM hängt demgegenüber stark von der Expertise der Bewertenden ab (2025a: 10962). Automatische LLM-basierte Metriken bleiben insgesamt deutlich hinter den menschlichen Bewertungsverfahren zurück (2025a: 10961). Obwohl GEMBA-MQM unter den automatischen Metriken am besten abschneidet, kann diese Metrik menschliche Übersetzungen und LLM-Ausgaben nur unzureichend voneinander unterscheiden (2025a: 10968). Darüber hinaus stellen die Autor*innen fest, dass LLM-Übersetzungen im Vergleich zu menschlichen Übersetzungen wörtlicher und weniger vielfältig ausfallen, wobei neuere Modelle erkennbar bessere Ergebnisse erzielen und ihre Ausgaben menschlichen Übersetzungen stärker angenähert sind (2025a: 10962).

Neben Studien zum Vergleich menschlicher und maschineller Übersetzungen liegen inzwischen auch Arbeiten vor, die sich mit dem Einsatz von Prompting bei literarischen Texten befassen. Van Egdom et al. (2024) untersuchen, inwieweit sich maschinell übersetzte literarische Texte durch promptgestütztes Post-Editing mit ChatGPT-4 verbessern lassen. Im Fokus stehen dabei insbesondere literarische Nuancen und autor*innenspezifische Stilmerkmale (2024: 10). Die Ergebnisse deuten darauf hin, dass promptgestütztes Post-Editing maschinelle Übersetzungen literarischer Texte nur in begrenztem Maße optimieren kann (2024: 16). Wie van Egdom et al. (2024: 17) anmerken, lassen sich unter bestimmten Bedingungen zwar leichte Verbesserungen erzielen, diese bleiben jedoch marginal. Überraschend ist dabei, dass spezifische, auf literarische Übersetzungen zugeschnittene Anweisungen keine besseren, sondern schwächere Ergebnisse ergaben.

Zusammenfassend lässt sich festhalten, dass die bisherige Forschung sowohl das Potenzial als auch die Grenzen maschineller und LLM-gestützter literarischer Übersetzung sichtbar

macht. Während aktuelle Systeme in Teilbereichen bessere Ergebnisse erzielen, bleiben insbesondere Fragen des Stils, der Kohäsion und der angemessenen Evaluation weiterhin zentral. Gerade im Bereich literarischer Übersetzung besteht daher weiterhin Bedarf an differenzierten Untersuchungen mit engem Bezug zum literarischen Text.

6 Untersuchungskorpus

6.1 Ausgewählte Erzählungen von Anton Tschechow

In der vorliegenden Arbeit werden drei kurze Erzählungen Anton Tschechows untersucht: *Der Tod eines Beamten* und *Der Dicke und der Dünne*, die beide 1883 erschienen, sowie *Ein Chamäleon* aus dem Jahr 1884. Die Texte stammen somit aus Tschechows früher Schaffensphase der 1880er Jahre. Aufgrund ihres vergleichsweise geringen Umfangs eignen sie sich besonders für eine detaillierte qualitative Analyse. Zugleich sind sie in einem ähnlichen zeitlichen und kulturellen Kontext entstanden, wodurch sie eine geeignete Grundlage für den systematischen Vergleich unterschiedlicher Übersetzungsvarianten bilden. Für die Analyse sind die ausgewählten Erzählungen auch deshalb relevant, weil sie durch eine präzise sprachliche Gestaltung, dialogische Passagen, erzählerische und beschreibende Elemente sowie kultur- und kontextspezifische Bezüge gekennzeichnet sind. Gerade solche Merkmale können für LLM-gestützte Übersetzungssysteme eine besondere Herausforderung darstellen, da neben der inhaltlichen Verständlichkeit auch stilistische Nuancen, Register, Figurenrede und Kohärenz im Zieltext berücksichtigt werden müssen.

Als Grundlage für die Analyse der russischen Originaltexte dient die 1950 erschienene Ausgabe von Tschechows *Собрание сочинений* (Gesammelte Werke). Für die Analyse der deutschen Übersetzungen wird die Ausgabe *In der Sommerfrische. Erzählungen 1880–1887* herangezogen. Den deutschen Fassungen liegen Übersetzungen von Vera Bischitzky (*Der Tod eines Beamten*, *Der Dicke und der Dünne*) sowie von Marianne Wiebe (*Ein Chamäleon*) zugrunde, zwei Übersetzerinnen russischsprachiger Literatur. Da die herangezogenen menschlichen Übersetzungen nicht von einer einzelnen Übersetzerin, sondern von zwei unterschiedlichen Übersetzerinnen stammen, sind mögliche Unterschiede in Stil, Register und sprachlicher Gestaltung bei der Interpretation der Ergebnisse zu berücksichtigen.

Im Folgenden wird zunächst die kurze Erzählung als literarische Form eingeordnet. Daran anschließend werden zentrale biographische Angaben zu Anton Tschechow dargestellt, bevor die ausgewählten Erzählungen inhaltlich charakterisiert werden.

6.2 Kurze Erzählung als Untersuchungsgegenstand

Für die literarische Einordnung von Tschechows Erzählungen ist es sinnvoll, sie im Kontext der russischsprachigen Literatur zu betrachten. Wie Lyudmila Parts (2024: 553) hervorhebt, ist die kurze Erzählung (engl. *short story*, russ. *рассказ*) eine Form, die sich nicht leicht eindeutig definieren lässt. Tschechow (1950d: 71) formuliert in einem Brief vom 22. Januar 1888: „В маленьких же рассказах лучше недосказать, чем пересказать [...]“.¹ Auch in einem Brief aus dem Jahr 1889 bringt Tschechow (1950e: 126) sein Ideal der Kürze mit der Formel „Краткость — сестра таланта.“ auf den Punkt.²

Während der englische Begriff *short story* die Länge des Textes betont, ist das russische *рассказ* mit dem Erzählen bzw. Sprechen verbunden. Gerade die Kürze und die Nähe zum mündlichen Erzählen ermöglichten der kurzen Erzählung die Einführung neuer, oft stigmatisierter Themen sowie Experimente mit Form und Sprache (Parts 2024: 553). Im Unterschied zum Roman richtet sich die kurze Erzählung nach Parts vor allem auf die moralische und emotionale Erfahrung des Individuums. In diesem Zusammenhang ist Anton Tschechow für die Entwicklung der modernen kurzen Erzählung von zentraler Bedeutung. Parts zufolge ist er der Autor, der den Übergang zu dem vollzieht, was heute als typische Form der modernen kurzen Erzählung gilt. Zugleich verliehen seine Erzählungen der Gattung in den 1880er und 1890er Jahren eine neue Form, Richtung und ein neues Prestige (2024: 557).

Loehlin (2010: 34) betont, dass Tschechows Erzählungen trotz der stilistischen Entwicklung seines Werks bestimmte konstante Merkmale aufweisen. Dazu zählen ein knapper Erzähleinstieg ohne lange Einführung, ein besonderes Interesse am Innenleben der Figuren, das durch gezielt ausgewählte Details vermittelt wird sowie sorgfältig gestaltete Schlüsse, die überraschend oder verhalten wirken können. Für Tschechows frühe Erzählungen hebt Loehlin besonders ihre Ökonomie hervor, die auch mit den formalen Bedingungen der Zeitschriftenpublikation zusammenhing. Die Notwendigkeit, rasch in die Handlung einzusteigen, führte dabei zu den für Tschechow charakteristischen knappen Anfängen, die er auch später beibehielt.

¹ „Bei kurzen Erzählungen ist es besser, etwas unausgesprochen zu lassen, als zu viel zu erzählen.“ (eigene Übersetzung, E. D.)

² „Die Kürze ist die Schwester des Talents.“ (eigene Übersetzung, E. D.)

6.3 Über den Autor

Anton Pawlowitsch Tschechow wurde 1860 in Russland, in der Stadt Taganrog geboren. Seine Kindheit war von schwierigen familiären Verhältnissen und einer angespannten wirtschaftlichen Situation geprägt, wobei auch die autoritäre Rolle des Vaters das Familienleben stark bestimmte. 1879 zog Tschechow nach Moskau, wo er ein Medizinstudium aufnahm. Bereits während des Studiums begann er, regelmäßig literarische Texte zu veröffentlichen. Diese schriftstellerische Tätigkeit diente zugleich dem Unterhalt seiner Familie. 1884 schloss er das Medizinstudium ab, und im selben Jahr erschien auch seine erste Sammlung von Kurzgeschichten (Bloom 2000: 11).

In den folgenden Jahren nahm Tschechows literarischer Erfolg zu, und 1887 wurde er mit dem Puschkin-Preis ausgezeichnet. Zugleich war sein Leben zunehmend von gesundheitlichen Problemen geprägt, da er bereits früh an Tuberkulose erkrankte. 1892 zog er nach Melichowo, nicht weit von Moskau, wo er sich sowohl literarisch als auch in medizinischen und sozialen Bereichen engagierte. Dort ließ er mehrere Schulen errichten, wirkte während einer Choleraepidemie ärztlich mit und beteiligte sich am Aufbau lokaler Einrichtungen für die Bevölkerung. In diese Lebensphase fällt zudem seine Arbeit an dramatischen Werken, mit *Die Möwe* entstand eines seiner bekanntesten Theaterstücke (Bloom 2000: 12).

In den späten 1890er Jahren verschlechterte sich Tschechows Gesundheitszustand weiter, sodass er 1897 wegen Tuberkulose behandelt werden musste und 1899 nach Jalta zog. Gleichzeitig festigte sich seine Bedeutung als Dramatiker: Nach einer zunächst verhalten aufgenommenen ersten Aufführung wurde *Die Möwe* in der Inszenierung des Moskauer Künstlertheaters erfolgreich rezipiert. In den folgenden Jahren entstanden weitere bedeutende Dramen. Sein Leben wurde jedoch schon 1904 durch seinen frühen Tod im Alter von 44 Jahren beendet (2000: 13).

6.4 Inhaltliche Charakterisierung der Erzählungen

6.4.1 Der Tod eines Beamten

In der Erzählung *Der Tod eines Beamten* schildert Tschechow die eskalierende Wirkung eines zunächst unbedeutenden Vorfalls. Der Staatsdiener Iwan Dmitrijewitsch Tschewjakow niest während einer Theateraufführung versehentlich auf den General Brisshalow. Obwohl der Ge-

neral den Vorfall zunächst nicht ernst nimmt und ihm keine weitere Bedeutung beimisst, beruhigt sich Tscherwjakow nicht. Er entschuldigt sich mehrfach und steigert sich zunehmend in die Vorstellung hinein, einen unverzeihlichen Fehler begangen zu haben. Auch nachdem der General wiederholt zu verstehen gibt, dass der Vorfall belanglos sei, gelingt es Tscherwjakow nicht, die Situation innerlich abzuschließen. Die wachsende Gereiztheit des Generals und Tscherwjakows übersteigerte Angst vor dessen Reaktion führen schließlich dazu, dass aus einer alltäglichen Begebenheit eine psychische und soziale Krise entsteht, die in seinem Zusammenbruch und Tscherwjakows Tod endet (Tschechow 1950a).

Inhaltlich thematisiert die Erzählung vor allem die innere Abhängigkeit des kleinen Beamten von hierarchischen Machtstrukturen. Tscherwjakows Reaktion steht in keinem Verhältnis zum tatsächlichen Anlass, macht jedoch deutlich, wie stark sein Denken von Angst, Unterordnung und Respekt vor Rang und Autorität geprägt ist. Auch die Namengebung ist in diesem Zusammenhang aufschlussreich. Der Familienname Tscherwjakow lässt sich mit dem russischen Wort *tscherwjak* in Verbindung bringen, das auf Deutsch „Wurm“ bedeutet. Auf diesen Bedeutungsaspekt weist auch eine Anmerkung in der von Vera Bischitzky übersetzten Ausgabe hin (Tschechow 2003a: 18). Die Namengebung unterstreicht damit die untergeordnete Stellung der Figur und verstärkt den Eindruck sozialer Kleinheit und innerer Unterwürfigkeit. Das Missverhältnis zwischen dem belanglosen Anlass und dem tragischen Ausgang verleiht der Erzählung eine humoristisch zugespitzte Wirkung. Tschechow verbindet dabei Momente der Komik mit sozialer Kritik. Die wiederholten Entschuldigungsversuche erscheinen zunächst lächerlich, machen zugleich aber die psychische Belastung sichtbar, die aus einem starren Verständnis von Rang und gesellschaftlicher Ordnung hervorgeht.

6.4.2 Der Dicke und der Dünne

In der Erzählung *Der Dicke und der Dünne* beschreibt Tschechow die zufällige Begegnung zweier ehemaliger Schulfreunde an einem Bahnhof. Zunächst verläuft das Wiedersehen herzlich und vertraut. Der „Dicke“ und der „Dünne“ erkennen einander sofort, begrüßen sich mit großer Freude und erinnern sich an gemeinsame Schulzeiten. Im weiteren Gespräch berichtet der „Dünne“ von seinem gegenwärtigen Leben, seiner Familie und seiner eher bescheidenen beruflichen Stellung. Als er jedoch erfährt, dass sein früherer Freund inzwischen einen deutlich höheren Rang erreicht hat, verändert sich sein Verhalten merklich. Aus dem zunächst ungewungenen und familiären Ton wird eine übertriebene, von Ehrfurcht und Unterwürfigkeit geprägte Ansprache. Auch seine Frau und sein Sohn passen sich dieser Situation sichtbar an. Der

„Dicke“ reagiert auf diese plötzliche Veränderung mit Unbehagen, sodass aus der anfänglichen Vertrautheit am Ende Distanz wird (Tschechow 1950b).

Im Mittelpunkt der Erzählung steht vor allem die Orientierung an Rang und gesellschaftlichem Status sowie die daraus folgende Unterordnung gegenüber sozialen Hierarchien. Die deutliche Veränderung im Verhalten des „Dünnen“ macht sichtbar, wie stark soziale Hierarchien persönliche Beziehungen beeinflussen können. Aus einer zunächst freundschaftlichen Begegnung wird innerhalb weniger Minuten eine Szene der Unterordnung, in der die anfängliche Vertrautheit durch demonstrative Ehrfurcht ersetzt wird. Der Kontrast zwischen früherer Nähe und der späteren Distanz trägt zur humoristisch zugespitzten Wirkung der Erzählung bei. Tschechow zeigt damit, wie tief die Orientierung an Rang und Status in das Denken und Auftreten seiner Figuren eingreift. Wie bereits in *Der Tod eines Beamten* wird auch hier sichtbar, in welchem Maß soziale Hierarchien zwischenmenschliche Beziehungen prägen können.

6.4.3 Ein Chamäleon

Mit der Erzählung *Ein Chamäleon* zeigt Tschechow einen Vorfall auf einem Marktplatz, bei dem ein kleiner Hund einen Passanten verletzt haben soll. Der Polizeiinspektor Otschumelow versucht zunächst, die Situation mit amtlicher Strenge zu behandeln, droht damit, den Hund töten zu lassen, und verlangt, den Besitzer des Tieres zur Verantwortung zu ziehen. Im Verlauf des Gesprächs stellt sich jedoch mehrfach die Frage, wem der Hund gehört. Je nachdem, ob das Tier einem gewöhnlichen Bürger oder einem General beziehungsweise dessen Umfeld zugeordnet wird, ändert Otschumelow seine Meinung rasch. Seine Einschätzungen schwanken fortlaufend zwischen Härte, Nachsicht und Unterwürfigkeit. Am Ende wird deutlich, dass nicht die objektive Klärung des Vorfalls, sondern die soziale Stellung des möglichen Besitzers über Otschumelows Verhalten entscheidet (Tschechow 1950c).

Die Erzählung macht deutlich, wie sehr behördliches Handeln von Rang und gesellschaftlichem Status beeinflusst wird. Otschumelow erscheint nicht als neutrale Amtsperson, sondern als Figur, deren Verhalten sich an Rang und Macht orientiert. Die wiederholten Meinungswechsel machen sichtbar, wie wenig seine Urteile auf sachlicher Prüfung beruhen. Dadurch entfaltet die Erzählung eine zugleich komische und kritische Wirkung. Tschechow zeigt, wie flexibel und widersprüchlich Macht ausgeübt wird, wenn sie nicht an Recht und Ordnung, sondern am gesellschaftlichen Status der Beteiligten orientiert.

Wie Chudakov (2000: 6) anmerkt, entwickeln sich Tschechows frühe humoristische Erzählungen nicht aus abstrakten Problemstellungen oder ausführlichen Figurenbiographien, sondern aus konkreten Alltagssituationen. Auch dann, wenn sie satirische oder sozialkritische Züge tragen, bleiben sie an äußerlich alltägliche Situationen gebunden. Der Alltag bildet bei Tschechow dabei nicht nur den Hintergrund des Geschehens, sondern den eigentlichen Kern der Handlung. Dies zeigt sich auch in den in der vorliegenden Arbeit untersuchten Erzählungen, die jeweils von einer konkreten alltäglichen Situation ausgehen und daraus soziale Spannungen sowie hierarchische Konstellationen entfalten.

7 Methode

Dieses Kapitel stellt das methodische Vorgehen der Untersuchung dar. Zunächst werden die erweiterte SCATE-Taxonomie nach Tezcan et al. (2019) sowie die ergänzende Key-Feature-Analyse nach Webster et al. (2020) erläutert. Auf dieser Grundlage wird der für die vorliegende Arbeit adaptierte Analyseansatz beschrieben. Anschließend werden das konkrete Untersuchungsdesign, die analysierten Übersetzungsvarianten, die ausgewählten Fehlerkategorien sowie die externe Überprüfung der Fehlerannotation dargestellt.

7.1 Analysegrundlage: erweiterte SCATE-Taxonomie nach Tezcan et al. (2019) und Key-Feature-Analyse nach Webster et al. (2020)

Für die qualitative Fehlerannotation der vorliegenden Arbeit ist insbesondere die Erweiterung der SCATE-Taxonomie nach Tezcan et al. (2019) relevant. Die Grundstruktur des SCATE-Modells wurde bereits im Kapitel 2.3.2.2 dargestellt. Die von Tezcan et al. vorgeschlagene Erweiterung zielt auf die Analyse literarischer maschineller Übersetzungen auf Dokumentenebene ab. Zu diesem Zweck wurde das Modell um die Kategorien Kohärenz sowie Stil & Register erweitert, um literaturspezifische Problembereiche annotieren zu können. Diese Erweiterungen dienen dazu, Fehler zu erfassen, die zwar nicht zwingend die Satzgrammatik betreffen, jedoch den Lesefluss sowie den literarischen Gesamteindruck des Zieltexts maßgeblich beeinflussen können (Tezcan et al. 2019: 40).

FLUENCY	ACCURACY
• coherence ○ logical problem ○ non-existing word ○ cultural reference ○ discourse marker ○ co-reference ○ inconsistency ○ verb tense • lexicon ○ lexical choice ○ wrong preposition • grammar & syntax ○ agreement ○ verb form ○ word order ○ extra word(s) ○ missing word(s) • style & register ○ disfluency ○ repetition ○ register ○ untranslated • spelling • other	• mistranslation ○ multiword ○ word sense ○ semantically unrelated ○ part-of-speech ○ partially translated ○ other • do not translate • untranslated • addition • omission • capitalisation & punctuation • other

Abb. 1: Die erweiterte SCATE-Fehlerannotation (Tezcan et al. 2019)

Für die Analyse maschineller Übersetzungen literarischer Texte werden in der Forschung kombinierte Ansätze vorgeschlagen, die eine SCATE-basierte Fehlerannotation mit der Untersuchung textueller Merkmale verbinden. Aufbauend auf der erweiterten SCATE-Fehlerannotation schlagen Webster et al. (2020) vor, diese durch eine zusätzliche Analyse ausgewählter textueller Merkmale zu ergänzen. Die Key-Feature-Analyse nach Webster et al. (2020: 5) ist nicht als Evaluierungsmethode im Sinne einer eigenständigen Qualitätsmetrik konzipiert, sondern dient der ergänzenden Analyse ausgewählter textueller Merkmale auf Dokumentebene. Dabei werden vier zentrale Übersetzungsmerkmale untersucht, nämlich lexikalische Vielfalt, Kohäsion sowie syntaktische und stilistische Unterschiede. Durch diese Kombination von Fehlerannotation und Merkmalsanalyse soll ein differenzierteres Bild der Übersetzungsqualität literarischer Texte erarbeitet werden.

Der in der vorliegenden Arbeit angewandte Analyseansatz orientiert sich an diesem kombinierten Vorgehen und verbindet eine erweiterte SCATE-Fehlerannotation mit einer für die Untersuchung adaptierten Key-Feature-Analyse nach Webster et al. (2020). Im Rahmen der SCATE-basierten Fehlerannotation werden die Bereiche Kohärenz sowie Stil & Register berücksichtigt. Aus der Key-Feature-Analyse werden selektiv die Aspekte lexikalische Vielfalt

und lokale Kohäsion aufgegriffen. Im Unterschied zu Webster et al. (2020) wird die Key-Feature-Analyse jedoch nicht vollständig computergestützt umgesetzt. Während die lexikalische Vielfalt quantitativ anhand der Type-Token-Ratio erfasst wird, erfolgt die Analyse der lokalen Kohäsion qualitativ und manuell. Auch die Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register wird manuell durchgeführt. Diese Auswahl orientiert sich an den Annahmen der vorliegenden Arbeit und dient dazu, diese anhand der analysierten Übersetzungen zu überprüfen. Durch die Kombination beider Ansätze werden die untersuchten Übersetzungen sowohl auf der Ebene konkreter Fehlerstellen als auch im Hinblick auf ausgewählte textuelle Merkmale beschrieben.

7.2 Methodisches Vorgehen

Die vorliegende Untersuchung ist als qualitative Analyse mit ergänzenden quantitativen Elementen konzipiert. Gegenstand der Analyse bilden LLM-gestützte Übersetzungen literarischer Texte, die mit humanen Übersetzungen verglichen werden. Die qualitative Analyse konzentriert sich auf die Fehlerarten in den Bereichen Kohärenz sowie Stil & Register und erfolgt auf Grundlage der erweiterten SCATE-Taxonomie. Ergänzend werden im Rahmen einer adaptierten Key-Feature-Analyse die textuellen Merkmale lexikalische Vielfalt und lokale Kohäsion untersucht.

Für jede der ausgewählten Erzählungen werden drei Übersetzungsvarianten herangezogen: eine Humanübersetzung sowie zwei LLM-gestützte Übersetzungen, erstellt mit DeepL Pro und ChatGPT-5.3. Die LLM-gestützten Übersetzungen aus dem Russischen ins Deutsche wurden im März 2026 erstellt. Für ChatGPT-5.3 wurde ein einheitlicher Prompt verwendet, um die Vergleichbarkeit der Ergebnisse sicherzustellen. Der Prompt lautete: „*Übersetze den folgenden Text ins Deutsche.*“ Ziel der Arbeit ist nicht die Untersuchung des Einflusses unterschiedlicher Prompt-Formulierungen, sondern die Analyse der resultierenden Übersetzungstexte im Vergleich zur jeweiligen Humanübersetzung. Die Beschränkung auf eine gemeinsame Zielsprache sowie auf Übersetzungen derselben Ausgangstexte stellt eine zentrale methodische Voraussetzung für die Vergleichbarkeit der untersuchten Texte dar.

Der quantitative Teil der Analyse bezieht sich auf die lexikalische Vielfalt der untersuchten Übersetzungen. Dazu werden quantitative Kennzahlen herangezogen, um Unterschiede zwischen den Übersetzungsvarianten sichtbar zu machen. Diese quantitative Analyse dient der Überprüfung der ersten Annahme der vorliegenden Arbeit, wonach LLM-gestützte Übersetzungen im Vergleich zu Humanübersetzungen eine geringere lexikalische Vielfalt aufweisen. Das

genaue Vorgehen zur Berechnung und Auswertung der lexikalischen Vielfalt wird in Kapitel 7.3 dargestellt. Die qualitative Analyse umfasst die Bereiche lokale Kohäsion, Kohärenz sowie Stil & Register. *Accuracy*-Fehler werden in der vorliegenden Arbeit nicht als eigene Analysekategorie berücksichtigt und nicht quantitativ ausgewertet. Die theoretische und methodische Einordnung der berücksichtigten Analysebereiche erfolgt in den folgenden Unterkapiteln.

Um die Nachvollziehbarkeit der manuellen Annotation in Bezug auf Kohärenz und Stil & Register zu erhöhen, wurde die erstellte Annotationstabelle zusätzlich von einer externen Kontrollperson überprüft. Bei der Kontrollperson handelt es sich um eine an der Universität Wien ausgebildete Übersetzerin, Untertitlerin, Dolmetscherin und Lektorin mit Deutsch als Erstsprache. Die Kontrollperson erhielt eine Tabelle mit den annotierten Problemstellen, den jeweiligen Kategorien und Kurzbegründungen. Zusätzlich wurden der Kontrollperson alle untersuchten Übersetzungsvarianten zur Verfügung gestellt, um bei Bedarf die Einordnung der jeweiligen Problemstelle im weiteren Kontext nachvollziehen zu können. Sie bewertete die Einordnung jeder Fehlerannotation anhand der Optionen „ja“, „teilweise“, „eher nein“ oder „nein“. Bei Bewertungen mit „teilweise“, „eher nein“ oder „nein“ war ein Kommentar vorgesehen, um die abweichende Einschätzung nachvollziehbar zu machen.

Für die finale quantitative Übersicht wurde folgendes Vorgehen festgelegt: Berücksichtigt werden ausschließlich Problemstellen, die nach der externen Überprüfung als eindeutig eingestuft wurden. Fälle mit „teilweise“, „eher nein“ oder „nein“ werden in der vollständigen Annotationstabelle dokumentiert, jedoch nicht in die finale Zählung einbezogen. Damit soll vermieden werden, dass uneindeutige oder eingeschränkt bestätigte Annotationen das quantitative Ergebnis verzerren. Bei Mehrfachzuordnungen wird eine Problemstelle in jeder zugeordneten Kategorie gezählt. Problemstellen, die beide LLM-gestützten Systeme betreffen, werden für jedes betroffene System separat berücksichtigt. Die vollständige Annotationstabelle mit den Rückmeldungen der Kontrollperson befindet sich im Anhang. Im Folgenden werden zunächst das quantitative Verfahren zur Erfassung der lexikalischen Vielfalt und anschließend die qualitative Analyse der lokalen Kohäsion sowie die manuelle Fehlerannotation in den Bereichen Kohärenz und Stil & Register erläutert.

7.3 Quantitative Analyse: lexikalische Vielfalt

Die lexikalische Vielfalt ist ein Merkmal literarischer Texte, das sich quantitativ erfassen lässt und Rückschlüsse auf Wortschatzvariation erlaubt. In der vorliegenden Untersuchung wird sie für den quantitativen Teil der adaptierten Key-Feature-Analyse herangezogen. Dabei wird die

lexikalische Vielfalt der LLM-gestützten Übersetzungen von DeepL Pro sowie ChatGPT-5.3 mit jener der Humanübersetzungen verglichen.

Zur Erfassung der lexikalischen Vielfalt wird in der vorliegenden Arbeit die Type-Token-Ratio (TTR) herangezogen. Die TTR beschreibt das Verhältnis der unterschiedlichen Wortformen (Types) zur Wortgesamtzahl (Tokens) innerhalb eines Textes. Wiederholungen desselben Wortes erhöhen dabei die Tokenzahl, nicht jedoch die Anzahl der Types (Webster et al. 2020: 5). Daraus folgt, dass ein hoher TTR-Wert auf eine größere lexikalische Vielfalt hinweist, während niedrigere Werte auf häufigere Wiederholungen verweisen (Tezcan et al. 2019: 42f.).

Ein wesentlicher methodischer Aspekt bei der Verwendung der TTR ist ihre Abhängigkeit von der Textlänge, da mit zunehmendem Textumfang die Wahrscheinlichkeit von Wortwiederholungen steigt und der TTR-Wert tendenziell sinkt (Webster et al. 2020: 5). Aus diesem Grund eignet sich die TTR primär für den Vergleich von Texten mit vergleichbarem Umfang. Da sich die vorliegende Analyse auf den Vergleich verschiedener Übersetzungsfassungen desselben Ausgangstextes beschränkt und die Übersetzungen in ihrer Länge weitgehend vergleichbar sind, ist der Einfluss der Textlänge im Rahmen der Analyse kontrolliert. Die TTR wird daher als orientierendes quantitatives Maß eingesetzt.

Die Berechnung der TTR erfolgte im April 2026 mithilfe des Textanalysetools Voyant Tools. Voyant Tools ist ein webbasiertes Open-Source-Analyseinstrument, das im Rahmen eines wissenschaftlichen Forschungsprojekts entwickelt wurde (Voyant Tools o.J.). In der vorliegenden Analyse wurde das Tool zur automatisierten Ermittlung statistischer Kennwerte wie der Anzahl von Tokens und Types eingesetzt, wodurch eine einheitliche Auswertung der untersuchten Übersetzungen gewährleistet wird.

Die quantitative Analyse beschränkt sich auf die deutschen Zieltexte, um die Vergleichbarkeit der Ergebnisse zu gewährleisten. Da lexikalische Vielfalt stark von sprachspezifischen Strukturen beeinflusst wird, würde ein sprachübergreifender Vergleich zu methodischen Verzerrungen führen. Durch die Beschränkung auf die deutschen Übersetzungen der Ausgangstexte bleiben die ermittelten TTR-Werte nachvollziehbar und überprüfbar.

7.4 Qualitative Analyse: lokale Kohäsion, Kohärenz, Stil & Register

7.4.1 Lokale Kohäsion

Im Vergleich zu anderen Textsorten weisen literarische Texte häufig ein höheres Maß an Kohäsion auf (vgl. Voigt & Jurafsky 2012: 18–25; Webster et al. 2020: 6). In der vorliegenden Arbeit wird die lokale Kohäsion als qualitativer Teil der adaptierten Key-Feature-Analyse untersucht. Im Fokus steht die Frage, inwiefern sich LLM-gestützte Übersetzungen und humane Übersetzungen literarischer Texte in Bezug auf ihre lokale Kohäsion unterscheiden, um die entsprechende Annahme der Arbeit zu überprüfen.

In der Textlinguistik wird Kohäsion als eines der zentralen Kriterien der Textualität verstanden (Beaugrande & Dressler 1981: 3ff.). Beaugrande und Dressler definieren Kohäsion als die Art und Weise, „wie die Komponenten des Oberflächentextes, d. h. die Worte, wie wir sie tatsächlich hören oder sehen, miteinander verbunden sind“ (1981:3f.). In Anlehnung an Webster et al. (2020: 2) wird lokale Kohäsion in der vorliegenden Arbeit als Kohäsion zwischen kleineren, benachbarten Textabschnitten verstanden, etwa zwischen aufeinanderfolgenden Sätzen oder Äußerungen. Webster et al. (2020) erfassen lokale Kohäsion quantitativ anhand lexikalischer und semantischer Überlappungen zwischen benachbarten Sätzen. Die Beschränkung auf benachbarte Textabschnitte wird in der vorliegenden Untersuchung übernommen. Die Analyse erfolgt jedoch nicht mithilfe automatischer Messverfahren, sondern qualitativ und manuell.

Aufbauend auf der textlinguistischen Konzeption von Kohäsion nach Beaugrande und Dressler (1981) werden in der vorliegenden Untersuchung jene Kohäsionsmittel berücksichtigt, die für die lokale Kohäsion unmittelbar benachbarter Textpassagen besonders relevant sind. Dazu zählen Formen der Rekurrenz und partiellen Rekurrenz, paraphrasierende Wiederaufnahmen, Pro-Formen sowie andere grammatische Formen der Wiederaufnahme. Es wird untersucht, ob diese Verknüpfungen in den Übersetzungen konsistent und nachvollziehbar realisiert werden.

Da pronominale Wiederaufnahmen Berührungspunkte mit der SCATE-Unterkategorie *Koreferenz* aufweisen können, werden entsprechende Fälle qualitativ beschrieben und nicht als zusätzlich quantitative Fehlerstatistik neben der SCATE-Annotation ausgewertet.

7.4.2 Kohärenz

Kohärenz betrifft nach Beaugrande und Dressler (1981) „die Funktionen, durch die die Komponenten der Textwelt, d.h. die Konstellation von Konzepten (Begriffen) und Relationen (Beziehungen), welche dem Oberflächentext zugrundeliegen, für einander gegenseitig zugänglich und relevant sind“ (1981: 5) ³. In der vorliegenden Untersuchung wird Kohärenz jedoch nicht als eigenständige textlinguistische Kategorie analysiert, sondern im Rahmen der erweiterten SCATE-Fehlerannotation erfasst.

In der erweiterten SCATE-Taxonomie ist Kohärenz der Dimension *Fluency* zugeordnet. Kohärenzfehler werden dann annotiert, wenn der Zieltext im Kontext des Textverlaufs inhaltlich oder logisch schwer nachvollziehbar ist, z. B. aufgrund einer fehlenden oder unklaren logischen Struktur.

Für die Annotation werden folgende Unterkategorien der Kohärenz berücksichtigt:

- *Logische Probleme (logical problems)* umfassen Fälle, in denen der Zieltext im Kontext des Textverlaufs inhaltlich widersprüchlich, verwirrend oder insgesamt schwer nachvollziehbar erscheint.
- Die Kategorie *nicht existierende Wörter (non-existing word)* erfasst Fälle, in denen im Deutschen ein Wort verwendet wird, das in der Zielsprache nicht existiert und keinen sinnvollen Bedeutungszusammenhang ergibt.
- *Kulturelle Referenzen (cultural reference)* werden annotiert, wenn ein Wort oder eine Phrase auf ein kulturspezifisches Element der Ausgangssprache verweist, dieses im Zieltext jedoch nicht verständlich ist oder keinen Sinn ergibt.
- *Diskursmarker (discourse markers)* werden annotiert, wenn Konnektoren oder verbindende Ausdrücke im Textverlauf eine logisch unplausible oder schwer nachvollziehbare Beziehung zwischen Aussagen ausdrücken.
- Die Kategorie *Koreferenz (co-reference)* betrifft fehlerhafte oder unklare Bezüge zwischen Entitäten, etwa wenn Pronomen nicht eindeutig oder inkorrekt auf zuvor im Text eingeführte Personen, Gegenstände oder andere Referenzgrößen verweisen.
- Die Kategorie *Inkonsistenz (inconsistency)* erfasst Fälle, in denen Bezeichnungen, Benennungen oder Schreibweisen im Verlauf des Textes uneinheitlich verwendet werden.

³ Beaugrande und Dressler (Beaugrande & Dressler 1981: 5) weisen darauf hin, dass die Begriffe Kohärenz und Kohäsion häufig vermischt oder gemeinsam verwendet werden. Die Autoren betonen jedoch die Notwendigkeit einer klaren Unterscheidung zwischen oberflächenstrukturellen Verknüpfungen (Kohäsion) und den dem Oberflächentext zugrundeliegenden Konzepten und Relationen (Kohärenz).

- *Verbform (verb tense)* betrifft Fälle, in denen die verwendete Zeitform des Verbs im Kontext des restlichen Satzes oder des Textverlaufs unplausibel oder logisch inkonsistent erscheint.

Die dargestellten Unterkategorien der Kohärenz basieren auf der erweiterten SCATE-Taxonomie nach Tezcan et al. (2019: 42). In der Analyse werden jene Kohärenz-Unterkategorien herangezogen, die im untersuchten Korpus auftreten. Auf diese Weise wird überprüft, ob und in welcher Hinsicht LLM-gestützte Übersetzungen im Vergleich zu den Humanübersetzungen Kohärenzprobleme aufweisen.

7.4.3 Stil & Register

In der Forschung zur literarischen maschinellen Übersetzung wird darauf hingewiesen, dass neuronale Übersetzungssysteme insbesondere im Bereich stilistischer Nuancen und registerbezogener Unterschiede Auffälligkeiten und Schwierigkeiten zeigen (Tezcan et al. 2019; Fonteyne et al. 2020; Webster et al. 2020). Stilistische und registerbezogene Aspekte gelten daher als besonders relevante Untersuchungskategorien bei der Analyse literarischer maschineller Übersetzungen.

In diesem Zusammenhang werden stilistische und registerbezogene Phänomene in der vorliegenden Arbeit im Rahmen der erweiterten SCATE-Fehlerannotation berücksichtigt. Diese Kategorien wurden von Tezcan et al. (2019) gezielt ergänzt, um die spezifischen Anforderungen literarischer Texte zu berücksichtigen. Stil- und Registerfehler werden dabei als Abweichungen verstanden, bei denen der Zieltext zwar grundsätzlich verständlich ist, jedoch eine idiomatische oder stilistisch passendere Ausdrucksweise in der Zielsprache möglich wäre (Webster et al. 2020: 11).

Innerhalb der erweiterten SCATE-Taxonomie sind stilistische und registerbezogene Auffälligkeiten der *Fluency*-Dimension zugeordnet. Für die Annotation werden mehrere Unterkategorien berücksichtigt, die unterschiedliche Formen solcher Abweichungen erfassen (Tezcan et al. 2019: 42):

- Zur Kategorie *Unflüssigkeit (disfluency)* zählen Fälle, in denen ein Satz oder Satzbestandteil zwar nicht grammatisch falsch ist, jedoch schwer lesbar wirkt oder in einer deutlich idiomatischeren Form hätte formuliert werden können.

- Die Kategorie *Wiederholung (repetition)* erfasst die wiederholte Verwendung identischer oder sehr ähnlicher Wörter oder Ausdrücke innerhalb eines Satzes oder in unmittelbarer Nähe, sofern diese Wiederholungen stilistisch auffällig sind.
- Unter *Register (register)* werden Fälle annotiert, in denen zwar eine bedeutungsgleiche Wortwahl vorliegt, das gewählte Wort oder der Ausdruck jedoch hinsichtlich des Formalitätsgrads oder der sprachlichen Varietät für den gegebenen Textzusammenhang nicht angemessen erscheint. Dazu zählen etwa unangemessen formelle oder informelle Ausdrucksweisen sowie der Gebrauch regionaler Varianten, die für den Zieltext nicht geeignet erscheinen.
- Die Kategorie *Nicht übersetzt (untranslated)* umfasst schließlich Textstellen, in denen Wörter oder Textfragmente trotz vorhandener Übersetzungsmöglichkeiten in der Zielsprache unverändert aus der Ausgangssprache übernommen wurden.

Die Analyse von Stil und Register erfolgt qualitativ auf Grundlage der deutschen Übersetzungsvarianten der drei untersuchten Erzählungen. Auf diese Weise wird überprüft, ob und in welcher Hinsicht LLM-gestützte Übersetzungen literarischer Texte im Vergleich zu den Humanübersetzungen Schwächen im Bereich stilistischer Gestaltung und Registerwahl aufweisen.

8 Übersetzungsanalyse

8.1 Quantitative Analyse: Type-Token-Ratio

Zur Überprüfung der ersten Annahme der vorliegenden Arbeit, wonach LLM-gestützte Übersetzungen eine geringere lexikalische Vielfalt als humane Übersetzungen aufweisen, wird im Folgenden eine quantitative Analyse auf Grundlage der Type-Token-Ratio (TTR) durchgeführt. Der Vergleich erfolgt zwischen der Humanübersetzung sowie den LLM-gestützten Übersetzungen von DeepL Pro und ChatGPT-5.3 für drei Erzählungen. Ein Vergleich mit dem russischen Ausgangstext wurde nicht vorgenommen, da sich die Analyse auf Unterschiede zwischen Übersetzungsvarianten derselben Zielsprache konzentriert. Da die Type-Token-Ratio grundsätzlich von der Textlänge beeinflusst wird und mit zunehmendem Textumfang tendenziell sinkt, wird dieser Zusammenhang bei der Interpretation der Ergebnisse berücksichtigt (Webster et al. 2020: 5).

Zunächst werden grundlegende Umfangsdaten der untersuchten Übersetzungen dargestellt. Tabelle 1 enthält die Token- und Type-Werte sowie die daraus berechnete Type-Token-

Ratio (TTR) der Humanübersetzung und der beiden LLM-gestützten Übersetzungen für die Erzählung *Der Tod eines Beamten*.

Tab. 1: Type-Token-Ratio der untersuchten Übersetzungen von *Der Tod eines Beamten*

Übersetzung	Types (t)	Tokens (n)	TTR (t/n)
Humane Übersetzung	448	984	0,455 (45,5 %)
DeepL Pro	374	901	0,415 (41,5 %)
ChatGPT-5.3	389	902	0,431 (43,1 %)

Für *Der Tod eines Beamten* ergibt die TTR-Berechnung, dass die Humanübersetzung mit einem TTR-Wert von 0,455 bzw. 45,5 % die höchste lexikalische Vielfalt erzielt. Die Übersetzung von ChatGPT-5.3 erreicht einen TTR-Wert von 0,431 bzw. 43,1 %, während die Übersetzung von DeepL Pro mit 0,415 bzw. 41,5 % den niedrigsten Wert demonstriert. In relativen Zahlen liegt die lexikalische Vielfalt der humanen Übersetzung damit um 2,4 Prozentpunkte über jener der ChatGPT-Übersetzung und um 4 Prozentpunkte über jener der DeepL-Pro-Übersetzung.

Die untersuchten Übersetzungen unterscheiden sich hinsichtlich ihrer Textlänge moderat. Während die Tokenzahlen der beiden LLM-gestützten Übersetzungen nahezu identisch sind, ist die Humanübersetzung etwas umfangreicher. Im vorliegenden Fall weist die Humanübersetzung die höchste Tokenzahl auf, erzielt jedoch gleichzeitig den höchsten TTR-Wert.

Tabelle 2 gibt die Type- und Token-Werte sowie die daraus berechneten TTR-Werte für die drei Übersetzungsfassungen der Erzählung *Der Dicke und der Dünne* wieder.

Tab. 2: Type-Token-Ratio der untersuchten Übersetzungen von *Der Dicke und der Dünne*

Übersetzung	Types (t)	Tokens (n)	TTR (t/n)
Humane Übersetzung	362	676	0,536 (53,6 %)
DeepL Pro	347	679	0,511 (51,1 %)
ChatGPT-5.3	357	679	0,526 (52,6 %)

Die Textlänge der drei Versionen unterscheidet sich bei der Erzählung minimal. Die Humanübersetzung umfasst 676 Tokens, während DeepL Pro und ChatGPT-5.3 jeweils 679 Tokens aufweisen. Da die Tokenzahlen nahezu identisch sind, kann der Unterschied zwischen den TTR-Werten hier nicht primär auf unterschiedliche Textlängen zurückgeführt werden. Die Ergebnisse der TTR-Berechnung für die Erzählung *Der Dicke und der Dünne* zeigen, dass die Humanübersetzung mit einem Wert von 0,536 bzw. 53,6 % die höchste lexikalische Vielfalt aufweist. ChatGPT-5.3 erreicht mit 0,526 bzw. 52,6 % einen leicht niedrigeren Wert, während

DeepL Pro mit 0,511 bzw. 51,1 % den niedrigsten TTR-Wert erzielt. Der Abstand zwischen der Humanübersetzung und ChatGPT-5.3 beträgt damit 1 Prozentpunkt. Zwischen der Humanübersetzung und DeepL Pro liegt die Differenz bei 2,5 Prozentpunkten. Insgesamt lässt sich feststellen, dass die Unterschiede zwischen den TTR-Werten relativ gering sind, insbesondere im Vergleich zwischen der Humanübersetzung und ChatGPT-5.3. Der niedrigere Wert bei DeepL Pro kann auf eine stärkere Wiederholung bereits verwendeter Wortformen hindeuten.

Anschließend werden die TTR-Werte für die Erzählung *Ein Chamäleon* betrachtet. Die entsprechenden Type- und Token-Werte sowie die daraus resultierenden TTR-Werte sind in Tabelle 3 dargestellt.

Tab. 3: Type-Token-Ratio der untersuchten Übersetzungen von *Ein Chamäleon*

Übersetzung	Types (t)	Tokens (n)	TTR (t/n)
Humane Übersetzung	508	1186	0,428 (42,8 %)
DeepL Pro	522	1201	0,435 (43,5 %)
ChatGPT-5.3	492	1114	0,442 (44,2 %)

Die Ergebnisse der TTR-Berechnung für die Erzählung *Ein Chamäleon* weichen von den zuvor beobachteten Tendenzen ab. Die Humanübersetzung umfasst 508 Types bei 1186 Tokens und erreicht damit einen TTR-Wert von 0,428 bzw. 42,8 %. DeepL Pro liegt mit 522 Types bei 1201 Tokens bei einem leicht höheren Wert von 0,435 bzw. 43,5 %. ChatGPT-5.3 erzielt mit 492 Types bei 1114 Tokens den höchsten TTR-Wert von 0,442 bzw. 44,2 %.

Damit erzielt die Humanübersetzung in dieser Erzählung den niedrigsten TTR-Wert, während die Humanübersetzungen der beiden zuvor untersuchten Erzählungen jeweils die höchste lexikalische Vielfalt aufwiesen. Der Abstand zwischen der Humanübersetzung und DeepL Pro beträgt 0,7 Prozentpunkte, zwischen der Humanübersetzung und ChatGPT-5.3 1,4 Prozentpunkte. Die Unterschiede sind damit zwar vergleichsweise gering, das Ergebnis widerspricht jedoch der Annahme, dass die Humanübersetzung eine höhere lexikalische Vielfalt als die LLM-gestützten Übersetzungen aufweist. Da die TTR durch die Textlänge beeinflusst wird, ist das Ergebnis bei *Ein Chamäleon* vorsichtig zu interpretieren. ChatGPT-5.3 weist mit 1114 Tokens die kürzeste Übersetzungsfassung auf, wodurch der vergleichsweise höhere TTR-Wert begünstigt sein kann. Der höhere Wert sollte daher nicht unmittelbar als Hinweis auf eine qualitativ bessere lexikalische Vielfalt interpretiert werden.

Zusammenfassend ergibt sich für die TTR-Analyse kein vollständig einheitliches Bild. In zwei der drei untersuchten Erzählungen erreicht die Humanübersetzung den höchsten TTR-

Wert. Dies gilt sowohl für *Der Tod eines Beamten* als auch für *Der Dicke und der Dünne*. In beiden Fällen erreichen die LLM-gestützten Übersetzungen niedrigere TTR-Werte, wobei ChatGPT-5.3 jeweils höhere Werte als DeepL Pro erzielt. Für diese beiden Erzählungen stützen die quantitativen Ergebnisse somit die erste Annahme.

Bei *Ein Chamäleon* ergibt sich hingegen ein abweichendes Ergebnis. Die Humanübersetzung weist hier den niedrigsten TTR-Wert auf, während ChatGPT-5.3 den höchsten Wert erreicht, gefolgt von DeepL Pro. Die Differenzen sind zwar gering, dennoch lässt sich die Annahme auf Grundlage der vorliegenden Daten nicht vollständig bestätigen. Außerdem ist zu berücksichtigen, dass die TTR textlängenabhängig ist und höhere Werte nicht automatisch als qualitative Verbesserung zu verstehen sind. Insbesondere in literarischen Texten können Wiederholungen kohäsiv oder stilistisch motiviert sein. Die erste Annahme wird daher durch die Ergebnisse nur teilweise gestützt.

8.2 Lokale Kohäsion

Die qualitative Analyse der lokalen Kohäsion ergab nur wenige eindeutige Problemstellen. In *Der Tod eines Beamten* und *Der Dicke und der Dünne* wurden keine lokalen Kohäsionsprobleme festgestellt. Die identifizierten Fälle konzentrieren sich auf *Ein Chamäleon* und werden im Folgenden anhand der relevanten Textstellen dargestellt.

Tab. 4: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Du willst mich beißen, verdammte Töle?« hört Otschumelow plötzlich. »Laßt ihn nicht laufen, Leute! Heutzutage darf man nicht beißen! Haltet ihn! Ah ... ah!« Man hört einen Hund winseln. Otschumelow blickt zur Seite und sieht: Aus dem Holzlager des Kaufmanns Pitschugin läuft eilig ein Hund, er hüpfte auf drei Pfoten und sieht sich immer wieder um. Ihm nach jagt ein Mann in einem gestärkten Kattunhemd und aufgeknöpfter Weste. Er läuft hinter ihm her, den Oberkörper vorgestreckt, fällt zu Boden und packt den Hund an den Hinterpfoten.
DeepL Pro	– Willst du also beißen, du <u>elendes Weib</u>? – hört Ochumelow plötzlich. – Jungs, lasst <u>sie</u> nicht durch! Heute darf man nicht beißen! Halt <u>sie</u> fest! Ah... ah! Man hört ein Hundegeheul. Ochumelow schaut zur Seite und sieht: Aus

	dem Holzschuppen des Kaufmanns Pichugin rennt <u>ein Hund</u> , auf drei Beinen hüpfend und sich umschauend. Hinter <u>ihr</u> jagt ein Mann in einem gestärkten Baumwollhemd und einer aufgeknöpften Weste. Er rennt <u>ihr</u> hinterher, beugt sich nach vorne, fällt zu Boden und packt <u>den Hund</u> an den Hinterpfoten.
ChatGPT-5.3	„Willst du etwa beißen, du <u>verfluchtes Vieh</u> ?“ hört Otschumelow plötzlich. „Leute, lasst <u>sie</u> nicht durch! Heute ist Beißen verboten! Haltet <u>sie</u> ! Ah... ah!“ Man hört das Jaulen eines Hundes. Otschumelow blickt zur Seite und sieht: aus dem Holzlager des Kaufmanns Pitschugin rennt ein Hund, auf drei Beinen hüpfend und sich umblickend. Hinter ihm jagt ein Mann in einem gestärkten Kattunhemd und einer aufgeknöpften Weste. Er läuft hinterher, beugt den Oberkörper vor, stürzt zu Boden und packt den Hund an den Hinterbeinen.

Eine erste Auffälligkeit im Bereich lokaler Kohäsion betrifft die Eingangsszene von *Ein Chamäleon*. In der Humanübersetzung wird der Hund zunächst als „*verdammte Töle*“ bezeichnet und anschließend mit „*ihn*“ wiederaufgenommen. Dadurch bleibt die Bezugskette konsistent. In den LLM-gestützten Übersetzungen wird diese Bezugskette dagegen uneinheitlich realisiert. ChatGPT-5.3 verwendet zunächst die Bezeichnung „*verfluchtes Vieh*“, führt die Bezugskette in den unmittelbar folgenden Äußerungen jedoch mit dem Pronomen „*sie*“ fort: „*lasst sie nicht durch*“ und „*Haltet sie*“. Da „*Vieh*“ im Deutschen ein sächliches Substantiv ist, wäre eine Wiederaufnahme mit „*es*“ näherliegend. Mit Blick auf den später explizit genannten Hund wäre auch „*ihn*“ möglich. Bei DeepL Pro ist die uneinheitliche Referenzführung über einen längeren Abschnitt hinweg erkennbar: Der Hund wird zunächst als „*elendes Weib*“ bezeichnet und mit „*sie*“ wiederaufgenommen. Nach der expliziten Einführung als „*ein Hund*“ wird die Referenz weiterhin mit „*ihr*“ fortgeführt: „*Hinter ihr jagt ein Mann*“ und „*Er rennt ihr hinterher*“. Dadurch wird die lokale Kohäsion in der Eingangsszene geschwächt, da die Bezugnahme auf den Hund in benachbarten Sätzen nicht konsistent wiederaufgenommen wird. Ein möglicher Grund für die uneinheitliche Referenzführung dürfte darin liegen, dass in dieser Szene die Bezugnahme auf den Hund im russischen Ausgangstext grammatisch feminin realisiert wird. Das russische Substantiv „*собака*“ („*Hund*“) ist feminin. Dementsprechend stehen auch darauf bezogene Adjektive und Pronomen im Russischen in femininer Form. Es ist daher möglich, dass sich beide LLM-gestützten Systeme an der russischen Referenzführung orientieren und die Bezugskette im Deutschen dadurch nicht konsistent realisiert wird.

Tab. 5: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Da brauchen wir nicht lange zu fragen«, sagt Otschumelow. »Das ist ein streunender Hund! Was reden wir da noch ... Wenn ich gesagt habe, daß er streunt, dann streunt er ... Beseitigen, und Schluß.« »Das ist nicht unsrer«, fährt Prochor fort.
DeepL Pro	– Da gibt’s nichts lange zu fragen, – sagt Ochumelow. – <u>Sie</u> ist ein Streuner! Da gibt’s nichts lange zu reden... Wenn er sagt, <u>sie</u> sei ein Streuner, dann ist <u>sie</u> eben ein Streuner... Erledigen, das ist alles. – <u>Der</u> gehört nicht uns, – fährt Prochor fort.
ChatGPT-5.3	Dann gibt es nichts weiter zu fragen,“ sagt Otschumelow. „Er ist ein Streuner! Wenn ich sage, er ist ein Streuner, dann ist er es auch... Töten, das ist alles.“ „Er ist nicht von uns,“ fährt Prochor fort.

Ein weiteres Beispiel betrifft die spätere Passage, in der die Zugehörigkeit des Hundes geklärt wird. In der Humanübersetzung wird der Hund zunächst als „*ein streunender Hund*“ bezeichnet. Anschließend wird diese Referenz mit „*er*“ und „*unsrer*“ weitergeführt. Die Bezugskette bleibt dadurch im lokalen Textzusammenhang nachvollziehbar. ChatGPT-5.3 stellt den Bezug ähnlich her und verwendet durchgehend maskuline Formen: „*Er ist ein Streuner*“, „*er ist es ...*“ und „*Er ist nicht von uns*“. DeepL Pro führt die Referenz dagegen uneinheitlich: Zunächst steht „*Sie ist ein Streuner*“, anschließend „*Wenn er sagt, sie sei ein Streuner*“, dann erneut „*sie*“ und im nächsten Satz „*Der gehört nicht uns*“. Der Hund wird somit in unmittelbar aufeinanderfolgenden Äußerungen mit wechselnden pronominalen Formen wiederaufgenommen. Dadurch wird die lokale Kohäsion in der Passage beeinträchtigt, weil die Bezugskette nicht konsistent fortgeführt wird. Wie bereits im vorigen Beispiel erläutert wurde, dürfte diese Problematik mit der grammatisch femininen Bezugnahme auf den Hund im russischen Ausgangstext zusammenhängen.

Tab. 6: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Das ist nicht unsrer«, fährt Prochor fort. »Das ist der vom Bruder des Generals, der neulich gekommen ist. Unser General ist kein Liebhaber von Windhunden. Sein Bruder ist ...«
--------------------	--

	»Ist denn sein Bruder gekommen? Wladimir Iwanytsch?« fragt Otschumelow, und ein Lächeln der Rührung breitet sich über sein Gesicht. »Sieh mal einer an! Und ich hab es nicht gewußt! Ist er zu Besuch gekommen?« »Ja ... « »Sieh mal an ... Hatte wohl Sehnsucht nach seinem Brüderchen ... Und ich habe das gar nicht gewußt. Das ist also sein Hündchen? Freut mich sehr ... Nimm es mit ... Ein liebes Hündchen ...
DeepL Pro	– Der gehört dem Bruder des Generals, der neulich angekommen ist. Unserer ist kein Fan von Windhunden. <u>Ihr</u> Bruder ist es aber... – Ist <u>ihr</u> Bruder denn angekommen? Wladimir Iwanowitsch? – fragt Ochumelow, und sein ganzes Gesicht erhellt sich in einem gerührten Lächeln. – Ach du meine Güte! Und ich wusste nichts davon! <u>Sind sie</u> zu Besuch gekommen? – Zu Besuch... – Ach du meine Güte... Wir haben den Bruder vermisst... Und ich wusste es gar nicht! Ist das also <u>ihr</u> Hündchen? Ich freue mich sehr... Nimm <u>sie</u>... Ein ganz hübsches Hündchen...
ChatGPT-5.3	„Er gehört dem Bruder des Generals, der neulich angekommen ist. Unser Herr mag keine Windhunde, aber sein Bruder schon...“ „Wie, der Bruder ist angekommen? Wladimir Iwanowitsch?“ fragt Otschumelow, und sein ganzes Gesicht erstrahlt vor Rührung. „Sieh mal einer an! Das wusste ich gar nicht! Zu Besuch gekommen?“ „Zu Besuch...“ „Ach, sieh mal einer an... Hat seinen Bruder vermisst... Und ich wusste nichts davon! Also ist das sein Hund? Sehr erfreut... Nimm ihn... Ein ganz netter Hund... so lebhaft... Hat ihm in den Finger geschnappt! Ha-ha-ha... Nun, warum zitterst du? Rrr... Rr... Er ist böse, der Schlingel... so ein kleines Hündchen...“

Die Passage über den Bruder des Generals weist in der DeepL-Pro-Übersetzung ebenfalls eine kohäsive Unstimmigkeit auf. In der Humanübersetzung wird die Bezugskette klar aufgebaut: Zunächst ist vom „*Bruder des Generals*“ die Rede, anschließend wird diese Referenz mit „*sein*

Bruder“, „*er*“ und „*sein Hündchen*“ weitergeführt. Dadurch bleibt nachvollziehbar, dass sich die Äußerungen auf den Bruder des Generals und dessen Hund beziehen. Auch ChatGPT-5.3 stellt diese Bezugskette mit „*sein Bruder*“ und „*sein Hund*“ eindeutig her. DeepL Pro verwendet dagegen mehrfach Pronomen, die keinen eindeutigen Anschluss an die vorherigen Referenzen herstellen. In der Formulierung „*Ihr Bruder ist es aber*“ bleibt unklar, auf welche Referenzgröße sich „*ihr*“ bezieht. Dasselbe gilt für „*Ist ihr Bruder denn angekommen?*“ und „*Ist das also ihr Hündchen?*“. Zusätzlich wird „*Hündchen*“ anschließend mit dem Pronomen „*sie*“ wiederaufgenommen: „*Nimm sie*“. Aufgrund des sächlichen Bezugswortes wäre an dieser Stelle jedoch eine Wiederaufnahme mit „*es*“ zu erwarten. Die Wahl des Pronomens dürfte damit zusammenhängen, dass das russische Substantiv „*собачонка*“, das DeepL Pro mit „*Hündchen*“ wiedergibt, ebenfalls feminin ist. Dadurch dürfte die feminine Referenzführung des russischen Ausgangstextes in die deutsche Übersetzung übertragen werden, ohne an das sächliche Bezugswort angepasst zu werden. Die lokale Bezugskette zwischen dem General, dessen Bruder und dem Hund wird auf diese Weise uneinheitlich und teilweise missverständlich geführt.

Insgesamt zeigt die Analyse, dass lokale Kohäsionsprobleme nicht in allen drei untersuchten Erzählungen auftreten, sondern sich auf *Ein Chamäleon* konzentrieren. Die relevanten Stellen betreffen vor allem die pronominale Wiederaufnahme von Referenzen in benachbarten Sätzen und Äußerungen. In mehreren Fällen dürfte die im russischen Ausgangstext grammatisch feminine Bezugnahme auf den Hund sowie auf weitere feminine Bezeichnungen für den Hund eine Rolle spielen. Die lokalen Kohäsionsprobleme beschränken sich jedoch nicht allein auf diese Genusproblematik, sondern betreffen auch unklare Anschlüsse an andere Referenzgrößen. In der Humanübersetzung werden die Bezugsketten auf den Hund, den Bruder des Generals und weitere Referenzgrößen konsistent geführt. In den LLM-gestützten Übersetzungen entstehen dagegen teilweise uneinheitliche Verknüpfungen zwischen benachbarten Textstellen und Sätzen. Besonders DeepL Pro weist mehrere lokale Bezugsketten auf, in denen dieselbe Bezugsgröße mit wechselnden oder unklaren Formen wiederaufgenommen wird. ChatGPT-5.3 zeigt in diesem Bereich weniger Auffälligkeiten, weist jedoch ebenfalls eine uneinheitliche pronominale Wiederaufnahme in der Eingangsszene von *Ein Chamäleon* auf. Die lokale Verständlichkeit geht dadurch nicht vollständig verloren. Die sprachliche Verknüpfung zwischen benachbarten Sätzen und Äußerungen wird jedoch dadurch beeinträchtigt. Da diese Beobachtungen teilweise Berührungspunkte mit der SCATE-Unterkategorie *Koreferenz* aufweisen, werden sie nicht als zusätzliche quantitative Fehlerstatistik behandelt. Stattdessen wurde qualitativ beschrieben, wie uneinheitliche Wiederaufnahmen die Verknüpfung benachbarter Sätze und Äußerungen beeinflussen.

8.3 Kohärenz

In diesem Unterkapitel werden ausgewählte Beispiele aus den untersuchten Übersetzungen betrachtet, die im Rahmen der erweiterten SCATE-Taxonomie dem Bereich Kohärenz zugeordnet wurden. Die Analyse dient der Überprüfung der dritten Annahme, der zufolge LLM-gestützte Übersetzungen im Vergleich zu humanen Übersetzungen eine geringere Kohärenz aufweisen. Aufgrund des Umfangs der Fehlerannotation werden im Analyseteil ausgewählte, für die jeweiligen Unterkategorien repräsentative Beispiele diskutiert. Die vollständige Annotation der Kohärenz- sowie Stil- und Registerfehler befindet sich im Anhang.

Tab. 7: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Der Tod eines Beamten* (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a)

Humane Übersetzung	Zu niesen ist schließlich nirgends und niemandem verwehrt. Bauern niesen und Polizisten, mitunter sogar Geheimräte. Alle tun es.
DeepL Pro	Niemandem und nirgendwo ist das Niesen verboten. Männer, Polizeichefs und manchmal sogar <u>Geheimratsecken</u> niesen. Alle niesen.
ChatGPT-5.3	Niesen ist niemandem und nirgends verboten. Es niesen Bauern, Polizeimeister und manchmal sogar wirkliche Staatsräte. Alle niesen.

In einer Passage, in der verschiedene Personengruppen genannt werden, übersetzt DeepL Pro wie folgt: „*Männer, Polizeichefs und manchmal sogar Geheimratsecken niesen.*“ Während im Ausgangstext sowie in der humanen Übersetzung ausschließlich Personengruppen genannt werden, führt die Einfügung des Substantivs „*Geheimratsecken*“ zu einer semantisch unpassenden Aufzählung, die die inhaltliche Nachvollziehbarkeit der Passage beeinträchtigt. Im *Langenscheidt Handwörterbuch Russisch* wird „*тайный советник*“ als „*Geheimrat*“ angegeben (Daum & Schenk 2015: 612). Im vorliegenden Kontext stellt der Ausdruck „*тайные советники*“ („*Geheimräte*“) daher keine mehrdeutige Bezeichnung dar, sondern bezeichnet Personen mit dem Rang eines Geheimrats. Die DeepL-Pro-Übersetzung „*Geheimratsecken*“ lässt sich aus diesem Grund nicht durch eine Mehrdeutigkeit des Ausgangsausdrucks erklären, sondern stellt eine für den Zieltext semantisch unpassende Lösung dar. Der Begriff „*Geheimratsecken*“ bezeichnet im Deutschen keinen Personentyp, sondern ein körperliches Merkmal. Dadurch wird die Aufzählung handelnder Subjekte im Text inhaltlich inkohärent. Die Rezipient*innen können die Aussage nicht mehr sinnvoll in den bestehenden Bedeutungszusammenhang integrieren. Diese Abweichung wird als Kohärenzfehler der Unterkategorie *logische Probleme* annotiert, da der Zieltext im vorliegenden Textzusammenhang inhaltlich schwer nachvollziehbar

ist. Zum Vergleich zeigt die ChatGPT-5.3-Übersetzung derselben Passage eine kohärente Umsetzung: „*Es niesen Bauern, Polizeimeister und manchmal sogar wirkliche Staatsräte.*“. Hier bleibt die Art der genannten Personengruppen konsistent, sodass die Aufzählung in den inhaltlichen Zusammenhang des Textes eingebettet ist.

Tab. 8: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Der Dicke und der Dünne* (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b)

Humane Übersetzung	»Bist du Beamter? Schon vorangekommen?« »Ja, mein Lieber, ich bin Beamter! Schon das zweite Jahr Kollegienassessor und hab den Stanislaw-Orden.
DeepL Pro	– Wo dienst du? Hast du es weit gebracht? – Ich diene, mein Lieber! Ich bin schon im zweiten Jahr Kollegialassessor und habe eine <u>Stanislava</u> .
ChatGPT-5.3	„Wo dienst du? Hast du es zu etwas gebracht?“ „Ich diene, mein Lieber! Seit dem zweiten Jahr bin ich Kollegienassessor und habe den Stanislausorden.

Ein Beispiel für eine problematische Wiedergabe einer kulturellen Referenz lässt sich in der DeepL-Pro-Übersetzung beobachten. Die Auszeichnung, die in der Humanübersetzung als „*Stanislaw-Orden*“ wiedergegeben wird, erscheint bei DeepL Pro als „*eine Stanislava*“. Diese Benennung ist im Deutschen nicht als Orden bzw. Auszeichnung identifizierbar. Die DeepL-Pro-Variante könnte damit zusammenhängen, dass im russischen Ausgangstext nicht die vollständige Ordensbezeichnung mit dem Wort „*Orden*“, sondern die verkürzte Form „*Станислава*“ („*Stanislawa*“) steht. DeepL Pro scheint diese Form offenbar als eigenständige Bezeichnung zu übernehmen, anstatt den Bezug auf den Stanislaw-Orden im Deutschen zu erkennen. Für deutschsprachige Rezipient*innen kann daher unklar bleiben, worauf sich die Aussage bezieht. ChatGPT-5.3 löst diese Stelle mit „*Stanislausorden*“ zielsprachlich nachvollziehbarer, da der Bezug auf eine Auszeichnung erhalten bleibt. Die DeepL-Pro-Stelle wurde daher im Rahmen der Kohärenzanalyse unter *kulturelle Referenz* gefasst.

Tab. 9: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Du willst mich beißen, verdammte Töle?« hört Otschumelow plötzlich. »Laßt ihn nicht laufen, Leute! Heutzutage darf man nicht beißen! Haltet ihn!
--------------------	---

DeepL Pro	– Willst du also beißen, du <u>elendes Weib</u> ? – hört Ochumelow plötzlich. – Jungs, lasst sie nicht durch! Heute darf man nicht beißen! Halt sie fest!
ChatGPT-5.3	Willst du etwa beißen, du verfluchtes Vieh?“ hört Otschumelow plötzlich. „Leute, lasst sie nicht durch! Heute ist Beißen verboten! Haltet sie!

Ein logisches Problem zeigt sich in der DeepL-Pro-Übersetzung der Anfangsszene von *Ein Chamäleon*. Während in der Humanübersetzung mit „*verdammte Töle*“ bereits eindeutig eine Beschimpfung des Hundes vorliegt, verwendet DeepL Pro die Formulierung „*du elendes Weib*“. Da die Textstelle vor der eindeutigen Einführung des Hundes steht, kann diese Formulierung zunächst den Eindruck erwecken, dass eine weibliche Person angesprochen wird. Erst im weiteren Verlauf wird deutlich, dass sich die Beschimpfung auf den Hund beziehen soll. Im Vergleich dazu bleibt der Bezug in der Humanübersetzung durch „*Töle*“ eindeutig auf ein Tier gerichtet. Auch ChatGPT-5.3 stellt mit „*verfluchtes Vieh*“ zunächst eine tierbezogene Beschimpfung her. In der DeepL-Pro-Übersetzung wird die Referenz zu Beginn der Szene dadurch unklar. In der Annotation wurde diese Stelle entsprechend als Kohärenzfehler der Unterkategorie *logische Probleme* berücksichtigt.

Tab. 10: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Das ist, scheint's, der von General Shigalow!« ruft jemand aus der Menge. »Von General Shigalow? Hm!
DeepL Pro	„Das ist, glaube ich, der von General <u>Zhigalov</u> !“, sagt jemand aus der Menge. – General <u>Schigalows</u> ? Hm!
ChatGPT-5.3	Das ist wohl der vom General Schigalow!“ sagt jemand aus der Menge. „Vom General Schigalow? Hm!

Ein weiterer kohärenzbezogener Fehler lässt sich in der DeepL-Pro-Übersetzung von *Ein Chamäleon* bei der Wiedergabe eines Eigennamens beobachten. In zwei unmittelbar aufeinanderfolgenden Sätzen wird derselbe Name unterschiedlich geschrieben: zunächst als „*Zhigalov*“, anschließend als „*Schigalows*“. Da beide Formen im selben lokalen Textabschnitt auftreten und sich auf dieselbe Person beziehen, entsteht eine Inkonsistenz in der Namensreferenz. Obwohl für die Rezipient*innen grundsätzlich erkennbar bleibt, dass dieselbe Person gemeint ist, stört die uneinheitliche Namenswiedergabe die lokale Kohärenz. Im Vergleich dazu gibt ChatGPT-

5.3 den Namen in beiden Sätzen einheitlich als „*Schigalow*“ wieder. Aufgrund der uneinheitlichen Namensreferenz wurde die DeepL-Pro-Stelle der Unterkategorie *Inkonsistenz* zugewiesen.

Tab. 11: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	Du hast dir wohl den Finger an einem Nagel aufgerissen, und dann kam dir die Idee, dabei was herauszuschlagen.
DeepL Pro	Du hast dir wohl den Finger mit einem Nagel aufgerissen, und dann kam dir die Idee, <u>sie abzureißen</u> .
ChatGPT-5.3	Du hast dir den Finger wohl an einem Nagel aufgerissen, und dann kam dir die Idee, etwas herauszuschlagen.

Eine weitere kohärenzrelevante Auffälligkeit betrifft die DeepL-Pro-Übersetzung von *Ein Chamäleon*. In der Humanübersetzung wird der Vorwurf, der Betroffene habe sich den Finger möglicherweise selbst verletzt, um daraus einen Vorteil zu ziehen, mit „*dabei was herauszuschlagen*“ idiomatisch wiedergegeben. ChatGPT-5.3 formuliert ähnlich mit „*etwas herauszuschlagen*“ und stellt den Zusammenhang damit ebenfalls verständlich her. DeepL Pro gibt die Stelle dagegen mit „*dann kam dir die Idee, sie abzureißen*“ wieder. Die Formulierung „*sie abzureißen*“ lässt sich vermutlich auf eine wörtliche Wiedergabe des im Ausgangstext verwendeten russischen Verbs „*copbamb*“ zurückführen. Für „*copbamb*“ werden im *Langenscheidt Handwörterbuch Russisch* unter anderem die Bedeutungen „*abreißen*“ und „*herunterreißen*“ angegeben (Daum & Schenk 2015: 584). Im vorliegenden Kontext ist das Verb nicht wörtlich zu verstehen, sondern übertragen im Sinne von „*etwas herausschlagen*“ bzw. „*aus der Situation einen Vorteil ziehen*“. In der DeepL-Pro-Übersetzung bleibt unklar, worauf sich das Pronomen „*sie*“ bezieht und was mit „*abzureißen*“ gemeint ist. Der Vorwurf wird dadurch im Zieltext nicht nachvollziehbar realisiert. Aus diesem Grund wurde für diese Stelle in der Annotation ein Kohärenzfehler der Unterkategorie *logische Probleme* erfasst.

Tab. 12: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	Würde man so einen Hund in Petersburg oder in Moskau antreffen, wißt ihr, was dann wäre? Dort würde man nicht erst ins Gesetz sehen, sondern augenblicklich – weg damit!
--------------------	--

DeepL Pro	Wenn so ein Hund in St. Petersburg oder Moskau auftauchen würde, weißt du, was dann passieren würde? Da würde man nicht erst im Gesetz nachschlagen, sondern sofort – <u>atme nicht mehr!</u>
ChatGPT-5.3	Wenn so ein Hund in Petersburg oder Moskau auftauchen würde, wisst ihr, was geschähe? Da würde man nicht ins Gesetz schauen, sondern sofort – aus!

Ein weiteres logisches Problem tritt in der DeepL-Pro-Übersetzung von *Ein Chamäleon* auf. In der Humanübersetzung wird die Aussage durch die elliptische Formulierung „weg damit!“ wiedergegeben. Gemeint ist, dass ein solcher Hund sofort beseitigt bzw. getötet würde. ChatGPT-5.3 formuliert mit „sofort – aus!“ ebenfalls knapp und zielsprachlich nachvollziehbar. DeepL Pro gibt die Stelle dagegen mit „sofort – atme nicht mehr!“ wieder. Diese Formulierung wirkt im Deutschen wie ein direkter Befehl an den Hund, mit dem Atmen aufzuhören. Dadurch wird die Aussage im Kontext schwer nachvollziehbar, da die Bedeutung einer sofortigen Beseitigung des Hundes nicht klar hergestellt wird. Die Stelle wurde daher als Kohärenzfehler der Unterkategorie *logische Probleme* erfasst.

Die folgende Tabelle gibt einen Überblick über die in der finalen Auswertung berücksichtigten Kohärenzfehler nach Unterkategorie und Übersetzungsvariante.

Tab. 13: Übersicht der in die finale Auswertung aufgenommenen Kohärenzfehler nach Unterkategorie und Übersetzungsvariante

Unterkategorie	Humanübersetzung	DeepL Pro	ChatGPT-5.3
logisches Problem	0	5	0
Inkonsistenz	0	6	0
kulturelle Referenz	0	1	1
Koreferenz	0	3	1
Gesamt	0	15	2

Tabelle 13 zeigt, dass die Kohärenzfehler in den untersuchten Übersetzungen deutlich häufiger bei DeepL Pro als bei ChatGPT-5.3 auftreten, während für die Humanübersetzungen keine Kohärenzfehler erfasst wurden. Insgesamt wurden 15 kohärenzbezogene Fehler bei DeepL Pro und zwei bei ChatGPT-5.3 identifiziert. Besonders auffällig ist die Unterkategorie der *Inkonsistenzen*, in der sechs Fälle ausschließlich auf DeepL Pro entfallen. Diese betreffen vor allem

uneinheitliche Wiedergaben von Eigennamen und Anredeformen, wodurch Figuren oder soziale Rollen im Zieltext nicht einheitlich dargestellt werden. ChatGPT-5.3 weist in dieser Unterkategorie dagegen keine Fehler auf. Auch bei den *logischen Problemen* zeigt sich ein eindeutiges Bild: Alle fünf identifizierten Fälle entfallen DeepL Pro. Diese Fehler betreffen vor allem Stellen, an denen der Zieltext den handlungslogischen Zusammenhang nicht eindeutig herstellt, wie beispielsweise durch semantisch unpassende Formulierungen oder unklare Bezüge innerhalb einer Passage.

In der Unterkategorie *kulturelle Referenz* wurde jeweils ein Fehler pro System gezählt. Dies weist darauf hin, dass sowohl DeepL Pro als auch ChatGPT-5.3 Schwierigkeiten mit historisch bzw. institutionell gebundenen Bezeichnungen haben können. Bei Koreferenzproblemen ist DeepL Pro mit drei Fällen stärker vertreten als ChatGPT-5.3 mit einem Fall. Der Unterschied fällt hier weniger deutlich aus als bei den logischen Problemen, macht jedoch sichtbar, dass die lokale Referenzführung in beiden Systemen problematisch sein kann.

Insgesamt zeigen die identifizierten Fehler und die quantitative Übersicht, dass Kohärenzprobleme in den LLM-gestützten Übersetzungen vor allem dort auftreten, wo handlungslogische Zusammenhänge, soziale Rollen, Eigennamen, kulturelle Referenzen und Koreferenzbezüge präzise hergestellt werden müssen. Dabei ist hervorzuheben, dass die untersuchten Kohärenzfehler nicht immer zu einem vollständigen Verlust der Verständlichkeit führen, jedoch die lokale Nachvollziehbarkeit des Zieltextes beeinträchtigen. Im Vergleich zur Humanübersetzung werden Bezugnahmen auf Figuren, soziale Rollen und Handlungszusammenhänge in den LLM-gestützten Übersetzungen teilweise weniger konsistent und weniger kontextangemessen hergestellt. Besonders DeepL Pro weist in der Fehlerannotation mehrere kohärenzrelevante Auffälligkeiten auf. ChatGPT-5.3 zeigt zwar in einigen Fällen stabilere Lösungen, erzeugt jedoch ebenfalls lokale Kohärenzprobleme. Die Zahlen stützen innerhalb der untersuchten Übersetzungen die Annahme, dass die untersuchten LLM-gestützten Übersetzungen im Bereich Kohärenz Schwächen gegenüber der Humanübersetzung aufweisen.

8.4 Stil & Register

In diesem Unterkapitel werden stilistische und registerbezogene Problemstellen in den untersuchten Übersetzungen analysiert. Die Analyse dient der Überprüfung der vierten Annahme, der zufolge LLM-gestützte Übersetzungen im Bereich stilistischer Nuancen und Registerwahl Schwächen gegenüber humanen Übersetzungen aufweisen. Im Folgenden werden ausgewählte

Beispiele zur Veranschaulichung der zentralen Ergebnisse diskutiert. Die vollständige Fehlerannotation ist im Anhang dokumentiert.

Tab. 14: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Der Tod eines Beamten* (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a)

Humane Übersetzung	Euer Exz'lenz! Wenn ich es wage, Euer Exz'lenz zu belästigen, dann einzig, ich versichere Sie, aus einem Gefühl der Reue!
DeepL Pro	„ <u>Eure Hoheit</u> ! Wenn ich es wage, Eure Hoheit zu stören, dann nur aus einem Gefühl der Reue, das kann ich Ihnen versichern!
ChatGPT-5.3	– Euer Exzellenz! Wenn ich mir erlaube, Sie zu behelligen, so einzig aus einem Gefühl, ich darf sagen, der Reue! Nicht absichtlich, Sie wissen es selbst!

Ein Registerproblem zeigt sich in der DeepL-Pro-Übersetzung von *Der Tod eines Beamten*. Während in der Humanübersetzung die Anrede „*Euer Exz'lenz*“ verwendet wird, gibt DeepL Pro diese Stelle mit „*Eure Hoheit*“ wieder. Diese Anrede verschiebt den sozialen Rahmen der Szene vom Amts- bzw. Beamtenbereich in den Bereich monarchischer Anreden. Dadurch wird der Rangbezug des angesprochenen Geheimrats registerbezogen falsch eingeordnet. Zugleich entsteht eine Kohärenzproblematik, da DeepL Pro an anderen Stellen „*Eure Exzellenz*“ verwendet und die Anredeform im Zieltext somit nicht einheitlich bleibt. Die Stelle wurde daher als Beispiel für eine Mehrfachzuordnung behandelt: Sie wurde als Fehler im Bereich Stil & Register mit der Unterkategorie *Register* sowie als Kohärenzfehler der Unterkategorie *Inkonsistenz* erfasst.

Tab. 15: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Der Tod eines Beamten* (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a)

Humane Übersetzung	Darf ich es denn wagen, mich lustig zu machen? Würden wir uns lustig machen, wo, bitteschön, bliebe dann die Achtung vor Respektspersonen?
DeepL Pro	Wie könnte ich lachen? Wenn wir lachen würden, gäbe es keinen Respekt mehr <u>vor Personen</u> ...
ChatGPT-5.3	Wie dürfte ich mir erlauben zu spotten? Wenn wir spotten, dann gäbe es ja überhaupt keinen Respekt mehr <u>vor Personen</u> ...

Ein weiteres Registerproblem betrifft die Abschwächung sozialer Hierarchie in *Der Tod eines Beamten*. Die Humanübersetzung verwendet in der oben zitierten Passage „*Respektspersonen*“

und macht damit deutlich, dass es sich nicht um beliebige Personen handelt, sondern um gesellschaftlich bzw. amtlich ranghöhere Personen, vor denen besondere Achtung gilt. DeepL Pro und ChatGPT-5.3 reduzieren diese hierarchische Komponente auf die allgemeine Bezeichnung „Personen“. Dadurch wird die Aussage neutralisiert und die soziale Rangordnung abgeschwächt. Die entsprechenden Stellen in den Übersetzungen von DeepL Pro und ChatGPT-5.3 wurden aus diesem Grund als Fehler im Bereich Stil & Register erfasst und der Unterkategorie *Register* zugewiesen.

Tab. 16: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Der Dicke und der Dünne* (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b)

Humane Übersetzung	»Porfiri!« rief der Dicke, als er den Dünnen erblickte. »Bist du's? Mein Bester! Wieviel Zeit ins Land gegangen ist!«
DeepL Pro	– Porfirij! – rief der Dicke aus, als er den Dünnen sah. – Bist du das? Mein Lieber! <u>Wie viele Winter, wie viele Jahre!</u>
ChatGPT-5.3	„Porfirij!“, rief der Dicke aus, als er den Dünnen erblickte. „Bist du es wirklich? Mein Lieber! <u>Wie viele Winter, wie viele Jahre!</u> “

Eine unflüssige Wiedergabe liegt auch bei der Begrüßungsformel in *Der Dicke und der Dünne* vor. In der Humanübersetzung wird die russische Begrüßungsformel mit „*Wieviel Zeit ins Land gegangen ist!*“ idiomatisch ins Deutsche übertragen. DeepL Pro und ChatGPT-5.3 geben die Stelle dagegen identisch mit „*Wie viele Winter, wie viele Jahre!*“ wieder. Dabei handelt es sich um eine wörtliche Übertragung der russischen Begrüßungsformel, die im Deutschen ungewöhnlich wirkt und als Begrüßungsformel nicht geläufig ist. In der Annotation wurde die Stelle sowohl bei DeepL Pro als auch bei ChatGPT-5.3 als Fehler im Bereich Stil & Register mit der Unterkategorie *Unflüssigkeit* erfasst.

Tab. 17: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Der Dicke und der Dünne* (Tschechow 2003b; DeepL Pro 2026b; ChatGPT-5.3 2026b)

Humane Übersetzung	»Euer Exzellenz, ich ... Ehm, sehr angenehm! Ein Freund, wie man so sagt, aus Kindertagen und sind plötzlich, ehm, zu solchen Würden gelangt! Hi hi hi.«
DeepL Pro	– Ich, Eure Exzellenz... Sehr erfreulich! Ein Freund, sozusagen aus Kindertagen, und plötzlich sind Sie zu solchen Würdenträgern geworden! Hi-hi.

ChatGPT-5.3	„Ich, Euer Exzellenz ... Sehr <u>angenehm-s</u> ! Ein Freund, könnte man sagen, aus Kindertagen, und nun haben Sie sich zu solch einem Würdenträger <u>emporgeschwungen-s</u> ! <u>Hi-hi-s</u> .“
-------------	---

Ein weiteres Beispiel für stilistisch-registerbezogene Unflüssigkeit lässt sich in der ChatGPT-5.3-Übersetzung von *Der Dicke und der Dünne* beobachten. Nachdem der Dünne den höheren Rang seines ehemaligen Schulfreundes erkennt, verändert sich seine Rede deutlich in Richtung Unterwürfigkeit. ChatGPT-5.3 versucht offenbar, ein im russischen Ausgangstext vorhandenes historisch markiertes Höflichkeits- bzw. Unterwürfigkeitselement nachzubilden, indem es Formen wie „*Sehr angenehm-s*“, „*emporgeschwungen-s*“ und „*Hi-hi-hi-s*“ bildet. Im Deutschen ist das angehängte „-s“ kein etabliertes Mittel, um eine solche Sprechweise auszudrücken. Die unterwürfige Rede wird dadurch zwar sichtbar hervorgehoben, aber zielsprachlich unnatürlich realisiert. DeepL Pro übernimmt das angehängte „-s“ dagegen nicht und vermeidet damit diese Form der zielsprachigen Unflüssigkeit. Aus diesem Grund wurde die ChatGPT-5.3-Stelle als Fehler im Bereich Stil & Register erfasst und der Unterkategorie *Unflüssigkeit* zugeordnet.

Tab. 18: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	Ich mache eine feine Arbeit. Sie sollen mir das zahlen, weil ich den Finger vielleicht eine Woche nicht bewegen kann ...
DeepL Pro	Ich habe nur kleine Arbeiten. Man soll mir das Geld geben, <u>denn mit diesem Finger werde ich vielleicht eine Woche lang keinen Finger rühren können...</u>
ChatGPT-5.3	Meine Arbeit ist fein. Man soll mir Schadenersatz zahlen, denn vielleicht kann ich diesen Finger eine Woche lang nicht bewegen...

Eine weitere zielsprachlich unflüssige Lösung liegt in der DeepL-Pro-Übersetzung von *Ein Chamäleon* vor. In der Humanübersetzung begründet der Betroffene seine Forderung nach Entschädigung damit, dass er den verletzten Finger möglicherweise eine Woche lang nicht bewegen könne. DeepL Pro gibt diese Stelle mit „*denn mit diesem Finger werde ich vielleicht eine Woche lang keinen Finger rühren können...*“ wieder. Dabei wird der konkret verletzte Finger mit der festen Redewendung „*keinen Finger rühren*“ verbunden. Diese Redewendung bedeutet im Deutschen, nichts zu tun, während sich „*mit diesem Finger*“ auf den tatsächlich verletzten Körperteil bezieht. Dadurch überschneiden sich die wörtliche und die übertragene Bedeutung von „*Finger*“ auf unnatürliche Weise. Die beabsichtigte Aussage bleibt zwar verständlich, die Formulierung wirkt jedoch zielsprachlich unflüssig. ChatGPT-5.3 vermeidet dieses Problem,

indem die Übersetzung näher an der wörtlichen Bedeutung bleibt: „*diesen Finger eine Woche lang nicht bewegen*“. Die DeepL-Pro-Stelle wurde daher als Fehler im Bereich Stil & Register erfasst und der Unterkategorie *Unflüssigkeit* zugeordnet.

Tab. 19: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Nun gut ... Wessen Hund ist das? Ich lasse das so nicht durchgehen. Ich werde euch zeigen, was es heißt, Hunde herumstreunen zu lassen!
DeepL Pro	–Gut... Wem gehört der Hund? <u>Das lasse ich nicht auf sich sitzen</u> . Ich werde euch zeigen, wie man Hunde laufen lässt!
ChatGPT-5.3	„Gut... Wessen Hund ist das? Ich lasse das nicht so! Ich werde euch zeigen, wie man Hunde frei herumlaufen lässt!

Ein anderer Fall von Unflüssigkeit betrifft ebenfalls die DeepL-Pro-Übersetzung und zeigt sich bei der Wiedergabe einer festen Wendung. In der Humanübersetzung lautet die Stelle: „*Ich lasse das so nicht durchgehen*“. DeepL Pro gibt sie dagegen mit „*Das lasse ich nicht auf sich sitzen*“ wieder. Diese Formulierung entspricht keiner üblichen deutschen Wendung und wirkt wie eine Vermischung zweier Ausdrücke: „*das lasse ich nicht auf mir sitzen*“ und „*das lasse ich nicht auf sich beruhen*“. Obwohl die beabsichtigte Bedeutung aus dem Kontext noch erkennbar bleibt, wirkt die Äußerung zielsprachlich unflüssig. Im Vergleich dazu vermeidet ChatGPT-5.3 mit „*Ich lasse das nicht so!*“ die Vermischung fester Wendungen. Die DeepL-Pro-Stelle wurde deshalb als Fehler im Bereich Stil & Register erfasst und der Unterkategorie *Unflüssigkeit* zugewiesen.

Tab. 20: Humanübersetzung und LLM-gestützte Übersetzungen einer ausgewählten Passage aus *Ein Chamäleon* (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3 2026c)

Humane Übersetzung	»Das ist nicht unsrer«, fährt Prochor fort. »Das ist der vom Bruder des Generals, der neulich gekommen ist. Unser General ist kein Liebhaber von Windhunden.
DeepL Pro	– Der gehört nicht uns, – fährt Prochor fort. – Der gehört dem Bruder des Generals, der neulich angekommen ist. Unserer <u>ist kein Fan</u> von Windhunden.
ChatGPT-5.3	„Er ist nicht von uns,“ fährt Prochor fort. „Er gehört dem Bruder des Generals, der neulich angekommen ist. Unser Herr mag keine Windhunde ...

Ein weiteres Registerproblem betrifft die DeepL-Pro-Übersetzung der Aussage über die Windhunde in *Ein Chamäleon*. In der Humanübersetzung wird die Aussage mit „*Unser General ist*

kein Liebhaber von Windhunden“ wiedergegeben. Diese Variante passt zum historischen Kontext sowie zur Figurenrede. DeepL Pro gibt die Stelle dagegen mit „Unserer ist kein Fan von Windhunden“ wieder. Diese Variante ist im Deutschen zwar verständlich, wirkt jedoch modern und umgangssprachlich, wodurch ein deutlicher Registerbruch entsteht, da die Ausdrucksweise nicht zum historischen Kontext der Erzählung passt. ChatGPT-5.3 vermeidet diese moderne Wendung und gibt die Aussage mit „Unser Herr mag keine Windhunde...“ stilistisch passender wieder. Ausschlaggebend für die Einordnung in die Unterkategorie *Register* ist hier die modern wirkende Ausdrucksweise, die nicht zum historischen Kontext der Figurenrede passt.

Zur Einordnung der qualitativen Analyse wird auch für den Bereich Stil & Register eine quantitative Übersicht erstellt. Tabelle 21 zeigt die in der finalen Auswertung berücksichtigten Fehler nach Unterkategorie und Übersetzungsvariante.

Tab. 21: Übersicht der in die finale Auswertung aufgenommenen Fehler im Bereich Stil & Register nach Unterkategorie und Übersetzungsvariante

Unterkategorie	Humanübersetzung	DeepL Pro	ChatGPT-5.3
Register	0	9	3
Unflüssigkeit	0	7	8
Wiederholung	0	1	1
Gesamt	0	17	12

Tabelle 21 zeigt, dass für die Humanübersetzungen keine Fehler im Bereich Stil & Register erfasst wurden, während in den LLM-gestützten Übersetzungen insgesamt 29 Fehler in die finale Auswertung aufgenommen wurden. Davon entfallen 17 auf DeepL Pro und 12 auf ChatGPT-5.3. Mit insgesamt 15 Fällen ist die Unterkategorie *Unflüssigkeit* am stärksten vertreten, gefolgt von *Register* mit 12 Fällen. Dies deutet darauf hin, dass beide LLM-gestützten Übersetzungssysteme Schwierigkeiten mit idiomatischen und zielsprachlich natürlich wirkenden Formulierungen haben können.

Die zweithäufigste Fehlerart bildet die Kategorie *Register*. Bei Registerfehlern lässt sich im Vergleich zur *Unflüssigkeit* ein deutlicherer Unterschied zwischen DeepL Pro und ChatGPT-5.3 erkennen. DeepL Pro weist neun Fälle auf, während bei ChatGPT-5.3 drei Fälle identifiziert wurden. Die Registerfehler betreffen vor allem Stellen, an denen soziale Hierarchie, historische Figurenrede oder der stilistische Ton der Aussage nicht angemessen wiedergegeben werden. Besonders auffällig sind dabei abgeschwächte soziale Rangmarkierungen und unpassende Anredeformen. Vereinzelt treten auch modern bzw. umgangssprachlich wirkende Wendungen auf.

Die Unterkategorie *Wiederholung* tritt mit insgesamt zwei Fällen deutlich seltener auf. Beide Fälle betreffen dieselbe Textstelle, da sowohl DeepL Pro als auch ChatGPT-5.3 dort eine wiederholte Wortwahl aufweisen. Obwohl diese Unterkategorie quantitativ schwach vertreten ist, ist sie für literarische Übersetzung relevant, da wiederholte oder vereinfachte Ausdrucksweisen die stilistische Wirkung des Zieltextes abschwächen können.

Insgesamt deuten die Ergebnisse darauf hin, dass Stil- und Registerprobleme in den untersuchten LLM-gestützten Übersetzungen mehrere Ebenen betreffen: die zielsprachliche Flüssigkeit, die stilistische Gestaltung der Figurenrede sowie die sprachliche Darstellung sozialer Rangverhältnisse. Während *Unflüssigkeiten* bei beiden Übersetzungssystemen in nahezu gleichem Umfang auftreten, ist DeepL Pro in der vorliegenden Auswertung besonders bei *Registerfehlern* auffällig. ChatGPT-5.3 bietet in mehreren Fällen stilistisch passendere Lösungen, vermeidet aber nicht immer unnatürlich wirkende oder künstliche Formulierungen. Die Ergebnisse stützen damit die Annahme, dass die untersuchten LLM-gestützten Übersetzungen im Bereich Stil & Register Schwächen gegenüber der Humanübersetzung aufweisen.

9 Ergebnisse

In diesem Kapitel werden die Ergebnisse der durchgeführten Analyse beschrieben. Die Darstellung orientiert sich an den vier Grundannahmen der Arbeit bezogen auf lexikalische Vielfalt, lokale Kohäsion, Kohärenz sowie Stil & Register. Die lexikalische Vielfalt und die lokale Kohäsion wurden im Rahmen einer für die vorliegende Arbeit adaptierten Key-Feature-Analyse untersucht, wobei die lexikalische Vielfalt quantitativ mithilfe von Type-Token-Ratio erfasst und die lokale Kohäsion qualitativ analysiert wurde. Kohärenz sowie Stil & Register wurden mithilfe einer manuellen Fehlerannotation auf Grundlage der erweiterten SCATE-Taxonomie erfasst.

Die Ergebnisse zur lexikalischen Vielfalt ergeben kein einheitliches Bild. Die TTR-Analyse zeigt, dass die Humanübersetzung in den Erzählungen *Der Tod eines Beamten* und *Der Dicke und der Dünne* jeweils die höchste lexikalische Vielfalt aufweist. Die LLM-gestützten Übersetzungen erreichen in beiden Fällen niedrigere TTR-Werte, wobei ChatGPT-5.3 näher an der Humanübersetzung als DeepL Pro liegt. Zugleich sind die Unterschiede zwischen den TTR-Werten insgesamt gering, insbesondere im Vergleich zwischen der Humanübersetzung und ChatGPT-5.3-Übersetzung.

Ein abweichendes Ergebnis zeigt sich bei *Ein Chamäleon*. Hier weist ChatGPT-5.3 den höchsten TTR-Wert auf, gefolgt von DeepL Pro, während die Humanübersetzung den niedrigsten Wert erreicht. Auch in diesem Fall fallen die Unterschiede jedoch gering aus. Die Ergebnisse zu *Ein Chamäleon* sprechen daher gegen eine durchgängige Bestätigung der Annahme, dass LLM-gestützte Übersetzungen eine geringere lexikalische Vielfalt als humane Übersetzungen aufweisen. Für die Interpretation der TTR-Werte ist jedoch zu berücksichtigen, dass eine höhere lexikalische Vielfalt nicht zwangsläufig mit einer höheren Übersetzungsqualität gleichzusetzen ist. Wiederholungen können in literarischen Texten kohäsiv oder stilistisch motiviert sein. Außerdem kann eine höhere lexikalische Vielfalt in LLM-gestützten Übersetzungen auch durch uneinheitliche Wortwahl, inkonsistente Benennungen oder fehlerhafte Varianten entstehen. Die erste Annahme der vorliegenden Arbeit wird durch die TTR-Analyse daher nur teilweise gestützt.

Die Analyse der lokalen Kohäsion zeigt, dass eindeutige Auffälligkeiten nur in begrenztem Umfang auftreten und ausschließlich in der Erzählung *Ein Chamäleon* identifiziert wurden. In *Der Tod eines Beamten* und *Der Dicke und der Dünne* wurden keine lokalen Kohäsionsprobleme festgestellt. Die identifizierten Stellen in *Ein Chamäleon* betreffen insbesondere pronominale Wiederaufnahmen und lokale Bezugsketten. Während in der Humanübersetzung keine Auffälligkeiten in Bezug auf lokale Kohäsion festgestellt wurden, zeigen sich in den LLM-gestützten Übersetzungen einzelne Problemstellen. DeepL Pro ist in diesem Bereich stärker betroffen als ChatGPT-5.3. Die Auffälligkeiten bestehen vor allem in der uneinheitlichen Wiederaufnahme derselben Referenzgröße innerhalb eines engen lokalen Zusammenhangs, insbesondere wenn Bezeichnungen in Form von Substantiven und anschließende pronominale Wiederaufnahmen nicht konsistent aufeinander abgestimmt sind. Bei der Einordnung dieser Ergebnisse ist zu berücksichtigen, dass sie teilweise Berührungspunkte mit der SCATE-Unterkategorie *Koreferenz* aufweisen. Sie werden daher nicht als zusätzliche quantitative Fehlerstatistik behandelt. Die zweite Annahme der vorliegenden Arbeit wird durch die qualitative Analyse grundsätzlich gestützt, da entsprechende Auffälligkeiten ausschließlich in den LLM-gestützten Übersetzungen auftreten. Die Aussagekraft dieses Ergebnisses bleibt jedoch begrenzt, da die identifizierten Fälle nicht über alle drei Erzählungen verteilt sind, sondern sich auf einzelne Passagen in *Ein Chamäleon* konzentrieren.

Im Bereich Kohärenz fällt der Unterschied zwischen Humanübersetzungen und LLM-gestützten Übersetzungen deutlicher als bei der lexikalischen Vielfalt und der lokalen Kohäsion aus. Während für die Humanübersetzungen keine Kohärenzfehler erfasst wurden, wurden in

den LLM-gestützten Übersetzungen insgesamt 17 kohärenzbezogene Fehler identifiziert. Der Großteil entfällt auf DeepL Pro mit 15 Fehlern, ChatGPT-5.3 weist demgegenüber zwei Fehler auf. Die häufigste Unterkategorie bilden *Inkonsistenzen* mit sechs Fällen, die ausschließlich bei DeepL Pro auftreten. Sie betreffen vor allem uneinheitliche Wiedergaben von Eigennamen und Anredeformen. Darauf folgen logische Probleme mit insgesamt fünf Fällen, die vollständig auf DeepL Pro entfallen. *Koreferenzprobleme* sowie *kulturelle Referenzen* treten seltener auf, zeigen jedoch ebenfalls, dass präzise Bezugnahmen und kontextgebundene Bezeichnungen für beide LLM-gestützten Übersetzungssysteme problematisch sein können.

Besonders auffällig ist bei DeepL Pro die uneinheitliche Wiedergabe von Eigennamen: Teilweise werden unterschiedliche Namensformen innerhalb desselben Textes verwendet, in einzelnen Fällen sogar in unmittelbarer Nähe zueinander. Damit zeigt sich, dass Kohärenzprobleme insbesondere dort entstehen, wo der Zieltext Figuren, soziale Rollen und Handlungszusammenhänge über mehrere Sätze hinweg stabil herstellen muss. Die Ergebnisse stützen somit die dritte Annahme der Arbeit, nach der LLM-gestützte Übersetzungen im untersuchten Korpus eine geringere Kohärenz als die Humanübersetzung aufweisen.

Im Bereich Stil & Register lassen sich ebenfalls markante Unterschiede zwischen den Humanübersetzungen und den LLM-gestützten Übersetzungen feststellen. Während in den Humanübersetzungen keine entsprechenden Fehler festgestellt wurden, treten in den LLM-gestützten Übersetzungen insgesamt 29 Fälle auf, davon 17 in DeepL Pro und 12 in ChatGPT-5.3. Die häufigste Auffälligkeit stellt die Unterkategorie *Unflüssigkeit* dar. Diese umfasst insgesamt 15 Fälle und betrifft insbesondere Formulierungen, die im Zieltext nicht idiomatisch oder stilistisch natürlich wirken. Unflüssige Formulierungen treten bei DeepL Pro und ChatGPT-5.3 nahezu gleich häufig auf. Eine weitere zentrale Fehlerkategorie bildet das *Register*. Hier lassen sich insgesamt 12 Fälle beobachten, wobei DeepL Pro mit neun Fällen deutlich stärker als ChatGPT-5.3 betroffen ist. Die Registerprobleme betreffen vor allem unangemessene soziale Markierungen, insbesondere im Bereich der Anredeformen sowie der Wiedergabe historisch oder institutionell geprägter Ausdrucksweisen. In mehreren Fällen kommt es zu einer Abschwächung der hierarchischen Distanz oder zu stilistisch nicht angemessenen Formulierungen. Die Unterkategorie *Wiederholung* tritt mit zwei Fällen vergleichsweise selten auf und betrifft bei beiden Übersetzungssystemen jeweils eine identische Textstelle. Obwohl die Anzahl gering ist, können solche Wiederholungen in literarischen Texten die stilistische Wirkung beeinträchtigen. Insgesamt zeigen die Ergebnisse, dass Stil- und Registerprobleme in den LLM-gestützten Übersetzungen sowohl die sprachliche Flüssigkeit als auch die stilistische Angemessenheit und die

Darstellung sozialer Hierarchien betreffen. Während Unflüssigkeiten bei beiden Systemen ähnlich häufig auftreten, weist insbesondere DeepL Pro im Bereich *Register* eine höhere Fehleranfälligkeit auf. ChatGPT-5.3 zeigt insgesamt etwas stabilere Ergebnisse, bleibt jedoch ebenfalls von stilistischen Inkonsistenzen betroffen. Die Ergebnisse stützen damit die Annahme, dass LLM-gestützte Übersetzungen im Bereich Stil & Register Schwächen im Vergleich zu Humanübersetzungen aufweisen.

Zusammenfassend lässt sich feststellen, dass die Unterschiede zwischen humanen und LLM-gestützten Übersetzungen vor allem in den Bereichen Kohärenz sowie Stil & Register deutlich hervortreten. In diesen beiden Bereichen weisen die LLM-gestützten Übersetzungen im untersuchten Korpus im Vergleich zu Humanübersetzungen Auffälligkeiten auf, wobei DeepL Pro insgesamt stärker betroffen als ChatGPT-5.3 ist. Die Resultate zur lexikalischen Vielfalt und zur lokalen Kohäsion fallen dagegen differenzierter aus: Die TTR-Analyse bestätigt die erste Annahme nur teilweise und Auffälligkeiten bezogen auf lokale Kohäsion treten nicht systematisch über alle untersuchten Übersetzungen hinweg, sondern nur in einzelnen Passagen von *Ein Chamäleon* auf.

10 Fazit

Ziel der vorliegenden Arbeit war es, die aus der Forschungsfrage abgeleiteten Annahmen zu überprüfen. Im Mittelpunkt stand die Frage, in welcher Hinsicht sich die LLM-gestützten Übersetzungen ausgewählter literarischer Texte von menschlichen Übersetzungen aus dem Russischen ins Deutsche unterscheiden. Dafür wurden drei kurze Erzählungen bzw. ihre Übersetzungen von Anton Tschechow analysiert: *Der Tod eines Beamten*, *Der Dicke und der Dünne* und *Ein Chamäleon*. Die lexikalische Vielfalt wurde quantitativ anhand der Type-Token-Ratio untersucht, während lokale Kohäsion qualitativ analysiert wurde. Kohärenz sowie Stil & Register wurden auf Grundlage der erweiterten SCATE-Taxonomie nach Tezcan et al. (2019) annotiert und ausgewertet.

Die Analyse zeigt, dass die Unterschiede zwischen Humanübersetzungen und LLM-gestützten Übersetzungen nicht in allen Untersuchungsbereichen gleichermaßen ausgeprägt sind. Die Annahme einer geringeren lexikalischen Vielfalt LLM-gestützter Übersetzungen konnte nicht bestätigt werden. In zwei der drei untersuchten Erzählungen wiesen die Humanübersetzungen die höchsten TTR-Werte auf. Bei *Ein Chamäleon* ergab sich jedoch ein abweichendes Ergebnis. Die TTR erwies sich damit als nur begrenzt aussagekräftiges Maß, da keine signifi-

kanten Unterschiede aus den TTR-Werten erkennbar sind. Außerdem soll berücksichtigt werden, dass höhere lexikalische Vielfalt nicht automatisch mit höherer Übersetzungsqualität gleichzusetzen ist.

Auch die Resultate zur lokalen Kohäsion fallen differenziert aus. Die Analyse zeigte einzelne Auffälligkeiten in den LLM-gestützten Übersetzungen, während in den Humanübersetzungen keine entsprechenden Problemstellen festgestellt wurden. Diese Auffälligkeiten konzentrierten sich jedoch auf einzelne Passagen von *Ein Chamäleon*. Die Annahme geringerer lokaler Kohäsion in LLM-gestützten Übersetzungen wird daher grundsätzlich gestützt, ihre Reichweite bleibt im vorliegenden Korpus jedoch begrenzt.

Stärker ausgeprägt sind die Unterschiede in den Bereichen Kohärenz sowie Stil & Register aus. Hier weisen die LLM-gestützten Übersetzungen gegenüber den Humanübersetzungen erkennbare Schwächen auf. Besonders relevant sind inkonsistente Benennungen, unklare Bezugnahmen, logische Brüche, unidiomatische Formulierungen sowie Verschiebungen im sozialen und historischen Register. Die Unterschiede betreffen somit vor allem solche Bereiche, in denen literarische Übersetzung nicht nur lexikalische Variation, sondern auch konsistente Bezugnahmen, stilistisch passende Figurenrede, soziale Rangmarkierung und stilistische Angemessenheit verlangt.

Insgesamt lässt sich festhalten, dass die untersuchten LLM-gestützten Übersetzungen den Humanübersetzungen nicht in jeder Hinsicht unterlegen sind, jedoch insbesondere in den Bereichen lokale Kohäsion, Kohärenz sowie Stil & Register deutliche Einschränkungen aufweisen. Im Vergleich zu den LLM-gestützten Übersetzungen wurden in den Humanübersetzungen im untersuchten Korpus keine entsprechenden Auffälligkeiten in den Bereichen lokale Kohäsion, Kohärenz sowie Stil & Register festgestellt.

Bei der Einordnung und Interpretation der Ergebnisse ist zu berücksichtigen, dass ihre Aussagekraft durch den Umfang des Korpus begrenzt ist. Ein weiterer methodischer Aspekt betrifft die TTR als Maß lexikalischer Vielfalt. Die TTR liefert Informationen über das Verhältnis von Types und Tokens, erlaubt jedoch keine direkte Aussage darüber, ob einzelne lexikalische Entscheidungen im jeweiligen Kontext passend sind. Auch die Fehlerannotation nach der erweiterten SCATE-Taxonomie ist mit einem gewissen interpretativen Anteil verbunden. Dies gilt insbesondere für stilistische und registerbezogene Bewertungen, bei denen Fragen der stilistischen Angemessenheit, der Figurenrede und angemessenen Registers eine Rolle spielen.

Die externe Überprüfung der Annotationen diene dazu, diese Einschätzungen abzusichern und die finale Auswertung auf eindeutig bestätigte Fälle zu beschränken.

Für weiterführende Untersuchungen wäre es sinnvoll, das Korpus zu erweitern und weitere kurze Erzählungen von Tschechow einzubeziehen. Dadurch ließe sich prüfen, ob sich die in der vorliegenden Analyse beobachteten Tendenzen auch anhand eines größeren und stärker variierten Textkorpus zeigen lassen. Außerdem könnten neben der Type-Token-Ratio textlängenunabhängige Maße lexikalischer Vielfalt herangezogen und durch eine qualitative Bewertung der Wortwahl ergänzt werden. Zudem wäre eine Erweiterung der manuellen Annotation denkbar, beispielsweise durch den Vergleich mehrerer Annotator*innen oder durch eine Gewichtung der annotierten Problemstellen. Darüber hinaus könnten weiterführende Untersuchungen die vollständige erweiterte SCATE-Taxonomie nach Tezcan et al. (2019) berücksichtigen und neben den in der vorliegenden Arbeit untersuchten Kategorien auch *Accuracy*-Fehler sowie weitere *Fluency*-Unterkategorien einbeziehen. Dadurch könnte ein umfassenderes Bild der Übersetzungsqualität LLM-gestützter literarischer Übersetzungen gewonnen werden. Besonders aufschlussreich wäre schließlich ein Vergleich verschiedener LLM-gestützten Systemen sowie unterschiedlicher Prompting-Strategien bzw. System- und Einstellungskonfigurationen. Da LLM-gestützte Übersetzungssysteme fortlaufend weiterentwickelt werden, wären solche Vergleiche auch mit Blick auf die zeitliche Veränderbarkeit der Ergebnisse relevant.

11 Bibliographie

Primärliteratur

ChatGPT-5.3 (2026a). Übersetzung von Anton Tschechows *Der Tod eines Beamten* aus dem Russischen ins Deutsche. Prompt: „Übersetze den folgenden Text ins Deutsche.“ Erstellt im Rahmen der vorliegenden Untersuchung im März 2026.

ChatGPT-5.3 (2026b). Übersetzung von Anton Tschechows *Der Dicke und der Dünne* aus dem Russischen ins Deutsche. Prompt: „Übersetze den folgenden Text ins Deutsche.“ Erstellt im Rahmen der vorliegenden Untersuchung im März 2026.

ChatGPT-5.3 (2026c). Übersetzung von Anton Tschechows *Ein Chamäleon* aus dem Russischen ins Deutsche. Prompt: „Übersetze den folgenden Text ins Deutsche.“ Erstellt im Rahmen der vorliegenden Untersuchung im März 2026.

DeepL Pro (2026a). Übersetzung von Anton Tschechows *Der Tod eines Beamten* aus dem Russischen ins Deutsche. Erstellt im Rahmen der vorliegenden Untersuchung im März 2026.

DeepL Pro (2026b). Übersetzung von Anton Tschechows *Der Dicke und der Dünne* aus dem Russischen ins Deutsche. Erstellt im Rahmen der vorliegenden Untersuchung im März 2026.

DeepL Pro (2026c). Übersetzung von Anton Tschechows *Ein Chamäleon* aus dem Russischen ins Deutsche. Erstellt im Rahmen der vorliegenden Untersuchung im März 2026.

Tschechow, Anton (1950a). Смерть чиновника [Der Tod eines Beamten]. In: Ermilov, Vladimir (Hg.) *Собрание сочинений. Рассказы 1880–1883* [Gesammelte Werke. Erzählungen 1880–1883]. Bd. I. Moskau: Prawda, 86–88.

Tschechow, Anton (1950b). Толстый и тонкий [Der Dicke und der Dünne]. In: Ermilov, Vladimir (Hg.) *Собрание сочинений. Рассказы 1880–1883* [Gesammelte Werke. Erzählungen 1880–1883]. Bd. I. Moskau: Prawda, 104–105.

Tschechow, Anton (1950c). Хамелеон [Ein Chamäleon]. In: Ermilov, Vladimir (Hg.) *Собрание сочинений. Рассказы 1883–1884* [Gesammelte Werke. Erzählungen 1883–1884]. Bd. II. Moskau: Prawda, 43–45.

- Tschechow, Anton (2003a). Der Tod eines Beamten. Aus dem Russischen übersetzt von Vera Bischitzky. In: *In der Sommerfrische. Erzählungen 1880–1887*. Düsseldorf/Zürich: Artemis & Winkler, 18–21.
- Tschechow, Anton (2003b). Der Dicke und der Dünne. Aus dem Russischen übersetzt von Vera Bischitzky. In: *In der Sommerfrische. Erzählungen 1880–1887*. Düsseldorf/Zürich: Artemis & Winkler, 25–27.
- Tschechow, Anton (2003c). Ein Chamäleon. Aus dem Russischen übersetzt von Marianne Wiebe. In: *In der Sommerfrische. Erzählungen 1880–1887*. Düsseldorf/Zürich: Artemis & Winkler, 43–47.

Sekundärliteratur

- Bai, Xiaojing & Yu, Shiwen (2023). Rule-Based Machine Translation. In: Chan, Sin-wai (ed.) *Routledge Encyclopedia of Translation Technology*. London/New York: Routledge, 193–207.
- Beaugrande, Robert-Alain de & Dressler, Wolfgang Ulrich (1981). *Einführung in die Textlinguistik*. Tübingen: Niemeyer.
- Bloom, Harold (2000). *Anton Chekhov*. Philadelphia, Pa.: Chelsea House Publishers.
- Brusasco, Paola (2022). Pragmatic and cognitive elements in literary machine translation: An assessment of an excerpt from J. Polzin’s *Brood* translated with Google, DeepL, and Microsoft Bing. In: Hadley, James Luke; Taivalkoski-Shilov, Kristiina; Teixeira, Carlos S. C. & Toral, Antonio (eds.) *Using Technologies for Creative-Text Translation*. New York/London: Routledge, 139–160.
- Castilho, Sheila; Doherty, Stephen; Gaspari, Federico & Moorkens, Joss (2018). Approaches to Human and Machine Translation Quality Assessment. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (eds.) *Translation Quality Assessment: From Principles to Practice*. Cham: Springer, 9–38.
- Chudakov, Alexander (2000). Dr Chekhov: a biographical essay (29 January 1860–15 July 1904). In: Gottlieb, Vera & Allain, Paul (eds.) *The Cambridge Companion to Chekhov*. Cambridge: Cambridge University Press, 3–16.

- Colina, Sonia (2020). Quality, translation. In: Baker, Mona & Saldanha, Gabriela (eds.) *Routledge Encyclopedia of Translation Studies*. 3rd. ed. London/New York: Routledge, 458–463.
- Daum, Edmund & Schenk, Werner (2015). *Langenscheidt Handwörterbuch Russisch: Russisch – Deutsch, Deutsch – Russisch*. Neubearbeitung. Stuttgart: PONS.
- DeepL (2021). How does DeepL work? https://www.deepl.com/en/blog/how-does-deepl-work#training_data (Stand: 01.03.2026).
- DeepL (2024). DeepL launches next generation LLM that outperforms GPT-4, Google, and Microsoft for translation quality. <https://www.deepl.com/en/press-release/deepl-launches-next-generation-llm-that-outperforms-gpt-4-google-and-microsoft-for-translation-quality> (Stand: 01.03.2026).
- DeepL (o.J.a). About Us. <https://www.deepl.com/en/about-us> (Stand: 01.03.2026).
- DeepL (o.J.b). About DeepL language models. <https://support.deepl.com/hc/en-us/articles/14241705319580-About-DeepL-language-models> (Stand: 01.03.2026).
- DeepL (o.J.c). About DeepL plans. <https://support.deepl.com/hc/en-us/articles/360019924499-About-DeepL-plans> (Stand: 01.03.2026).
- DeepL (o.J.d). Privacy Policy DeepL. <https://www.deepl.com/en/privacy> (Stand: 01.03.2026).
- DeepL (o.J.e). Publisher. <https://www.deepl.com/en/publisher> (Stand: 01.03.2026).
- Dorst, Aletta G. (2023). Metaphor in literary machine translation: style, creativity and literariness. In: Rothwell, Andrew; Way, Andy & Youdale, Roy (eds.) *Computer-Assisted Literary Translation*. New York/London: Routledge, 173–186.
- Du, Shuxiang; Guerberofo Arenas, Ana; Toral, Antonio; Gerrits, Kyo & Borillo, Josep Marco (2025). Optimising ChatGPT for creativity in literary translation: A case study from English into Dutch, Chinese, Catalan and Spanish. In: Bouillon, Pierrette; Gerlach, Johanna; Girletti, Sabrina; Volkart, Lise; Rubino, Raphael; Sennrich, Rico; Farinha, Ana C.; Gaido, Marco; Daems, Joke; Kenny, Dorothy; Moniz, Helena & Szoc, Sara (eds.) *Proceedings of Machine Translation Summit XX: Volume I*. Geneva: European Association for Machine Translation, 578–591.
- Fonteyne, Margot; Tezcan, Arda & Macken, Lieve (2020). Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on

- Document Level. In: Calzolari, Nicoletta; Béchet, Frédéric; Blache, Philippe; Choukri, Khalid; Cieri, Christopher; Declerck, Thierry; Goggi, Sara; Isahara, Hitoshi; Maegaard, Bente; Mariani, Joseph; Mazo, Hélène; Moreno, Asuncion; Odijk, Jan & Piperidis, Stelios (eds.) *Proceedings of the 12th Conference on Language Resources and Evaluation*. Marseille: European Language Resources Association, 3790–3798.
- Gao, Ruiyao; Lin, Yumeng; Zhao, Nan & Cai, Zhenguang G. (2024). Machine translation of Chinese classical poetry: a comparison among ChatGPT, Google Translate, and DeepL Translator. *Humanities & Social Sciences Communications* 11 (1), 1–10.
- Hadley, James Luke (2023). MT and CAT: Challenges, Irrelevancies, or Opportunities for Literary Translation? In: Rothwell, Andrew; Way, Andy & Youdale, Roy (eds.) *Computer-Assisted Literary Translation*. New York/London: Routledge, 91–105.
- Hadley, James Luke; Taivalkoski-Shilov, Kristiina; Teixeira, Carlos S. C. & Toral, Antonio (2022). Introduction. In: Hadley, James Luke; Taivalkoski-Shilov, Kristiina; Teixeira, Carlos S. C. & Toral, Antonio (eds.) *Using Technologies for Creative-Text Translation*. New York/London: Routledge, 1–17.
- House, Juliane (2015). *Translation quality assessment: past and present*. New York: Routledge.
- House, Juliane (2024). *Translation: The Basics*. 2nd. ed. London/New York: Routledge.
- Hutchins, W. John (1997). From First Conception to First Demonstration: the Nascent Years of Machine Translation, 1947-1954. A Chronology. *Machine Translation* 12 (3), 195–252.
- Hutchins, W. John (2004). The Georgetown-IBM Experiment Demonstrated in January 1954. In: Frederking, Robert E. & Taylor, Kathryn B (eds.) *Machine Translation: From Real Users to Research*. Berlin/Heidelberg: Springer, 102–114.
- Hutchins, W. John (2023). Machine Translation: History of Research and Applications. In: Chan, Sin-wai (ed.) *Routledge Encyclopedia of Translation Technology*. London/New York: Routledge, 128–144.
- Hutchins, W. John & Somers, Harold L. (1992). *An Introduction to Machine Translation*. London: Academic Press.
- Karakanta, Alina; Nas, Mayra & Dorst, Aletta G. (2025). Metaphors in Literary Machine Translation: Close but no cigar? In: Bouillon, Pierrette; Gerlach, Johanna; Girletti, Sabrina; Volkart, Lise; Rubino, Raphael; Sennrich, Rico; Farinha, Ana C.; Gaido, Marco; Daems,

- Joke; Kenny, Dorothy; Moniz, Helena & Szoc, Sara (eds.) *Proceedings of Machine Translation Summit XX: Volume 1*. Geneva: European Association for Machine Translation, 276–286.
- Karpinska, Marzena & Iyyer, Mohit (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. In: Koehn, Philipp; Haddow, Barry; Kocmi, Tom & Monz, Christof (eds.) *Proceedings of the Eighth Conference on Machine Translation (WMT)*. Singapore: Association for Computational Linguistics, 419–451.
- Kenny, Dorothy (2019). Machine translation. In: Rawling, Piers & Wilson, Philip (eds.) *The Routledge Handbook of Translation and Philosophy*. London/New York: Routledge, 428–445.
- Kenny, Dorothy (2022). Human and machine translation. In: Kenny, Dorothy (ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, 23–49.
- Kenny, Dorothy (2025). Technologies and the Future of Translation: Two Perspectives. In: Bassnett, Susan & Johnston, David (eds.) *Debates in Translation Studies*. London/New York: Routledge, 91–105.
- Koehn, Philipp (2020). *Neural Machine Translation*. Cambridge: Cambridge University Press.
- Koller, Werner & Henjum, Kjetil Berg (2020). *Einführung in die Übersetzungswissenschaft*. 9. Aufl. Tübingen: Narr Francke Attempto.
- Kong, Delu & Macken, Lieve (2025). Can Peter Pan Survive MT? A Stylometric Study of LLMs, NMTs, and HTs in Children’s Literature Translation. In: Vanroy, Bram; Lefer, Marie-Aude; Macken, Lieve; Ruffo, Paola; Guerberof Arenas, Ana & Hansen, Damien (eds.) *Proceedings of the Second Workshop on Creative-text Translation and Technology (CTT)*. Geneva: European Association for Machine Translation, 52–70.
- Lauscher, Susanne (2000). Translation Quality Assessment: Where Can Theory and Practice Meet? *Translator* 6 (2), 149–168.
- Lavie, Alon & Denkowski, Michael J. (2009). The METEOR metric for automatic evaluation of machine translation. *Machine Translation* 23 (2–3), 105–115.

- Liu, Yang & Zhang, Min (2023). Statistical Machine Translation. In: Chan, Sin-wai (ed.) *Routledge Encyclopedia of Translation Technology*. London/New York: Routledge, 208–218.
- Loehlin, James N. (2010). *The Cambridge Introduction to Chekhov*. Cambridge: Cambridge University Press.
- Lommel, Arle; Uszkoreit, Hans & Burchardt, Aljoscha (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica* 12, 455–463.
- Lommel, Arle; Gladkoff, Serge; Melby, Alan; Wright, Sue Ellen; Strandvik, Ingemar; Gasova, Katerina; Vaasa, Angelika; Benzo, Andy; Marazzato Sparano, Romina; Foresi, Monica; Innis, Johani; Han, Lifeng & Nenadic, Goran (2024). The Multi-Range Theory of Translation Quality Measurement: MQM scoring models and Statistical Quality Control. In: Martindale, Marianna; Campbell, Janice; Savenkov, Konstantin & Goel, Shivali (eds.) *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 2: Presentations)*. Chicago: Association for Machine Translation in the Americas, 75–94.
- Manning, Christopher D. & Schütze, Hinrich (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: MIT Press.
- Matusov, Evgeny (2019). The Challenges of Using Neural Machine Translation for Literature. In: Hadley, James; Popović, Maja; Afli, Haithem & Way, Andy (eds.) *Proceedings of the Qualities of Literary Machine Translation*. Dublin: European Association for Machine Translation, 10–19.
- Moorkens, Joss; Way, Andy & Lankford, Séamus (2025). *Automating Translation*. London/New York: Routledge.
- Newmark, Peter (1991). *About Translation*. Clevedon: Multilingual Matters.
- Nida, Eugene A. (1964). *Toward a science of translating: with special reference to principles and procedures involved in Bible translating*. Leiden: E. J. Brill.
- Nida, Eugene A. & Taber, Charles R. (1969). *Theorie und Praxis des Übersetzens: unter besonderer Berücksichtigung der Bibelübersetzung*. Stuttgart: Weltbund der Bibelgesellschaften.

- OpenAI (2026). GPT-5.3 Instant: Smoother, more useful everyday conversations <https://openai.com/index/gpt-5-3-instant/> (Stand: 05.03.2026).
- OpenAI (o.J.a). About. <https://openai.com/about/> (Stand: 01.03.2026).
- OpenAI (o.J.b). ChatGPT Capabilities Overview. <https://help.openai.com/en/articles/9260256-chatgpt-capabilities-overview> (Stand: 01.03.2026).
- OpenAI (o.J.c). Our structure. <https://openai.com/our-structure/> (Stand: 01.03.2026).
- OpenAI Developers (o.J.). Key concepts. <https://developers.openai.com/api/docs/concepts> (Stand: 01.03.2026).
- Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In: Isabelle, Pierre; Charniak, Eugene & Lin, Dekang (eds.) *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania: Association for Computational Linguistics, 311–318.
- Parts, Lyudmila (2024). The Short Story. In: Franklin, Simon; Reich, Rebecca & Widdis, Emma (eds.) *The New Cambridge History of Russian Literature*. Cambridge: Cambridge University Press, 553–569.
- Pérez-Ortiz, Juan Antonio; Forcada, Mikel L. & Sánchez-Martínez, Felipe (2022). How neural machine translation works. In: Kenny, Dorothy (ed.) *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, 141–164.
- Prunč, Erich (2013). *Entwicklungslinien der Translationswissenschaft: Von den Asymmetrien der Sprachen zu den Asymmetrien der Macht*. 3., erweiterte und verbesserte Auflage. Berlin: Frank & Timme.
- Reiß, Katharina (1971). *Möglichkeiten und Grenzen der Übersetzungskritik: Kategorien und Kriterien für eine sachgerechte Beurteilung von Übersetzungen*. München: Max Hueber.
- Reiß, Katharina & Vermeer, Hans J. (1984). *Grundlegung einer allgemeinen Translationstheorie* (Linguistische Arbeiten; 147). Tübingen: Max Niemeyer.

- Schäffner, Christina (1998). From ‘Good’ to ‘Functionally Appropriate’: Assessing Translation Quality. In: Schäffner, Christina (ed.) *Translation and Quality*. Clevedon: Multilingual Matters, 1–5.
- Schwartz, Lane (2018). The history and promise of machine translation. In: Lacruz, Isabel & Jääskeläinen, Riitta (eds.) *Innovation and Expansion in Translation Process Research*. Amsterdam/Philadelphia: John Benjamins, 161–190.
- Snover, Matthew; Dorr, Bonnie; Schwartz, Richard; Micciulla, Linnea & Makhoul, John (2006). *A Study of Translation Edit Rate with Targeted Human Annotation*. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. Cambridge, Massachusetts: Association for Machine Translation in the Americas, 223–231.
- Stein, Daniel (2009). Maschinelle Übersetzung – ein Überblick. *Journal for Language Technology and Computational Linguistics* 24 (3), 5–18.
- Taivalkoski-Shilov, Kristiina (2019). Ethical issues regarding machine(-assisted) translation of literary texts. *Perspectives* 27 (5), 689–703.
- Tezcan, Arda; Hoste, Véronique & Macken, Lieve (2017). SCATE Taxonomy and Corpus of Machine Translation Errors. In: Corpas Pastor, Gloria & Durán-Muñoz, Isabel (eds.) *Trends in E-Tools and Resources for Translators and Interpreters*. Leiden/Boston: Brill/Rodopi, 219–244.
- Tezcan, Arda; Daems, Joke & Macken, Lieve (2019). When a ‘sport’ is a person and other issues for NMT of novels. In: Hadley, James; Popović, Maja; Afli, Haithem & Way, Andy (eds.) *Proceedings of the Qualities of Literary Machine Translation*. Dublin: European Association for Machine Translation, 40–49.
- Toral, Antonio & Way, Andy (2015). Machine-assisted translation of literary text: A case study. *Translation Spaces* 4 (2), 240–267.
- Toury, Gideon (1995). *Descriptive Translation Studies and beyond*. Amsterdam: Benjamins.
- Tschechow, Anton (1950d). И. Л. Леонтьеву (Щеглову) [an I. L. Leontjew (Schtscheglow)]. In: Ermilov, Vladimir (Hg.) *Собрание сочинений. Письма* [Gesammelte Werke. Briefe]. Bd. XII. Moskau: Prawda, 70–71.

- Tschechow, Anton (1950e). Ал. П. Чехову [an Al. P. Tschechow]. In: Ermilov, Vladimir (Hg.) *Собрание сочинений. Письма* [Gesammelte Werke. Briefe]. Bd. XII. Moskau: Prawda, 126.
- UrhG (1936). *Urheberrechtsgesetz. Dauer des Urheberrechtes. Werke der Literatur, der Tonkunst und der bildenden Künste*. § 60 Abs. 1. BGBl. Nr. 111/1936 zuletzt geändert durch BGBl. I Nr. 99/2015. <https://ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10001848&FassungVom=2026-01-14&Artikel=&Paragraf=60&Anlage=&Uebergangsrecht=> (Stand: 02.02.2026).
- Van Egdom, Gys-Walt; Declercq, Christophe & Kusters, Onno (2024). “Can make mistakes”: Prompting ChatGPT to Enhance Literary MT output. In: Vanroy, Bram; Lefer, Marie-Aude; Macken, Lieve & Ruffo, Paola (eds.) *Proceedings of the 1st Workshop on Creative-text Translation and Technology*. Sheffield, UK: European Association for Machine Translation, 10–20.
- Voigt, Rob & Jurafsky, Dan (2012). Towards a literary machine translation: The Role of Referential Cohesion. In: Elson, David; Kazantseva, Anna; Mihalcea, Rada & Szpakowicz, Stan (eds.) *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Montréal: Association for Computational Linguistics, 18–25.
- Voyant Tools (o.J.). About. <https://voyant-tools.org/docs/tutorial-about.html> (Stand: 10.04.2026).
- Walter, Katharina (2026). Jeopardizing Linguistic Diversity: How AI-generated Translations Neutralize Vernacular Irish English. In: Walter, Katharina & Agnetta, Marco (eds.) *Applying Artificial Intelligence in Translation: Possibilities, Processes and Phenomena*. New York/London: Routledge, 29–45.
- Wang, Longyue; Tu, Zhaopeng; Gu, Yan; Liu, Siyou; Yu, Dian; Ma, Quingsong; Lyu, Chenyang; Zhou, Liting; Liu, Chao-Hong; Ma, Yufeng; Chen, Weiyu; Graham, Yvette; Webber, Bonnie; Koehn, Philipp; Way, Andy; Yuan, Yulin & Shi, Shuming (2023). Findings of the WMT 2023 Shared Task on Discourse-Level Literary Translation: A Fresh Orb in the Cosmos of LLMs. In: Koehn, Philipp; Haddow, Barry; Kocmi, Tom; Monz, Christof (eds.) *Proceedings of the Eighth Conference on Machine Translation*. Singapore: Association for Computational Linguistics, 55–67.
- Webster, Rebecca; Fonteyne, Margot; Tezcan, Arda; Macken, Lieve & Daems, Joke (2020). Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw

Neural Machine Translation Outputs in a Corpus of English Literary Classics. *Informat-ics* 7 (3), 1–21.

Zhang, Ran; Zhao, Wei & Eger, Steffen (2025a). How Good Are LLMs for Literary Translation, Really? Literary Translation Evaluation with Humans and LLMs. In: Chiruzzo, Luis; Ritter, Alan & Wang, Lu (eds.) *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Albuquerque, New Mexico: Association for Computational Linguistics, 10961–10988.

Zhang, Ran; Zhao, Wei; Macken, Lieve & Eger, Steffen (2025b). LiTransProQA: An LLM-based Literary Translation Evaluation Metric with Professional Question Answering. In: Christodoulopoulos, Christos; Chakraborty, Tanmoy; Rose, Carolyn & Peng, Violet (eds.) *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*. Suzhou, China: Association for Computational Linguistics, 29099–29121.

Anhang

Anhang A: Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register zu *Der Tod eines Beamten* (Tschechow 2003a; DeepL Pro 2026a; ChatGPT-5.3 2026a)

Fall 1			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Eines lieblichen Abends saß der nicht minder liebreizende Staatsdiener Iwan Dmitritsch Tschervjakow in der zweiten Reihe des Parketts und betrachtete »Die Glocken von Corneville« durch sein Opernglas.	An einem <u>schönen</u> Abend saß der nicht minder <u>schöne</u> Henker Ivan Dmitrich Chervyakov in der zweiten Reihe und schaute durch sein Fernglas auf die „Glocken von Kornevil“.	An einem <u>schönen</u> Abend saß ein nicht minder <u>schöner</u> Exekutor namens Iwan Dmitrijewitsch Tschervjakow in der zweiten Reihe der Sessel und blickte durch ein Opernglas auf die „Glocken von Corneville“.	schön – schön (DeepL Pro, ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Wiederholung	Wiederholung von „schön“ statt differenzierter Adjektivwahl. Dadurch stilistische Nivellierung und abgeschwächte expressive Wirkung.	ja	Ja, Wiederholung von „schön“ nivelliert die stilistische Differenzierung. Es wirkt wie eine stilistische Monotonie.
Fall 2			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Eines lieblichen Abends saß der nicht minder liebreizende Staatsdiener Iwan Dmitritsch Tschervjakow in der zweiten Reihe des Parketts und betrachtete »Die Glocken von Corneville« durch sein Opernglas.	An einem schönen Abend saß der nicht minder schöne Henker Ivan Dmitrich Chervyakov in der zweiten Reihe und schaute durch sein Fernglas auf die „Glocken von Kornevil“.	An einem schönen Abend saß ein nicht minder schöner Exekutor namens Iwan Dmitrijewitsch Tschervjakow in der zweiten Reihe der Sessel und blickte durch ein Opernglas auf die „Glocken von Corneville“.	in der zweiten Reihe der Sessel (ChatGPT-5.3)

Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „in der zweiten Reihe der Sessel“ ist im Deutschen ungebräuchlich. Gemeint ist die zweite Reihe im Parkett bzw. im Zuschauerraum. Die räumliche Zuordnung bleibt zwar verständlich, die Formulierung wirkt jedoch zielsprachlich unnatürlich.	ja	Es wirkt nicht flüssig und stilistisch merkwürdig.
Fall 3			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Er nieste, wie man sieht. Zu niesen ist schließlich nirgends und niemandem verwehrt. Bauern niesen und Polizisten, mitunter sogar Geheimräte. Alle tun es.	Er nieste, wie Sie sehen. Niemandem und nirgendwo ist das Niesen verboten. Männer , Polizeichefs und manchmal sogar Geheimratsecken niesen. Alle niesen.	Er nieste, wie Sie sehen. Niesen ist niemandem und nirgends verboten. Es niesen Bauern , Polizeimeister und manchmal sogar wirkliche Staatsräte. Alle niesen.	Männer (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	„Männer“ ist eine sehr allgemeine Personenbezeichnung. Im Vergleich zur HÜ „Bauern“ wird der soziale Kontrast der Aufzählung abgeschwächt.	ja	
Fall 4			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Er nieste, wie man sieht. Zu niesen ist schließlich nirgends und niemandem verwehrt. Bauern niesen	Er nieste, wie Sie sehen. Niemandem und nirgendwo ist das Niesen verboten. Männer, Polizeichefs	Er nieste, wie Sie sehen. Niesen ist niemandem und nirgends verboten. Es	Geheimratsecken (DeepL Pro)

und Polizisten, mit- unter sogar Geheim- räte . Alle tun es.	und manchmal sogar <u>Geheimratsecken</u> niesen. Alle nieseen.	niesen Bauern, Poli- zeimeister und manchmal sogar wirkliche Staatsräte . Alle nieseen.	
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – logisches Problem	„Geheimratsecken“ bezeichnet im Deut- schen keine Personen- gruppe, sondern eine Form des Haaransat- zes. Dadurch er- scheint ein Körper- merkmal als Subjekt des Niesens, was ei- nen logischen Bruch in der Aufzählung er- zeugt.	ja	Semantisch ist der Satz problema- tisch.
Fall 5			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
In jenem alten Herrn erkannte Tscher- wjakow den Zivil- General Brisshalow vom Verkehrsminis- terium. [...]	In dem alten Mann er- kannte Chervyakow den Generalmajor <u>Brizhalov</u> , der im Verkehrsministerium tätig war. [...]	In dem alten Herrn erkannte Tscher- wjakow den Zivilge- neral Brisžalow aus dem Verkehrsminis- terium. [...]	Brizhalov / Brizjalow (DeepL Pro)
In der Pause trat er auf Brisshalow zu...	In der Pause ging er zu <u>Brizjalow</u> ...	In der Pause ging er zu Brisžalow ...	
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Inkonsistenz	In der DeepL-Pro- Übersetzung wird derselbe Eigenname zunächst als „Briz- halov“ und an späte- ren Stellen als „Briz- jalow“ wiedergege- ben. Die Bezugnahme auf die Figur bleibt zwar nachvollziehbar, die lexikalische Wie- deraufnahme erfolgt	ja	

	jedoch nicht konsistent.		
Fall 6			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Euer Exz'lenz! Wenn ich es wage, Euer Exz'lenz zu belästigen, dann einzig, ich versichere Sie, aus einem Gefühl der Reue!	Eure Hoheit! Wenn ich es wage, Eure Hoheit zu stören, dann nur aus einem Gefühl der Reue, das kann ich Ihnen versichern!	– Euer Exzellenz! Wenn ich mir erlaube, Sie zu behelligen, so einzig aus einem Gefühl, ich darf sagen, der Reue! Nicht absichtlich, Sie wissen es selbst!	Eure Hoheit (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register + Kohärenz – Inkonsistenz	DeepL Pro verwendet an anderen Stellen „Eure Exzellenz“. Der Wechsel zu „Eure Hoheit“ verschiebt den Rangbezug vom Amts-/Beamtenbereich in den Bereich monarchischer Anreden und ist dadurch registerbezogen unpassend. Zugleich entsteht eine Inkonsistenz der Anredeform.	ja	Der soziale Rahmen wird durch die Anrede eindeutig verschoben.
Fall 7			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Einen Brief schrieb er dem General allerdings nicht. So viel er auch nachdachte, er brachte diesen Brief einfach nicht zustande.	Er schrieb dem General keinen Brief. Er dachte und dachte, aber <u>ihm fiel einfach kein Brief ein.</u>	Einen Brief schrieb er dem General nicht. Er dachte und dachte und konnte sich diesen Brief einfach nicht ausdenken.	ihm fiel einfach kein Brief ein (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „ihm fiel einfach kein Brief ein“ wirkt im	ja	Die Formulierung wirkt unnatürlich.

	Deutschen kollokativ ungewohnt. Geläufiger wäre etwa „ihm fiel nicht ein, was er schreiben sollte“ oder „er brachte den Brief nicht zustande“.		
Fall 8			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Darf ich es denn wagen, mich lustig zu machen? Würden wir uns lustig machen, wo, bitteschön, bliebe dann die Achtung vor Respektpersonen ?«	Wie könnte ich lachen? Wenn wir lachen würden, gäbe es keinen Respekt mehr <u>vor Personen</u> ...	Wie dürfte ich mir erlauben zu spotten? Wenn wir spotten, dann gäbe es ja überhaupt keinen Respekt mehr <u>vor Personen</u> ...	vor Personen (DeepL Pro, ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	„Personen“ ist im Deutschen sehr allgemein und neutral. Im Vergleich zur HÜ „Respektpersonen“ geht dadurch die soziale und hierarchische Markierung der Aussage verloren. Der Bezug auf ranghöhere bzw. besonders zu achtende Personen wird abgeschwächt.	ja	Die Hierarchie der Aussage wird abgeschwächt.

**Anhang 2: Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register zu
Der Dicke und der Dünne (Tschechow 2003b; DeepL Pro 2026b; Chat-
GPT-5.3 2026b)**

Fall 1			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Porfiri!« rief der Dicke, als er den Dünnen erblickte. »Bist du's? Mein Bester! Wieviele Zeit ins Land gegangen ist! «	– Porfirij! – rief der Dicke aus, als er den Dünnen sah. – Bist du das? Mein Lieber! <u>Wie viele Winter, wie viele Jahre!</u>	„Porfirij!“, rief der Dicke aus, als er den Dünnen erblickte. „Bist du es wirklich? Mein Lieber! <u>Wie viele Winter, wie viele Jahre!</u> “	Wie viele Winter, wie viele Jahre! (DeepL Pro, ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung wirkt im Deutschen ungebräuchlich und als Begrüßungsformel nicht geläufig.	ja	Die Formulierung wirkt sehr „übersetzt“/nicht idiomatisch.
Fall 2			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Na, schau mich doch mal an! Noch immer der gleiche Adonis wie früher! Der gleiche fescbe Kerl und Dandy! Ach, du lieber Himmel! Erzähl mal! Bist du reich?	Nun, sieh mich doch mal genau an! Ein genauso hübscher Kerl wie früher! <u>Ein genauso lebenswerter und eleganter Mann!</u> Ach du Gott! Na, wie geht es dir? Bist du reich?	Nun, sieh mich nur richtig an! Noch immer derselbe Schönl-ing wie früher! Die-selbe kleine Seele und derselbe Geck! Ach, du lieber Gott! Na, wie geht's dir?	Ein genauso liebenswerter und eleganter Mann! (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	DeepL Pro glättet die scherzhaft überzeichnete Selbstbeschreibung zu einer neutraleren Charakterisierung. Die expressive Wirkung der Figurenrede wird dadurch abgeschwächt.	ja	Die leicht ironische Selbstinszenierung wird dadurch gemildert.

Fall 3			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Na, schau mich doch mal an! Noch immer der gleiche Adonis wie früher! Der gleiche fescche Kerl und Dandy! Ach, du lieber Himmel! Erzähl mal! Bist du reich?	Nun, sieh mich doch mal genau an! Ein genauso hübscher Kerl wie früher! Ein genauso liebenswerter und eleganter Mann! Ach du Gott! Na, wie geht es dir? Bist du reich?	Nun, sieh mich nur richtig an! Noch immer derselbe Schöning wie früher! <u>Die-selbe kleine Seele und derselbe Geck!</u> Ach, du lieber Gott! Na, wie geht's dir?	dieselbe kleine Seele (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „dieselbe kleine Seele“ wirkt im Deutschen ungebräuchlich und semantisch unklar. Die scherzhaft-vertrauliche Charakterisierung der Figur bleibt dadurch unpräzise.	ja	
Fall 4			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Und dies mein Sohn Nathanail , Quartaner. [...] Nathanail überlegte einen Augenblick und zog dann die Mütze.	Und das ist mein Sohn, <u>Nathanael</u> , Schüler der dritten Klasse. [...] <u>Nafanail</u> dachte kurz nach und versteckte sich hinter dem Rücken seines Vaters.	Und das ist mein Sohn, Nafanail , Schüler der dritten Klasse. [...] Nafanail überlegte einen Augenblick und versteckte sich hinter dem Rücken seines Vaters.	Nathanael / Nafanail (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Inkonsistenz	Dieselbe Figur wird im Zieltext zunächst	ja	

	als „Nathanael“ bezeichnet, an späteren Stellen jedoch als „Nafanail“ wiedergegeben. Dadurch wird die Namensreferenz im Zieltext uneinheitlich dargestellt. Der Bezug auf dieselbe Figur bleibt zwar aus dem Kontext erschließbar, die lexikalische Wiederaufnahme erfolgt jedoch nicht konsistent. Die Form „Nafanja“, die ebenfalls in der DeepL-Pro-Übersetzung vorkommt, wird nicht als Fehler gewertet, da sie als Kose- bzw. Kurzform zu „Nafanail“ verstanden werden kann.		
Fall 5			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Bist du Beamter? Schon vorangekommen?« »Ja, mein Lieber, ich bin Beamter! Schon das zweite Jahr Kollegienassessor und hab den Stanislaw-Orden .	– Wo dienst du? Hast du es weit gebracht? – Ich diene, mein Lieber! Ich bin schon im zweiten Jahr Kollegialassessor und habe eine <u>Stanislava</u> .	„Wo dienst du? Hast du es zu etwas gebracht?“ „Ich diene, mein Lieber! Seit dem zweiten Jahr bin ich Kollegienassessor und habe den Stanislaus-orden .	Stanislava (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – kulturelle Referenz	„Eine Stanislava“ ist im Deutschen nicht als Auszeichnung erkennbar. Dadurch bleibt unklar, worauf sich die Figur bezieht.	ja	Die Referenz ist für deutschsprachige Personen nicht klar.

Fall 6			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Zuerst hab ich ja in der Bezirksverwaltung gedient, jetzt aber bin ich hierher versetzt worden, als Ressort-Vorsteher im selben Amt ... Werd nun hier dienen.	Ich habe, weißt du, im Ministerium gedient, und jetzt bin ich als Verpflegungsleiter derselben Behörde hierher versetzt worden... Hier werde ich dienen.	Früher war ich im Departement, und jetzt bin ich hierher versetzt worden, als <u>Tischchef</u> in derselben Behörde ... Hier werde ich also dienen.	Tischchef (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – kulturelle Referenz	„Tischchef“ ist im Deutschen nicht als geläufige Amts- oder Behördenbezeichnung erkennbar. Die historisch-institutionelle Funktion der Figur bleibt dadurch unklar.	ja	Die Amtsfunktion ist nicht verständlich.
Fall 7			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Euer Exzellenz, ich ... Ehm, sehr angenehm ! Ein Freund, wie man so sagt, aus Kindertagen und sind plötzlich, ehm, zu solchen Würden gelangt! Hi hi hi. «	– Ich, Eure Exzellenz... Sehr erfreulich ! Ein Freund, sozusagen aus Kindertagen, und plötzlich sind Sie zu solchen Würdenträgern geworden! Hi-hi.	„Ich, Euer Exzellenz ... Sehr <u>angenehm-s</u>! Ein Freund, könnte man sagen, aus Kindertagen, und nun haben Sie sich zu solch einem Würdenträger <u>emporgeschwungen-s</u>! <u>Hi-hi-s.</u> “	angenehm-s; emporgeschwungen-s; hi-hi-s (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Das angehängte „-s“ ist im Deutschen nicht konventionalisiert und wirkt daher erklärungsbedürftig. Die Formen wirken künstlich und stören die	ja	

	zielsprachliche Flüssigkeit.		
Fall 8			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Ein Freund, wie man so sagt, aus Kindertagen und sind plötzlich, ehm, zu solchen Würden gelangt! Hi hi hi.«	Ein Freund, sozusagen aus Kindertagen, und plötzlich sind Sie <u>zu solchen Würdenträgern geworden!</u> Hi-hi.	Ein Freund, könnte man sagen, aus Kindertagen, und nun haben Sie sich zu solch einem Würdenträger emporgeschwungen-s! Hi-hi-s.“	zu solchen Würdenträgern geworden (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – logisches Problem	Die Pluralform „Würdenträgern“ passt nicht zur angesprochenen Einzelfigur. Dadurch entsteht ein lokales Bezugsproblem, auch wenn der soziale Aufstieg der Figur grundsätzlich verständlich bleibt.	teilweise	Der Plural passt hier nicht. Ich würde aber eher von einem semantischen/grammatischen Fehler sprechen.
Fall 9			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Das lange Kinn seiner Frau wurde noch länger; Nathanail nahm Haltung an und schloß sämtliche Knöpfe seiner Uniform ...	Das lange Kinn seiner Frau wurde noch länger; Nafanail <u>streckte sich zur Parade</u> und knöpfte alle Knöpfe seiner Uniform zu...	Das lange Kinn seiner Frau wurde noch länger; Nafanail nahm Haltung an und knöpfte alle Knöpfe seiner Uniform zu ...	streckte sich zur Parade (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „streckte sich zur Parade“ ist im Deutschen ungebräuchlich. Gemeint ist, dass die Figur Haltung annimmt bzw.	ja	Sehr unidiomatisch

	sich stramm hin- stellt.		
Fall 10			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Aber Gott behüte ... Ich bitte Sie ...« kicherte der Dünne und sackte noch mehr zusammen.	– Gnaden Sie mir... <u>Was sagen Sie da...</u> – kicherte der Dünne und zog sich noch mehr zusammen.	„Verzeihen Sie ... <u>Was reden Sie da-s</u> ...“, kicherte der Dünne und zog sich noch mehr zusam- men.	Was sagen Sie da ... (DeepL Pro) Was reden Sie da-s ... (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	Beide Formulierung- en können im Deut- schen als zurückwei- sender Einwand ver- standen werden und passen daher nicht zur unterwürfigen Haltung der Figur.	ja	Die unterwürfige Haltung wurde sprachlich nicht ideal getroffen. Außerdem wirkt das angehängte „s“ bei „da-s“ nicht idioma- tisch.
Fall 11			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Der Dicke wollte et- was entgegnen, im Gesicht des Dünnen aber spiegelte sich derart viel Ehrfurcht, Klebrigkeit und un- terwürfige Säuerlich- keit, daß dem Ge- heimrat übel wurde.	Der Dicke wollte ge- rade etwas erwidern, doch auf dem Ge- sicht des Dünnen stand so viel Ehr- furcht, <u>Liebenswür- digkeit</u> und unter- würfige Säure ge- schrieben, dass dem Geheimrat übel wurde.	Der Dicke wollte ge- rade etwas erwidern, doch im Gesicht des Dünnen stand so viel Ehrfurcht, <u>Süßlich- keit</u> und unterwür- fige Säure geschrie- ben, dass dem Ge- heimrat übel wurde.	Liebenswürdigkeit (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	„Liebenswürdigkeit“ ist positiv konnotiert und schwächt die ab- wertende Beschrei- bung der unterwürfi- gen Haltung ab.	ja	

**Anhang 3: Fehlerannotation in den Bereichen Kohärenz sowie Stil & Register zu
Ein Chamäleon (Tschechow 2003c; DeepL Pro 2026c; ChatGPT-5.3
2026c)**

Fall 1			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Über den Marktplatz geht der Polizeiinspektor Otschumelow in einem neuen Uniformmantel, ein Bündel in der Hand. [...] Otschumelow macht eine halbe Drehung nach links und geht auf die Menge zu.	Über den Marktplatz schreitet der Polizeiaufseher <u>Ochumelow</u> in einem neuen Mantel und mit einem Bündel in der Hand. [...] <u>Otschumelow</u> macht eine halbe Drehung nach links und schreitet auf die Ansammlung zu.	Über den Marktplatz geht der Polizeiaufseher Otschumelow in einem neuen Mantel und mit einem Bündel in der Hand. [...] Otschumelow dreht sich halb nach links und geht auf die Menge zu.	Ochumelow / Otschumelow (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Inkonsistenz	Derselbe Eigenname wird in der DeepL-Pro-Übersetzung uneinheitlich als „Ochumelow“ und „Otschumelow“ wiedergegeben. Die Bezugnahme auf die Figur bleibt nachvollziehbar, die lexikalische Wiederaufnahme erfolgt jedoch nicht konsistent.	ja	
Fall 2			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Du willst mich beißen, verdammte Töle? « hört Otschumelow plötzlich. »Laßt ihn nicht lau-	– Willst du also beißen, du <u>elendes Weib?</u> – hört Ochumelow plötzlich. – Jungs, lasst sie nicht durch! Heute darf	Willst du etwa beißen, du verfluchtes Vieh? « hört Otschumelow plötzlich. „Leute, lasst sie nicht	elendes Weib (DeepL Pro)

fen, Leute! Heutzutage darf man nicht beißen! Haltet ihn!	man nicht beißen! Halt sie fest!	durch! Heute ist Beißen verboten! Haltet sie!	
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – logisches Problem	Da die Szene an dieser Stelle erst eingeführt wird, deutet „du elendes Weib“ darauf hin, dass eine weibliche Person angesprochen wird. Erst im weiteren Verlauf wird deutlich, dass sich die Beschimpfung auf einen Hund beziehen soll.	ja	Eindeutig eine falsche Referenzannahme.
Fall 3			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Du willst mich beißen, verdammte Töle?« hört Otschumelow plötzlich. »Laßt ihn nicht laufen, Leute! Heutzutage darf man nicht beißen! Haltet ihn! Ah ... ah!« Man hört einen Hund winseln. Otschumelow blickt zur Seite und sieht: Aus dem Holzlager des Kaufmanns Pitschugin läuft eilig ein Hund, er hüpfte auf drei Pfoten und sieht sich immer wieder um. Ihm nach jagt ein Mann in einem gestärkten Kattunhemd und aufgeknöpfter Weste. Er läuft hinter ihm her, den Oberkörper vorgestreckt, fällt zu Boden und packt den	– Willst du also beißen, du elendes Weib? – hört Otschumelow plötzlich. – Jungs, lasst sie nicht durch! Heute darf man nicht beißen! Halt sie fest! Ah... ah! Man hört ein Hundegeheul. Otschumelow schaut zur Seite und sieht: Aus dem Holzschuppen des Kaufmanns Pitschugin rennt ein Hund, auf drei Beinen hüpfend und sich umschauend. Hinter ihr jagt ein Mann in einem gestärkten Baumwollhemd und einer aufgeknöpften Weste. Er rennt ihr hinterher, beugt sich nach vorne, fällt zu Boden und packt den	Willst du etwa beißen, du verfluchtes Vieh?« hört Otschumelow plötzlich. „Leute, lasst sie nicht durch! Heute ist Beißen verboten! Haltet sie! Ah... ah!“ Man hört das Jaulen eines Hundes. Otschumelow blickt zur Seite und sieht: aus dem Holzlager des Kaufmanns Pitschugin rennt ein Hund, auf drei Beinen hüpfend und sich umblickend. Hinter ihm jagt ein Mann in einem gestärkten Kattunhemd und einer aufgeknöpften Weste. Er läuft hinterher, beugt den Oberkörper vor, stürzt zu Boden und packt den Hund an den Hinterbeinen.	verfluchtes Vieh / sie / sie (ChatGPT-5.3)

Hund an den Hinterpfoten.	Hund an den Hinterpfoten.		
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Koreferenz	Das weibliche Pronomen „sie“ ist im Kontext nicht eindeutig nachvollziehbar. Nach dem sächlichen Substantiv „Vieh“ wäre eher „es“ zu erwarten. Bezogen auf „Hund“ wäre auch „ihn“ möglich. Durch den Wechsel zu „sie“ wird die Referenz auf den Hund bereits zu Beginn der Szene uneinheitlich geführt.	ja	
Fall 4			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Du willst mich beißen, verdammte Töle?« hört Otschumelow plötzlich. »Laßt ihn nicht laufen, Leute! Heutzutage darf man nicht beißen! Haltet ihn! Ah ... ah!« Man hört einen Hund winseln. Otschumelow blickt zur Seite und sieht: Aus dem Holzlager des Kaufmanns Pitschugin läuft eilig ein Hund, er hüpfte auf drei Pfoten und sieht sich immer wieder um. Ihm nach jagt ein Mann in einem gestärkten Kattunhemd und aufgeknöpfter Weste. Er läuft hinter ihm her,	– Willst du also beißen, du elendes Weib? – hört Otschumelow plötzlich. – Jungs, lasst sie nicht durch! Heute darf man nicht beißen! Halt sie fest! Ah... ah! Man hört ein Hundegeheul. Otschumelow schaut zur Seite und sieht: Aus dem Holzschuppen des Kaufmanns Pitschugin rennt ein Hund, auf drei Beinen hüpfend und sich umschauend. Hinter ihm jagt ein Mann in einem gestärkten Baumwollhemd und einer aufgeknöpften Weste. Er rennt ihm hinterher, beugt sich nach vorne, fällt zu Boden und packt den	Willst du etwa beißen, du verfluchtes Vieh?« hört Otschumelow plötzlich. „Leute, lasst sie nicht durch! Heute ist Beißen verboten! Haltet sie! Ah... ah!“ Man hört das Jaulen eines Hundes. Otschumelow blickt zur Seite und sieht: aus dem Holzlager des Kaufmanns Pitschugin rennt ein Hund, auf drei Beinen hüpfend und sich umblickend. Hinter ihm jagt ein Mann in einem gestärkten Kattunhemd und einer aufgeknöpften Weste. Er läuft hinterher, beugt den Oberkörper vor, stürzt zu Boden und	elendes Weib / sie / sie / ein Hund / ihr / ihr / den Hund (DeepL Pro)

den Oberkörper vorgestreckt, fällt zu Boden und packt den Hund an den Hinterpfoten.	<u>Hund</u> an den Hinterpfoten.	packt den Hund an den Hinterbeinen.	
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Koreferenz	DeepL Pro führt die Bezugskette auf den Hund in der Eingangsszene uneinheitlich. Der Hund wird zunächst als „elendes Weib“ angesprochen und anschließend mit „sie“ wiederaufgenommen. Nach der expliziten Einführung als „ein Hund“ wird die Referenz weiterhin mit „ihr“ fortgeführt. Dadurch wird dieselbe Bezugsgröße über mehrere Sätze hinweg nicht konsistent wiederaufgenommen.	ja	
Fall 5			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Du willst mich beißen, verdammte Töle?« hört Otschumelow plötzlich. » Laßt ihn nicht laufen , Leute! Heutzutage darf man nicht beißen! Haltet ihn!	– Willst du also beißen, du elendes Weib? – hört Ochumelow plötzlich. – Jungs, lasst sie nicht durch! Heute darf man nicht beißen! Halt sie fest!	Willst du etwa beißen, du verfluchtes Vieh?“ hört Otschumelow plötzlich. „Leute, lasst sie nicht durch! Heute ist Beißen verboten! Haltet sie!	lasst sie nicht durch (DeepL Pro, ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	„Durchlassen“ wirkt im Kontext der Verfolgungsszene unpassend. Gemeint ist, den Hund nicht entkommen zu lassen.	ja	

Fall 6			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Da gibt es Ärger, Euer Wohlgeboren « sagt der Schutzpolizist.	– Das ist ja ein ziemliches Durcheinander, <u>Eure Exzellenz!</u> – sagt der Stadtpolizist.	„Scheint eine Unordnung zu sein, Euer Wohlgeboren! ..“ sagt der Polizist.	Eure Exzellenz (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register + Kohärenz – Inkonsistenz	„Eure Exzellenz“ ist für den Rang des Polizeiinspektors registerbezogen überhöht. Durch den späteren Wechsel zu „Eure Ehren“ wird die Anredeform zudem inkonsistent. Der Rang des Polizeiinspektors wird dadurch sprachlich uneinheitlich eingeordnet.	ja	Die Einordnung über Rang und Register ist gut begründet.
Fall 7			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Ich gehe, Euer Wohlgeboren , tue niemand was ...«	– Ich gehe nur, <u>Eure Ehren</u> , ich störe niemanden...	„Ich ging meines Weges, Euer Wohlgeboren , habe niemanden angerührt...“	Eure Ehren (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	„Eure Ehren“ ist im Deutschen in diesem Kontext keine passende Anredeform für eine polizeiliche Amtsperson. Die Form wirkt ungebrauchlich und registerbezogen unpassend.	ja	Wirkt im Deutschen auch ungebrauchlich.

Fall 8			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Das ist, scheint's, der von General Shigalow!« ruft jemand aus der Menge. »Von General Shigalow? Hm!	„Das ist, glaube ich, der von General <u>Zhigalow!</u>“, sagt jemand aus der Menge. – General <u>Schigalows?</u> Hm!	Das ist wohl der vom General Schigalow!“ sagt jemand aus der Menge. „Vom General Schigalow? Hm!	Zhigalow / Schigalow (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Inkonsistenz	Derselbe Eigenname wird innerhalb kurzer Distanz unterschiedlich wiedergegeben. Dadurch wird die Namensreferenz im Zieltext uneinheitlich geführt, auch wenn grundsätzlich erkennbar bleibt, dass dieselbe Person gemeint ist.	ja	
Fall 9			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Ich mache eine feine Arbeit. Sie sollen mir das zahlen, weil ich den Finger vielleicht eine Woche nicht bewegen kann ...	Ich habe nur kleine Arbeiten. Man soll mir das Geld geben, <u>denn mit diesem Finger werde ich vielleicht eine Woche lang keinen Finger rühren können...</u>	Meine Arbeit ist fein. Man soll mir Schadenersatz zahlen, denn vielleicht kann ich diesen Finger eine Woche lang nicht bewegen...	denn mit diesem Finger werde ich vielleicht eine Woche lang keinen Finger rühren können... (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Der konkret verletzte Finger wird mit der festen Redewendung „keinen Finger rühren“ verbunden. Dadurch kollidieren die wörtliche und die	ja	

	übertragene Bedeutung von „Finger“, was den Satz unflüssig wirken lässt.		
Fall 10			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Das, Euer Wohlgeboren, gibt es nicht im Gesetz, daß man sich das von so einem Luder gefallen lassen muß ...	Das, Eure Ehren, steht nicht im Gesetz, dass man sich von einem Tier etwas gefallen lassen muss...	So etwas steht im Gesetz nicht, <u>dass man von einem Tier leiden soll...</u>	dass man von einem Tier leiden soll... (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „von einem Tier leiden soll“ ist im Deutschen ungebrauchlich. Gemeint ist, dass man sich von einem Tier etwas gefallen lassen muss bzw. durch ein Tier Schaden erleidet.	ja	
Fall 11			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Nun gut ... Wessen Hund ist das? Ich lasse das so nicht durchgehen. Ich werde euch zeigen, was es heißt, Hunde herumstreunen zu lassen!	–Gut... Wem gehört der Hund? <u>Das lasse ich nicht auf sich sitzen.</u> Ich werde euch zeigen, wie man Hunde laufen lässt!	„Gut... Wessen Hund ist das? Ich lasse das nicht so! Ich werde euch zeigen, wie man Hunde frei herumlaufen lässt!	Das lasse ich nicht auf sich sitzen (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „Das lasse ich nicht auf sich sitzen“ ist im Deutschen nicht geläufig und wirkt	ja	Ich stimme der Analyse der vermischten Redewendung zu.

	wie eine Vermischung der Wendungen „das lasse ich nicht auf mir sitzen“ und „das lasse ich nicht auf sich beruhen“. Die intendierte Bedeutung bleibt erkennbar, der Satz wirkt jedoch Zielsprachlich unnatürlich.		
Fall 12			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Du hast dir wohl den Finger an einem Nagel aufgerissen, und dann kam dir die Idee, dabei was herauszuschlagen.	Du hast dir wohl den Finger mit einem Nagel aufgerissen, und dann kam dir die Idee, <u>sie abzureißen.</u>	Du hast dir den Finger wohl an einem Nagel aufgerissen, und dann kam dir die Idee, etwas herauszuschlagen.	sie abzureißen (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – logisches Problem	Die Formulierung „sie abzureißen“ ist im gegebenen Kontext nicht eindeutig nachvollziehbar. Es bleibt unklar, worauf sich „sie“ bezieht und was genau abgerissen werden soll.	ja	Unklare Referenz
Fall 13			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Euer Wohlgeboren sind ein kluger Herr und verstehen, wer lügt und wer nach seinem Gewissen, wie vor Gott ...	<u>Eure Ehren sind kluge Herren</u> und verstehen, wer lügt und wer nach bestem Gewissen spricht, wie vor Gott...	Seine Hochwohlgeboren ist ein kluger Herr und versteht, wer lügt und wer ehrlich ist, wie vor Gott...	Eure Ehren sind kluge Herren (DeepL Pro)

Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „Eure Ehren sind kluge Herren“ wirkt im Deutschen als Anredekonstruktion unnatürlich, da eine einzelne Amtsperson mit einer Form im Plural angesprochen wird.	ja	Die Anredekonstruktion wirkt im Deutschen sehr ungewöhnlich/unnatürlich.
Fall 14			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Euer Wohlgeboren sind ein kluger Herr und verstehen, wer lügt und wer nach seinem Gewissen, wie vor Gott ...	Eure Ehren sind kluge Herren und verstehen, wer lügt und wer nach bestem Gewissen spricht, wie vor Gott...	<u>Seine Hochwohlgeboren ist ein kluger Herr</u> und versteht, wer lügt und wer ehrlich ist, wie vor Gott...	Seine Hochwohlgeboren ist ein kluger Herr (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register + Stil & Register – Unflüssigkeit	Obwohl eine Person direkt angesprochen wird, verwendet ChatGPT-5.3 die Form „Seine Hochwohlgeboren“. Diese Anrede wirkt im Deutschen ungebrauchlich und ist grammatisch nicht passend eingebunden.	ja	
Fall 15			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Würde man so einen Hund in Petersburg oder in Moskau antreffen, wißt ihr, was dann wäre? Dort würde man nicht erst ins Gesetz sehen,	Wenn so ein Hund in St. Petersburg oder Moskau auftauchen würde, weißt du, was dann passieren würde? Da würde man nicht erst im	Wenn so ein Hund in Petersburg oder Moskau auftauchen würde, wisst ihr, was geschähe? Da würde man nicht ins Gesetz	atme nicht mehr (DeepL Pro)

sondern augenblicklich – weg damit!	Gesetz nachschlagen, sondern sofort – <u>atme nicht mehr!</u>	schauen, sondern sofort – aus!	
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – logisches Problem	Die Formulierung „atme nicht mehr!“ wirkt im Deutschen wie ein direkter Befehl an das Tier. In der Szene geht es jedoch um eine Aufforderung zur Beseitigung des Hundes. Dadurch erscheint die Äußerung logisch unstimmig und sprachlich auffällig.	ja	
Fall 16			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Er ist vielleicht teuer, und wenn jeder Schweinehund ihm eine Zigarre in die Schnauze steckt, dann ist er bald verdorben.	Sie ist vielleicht teuer, aber wenn jeder Schwein ihr eine Zigarre in die Nase steckt, dauert es nicht lange, bis sie <u>ruiniert</u> ist.	Vielleicht ist er wertvoll, und wenn jeder Schwein ihm mit einer Zigarette in die Nase sticht, ist er schnell verdorben.	ruiniert (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	„Ruiniert“ ist in Bezug auf einen Hund im Deutschen ungewöhnlich und wirkt in diesem Kontext unpassend. Gemeint ist, dass der Hund durch solches Verhalten verdorben oder geschädigt wird.	teilweise	„Ruiniert“ klingt zwar ungewöhnlich, aber nicht falsch bzw. ich würde es nicht als schweren Fehler kategorisieren.

Fall 17			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Er ist vielleicht teuer, und wenn jeder Schweinehund ihm eine Zigarre in die Schnauze steckt , dann ist er bald verdorben.	Sie ist vielleicht teuer, aber wenn jeder Schwein ihr eine Zigarre in die Nase steckt , dauert es nicht lange, bis sie ruiniert ist.	Vielleicht ist er wertvoll, und wenn jeder Schwein ihm mit einer Zigarette in die Nase sticht , ist er schnell verdorben.	sticht (ChatGPT-5.3)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Unflüssigkeit	Die Formulierung „mit einer Zigarette in die Nase sticht“ ist im Deutschen kollokativ unpassend, da „stechen“ typischerweise mit einem spitzen Gegenstand verbunden wird. In Bezug auf eine Zigarette wäre eine Formulierung wie „in die Nase/Schnauze stecken“ oder „an die Nase halten“ passender.	ja	Die Kollokation passt nicht zu dieser Stelle und verursacht ein anderes Bild.
Fall 18			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Da brauchen wir nicht lange zu fragen«, sagt Otschumelow. »Das ist ein streunender Hund! Was reden wir da noch ... Wenn ich gesagt habe, daß er streunt, dann streunt er ... Beseitigen, und Schluß.« »Das ist nicht unsrer«, fährt Prochor fort. »Das ist der	– Da gibt's nichts lange zu fragen, – sagt Ochumelow. – Sie ist ein Streuner! Da gibt's nichts lange zu reden... Wenn er sagt, sie sei ein Streuner, dann ist sie eben ein Streuner... Erledigen, das ist alles. – Der gehört nicht uns, – fährt Prochor fort. – Der gehört	Dann gibt es nichts weiter zu fragen,“ sagt Otschumelow. „ Er ist ein Streuner! Wenn ich sage, er ist ein Streuner, dann ist er es auch... Töten, das ist alles.“ „ Er ist nicht von uns,“ fährt Prochor fort. „ Er gehört dem Bruder des Generals, der neulich angekommen ist.	Sie... der... (DeepL Pro)

vom Bruder des Generals, der neulich gekommen ist.	dem Bruder des Generals, der neulich angekommen ist.		
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Koreferenz	DeepL Pro referiert auf denselben Hund zunächst mit „sie“ und anschließend mit „der“. Dadurch wird die Bezugskette im Zieltext uneinheitlich geführt. Die Referenz auf den Hund bleibt zwar verständlich, ist aber lokal inkonsistent.	ja	
Fall 19			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
»Das ist nicht unser«, fährt Prochor fort. »Das ist der vom Bruder des Generals, der neulich gekommen ist. Unser General ist kein Liebhaber von Windhunden.	– Der gehört nicht uns, – fährt Prochor fort. – Der gehört dem Bruder des Generals, der neulich angekommen ist. Unserer <u>ist kein Fan</u> von Windhunden.	„Er ist nicht von uns,“ fährt Prochor fort. „Er gehört dem Bruder des Generals, der neulich angekommen ist. Unser Herr mag keine Windhunde ...	kein Fan (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Stil & Register – Register	„Kein Fan von Windhunden“ ist im Deutschen zwar verständlich, wirkt aber modern und umgangssprachlich. Die Formulierung passt daher registerbezogen nicht zur historischen Figurenrede.	ja	

Fall 20			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Unser General ist kein Liebhaber von Windhunden. Sein Bruder ist ...« »Ist denn sein Bruder gekommen? Wladimir Iwanysch?« fragt Otschumelow, und ein Lächeln der Rührung breitet sich über sein Gesicht. »Sieh mal einer an! Und ich hab es nicht gewußt! Ist er zu Besuch gekommen?« »Ja ... « »Sieh mal an ... Hatte wohl Sehnsucht nach seinem Brüderchen ... Und ich habe das gar nicht gewußt. Das ist also sein Hündchen?	Unserer ist kein Fan von Windhunden. Ihr Bruder ist es aber... – Ist ihr Bruder denn angekommen? Wladimir Iwanowitsch? – fragt Otschumelow, und sein ganzes Gesicht erhellt sich in einem gerührten Lächeln. – Ach du meine Güte! Und ich wusste nichts davon! Sind sie zu Besuch gekommen? – Zu Besuch... – Ach du meine Güte... Wir haben den Bruder vermisst... Und ich wusste es gar nicht! Ist das also ihr Hündchen?	Unser Herr mag keine Windhunde, aber sein Bruder schon...“ „Wie, der Bruder ist angekommen? Wladimir Iwanowitsch?“ fragt Otschumelow, und sein ganzes Gesicht erstrahlt vor Rührung. „Sieh mal einer an! Das wusste ich gar nicht! Zu Besuch gekommen?“ „Zu Besuch...“ „Ach, sieh mal einer an... Hat seinen Bruder vermisst... Und ich wusste nichts davon! Also ist das sein Hund?	ihr Bruder / ihr Bruder / ihr Hündchen / Sind sie zu Besuch gekommen? (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – Koreferenz	DeepL Pro führt die Referenz auf den Bruder des Generals uneinheitlich und missverständlich. Die Formen „ihr Bruder“, „Sind sie zu Besuch gekommen?“ und „ihr Hündchen“ machen nicht eindeutig nachvollziehbar, auf welche Person sich die Pronomen beziehen. Dadurch entsteht ein	ja	Sehr unverständliche und verwirrende Formulierung.

	lokales Bezugsproblem.		
Fall 21			
HÜ	DeepL Pro	ChatGPT-5.3	Problemstelle
Das ist also sein Hündchen? Freut mich sehr ... Nimm es mit ... Ein liebes Hündchen ... So flink ... Beißt den da in den Finger.	Ist das also ihr Hündchen? Ich freue mich sehr... Nimm sie... Ein ganz hübsches Hündchen... So flink... <u>Nimm das hier am Finger!</u>	Also ist das sein Hund? Sehr erfreut... Nimm ihn... Ein ganz netter Hund... so lebhaft... Hat ihm in den Finger geschnappt!	Nimm das hier am Finger! (DeepL Pro)
Fehlerkategorie	Kurzbegründung	Bewertung	Kommentar Kontrollperson
Kohärenz – logisches Problem	Die Formulierung ist im Deutschen kaum nachvollziehbar. Es bleibt unklar, was mit „das hier am Finger“ gemeint ist und wer was nehmen soll. Im Kontext geht es darum, dass der Hund jemanden in den Finger gebissen/geschnappt hat. DeepL Pro stellt diesen Zusammenhang nicht verständlich her.	ja	Die Formulierung ist nicht mehr verständlich.

Humanübersetzung: *Der Tod eines Beamten* (Tschechow 2003a)

Eines lieblichen Abends saß der nicht minder liebreizende Staatsdiener Iwan Dmitritsch Tscherwjakow* in der zweiten Reihe des Parketts und betrachtete »Die Glocken von Corneville« durch sein Opernglas.** Und wie er so schaute, schwebte er auf den Gipfeln der Glückseligkeit. Plötzlich aber ... Auf dieses »plötzlich aber« stößt man in Erzählungen immer wieder. Und die Autoren tun Recht daran, steckt das Leben doch voller Überraschungen! Plötzlich aber verzog sich sein Gesicht, die Augen quollen hervor, sein Atem stockte ... er ließ das Binokel sinken, krümmte sich und ... hatschi!!! Er nieste, wie man sieht. Zu niesen ist schließlich nirgends und niemandem verwehrt. Bauern niesen und Polizisten, mitunter sogar Geheimräte. Alle tun es. Tscherwjakow geriet denn auch keinesfalls in Verlegenheit, trocknete sich mit einem Tüchlein ab und blickte als höflicher Mensch, der er war, in die Runde, ob er durch sein Niesen nicht etwa jemandem zu nahe getreten sei. Nun aber begann er doch unruhig zu werden. Er sah nämlich, wie ein vor ihm in der ersten Reihe des Parketts sitzender alter Herr sorgfältig mit dem Handschuh über Glatze und Nacken fuhr und etwas vor sich hin murmelte. In jenem alten Herrn erkannte Tscherwjakow den Zivil-General Brisshalow vom Verkehrsministerium.

»Ich habe ihn besprüht!« ging es Tscherwjakow durch den Kopf. »Zwar ist er nicht mein Vorgesetzter, sondern aus einem anderen Ressort, aber unangenehm ist es trotzdem. Ich muß mich entschuldigen.«

Tscherwjakow räusperte sich, neigte den Oberkörper nach vorn und flüsterte dem General ins Ohr:

»Euer Exz'lenz verzeihen, ich habe Sie besprüht ... Es war keine Absicht ...«

»Schon gut, schon gut ...«

»Um Himmels Willen, verzeihen Sie. Ich ... ich wollte es nicht!«

»So geben Sie doch Ruhe! Ich kann ja nichts hören!«

Tscherwjakow verlor die Fassung, lächelte dümmlich und wandte seinen Blick wieder der Bühne zu. Er sah zwar hin, von Glückseligkeit aber konnte nun nicht mehr die Rede sein.

* Anmerkung der Übersetzerin aus Tschechow 2003a: Von tscherw, russ. Wurm.

** Anmerkung der Übersetzerin aus Tschechow 2003a: Operette von Robert Planquette (1877)

Unruhe hatte ihn gepackt. In der Pause trat er auf Brisschalow zu, lief neben ihm her und murmelte, nachdem er sich ein Herz gefaßt hatte:

»Ich habe Sie besprüht, Euer Exz'lenz ... Verzeihen Sie ... Es verhält sich folgendermaßen ... nicht, daß ich etwa ...«

»Lassen Sie es doch gut sein ... Ich hatte die Sache schon vergessen, Sie aber fangen immer wieder damit an!« sagte der General und zuckte ungeduldig mit der Unterlippe.

»Vergessen hat er's, der Groll aber ist ihm ins Gesicht geschrieben«, dachte Tscherwjakow bei sich, als er den General mißtrauisch betrachtete. »Und reden möchte er auch nicht. Ich muß ihm doch erklären, daß ich es keineswegs beabsichtigt habe ... daß es ein Naturgesetz ist, sonst denkt er noch, ich hätte ihn bespucken wollen. Und wenn er es jetzt nicht denkt, dann wird er es später denken!«

Zu Hause angelangt, erzählte Tscherwjakow seiner Frau von der Entgleisung. Diese nahm das Vorgefallene allzu leichtfertig auf, wie ihm schien; zunächst erschrak sie, beruhigte sich aber sogleich, nachdem sie erfahren hatte, daß Brisschalow einem anderen Amt angehörte.

»Geh aber trotzdem zu ihm und entschuldige dich«, sagte sie. »Sonst meint er noch, du kannst dich in der Öffentlichkeit nicht benehmen!«

»Das ist es ja eben! Ich habe mich entschuldigt, er aber benahm sich irgendwie merkwürdig. Nicht ein vernünftiges Wort hat er gesagt. Und es war auch gar keine Zeit zum Reden.«

Anderntags zog Tscherwjakow die neue Galauniform an, ließ sich das Haar schneiden und ging zu Brisschalow, die Angelegenheit aufzuklären ... Als er den Empfangsraum des Generals betrat, erblickte er dort zahlreiche Bittsteller und inmitten der Bittsteller auch den General selbst, der bereits die Gesuche entgegenzunehmen begann. Nachdem er einige der Bittsteller angehört hatte, richtete der General seinen Blick auch auf Tscherwjakow.

»Gestern in der Arkadija, falls Euer Exz'lenz sich zu erinnern geruhen«, begann der Beamte seinen Vortrag, »habe ich geniest und Sie ... versehentlich besprüht ... Entsch ...«

»Eine solche Lappalie aber auch ... Du lieber Himmel! Und was wünschen Sie?« wandte sich der General dem nächsten Bittsteller zu.

»Er möchte nicht mit mir sprechen«, dachte Tscherwjakow und erbleichte. »Folglich ist er böse auf mich ... Nein, das darf ich auf keinen Fall auf sich beruhen lassen ... Ich werde es ihm erklären ...«

Als der General die Unterredung mit dem letzten Bittsteller beendet hatte und sich in die inneren Gemächer begeben wollte, lief ihm Tscherwjakow hinterher und murmelte:

»Euer Exz'lenz! Wenn ich es wage, Euer Exz'lenz zu belästigen, dann einzig, ich versichere Sie, aus einem Gefühl der Reue! ... Es geschah nicht absichtlich, Sie belieben es doch zu wissen!«

Der General zog ein weinerliches Gesicht und winkte ab.

»Sie machen sich wohl über mich lustig, Wertester!« sagte er und verschwand hinter der Tür.

»Was heißt hier lustig machen«, dachte Tscherwjakow. »Von Lustigmachen kann überhaupt nicht die Rede sein! Ein General, begreift aber rein gar nichts! Wenn das so ist, werde ich diesen Wichtigtuer nicht weiter um Verzeihung bitten! Soll ihn doch der Teufel holen! Ich werde ihm einen Brief schreiben, hingehen aber werde ich nicht mehr! Bei Gott, das werde ich nicht tun!«

Dies waren Tscherwjakows Gedanken auf dem Heimweg. Einen Brief schrieb er dem General allerdings nicht. So viel er auch nachdachte, er brachte diesen Brief einfach nicht zustande. Es blieb ihm nichts übrig, als am folgenden Tag noch einmal persönlich vorzusprechen:

»Ich habe Euer Exz'lenz gestern nicht deshalb belästigt«, murmelte er, als der General seinen fragenden Blick zu ihm hob, »um mich lustig zu machen, wie Sie zu bemerken geruhten. Ich entschuldigte mich dafür, daß ich Sie beim Niesen besprühte ... Mich lustig zu machen hätte ich mir nie erlaubt. Darf ich es denn wagen, mich lustig zu machen? Würden wir uns lustig machen, wo, bitteschön, bliebe dann die Achtung vor Respektpersonen?«

»Hinaus!!« schrie plötzlich der General, der blau angelaufen war und bebte.

»Wie belieben?« fragte Tscherwjakow flüsternd und verging vor Entsetzen.

»Hinaus!!« wiederholte der General und stampfte auf.

In Tscherwjakows Innerem riß etwas entzwei. Ohne das geringste zu sehen oder zu hören, wick er zur Tür zurück, trat auf die Straße und schleppte sich mühsam vorwärts ... Zuhause angelangt, legte er sich, ohne die Galauniform abzulegen, auf den Diwan und ... verblich.

DeepL Pro-Übersetzung: *Der Tod eines Beamten* (DeepL Pro 2026a)

An einem schönen Abend saß der nicht minder schöne Henker Ivan Dmitrich Chervyakov in der zweiten Reihe und schaute durch sein Fernglas auf die „Glocken von Kornevil“. Er schaute und fühlte sich wie im siebten Himmel. Aber plötzlich ... In Erzählungen kommt dieses „aber plötzlich“ oft vor. Die Autoren haben Recht: Das Leben ist so voller Überraschungen! Aber plötzlich verzog sich sein Gesicht, seine Augen rollten, sein Atem stockte ... er nahm das Fernglas von den Augen, beugte sich vor und ... apshi!!! Er nieste, wie Sie sehen. Niemandem und nirgendwo ist das Niesen verboten. Männer, Polizeichefs und manchmal sogar Geheimrats-ecken niesen. Alle niesen. Chervyakov war überhaupt nicht verlegen, wischte sich mit einem Taschentuch ab und sah sich wie ein höflicher Mensch um: Hatte er jemanden mit seinem Niesen gestört? Aber jetzt musste er sich doch schämen. Er sah, dass der alte Mann, der vor ihm in der ersten Reihe saß, sich fleißig mit einem Handschuh die Glatze und den Hals abwischte und etwas vor sich hin murmelte. In dem alten Mann erkannte Chervyakov den Generalmajor Brizhalov, der im Verkehrsministerium tätig war.

„Ich habe ihn bespritzt!“, dachte Chervyakov. „Er ist zwar nicht mein Vorgesetzter, aber es ist trotzdem unangenehm. Ich muss mich entschuldigen.“

Chervyakov hustete, beugte sich vor und flüsterte dem General ins Ohr:

„Entschuldigen Sie, Eure Exzellenz, ich habe Sie bespritzt ... das war ein Versehen ...“

„Schon gut, schon gut ...“

„Um Gottes willen, entschuldigen Sie bitte. Ich wollte das nicht!“

„Ach, bleiben Sie bitte sitzen! Lassen Sie mich zuhören!“

Chervyakov war verlegen, lächelte albern und begann, auf die Bühne zu schauen. Er schaute, aber empfand keine Glückseligkeit mehr. Eine Unruhe begann ihn zu quälen. In der Pause ging er zu Brizjalow, stand neben ihm und murmelte, nachdem er seine Scheu überwunden hatte:

– Ich habe Sie bespritzt, Eure Exzellenz... Entschuldigen Sie... Ich wollte doch nicht...

– Ach, kommen Sie... Ich habe es schon vergessen, und Sie reden immer noch davon! – sagte der General und bewegte ungeduldig seine Unterlippe.

„Er hat es vergessen, aber in seinen Augen liegt Boshaftigkeit“, dachte Chervyakov und warf dem General einen misstrauischen Blick zu. „Und er will nicht darüber reden. Ich sollte ihm

erklären, dass ich das überhaupt nicht wollte ... dass es ein Naturgesetz ist, sonst denkt er noch, ich wollte ihn bespucken. Jetzt denkt er es nicht, aber später wird er es denken!“

Als er nach Hause kam, erzählte Chervyakov seiner Frau von seiner Unwissenheit. Seine Frau schien ihm gegenüber das Geschehene zu leichtfertig zu nehmen; sie erschrak nur, und als sie erfuhr, dass Brizjalow ein „Fremder“ war, beruhigte sie sich.

„Geh trotzdem hin und entschuldige dich“, sagte sie. „Er wird denken, dass du dich in der Öffentlichkeit nicht benehmen kannst!“

„Genau das ist es! Ich habe mich entschuldigt, aber er hat irgendwie seltsam reagiert ... Er hat kein einziges vernünftiges Wort gesagt. Und es war auch keine Zeit zum Reden.“

Am nächsten Tag zog Chervyakov einen neuen Frack an, ließ sich die Haare schneiden und ging zu Brizjalow, um ihm alles zu erklären... Als er das Empfangszimmer des Generals betrat, sah er dort viele Bittsteller und zwischen ihnen den General selbst, der bereits mit der Entgegennahme der Bitten begonnen hatte. Nachdem er einige Bittsteller befragt hatte, hob der General den Blick und sah Chervyakov an.

„Gestern im „Arkadien“, wenn Sie sich erinnern, Eure Exzellenz“, begann der Vollstrecker zu berichten, „habe ich geniest und ... versehentlich bespritzt ... Entschuldigung ...“

„Was für eine Kleinigkeit ... Gott weiß was! Was wünschen Sie?“, wandte sich der General an den nächsten Bittsteller.

„Er will nicht reden!“, dachte Chervyakov und wurde blass. „Er ist also wütend ... Nein, das kann man so nicht stehen lassen ... Ich werde es ihm erklären ...“

Als der General das Gespräch mit dem letzten Bittsteller beendet hatte und sich in seine Gemächer begab, trat Chervyakov hinter ihm her und murmelte:

„Eure Hoheit! Wenn ich es wage, Eure Hoheit zu stören, dann nur aus einem Gefühl der Reue, das kann ich Ihnen versichern! Es war keine Absicht, das wissen Sie selbst!“

Der General machte ein weinerliches Gesicht und winkte ab.

„Sie machen sich doch nur lustig, mein Herr!“, sagte er und versteckte sich hinter der Tür.

„Was soll das für ein Spott sein?“, dachte Chervyakov. „Das ist überhaupt kein Spott! Der General versteht das einfach nicht! Wenn das so ist, werde ich mich nicht mehr bei diesem Angeber entschuldigen! Zum Teufel mit ihm! Ich werde ihm einen Brief schreiben, aber ich werde nicht hingehen! Bei Gott, ich werde nicht hingehen!“

So dachte Chervyakov, als er nach Hause ging. Er schrieb dem General keinen Brief. Er dachte und dachte, aber ihm fiel einfach kein Brief ein. Am nächsten Tag musste er selbst hingehen, um sich zu erklären.

„Ich habe Sie gestern gestört“, murmelte er, als der General ihn fragend ansah, „nicht, um zu lachen, wie Sie gesagt haben. Ich habe mich dafür entschuldigt, dass ich beim Niesen gespritzt habe ... aber ich hatte nicht vor zu lachen. Wie könnte ich lachen? Wenn wir lachen würden, gäbe es keinen Respekt mehr vor Personen ...

„Verschwinden Sie!“, brüllte plötzlich der General, der blau angelaufen war und zitterte.

„Was?“, fragte Chervyakov flüsternd, vor Schreck ganz benommen.

„Hau ab!“ wiederholte der General und stampfte mit den Füßen.

In Chervyakovs Bauch löste sich etwas. Ohne etwas zu sehen oder zu hören, wich er zur Tür zurück, ging auf die Straße hinaus und schlurfte davon... Als er mechanisch nach Hause kam, legte er sich, ohne seinen Uniformrock auszuziehen, auf das Sofa und... starb.

ChatGPT-5.3-Übersetzung: *Der Tod eines Beamten* (ChatGPT-5.3 2026a)

An einem schönen Abend saß ein nicht minder schöner Exekutor namens Iwan Dmitrijewitsch Tscherwjakow in der zweiten Reihe der Sessel und blickte durch ein Opernglas auf die „Glocken von Corneville“. Er schaute und fühlte sich auf dem Gipfel der Glückseligkeit. Doch plötzlich ... In Erzählungen kommt dieses „doch plötzlich“ häufig vor. Die Autoren haben recht: Das Leben ist so voller Überraschungen! Doch plötzlich verzog sich sein Gesicht, die Augen rollten nach oben, der Atem stockte ... er nahm das Opernglas von den Augen, beugte sich vor und ... hatschi!!! Er nieste, wie Sie sehen. Niesen ist niemandem und nirgends verboten. Es niesen Bauern, Polizeimeister und manchmal sogar wirkliche Staatsräte. Alle niesen. Tscherwjakow war keineswegs verlegen, wischte sich mit dem Taschentuch ab und schaute sich als höflicher Mensch um: ob er wohl mit seinem Niesen jemanden beunruhigt habe? Doch da musste er verlegen werden. Er sah, dass ein alter Herr, der vor ihm in der ersten Reihe saß, sich eifrig mit dem Handschuh die Glatze und den Nacken abwischte und etwas vor sich hinhin murmelte. In dem alten Herrn erkannte Tscherwjakow den Zivilgeneral Brisžalow aus dem Verkehrsministerium.

„Ich habe ihn angespritzt!“, dachte Tscherwjakow. „Nicht mein Vorgesetzter, ein Fremder, aber trotzdem peinlich. Ich muss mich entschuldigen.“

Tscherwjakow räusperte sich, beugte den Oberkörper vor und flüsterte dem General ins Ohr:

- Verzeihen Sie, Euer Exzellenz, ich habe Sie angespritzt ... ich habe es nicht absichtlich ...
- Schon gut, schon gut ...
- Um Gottes willen, verzeihen Sie. Ich habe ja ... ich wollte doch nicht!
- Ach, sitzen Sie doch! Lassen Sie mich zuhören!

Tscherwjakow war verlegen, lächelte dumm und begann wieder auf die Bühne zu schauen. Er schaute, doch die Glückseligkeit war verschwunden. Unruhe begann ihn zu quälen. In der Pause ging er zu Brisžalow, lief ein wenig neben ihm her und murmelte, nachdem er seine Scheu überwunden

hatte:

- Ich habe Sie angespritzt, Euer Exzellenz ... Verzeihen Sie ... ich habe ja ... nicht etwa ...
- Ach, lassen Sie doch ... Ich habe es schon vergessen, und Sie kommen immer wieder darauf zurück! – sagte der General und verzog ungeduldig die Unterlippe.

„Vergessen – und dabei dieser spöttische Blick“, dachte Tscherwjakow und sah den General misstrauisch an. „Er will nicht reden. Ich müsste ihm erklären, dass ich wirklich nichts wollte

... dass das ein Naturgesetz ist, sonst denkt er noch, ich wollte ihn ansucken. Jetzt denkt er es vielleicht nicht, aber später denkt er es bestimmt!“

Zu Hause erzählte Tscherwjakow seiner Frau von seinem Missgeschick. Die Frau nahm die Sache, wie es ihm schien, allzu leicht; sie erschrak nur kurz, beruhigte sich aber, als sie erfuhr, dass Brisžalow „fremd“ sei.

– Aber geh trotzdem hin und entschuldige dich, – sagte sie. – Er könnte denken, du weißt dich in der Öffentlichkeit nicht zu benehmen!

– Genau das ist es ja! Ich habe mich entschuldigt, aber er war so merkwürdig ... Kein vernünftiges Wort hat er gesagt. Und Zeit zum Reden war auch nicht.

Am nächsten Tag zog Tscherwjakow seine neue Uniform an, ließ sich die Haare schneiden und ging zu Brisžalow, um alles zu erklären ... Als er das Vorzimmer des Generals betrat, sah er dort viele Bittsteller und unter ihnen auch den General selbst, der bereits mit der Entgegennahme der Gesuche begonnen hatte. Nachdem er mehrere Bittsteller angehört hatte, hob der General den Blick und sah Tscherwjakow an.

– Gestern in der „Arkadia“, wenn Sie sich erinnern, Euer Exzellenz, – begann der Exekutor zu berichten, – habe ich geniest und ... Sie versehentlich angespritzt ... Ich wollte mich ent...

– Was für Bagatellen ... Gott weiß was! Was wünschen Sie? – wandte sich der General an den nächsten Bittsteller.

„Er will nicht reden!“, dachte Tscherwjakow und wurde bleich. „Er ist also böse ... Nein, so kann man das nicht lassen ... Ich werde es ihm erklären ...“

Als der General das Gespräch mit dem letzten Bittsteller beendet hatte und sich in die inneren Gemächer begeben wollte, trat Tscherwjakow hinter ihm her und murmelte:

– Euer Exzellenz! Wenn ich mir erlaube, Sie zu behelligen, so einzig aus einem Gefühl, ich darf sagen, der Reue! Nicht absichtlich, Sie wissen es selbst!

Der General machte ein weinerliches Gesicht und winkte ab.
– Sie machen sich ja einfach lustig, mein Herr! – sagte er und verschwand hinter der Tür.

„Was für ein Spott soll das sein?“, dachte Tscherwjakow. „Hier ist doch gar kein Spott! Ein General – und versteht das nicht! Wenn das so ist, werde ich mich vor diesem Aufschneider nicht mehr entschuldigen! Zum Teufel mit ihm! Ich schreibe ihm einen Brief, aber hingehen werde ich nicht mehr! Bei Gott, ich werde nicht!“

So dachte Tscherwjakow auf dem Heimweg. Einen Brief schrieb er dem General nicht. Er dachte und dachte und konnte sich diesen Brief einfach nicht ausdenken. Also musste er am nächsten Tag selbst hingehen, um alles zu erklären.

– Ich bin gestern gekommen, um Euer Exzellenz zu behelligen, – murmelte er, als der General ihn fragend ansah, – nicht um zu spotten, wie Sie zu sagen geruhen. Ich entschuldigte mich dafür, dass ich beim Niesen gespritzt habe ... und an Spott habe ich gar nicht gedacht. Wie dürfte ich mir erlauben zu spotten? Wenn wir spotten, dann gäbe es ja überhaupt keinen Respekt mehr vor Personen ...

– Raus!! – brüllte plötzlich der blau angelaufene und zitternde General.

– Wie bitte? – fragte Tscherwjakow flüsternd, vor Entsetzen wie gelähmt.

– Raus!! – wiederholte der General und stampfte mit den Füßen.

In Tscherwjakows Bauch riss etwas. Ohne etwas zu sehen oder zu hören, taumelte er zur Tür, ging auf die Straße hinaus und schleppte sich davon ... Zu Hause angekommen, legte er sich mechanisch, ohne die Uniform auszuziehen, auf das Sofa und ... starb.

Humanübersetzung: *Der Dicke und der Dünne* (Tschechow 2023b)

Auf einem Bahnhof der Nikolajewer Eisenbahnlinie trafen sich zwei Bekannte: der eine dick, der andere dünn. Der Dicke hatte gerade im Bahnrestaurants zu Mittag gespeist, und seine fettbenetzten Lippen glänzten wie reife Kirschen. Er roch nach Sherry und Fleur d'orange. Der Dünne dagegen war soeben aus dem Zug gestiegen, bepackt mit Koffern, Bündeln und Schachteln. Er roch nach Schinken und Kaffeesatz. Hinter seinem Rücken lugten eine magere Frau mit langem Kinn hervor, seine Gattin, und ein hochaufgeschossener Gymnasiast, der ein Auge zukniff, sein Sohn.

»Porfiri!« rief der Dicke, als er den Dünnen erblickte. »Bist du's? Mein Bester! Wieviel Zeit ins Land gegangen ist!«

»Du meine Güte!« reagierte der Dünne verblüfft. »Mischa! Freund meiner. Kindejahre! Wo kommst du denn her?«

Die Freunde küßten sich drei Mal und betrachteten einander mit Tränen in den Augen. Beide waren angenehm überrascht.

»Mein Lieber!« begann der Dünne, nachdem sie sich abgeküßt hatten. »Damit habe ich nicht gerechnet! Eine Überraschung aber auch! Na, schau mich doch mal an! Noch immer der gleiche Adonis wie früher! Der gleiche fescche Kerl und Dandy! Ach, du lieber Himmel! Erzähl mal! Bist du reich? Verheiratet? Ich bin bereits verheiratet, wie du siehst ... Dies hier ist meine Frau Luise, geborene Wanzenbach ... lutheranisch ... Und dies mein Sohn Nathanail, Quartaner. Dies, Nathanja, ist der Freund meiner Kindheit! Wir sind zusammen aufs Gymnasium gegangen!«

Nathanail überlegte einen Augenblick und zog dann die Mütze.

»Wir sind zusammen aufs Gymnasium gegangen!« fuhr der Dünne fort.

»Weißt du noch, wie du geneckt wurdest? Dich nannten sie Herostratos, weil du mit der Zigarette ein Loch ins Klassenbuch gebrannt hast, und mich Ephialtes, weil ich gern petzte. Ha ha ... Wir waren Kinder! Hab keine Angst, Nathanja! Geh hin zu ihm ... Und dies ist meine Frau, geborene Wanzenbach ... lutheranisch.«

Nathanail überlegte einen Moment und versteckte sich dann hinter dem Rücken seines Vaters.

»Na, wie geht's dir so, mein Freund?« fragte der Dicke und schaute seinen Schulfreund entzückt an. »Bist du Beamter? Schon vorangekommen?«

»Ja, mein Lieber, ich bin Beamter! Schon das zweite Jahr Kollegienassessor und hab den Stanislaw-Orden. Das Gehalt ist dürftig ... na ja, was soll's! Meine Frau gibt Musikstunden, und ich mache privat Zigarrenetuis aus Holz. Vorzügliche Zigarrenetuis! Ich verkaufe sie zu einem Rubel das Stück. Nimmt mir jemand zehn ab oder mehr, gebe ich natürlich Rabatt, klare Sache. Irgendwie kommen wir zurecht. Zuerst hab ich ja in der Bezirksverwaltung gedient, jetzt aber bin ich hierher versetzt worden, als Ressort-Vorsteher im selben Amt ... Werd nun hier dienen. Und was ist mit dir? Bist wohl schon Staatsrat, was?«

»Nein, mein Lieber, geh ruhig noch höher«, sagte der Dicke. »Hab mich schon bis zum Geheimrat emporgedient ... Hab auch zwei Ordenssterne.«

Der Dünne erbleichte plötzlich, versteinerte, doch kurz darauf verzog sich sein Gesicht zu einem breiten Lächeln; es schien, als sprühten sein Gesicht und seine Augen Funken. Er selbst sackte zusammen, krümmte sich und schrumpfte gleichsam ein ... Auch seine Koffer, Bündel und Schachteln sackten zusammen und überzogen sich mit Runzeln ... Das lange Kinn seiner Frau wurde noch länger; Nathanail nahm Haltung an und schloß sämtliche Knöpfe seiner Uniform ...

»Euer Exzellenz, ich ... Ehm, sehr angenehm! Ein Freund, wie man so sagt, aus Kindertagen und sind plötzlich, ehm, zu solchen Würden gelangt! Hi hi hi.«

»Nun mach aber mal nen Punkt!« verzog der Dicke das Gesicht. »Was soll dieser Tonfall? Wir sind doch Jugendfreunde, und was soll überhaupt dieses Getue mit den Rängen!«

»Aber Gott behüte ... Ich bitte Sie ...« kicherte der Dünne und sackte noch mehr zusammen. »Die gnädige Aufmerksamkeit Eurer Exzellenz ... ist gleichsam belebender Tau ... Dies hier, Euer Exzellenz, ist mein Sohn Nathanail und dies meine Gattin Luise, lutheranisch, gewissermaßen ...«

Der Dicke wollte etwas entgegnen, im Gesicht des Dünnen aber spiegelte sich derart viel Ehrfurcht, Klebrigkeit und unterwürfige Säuerlichkeit, daß dem Geheimrat übel wurde. Er wandte sich vom Dünnen ab und reichte ihm zum Abschied die Hand.

Der Dünne drückte ihm drei Finger, verneigte sich tief und kicherte wie ein Chinese: »hi hi hi«. Seine Frau lächelte. Nathanail scharrte mit dem Fuß und ließ die Mütze fallen. Alle drei waren angenehm überrascht.

DeepL-Pro-Übersetzung: *Der Dicke und der Dünne* (DeepL Pro 2026b)

Am Bahnhof der Nikolajew-Eisenbahn trafen sich zwei Freunde: der eine dick, der andere dünn. Der Dicke hatte gerade am Bahnhof zu Mittag gegessen, und seine mit Öl überzogenen Lippen glänzten wie reife Kirschen. Er roch nach Sherry und Fleur d'Oranger. Der Dünne hingegen war gerade aus dem Waggon gestiegen und mit Koffern, Bündeln und Kartons beladen. Er roch nach Schinken und Kaffeesatz. Hinter seinem Rücken lugten eine schlanke Frau mit langem Kinn – seine Frau – und ein großer Gymnasiast mit zusammengekniffenen Augen – sein Sohn – hervor.

– Porfirij! – rief der Dicke aus, als er den Dünnen sah. – Bist du das? Mein Lieber! Wie viele Winter, wie viele Jahre!

– Meine Güte! – staunte der Dünne. – Mischa! Mein Freund aus Kindertagen! Wo kommst du denn her?

Die Freunde umarmten sich dreimal und sahen sich mit tränengefüllten Augen an. Beide waren angenehm überwältigt.

– Mein Lieber! – begann der Dünne nach dem Kuss. – Das hätte ich nicht erwartet! Was für eine Überraschung! Nun, sieh mich doch mal genau an! Ein genauso hübscher Kerl wie früher! Ein genauso liebenswerter und eleganter Mann! Ach du Gott! Na, wie geht es dir? Bist du reich? Verheiratet? Ich bin bereits verheiratet, wie du siehst... Das ist meine Frau, Louise, geborene Wanzenbach... eine Lutheranerin... Und das ist mein Sohn, Nathanael, Schüler der dritten Klasse. Das ist, Nathanael, mein Freund aus Kindertagen! Wir haben zusammen am Gymnasium gelernt!

Nathanael dachte kurz nach und nahm seine Mütze ab.

– Wir sind zusammen auf das Gymnasium gegangen! – fuhr der Dünne fort.

– Erinnerst du dich, wie man dich gehänselt hat? Man hat dich Herostratus genannt, weil du das staatliche Buch mit einer Zigarette durchgebrannt hast, und mich Ephialtes, weil ich so gern petzen habe. Ha-ha... Wir waren Kinder! Hab keine Angst, Nafanja! Komm näher zu ihm... Und das ist meine Frau, geborene Wanzenbach... eine Lutheranerin.

Nafanail dachte kurz nach und versteckte sich hinter dem Rücken seines Vaters.

– Na, wie geht's dir, mein Freund? – fragte der Dicke und blickte seinen Freund begeistert an. – Wo dienst du? Hast du es weit gebracht?

– Ich diene, mein Lieber! Ich bin schon im zweiten Jahr Kollegialassessor und habe eine Stanislava. Das Gehalt ist schlecht... na ja, was soll's! Meine Frau gibt Musikunterricht, ich fertige privat Zigarettenschachteln aus Holz an. Ausgezeichnete Zigarettenschachteln! Ich verkaufe sie für einen Rubel pro Stück. Wenn jemand zehn Stück oder mehr nimmt, gibt es, verstehst du, einen Rabatt. Wir schlagen uns so durch. Ich habe, weißt du, im Ministerium gedient, und jetzt bin ich als Verpflegungsleiter derselben Behörde hierher versetzt worden... Hier werde ich dienen. Na, und wie geht es dir? Du bist sicher schon Beamter? Was?

– Nein, mein Lieber, schau mal genauer hin, – sagte der Dicke. – Ich habe es schon bis zum Geheimrat gebracht... Ich habe zwei Sterne.

Der Dünne wurde plötzlich blass, erstarrte, doch bald verzog sich sein Gesicht zu einem breitesten Lächeln; es schien, als sprühten Funken aus seinem Gesicht und seinen Augen. Er selbst schrumpfte zusammen, krümmte sich, wurde schmaler... Seine Koffer, Bündel und Kartons schrumpften zusammen, verzogen sich... Das lange Kinn seiner Frau wurde noch länger; Nafanail streckte sich zur Parade und knöpfte alle Knöpfe seiner Uniform zu...

– Ich, Eure Exzellenz... Sehr erfreulich! Ein Freund, sozusagen aus Kindertagen, und plötzlich sind Sie zu solchen Würdenträgern geworden! Hi-hi.

– Ach, hör doch auf! – verzog der Dicke das Gesicht. – Was soll dieser Ton? Wir sind Freunde aus Kindertagen – wozu also diese Ehrerbietung!

– Gnaden Sie mir... Was sagen Sie da... – kicherte der Dünne und zog sich noch mehr zusammen. – Die gnädige Aufmerksamkeit Eurer Exzellenz... ist so etwas wie lebensspendender Nektar... Das hier, Eure Exzellenz, ist mein Sohn Nafanail... meine Frau Louise, eine Lutherin, gewissermaßen...

Der Dicke wollte gerade etwas erwidern, doch auf dem Gesicht des Dünnen stand so viel Ehrfurcht, Liebenswürdigkeit und unterwürfige Säure geschrieben, dass dem Geheimrat übel wurde. Er wandte sich vom Dünnen ab und reichte ihm zum Abschied die Hand.

Der Dünne schüttelte drei Finger, verbeugte sich mit dem ganzen Oberkörper und kicherte wie ein Chinese: „Hi-hi-hi“. Seine Frau lächelte. Nafanail scharrte mit dem Fuß und ließ seine Mütze fallen. Alle drei waren angenehm verblüfft.

ChatGPT-5.3-Übersetzung: *Der Dicke und der Dünne* (ChatGPT-5.3 2026b)

Am Bahnhof der Nikolai-Eisenbahn trafen sich zwei Freunde: der eine dick, der andere dünn. Der Dicke hatte gerade im Bahnhof zu Mittag gegessen, und seine von Öl glänzenden Lippen schimmerten wie reife Kirschen. Er roch nach Sherry und Orangenblüten. Der Dünne hingegen war eben erst aus dem Waggon gestiegen und mit Koffern, Bündeln und Hutschachteln bepackt. Von ihm roch es nach Schinken und Kaffeesatz. Hinter seinem Rücken lugte eine magere Frau mit langem Kinn hervor – seine Ehefrau – und ein hochgewachsener Gymnasiast mit zusammengekniffenem Auge – sein Sohn.

„Porfirij!“, rief der Dicke aus, als er den Dünnen erblickte. „Bist du es wirklich? Mein Lieber! Wie viele Winter, wie viele Jahre!“

„Ach du meine Güte!“, staunte der Dünne. „Mischa! Freund meiner Kindheit! Wo kommst du denn her?“

Die Freunde küssten sich dreimal und blickten einander mit tränenvollen Augen an. Beide waren angenehm überwältigt.

„Mein Lieber!“, begann der Dünne nach den Küssen. „Damit hätte ich nicht gerechnet! Was für eine Überraschung! Nun, sieh mich nur richtig an! Noch immer derselbe Schönling wie früher! Dieselbe kleine Seele und derselbe Geck! Ach, du lieber Gott! Na, wie geht’s dir? Reich? Verheiratet? Ich bin schon verheiratet, wie du siehst ... Das hier ist meine Frau, Luise, geborene Wanzenbach ... evangelisch-lutherisch ... Und das ist mein Sohn, Nafanail, Schüler der dritten Klasse. Das hier, Nafanja, ist ein Freund meiner Kindheit! Wir haben zusammen im Gymnasium gelernt!“

Nafanail überlegte einen Augenblick und nahm die Mütze ab.

„Wir haben zusammen im Gymnasium gelernt!“, fuhr der Dünne fort.

„Weißt du noch, wie man dich gehänselt hat? Man nannte dich Herostratos, weil du ein staatliches Büchlein mit einer Zigarette durchgebrannt hattest, und mich Ephialtes, weil ich so gern petzte. Ho-ho ... Wir waren Kinder! Hab keine Angst, Nafanja! Komm näher zu ihm ... Und das ist meine Frau, geborene Wanzenbach ... evangelisch-lutherisch.“

Nafanail überlegte einen Augenblick und versteckte sich hinter dem Rücken seines Vaters.

„Nun, wie lebst du, Freund?“, fragte der Dicke und sah seinen Freund begeistert an. „Wo dienst du? Hast du es zu etwas gebracht?“

„Ich diene, mein Lieber! Seit dem zweiten Jahr bin ich Kollegienassessor und habe den Stanislausorden. Das Gehalt ist schlecht ... nun ja, Gott sei's drum! Meine Frau gibt Musikunterricht, und ich fertige nebenbei aus Holz Zigarettenetuis an. Ausgezeichnete Etuis! Ich verkaufe sie für einen Rubel das Stück. Wer zehn oder mehr nimmt, bekommt, verstehst du, Rabatt. Wir schlagen uns so einigermaßen durch. Früher war ich im Departement, und jetzt bin ich hierher versetzt worden, als Tischchef in derselben Behörde ... Hier werde ich also dienen. Nun, und du? Bist du wohl schon Staatsrat?“

„Nein, mein Lieber, greif höher“, sagte der Dicke. „Ich habe es bereits bis zum Geheimrat gebracht ... Ich trage zwei Sterne.“

Der Dünne erbleichte plötzlich, erstarrte, doch bald verzog sich sein Gesicht zu einem breiten Lächeln; es schien, als sprühten Funken aus seinem Gesicht und seinen Augen. Er selbst schrumpfte zusammen, krümmte sich, wurde schmaler ... Seine Koffer, Bündel und Hutschachteln schrumpften ebenfalls zusammen und verzogen sich ... Das lange Kinn seiner Frau wurde noch länger; Nafanail nahm Haltung an und knöpfte alle Knöpfe seiner Uniform zu ...

„Ich, Euer Exzellenz ... Sehr angenehm-s! Ein Freund, könnte man sagen, aus Kindertagen, und nun haben Sie sich zu solch einem Würdenträger emporgeschwungen-s! Hi-hi-s.“

„Nun hör aber auf!“, verzog der Dicke das Gesicht. „Wozu dieser Ton? Wir sind Freunde aus Kindertagen – und was soll hier plötzlich diese Rangverehrung!“

„Verzeihen Sie ... Was reden Sie da-s ...“, kicherte der Dünne und zog sich noch mehr zusammen. „Die huldvolle Aufmerksamkeit Eurer Exzellenz ... ist gleichsam wie belebendes Wasser ... Das hier, Euer Exzellenz, ist mein Sohn Nafanail ... meine Frau Luise, evangelisch-lutherisch, gewissermaßen ...“

Der Dicke wollte gerade etwas erwidern, doch im Gesicht des Dünnen stand so viel Ehrfurcht, Süßlichkeit und unterwürfige Säure geschrieben, dass dem Geheimrat übel wurde. Er wandte sich vom Dünnen ab und reichte ihm zum Abschied die Hand.

Der Dünne drückte drei Finger, verbeugte sich mit dem ganzen Oberkörper und kicherte wie ein Chinese: „Hi-hi-hi.“ Die Frau lächelte. Nafanail scharrte mit dem Fuß und ließ seine Mütze fallen. Alle drei waren angenehm überwältigt.

Humanübersetzung: *Ein Chamäleon* (Tschechow 2023c)

Über den Marktplatz geht der Polizeiinspektor Otschumelow in einem neuen Uniformmantel, ein Bündel in der Hand. Hinter ihm trabt ein Schutzpolizist mit einem Sieb, das bis zum Rand mit konfiszierten Stachelbeeren gefüllt ist. Rundherum Stille ... Auf dem Platz keine Menschenseele ... Die geöffneten Türen der Läden und Schenken blicken traurig in Gottes schöne Welt wie hungrige Mäuler; nicht einmal Bettler gibt es in ihrer Nähe.

»Du willst mich beißen, verdammte Töle?« hört Otschumelow plötzlich. »Laßt ihn nicht laufen, Leute! Heutzutage darf man nicht beißen! Haltet ihn! Ah ... ah!«

Man hört einen Hund winseln. Otschumelow blickt zur Seite und sieht: Aus dem Holzlager des Kaufmanns Pitschugin läuft eilig ein Hund, er hüpfte auf drei Pfoten und sieht sich immer wieder um. Ihm nach jagt ein Mann in einem gestärkten Kattunhemd und aufgeknöpfter Weste. Er läuft hinter ihm her, den Oberkörper vorgestreckt, fällt zu Boden und packt den Hund an den Hinterpfoten. Wieder hört man den Hund winseln und den Schrei: »Laßt ihn nicht laufen!« Aus den Läden beugen sich verschlafene Gesichter, und schnell hat sich neben dem Holzlager, wie aus der Erde gewachsen, eine Menschenmenge versammelt.

»Da gibt es Ärger, Euer Wohlgeboren« sagt der Schutzpolizist.

Otschumelow macht eine halbe Drehung nach links und geht auf die Menge zu. Neben dem Lagertor sieht er den oben beschriebenen Mann mit der aufgeknöpften Weste, er hat die rechte Hand hochgestreckt und zeigt der Menge seinen blutigen Finger. Auf seinem Säufergesicht scheint geschrieben zu stehen: »Na warte, ich krieg dich schon, du Schurke!« und der Finger selbst sieht aus wie ein Siegesbanner. In dieser Menschenmenge erkennt Otschumelow den Goldschmiedemeister Chrjukin. Mitten in der Menschenmenge sitzt mit gespreizten Vorderbeinen und am ganzen Körper zitternd der Stein des Anstoßes, ein junger weißer Windhund mit spitzer Schnauze und einem gelben Fleck auf dem Rücken. In seinen tränenden Augen liegt ein Ausdruck von Traurigkeit und Angst.

»Was ist hier los?« fragt Otschumelow, sich durch die Menge drängend. »Warum seid ihr hier? Was hältst du den Finger hoch? Wer hat geschrien?«

»Ich gehe, Euer Wohlgeboren, tue niemand was ...« beginnt Chrjukin und hustet in die Faust, »wegen Holz zu Mitri Mitritsch, und plötzlich beißt mich dieses Biest so mir nichts, dir nichts in den Finger ... Entschuldigen Sie, ich bin ein Mensch, der arbeitet ... Ich mache eine feine Arbeit. Sie sollen mir das zahlen, weil ich den Finger vielleicht eine Woche nicht bewegen

kann ... Das, Euer Wohlgeboren, gibt es nicht im Gesetz, daß man sich das von so einem Luder gefallen lassen muß ... Wenn jeder beißen würde, dann wär's besser, nicht auf dieser Welt zu leben ...«

»Hm! ... Gut ...« sagt Otschumelow streng, hüstelt und zieht die Brauen hoch. »Nun gut ... Wessen Hund ist das? Ich lasse das so nicht durchgehen. Ich werde euch zeigen, was es heißt, Hunde herumstreunen zu lassen! Es wird Zeit, solchen Herrschaften Aufmerksamkeit zu schenken, die sich nicht an die Verordnungen halten wollen! Wenn dieser Schurke bestraft wird, dann wird er von mir hören, was das heißt, Hunde und anderes streunendes Vieh ... Ich werde es ihm schon zeigen ... Jeldyrin«, wendet sich der Inspektor an den Schutzmann, »krieg heraus, wem der Hund gehört und mach ein Protokoll! Und der Hund muß getötet werden. Unverzüglich! Wahrscheinlich ist er tollwütig ... Wem gehört der Hund, frage ich?«

»Das ist, scheint's, der von General Shigalow!« ruft jemand aus der Menge.

»Von General Shigalow? Hm! ... Nimm mir den Mantel ab, Jeldyrin ... Schrecklich, wie heiß es ist! Anzunehmen, daß Regen kommt ... Eins nur verstehe ich nicht: Wie konnte er dich beißen?« wendet sich Otschumelow an Chrjukin. »Wie kommt er denn an deinen Finger? Er ist klein, und du bist doch so ein kräftiger Kerl! Du hast dir wohl den Finger an einem Nagel aufgerissen, und dann kam dir die Idee, dabei was herauszuschlagen. Du bist doch ... einer von denen! Ich kenne euch, ihr Teufel!«

»Er hat ihm, Euer Wohlgeboren, eine Zigarre in die Schnauze gesteckt, so zum Spaß, und der, nicht faul, hat zugeschnappt ... Ein zänkischer Mensch, Euer Wohlgeboren!«

»Du lügst, du Einäugiger! Du hast es nicht gesehen, warum lügst du also? Euer Wohlgeboren sind ein kluger Herr und verstehen, wer lügt und wer nach seinem Gewissen, wie vor Gott ... Und wenn ich lüge, dann soll der Friedensrichter entscheiden. Bei ihm steht es im Gesetz ... Heutzutage sind alle gleich ... Mein Bruder ist selbst bei den Gendarmen ... wenn Ihr's wissen wollt ...«

»Schluß mit dem Geschwätz!«

»Nein, das ist nicht der vom General ...« bemerkt der Schutzpolizist sinnend. »Der General hat nicht solche. Er hat mehr Vorstehhunde ...«

»Weißt du das sicher?«

»Sicher, Euer Wohlgeboren ...«

»Ich weiß das auch. Der General hat teure Hunde, solche von Rasse, aber weiß der Teufel, was der hier für einer ist! Kein Fell, nicht anzusehen ... Ein einziges Greuel ... Solch einen Hund zu halten?! Wo habt ihr euren Verstand? Würde man so einen Hund in Petersburg oder in Moskau antreffen, wißt ihr, was dann wäre? Dort würde man nicht erst ins Gesetz sehen, sondern augenblicklich – weg damit! Du, Chirjuhin, hast gelitten, laß die Sache so nicht durchgehen! Es soll allen eine Lehre sein! Es wird Zeit ...«

»Aber vielleicht ist das doch der vom General ...« denkt der Schutzpolizist laut. »Man kann es ihm nicht von der Schnauze ablesen ... Neulich habe ich so einen bei ihm auf dem Hof gesehen.«

»Es stimmt, es ist der vorn General!« ertönt eine Stimme aus der Menge.

»Hm ...Hilf mir mal, lieber Jeldyrin, in den Mantel ...Es ist irgendwie windig geworden ... Mich fröstelt ... Du bringst ihn zum General und fragst dort. Sagst, ich hätte ihn gefunden und schicke ihn ... Und sag, man solle ihn nicht auf die Straße lassen ... Er ist vielleicht teuer, und wenn jeder Schweinehund ihm eine Zigarre in die Schnauze steckt, dann ist er bald verdorben. Ein Hund ist eine empfindsame Kreatur ... Und du, Trottel, nimm die Hand herunter. Kein Grund, deinen dummen Finger hochzuhalten! Bist selber schuld!«

»Da kommt der Koch des Generals, fragen wir den ... He, Prochor! Komm mal her, mein Lieber! Guck dir den Hund an ... Ist das Eurer?«

»Was fällt dir ein! Solche haben wir nie im Leben gehabt!«

»Da brauchen wir nicht lange zu fragen«, sagt Otschumelow. »Das ist ein streunender Hund! Was reden wir da noch ... Wenn ich gesagt habe, daß er streunt, dann streunt er ... Beseitigen, und Schluß.«

»Das ist nicht unsrer«, fährt Prochor fort. »Das ist der vom Bruder des Generals, der neulich gekommen ist. Unser General ist kein Liebhaber von Windhunden. Sein Bruder ist ...«

»Ist denn sein Bruder gekommen? Wladimir Iwanytsch?« fragt Otschumelow, und ein Lächeln der Rührung breitet sich über sein Gesicht. »Sieh mal einer an! Und ich hab es nicht gewußt! Ist er zu Besuch gekommen?«

»Ja ... «

»Sieh mal an ... Hatte wohl Sehnsucht nach seinem Brüderchen ... Und ich habe das gar nicht gewußt. Das ist also sein Hündchen? Freut mich sehr ... Nimm es mit ... Ein liebes

Hündchen ... So flink ... Beißt den da in den Finger. Ha ha ha ... Na, was zitterst du so? Rrr ... Rrr ... Ist böse, der Schelm ... So ein niedliches Kerlchen ...«

Prochor ruft den Hund und entfernt sich mit ihm von dem Holzlager ... Die Menge lacht über Chrjukin.

»Du kriegst auch noch deinen Teil!« droht ihm Otschumelow, schlägt den Mantel um sich und setzt seinen Weg über den Marktplatz fort.

DeepL-Pro-Übersetzung: *Ein Chamäleon* (DeepL Pro 2026c)

Über den Marktplatz schreitet der Polizeiaufseher Ochumelow in einem neuen Mantel und mit einem Bündel in der Hand. Hinter ihm schreitet ein rothaariger Stadtpolizist mit einem Korb, der bis zum Rand mit beschlagnahmten Stachelbeeren gefüllt ist. Rundherum herrscht Stille... Auf dem Platz ist keine Menschenseele zu sehen... Die offenen Türen der Läden und Kneipen starren trostlos ins Tageslicht wie hungrige Mäuler; nicht einmal Bettler sind in ihrer Nähe.

– Willst du also beißen, du elendes Weib? – hört Ochumelow plötzlich. – Jungs, lasst sie nicht durch! Heute darf man nicht beißen! Halt sie fest! Ah... ah!

Man hört ein Hundegeheul. Ochumelow schaut zur Seite und sieht: Aus dem Holzschuppen des Kaufmanns Pichugin rennt ein Hund, auf drei Beinen hüpfend und sich umschauend. Hinter ihr jagt ein Mann in einem gestärkten Baumwollhemd und einer aufgeknöpften Weste. Er rennt ihr hinterher, beugt sich nach vorne, fällt zu Boden und packt den Hund an den Hinterpfoten. Erneut ist ein Hundegeheul und der Ruf zu hören: „Nicht loslassen!“ Aus den Läden strecken sich verschlafene Gesichter hervor, und bald versammelt sich in der Nähe des Holzlagers, als wäre sie aus dem Boden gewachsen, eine Menschenmenge.

– Das ist ja ein ziemliches Durcheinander, Eure Exzellenz! – sagt der Stadtpolizist.

Otschumelow macht eine halbe Drehung nach links und schreitet auf die Ansammlung zu. Direkt am Tor des Lagerhauses, so sieht er, steht der oben beschriebene Mann in der aufgeknöpften Weste und zeigt der Menge mit erhobener rechter Hand seinen blutigen Finger. Auf seinem halbtrunkenen Gesicht steht förmlich geschrieben: „Den reiße ich dir schon ab, du Schurke!“ – und selbst der Finger sieht aus wie ein Siegeszeichen. In diesem Mann erkennt Ochumelow den Goldschmied Chrjukin. In der Mitte der Menge sitzt der eigentliche Auslöser des Skandals auf dem Boden, die Vorderbeine gespreizt und am ganzen Körper zitternd – ein weißer Windhundwelpen mit spitzer Schnauze und einem gelben Fleck auf dem Rücken. In seinen tränenden Augen liegt ein Ausdruck von Sehnsucht und Entsetzen.

– Was führt dich hierher? – fragt Ochumelow, als er sich durch die Menge drängt. – Warum bist du hier? Warum hast du den Finger... Wer hat geschrien?

– Ich gehe nur, Eure Ehren, ich störe niemanden... – beginnt Chrjukin und hustet in seine Faust.

– Es ging um das Brennholz mit Mitri Mitritsch – und plötzlich hat diese Schurkin mir aus heiterem Himmel in den Finger gebissen... Entschuldigen Sie bitte, ich bin ein arbeitender Mann... Ich habe nur kleine Arbeiten. Man soll mir das Geld geben, denn mit diesem Finger

werde ich vielleicht eine Woche lang keinen Finger rühren können... Das, Eure Ehren, steht nicht im Gesetz, dass man sich von einem Tier etwas gefallen lassen muss... Wenn jeder beißt, dann ist es besser, gar nicht mehr auf der Welt zu leben...

– Hm!.. Gut... – sagt Ochumelow streng, hustet und zieht die Augenbrauen hoch. – Gut... Wem gehört der Hund? Das lasse ich nicht auf sich sitzen. Ich werde euch zeigen, wie man Hunde laufen lässt! Es ist an der Zeit, solchen Herren, die sich den Verordnungen nicht unterwerfen wollen, Beachtung zu schenken! Sobald man diesen Schurken mit einer Geldstrafe belegt, wird er von mir erfahren, was ein Hund und anderes streunendes Vieh bedeutet! Ich werde ihm die Hölle zeigen! „Eldyrin“, wendet sich der Aufseher an den Stadtpolizisten, „finde heraus, wem dieser Hund gehört, und erstelle ein Protokoll! Und der Hund muss getötet werden. Sofort! Er ist wahrscheinlich tollwütig... Wem gehört dieser Hund, frage ich?“

„Das ist, glaube ich, der von General Zhigalov!“, sagt jemand aus der Menge.

– General Schigalows? Hm! ... Hilf mir mal, Eldyrin, den Mantel auszuziehen ... Schrecklich, wie heiß es ist! Es wird wohl bald regnen ... Nur eines verstehe ich nicht: Wie konnte sie dich beißen? – wendet sich Ochumelow an Chrukin. – Wie soll sie denn an deinen Finger kommen? Sie ist doch klein, und du bist ja ein ganzer Hüne! Du hast dir wohl den Finger mit einem Nagel aufgerissen, und dann kam dir die Idee, sie abzureißen. Ihr seid doch ... ein bekanntes Völkchen! Ich kenne euch, ihr Teufel!

– Er, Eure Ehren, hat ihr zum Spaß eine Zigarette ins Gesicht gesteckt, und sie – nicht dumm sein und zubeißen... Ein verrückter Mensch, Eure Ehren!

– Du lügst, du Schurke! Du hast es nicht gesehen, also warum lügst du? Eure Ehren sind kluge Herren und verstehen, wer lügt und wer nach bestem Gewissen spricht, wie vor Gott... Und wenn ich lüge, soll der Friedensrichter darüber entscheiden. So steht es im Gesetz... Heutzutage sind alle gleich... Ich habe selbst einen Bruder bei der Gendarmerie... falls Sie es wissen wollen...

– Keine Argumente!

– Nein, das ist kein Generalhund... – bemerkt der Stadtpolizist nachdenklich. – Der General hat keine solchen. Bei ihm sind es eher Jagdhunde...

– Weißt du das ganz sicher?

– Ganz sicher, Eure Exzellenz...

– Das weiß ich selbst. Der General hat teure, reinrassige Hunde, und dieser hier – der Teufel weiß, was das ist! Kein Fell, kein Aussehen... nur Niedertracht... Und so einen Hund zu halten?! Wo ist denn dein Verstand? Wenn so ein Hund in St. Petersburg oder Moskau auftauchen würde, weißt du, was dann passieren würde? Da würde man nicht erst im Gesetz nachschlagen, sondern sofort – atme nicht mehr! Du, Chrljukin, bist geschädigt worden, und lass die Sache nicht so auf sich beruhen... Man muss ihm eine Lektion erteilen! Es ist Zeit...

– Ach, vielleicht doch vom General... – denkt der Stadtpolizist laut. – Das steht ihr nicht ins Gesicht geschrieben... Neulich habe ich so eine in seinem Hof gesehen.

– Natürlich, die des Generals! – sagt eine Stimme aus der Menge.

– Hm!.. Zieh mir mal den Mantel über, Bruder Eldyrin... Es weht ein Wind... Mir ist kalt... Du bringst sie zum General und fragst dort nach. Sag, dass ich sie gefunden und geschickt habe... Und sag, dass man sie nicht auf die Straße lässt... Sie ist vielleicht teuer, aber wenn jeder Schwein ihr eine Zigarre in die Nase steckt, dauert es nicht lange, bis sie ruiniert ist. Ein Hund ist ein zartes Wesen... Und du, Dummkopf, nimm die Hand runter! Du musst nicht deinen blöden Finger hinhalten! Du bist selbst schuld!..

– Der General-Koch kommt, wir fragen ihn... Hey, Prochor! Komm mal her, mein Lieber! Schau dir den Hund an... Gehört er dir?

– Quatsch! So einen hatten wir hier noch nie!

– Da gibt's nichts lange zu fragen, – sagt Ochumelow. – Sie ist ein Streuner! Da gibt's nichts lange zu reden... Wenn er sagt, sie sei ein Streuner, dann ist sie eben ein Streuner... Erledigen, das ist alles.

– Der gehört nicht uns, – fährt Prochor fort. – Der gehört dem Bruder des Generals, der neulich angekommen ist. Unserer ist kein Fan von Windhunden. Ihr Bruder ist es aber...

– Ist ihr Bruder denn angekommen? Wladimir Iwanowitsch? – fragt Ochumelow, und sein ganzes Gesicht erhellt sich in einem gerührten Lächeln. – Ach du meine Güte! Und ich wusste nichts davon! Sind sie zu Besuch gekommen?

– Zu Besuch...

– Ach du meine Güte... Wir haben den Bruder vermisst... Und ich wusste es gar nicht! Ist das also ihr Hündchen? Ich freue mich sehr... Nimm sie... Ein ganz hübsches Hündchen... So flink... Nimm das hier am Finger! Ha-ha-ha... Na, warum zitterst du? Rrr... Rr... Der ist sauer, der Schlingel... so ein kleiner Kerl...

Prochor ruft den Hund und geht mit ihm vom Holzschuppen weg... Die Menge lacht über Chrujkin.

– Ich werde dich noch kriegen! – droht ihm Ochumelow und setzt, in seinen Mantel gehüllt, seinen Weg über den Marktplatz fort.

ChatGPT-5.3-Übersetzung: *Ein Chamäleon* (ChatGPT-5.3 2026c)

Über den Marktplatz geht der Polizeiaufseher Otschumelow in einem neuen Mantel und mit einem Bündel in der Hand. Hinter ihm schreitet ein rothaariger Polizist mit einem Sieb, randvoll mit beschlagnahmten Stachelbeeren. Ringsum herrscht Stille... Auf dem Platz ist keine Menschenseele... Die offenen Türen der Läden und Schenken blicken trüb ins Tageslicht wie hungrige Mäuler; nicht einmal Bettler sind in ihrer Nähe.

„Willst du etwa beißen, du verfluchtes Vieh?“ hört Otschumelow plötzlich. „Leute, lasst sie nicht durch! Heute ist Beißen verboten! Haltet sie! Ah... ah!“

Man hört das Jaulen eines Hundes. Otschumelow blickt zur Seite und sieht: aus dem Holzlager des Kaufmanns Pitschugin rennt ein Hund, auf drei Beinen hüpfend und sich umblickend. Hinter ihm jagt ein Mann in einem gestärkten Kattunhemd und einer aufgeknöpften Weste. Er läuft hinterher, beugt den Oberkörper vor, stürzt zu Boden und packt den Hund an den Hinterbeinen. Wieder ertönt ein Hundeschrei und ein Ruf: „Lasst sie nicht durch!“ Aus den Läden stecken verschlafene Gesichter hervor, und bald sammelt sich um das Holzlager, wie aus dem Boden gewachsen, eine Menschenmenge.

„Scheint eine Unordnung zu sein, Euer Wohlgeboren!..“ sagt der Polizist.

Otschumelow dreht sich halb nach links und geht auf die Menge zu. Direkt am Tor des Lagers sieht er den beschriebenen Mann in der aufgeknöpften Weste stehen; er hebt die rechte Hand und zeigt der Menge einen blutigen Finger. Auf seinem halb betrunkenen Gesicht scheint geschrieben zu stehen: „Warte nur, ich werde dich schon dran kriegen, du Gauner!“ – und der Finger selbst sieht wie ein Siegeszeichen aus. In diesem Mann erkennt Otschumelow den Goldschmied Chryukin. Im Zentrum der Menge sitzt, die Vorderbeine gespreizt und am ganzen Körper zitternd, der eigentliche Verursacher des Skandals – ein weißer Windhundwelp mit spitzer Schnauze und einem gelben Fleck auf dem Rücken. In seinen tränenden Augen liegt Ausdruck von Kummer und Angst.

„Was geht hier vor?“ fragt Otschumelow und drängt sich in die Menge. „Warum seid ihr hier? Wozu der Finger? Wer hat geschrien?“

„Ich ging meines Weges, Euer Wohlgeboren, habe niemanden angerührt...“ beginnt Chryukin und hustet in die Faust. „Wegen des Holzes mit Mitrij Mitritsch... und plötzlich beißt mich dieses Biest ohne jeden Grund in den Finger... Verzeihen Sie, ich bin ein arbeitender Mensch... Meine Arbeit ist fein. Man soll mir Schadenersatz zahlen, denn vielleicht kann ich diesen Finger

eine Woche lang nicht bewegen... So etwas steht im Gesetz nicht, dass man von einem Tier leiden soll... Wenn jeder beißen darf, dann lohnt es sich nicht mehr zu leben...“

„Hm!.. Gut...“ sagt Otschumelow streng, hustet und zieht die Augenbrauen hoch. „Gut... Wessen Hund ist das? Ich lasse das nicht so! Ich werde euch zeigen, wie man Hunde frei herumlaufen lässt! Es ist an der Zeit, solchen Herren Aufmerksamkeit zu schenken, die sich nicht an Vorschriften halten! Wenn man ihn bestraft, den Schuft, wird er schon erfahren, was ein Hund und anderes herumstreunendes Vieh bedeutet! Ich werde ihm schon zeigen, was Sache ist!.. Jeldyrin,“ wendet sich der Aufseher an den Polizisten, „finde heraus, wem dieser Hund gehört, und fertige ein Protokoll an! Den Hund muss man töten. Sofort! Er ist bestimmt tollwütig... Wessen Hund ist das, frage ich?“

„Das ist wohl der vom General Schigalow!“ sagt jemand aus der Menge.

„Vom General Schigalow? Hm!.. Jeldyrin, zieh mir den Mantel aus... Es ist furchtbar heiß! Wahrscheinlich kommt Regen... Eines verstehe ich nur nicht: Wie konnte er dich beißen?“ wendet sich Otschumelow an Chryukin. „Kann er überhaupt an deinen Finger heran? Er ist doch klein, und du bist so ein kräftiger Kerl! Du hast dir den Finger wohl an einem Nagel aufgerissen, und dann kam dir die Idee, etwas herauszuschlagen. Ich kenne euch doch, ihr Teufel!“

„Er hat ihm, Euer Wohlgeboren, zum Spaß eine Zigarette ins Maul gesteckt, und der Hund hat – nicht dumm – zugebissen... Ein streitsüchtiger Mensch, Euer Wohlgeboren!“

„Du lügst, du Schielender! Wenn du es nicht gesehen hast, warum lügst du dann? Seine Hochwohlgeboren ist ein kluger Herr und versteht, wer lügt und wer ehrlich ist, wie vor Gott... Und wenn ich lüge, soll der Friedensrichter entscheiden... Im Gesetz steht... Jetzt sind alle gleich... Ich habe selbst einen Bruder bei der Gendarmerie...“

„Nicht diskutieren!“

„Nein, das ist kein Hund vom General...“ bemerkt der Polizist nachdenklich. „Der General hat solche nicht. Er hat meist Jagdhunde...“

„Bist du dir sicher?“

„Ganz sicher, Euer Wohlgeboren...“

„Ich weiß es auch. Beim General sind die Hunde teuer, reinrassig, und dieser hier – weiß der Teufel was! Kein Fell, kein Aussehen... nichts als ein elendes Tier... Und so einen Hund hält man? Wo habt ihr euren Verstand? Wenn so ein Hund in Petersburg oder Moskau auftauchen

würde, wisst ihr, was geschähe? Da würde man nicht ins Gesetz schauen, sondern sofort – aus! Du, Chryukin, bist geschädigt, und lass das nicht auf sich beruhen... Man muss ihn bestrafen! Es ist Zeit...”

„Vielleicht ist er doch vom General...” sagt der Polizist laut. „Man sieht es ihm nicht an... Neulich habe ich im Hof des Generals so einen gesehen.“

„Natürlich vom General!“ ruft eine Stimme aus der Menge.

„Hm!.. Zieh mir den Mantel wieder an, Bruder Jeldyrin... Es weht etwas Wind... Mir ist kalt... Bring ihn zum General und frage dort nach. Sag, ich habe ihn gefunden und geschickt... Und sag, sie sollen ihn nicht auf die Straße lassen... Vielleicht ist er wertvoll, und wenn jeder Schwein ihm mit einer Zigarette in die Nase sticht, ist er schnell verdorben. Ein Hund ist ein empfindliches Wesen... Und du, Dummkopf, nimm die Hand herunter! Zeig deinen blöden Finger nicht herum! Du bist selbst schuld!..“

„Der Koch des Generals kommt, den fragen wir... He, Prochor! Komm mal her! Schau dir den Hund an... Gehört er euch?“

„Unsinn! So einen haben wir nie gehabt!“

„Dann gibt es nichts weiter zu fragen,“ sagt Otschumelow. „Er ist ein Streuner! Wenn ich sage, er ist ein Streuner, dann ist er es auch... Töten, das ist alles.“

„Er ist nicht von uns,“ fährt Prochor fort. „Er gehört dem Bruder des Generals, der neulich angekommen ist. Unser Herr mag keine Windhunde, aber sein Bruder schon...“

„Wie, der Bruder ist angekommen? Wladimir Iwanowitsch?“ fragt Otschumelow, und sein ganzes Gesicht erstrahlt vor Rührung. „Sieh mal einer an! Das wusste ich gar nicht! Zu Besuch gekommen?“

„Zu Besuch...”

„Ach, sieh mal einer an... Hat seinen Bruder vermisst... Und ich wusste nichts davon! Also ist das sein Hund? Sehr erfreut... Nimm ihn... Ein ganz netter Hund... so lebhaft... Hat ihm in den Finger geschnappt! Ha-ha-ha... Nun, warum zitterst du? Rrr... Rr... Er ist böse, der Schlingel... so ein kleines Hündchen...”

Prochor ruft den Hund und geht mit ihm vom Holzlager weg... Die Menge lacht über Chryukin.

„Ich kriege dich noch!“ droht Otschumelow ihm und setzt, sich in den Mantel hüllend, seinen Weg über den Marktplatz fort.

Abstract

Die vorliegende Arbeit untersucht literarische maschinelle Übersetzung und geht der Forschungsfrage nach, wie sich LLM-gestützte Übersetzungen ausgewählter Erzählungen Anton Tschechows von Humanübersetzungen aus dem Russischen ins Deutsche unterscheiden. Das Untersuchungskorpus bilden drei kurze Erzählungen: *Der Tod eines Beamten*, *Der Dicke und der Dünne* sowie *Ein Chamäleon*. Berücksichtigt werden die russischen Ausgangstexte, jeweils eine Humanübersetzung ins Deutsche sowie je eine mit DeepL Pro und ChatGPT-5.3 Instant erstellte Übersetzungsvariante.

Ausgehend von der Forschungsfrage werden vier Annahmen formuliert. Diese beziehen sich auf lexikalische Vielfalt, lokale Kohäsion, Kohärenz sowie Stil & Register. Es wird angenommen, dass LLM-gestützte Übersetzungen im Vergleich zu Humanübersetzungen eine geringere lexikalische Vielfalt und eine geringere lokale Kohäsion aufweisen. Darüber hinaus wird davon ausgegangen, dass sie weniger kohärent sind und Schwächen bei stilistischen Nuancen und Registerwahl aufweisen. Die methodische Grundlage bildet ein kombinierter Analyseansatz, der eine erweiterte SCATE-Fehlertaxonomie nach Tezcan et al. (2019) mit einer für die vorliegende Arbeit adaptierten Key-Feature-Analyse nach Webster et al. (2020) verbindet.

Die Analyseergebnisse fallen je nach Untersuchungsbereich unterschiedlich aus. Die Annahme einer geringeren lexikalischen Vielfalt von LLM-gestützten Übersetzungen lässt sich nicht vollständig bestätigen. Die Annahme bezüglich zur lokalen Kohäsion wird im Rahmen des untersuchten Korpus grundsätzlich gestützt, ihre Reichweite bleibt jedoch begrenzt. In den Bereichen Kohärenz sowie Stil & Register zeigen die LLM-gestützten Übersetzungen hingegen Auffälligkeiten im Vergleich zu den Humanübersetzungen.