

Disputation: „Zwischen *Plain Language* und Einfacher Sprache: eine Korpusanalyse von Zusammenfassungen der Ergebnisse klinischer Studien für Laien im Englischen, Deutschen und Italienischen“ *Giulia Pedrini, Università di Trieste, Italien*¹

1 Einleitung

Die Verwendung Leichter und Einfacher Sprache in der Gesundheitskommunikation gewinnt zunehmend an Bedeutung und wird von zahlreichen Forscher:innen untersucht (z. B. Maaß/Rink 2017 und 2018, Ahrens et al. 2022, Pedrini 2022, Ross/Magris 2022). Die hier vorgestellte Dissertation (Pedrini 2024) stellt eine mehrsprachige korpuslinguistische Analyse medizinischer Texte dar, die nach der entsprechenden Gesetzgebung in Einfacher Sprache verfasst werden sollten, und zwar die Zusammenfassungen der Ergebnisse klinischer Studien für Laien. Während andere ähnliche Textsorten der fachexternen Kommunikation, wie Beipackzettel und Einwilligungserklärungen, schon breit erforscht wurden, sind diese Dokumente kaum untersucht worden, da sie eine sehr innovative Textsorte darstellen, die in einer europäischen Verordnung (536/2014) verankert ist. Da sie auf Englisch verfasst werden, aber mindestens in alle Sprachen der Mitgliedstaaten, die an der klinischen Studie teilgenommen haben, übersetzt werden müssen, haben sie eine neue Nische im Übersetzungsmarkt geschaffen.

2 Theoretischer Hintergrund

Die vorliegende Dissertation folgt einem interdisziplinären Ansatz. Sie positioniert sich in der Gesundheitskommunikation, also im Bereich der Fachsprachen, sowie in der kontrastiven Linguistik und Übersetzungswissenschaft. Der Hauptfokus der Arbeit liegt aber auf der Barrierefreiheit.

Wie unter 1 angekündigt, gewinnen Leichte und Einfache Sprache im medizinischen Bereich zunehmend an Brisanz. Dies ist darauf zurückzuführen, dass es um ein hochspezialisiertes Gebiet geht, mit dem Laien ständig konfrontiert werden. Sprachvereinfachung kann so dazu beitragen, Inhalte verständlich zu machen, die ihnen ansonsten unzugänglich wären. Leichte und Einfache Sprache (die in verschiedenen Ländern unterschiedlich benannt werden, wie z.B. *easy-to-read* oder *easy-read* und *plain language* im englischsprachigen Raum sowie *linguaggio facile* oder *lingua facile* und *plain language* – seltener *lingua semplice* – im italienischsprachigen Raum) lassen sich in einem Kontinuum einordnen; an einem Ende des Kontinuums findet man die Fachsprachen als Ausdruck der schwierigsten Varietäten, am anderen Leichte Sprache; Einfache Sprache lässt sich irgendwo zwischen den zwei Polen dieses Kontinuums verorten und spielt eine Schlüsselrolle in der fachexternen Kommunikation (Hansen-Schirra/Maaß 2020: 18).

Die im Rahmen der Dissertation analysierte Textsorte, die von der obengenannten europäischen Verordnung zur Gewährleistung von Transparenz vorgesehen ist, stellt ein treffendes Beispiel für die Verwendung von Einfacher Sprache im medizinischen Bereich dar. Da Versionen in unterschiedlichen Sprachen notwendig sind, stellt diese Textsorte auch vor andere Herausforderungen, nämlich wie man interlinguale Übersetzungen von intralingual angepassten Texten anfertigen kann. Hierbei geht es um eine wenig erforschte Nische, die aber dank solcher Texte sowie der immer mehr zunehmenden Verbreitung von vereinfachten Sprachvarietäten in

¹ Die Dissertation wurde im Rahmen des Doktoratsstudiums „Studi linguistici e letterari“ der Università degli Studi di Udine und der Università degli Studi di Trieste verfasst.

zweisprachigen Gebieten wie Südtirol an Bedeutung gewinnen wird. Somit sind solche Analysen nicht nur für die Praxis, sondern auch für die Übersetzungswissenschaft von Interesse.

3 Forschungsfragen, Korpus und Methodik

Ziel der Analyse war, vier Hauptforschungsfragen zu beantworten. Erstens sollte herausgefunden werden, wo auf dem Kontinuum der vereinfachten Sprachvarietäten sich die Zusammenfassungen der Ergebnisse klinischer Studien für Laien befinden, um damit auch festzustellen, ob sie überhaupt als mögliches Beispiel von Einfacher Sprache berücksichtigt werden können. Zweitens sollten die Hauptmerkmale der Zusammenfassungen der Ergebnisse klinischer Studien für Laien untersucht werden, sowohl in den Ausgangstexten als auch in den Zieltexten. In diesem Zusammenhang wurde der Frage nachgegangen, ob die Zieltexte weiter vereinfacht oder im Vergleich zu den entsprechenden Ausgangstexten verkompliziert werden. Drittens wurden die existierenden Richtlinien (Recommendations of the Expert Group on Clinical Trials 2018, Good Lay Summary Practice 2021) berücksichtigt; hier lautete die Forschungsfrage, ob die Texte die Regeln der Einfachen Sprache beachten. Viertens sollte festgestellt werden, inwiefern die existierenden Richtlinien im Lichte der vorliegenden Analyseergebnisse verbessert werden könnten.

Um diese Forschungsfragen zu beantworten, wurde ein dreisprachiges Parallelkorpus mit englischen Ausgangstexten und den entsprechenden deutschen und italienischen Zieltexten erstellt. Die Dokumente wurden aus den Portalen der Pharmakonzerne Bayer, Novartis, GSK und Roche heruntergeladen, da zur Zeit der Textsammlung und Korpuserstellung das Informationssystem für klinische Prüfungen, d. h. das europäische Portal, auf dem zukünftig die Dokumente zu den Studien hochgeladen und zur Verfügung gestellt werden sollen, noch nicht verfügbar war. Das Informationssystem wurde erst im Jahr 2022 eingeführt und enthält bislang noch keine Zusammenfassungen. Insgesamt bestand das Korpus aus 180 von 2016 bis 2022 verfassten Dokumenten, die insgesamt 449.689 Tokens umfassen. Im Vergleich zu Allgemeinsprachekorpora ging es um ein relativ kleines Korpus, aber da es ein Fachsprachenkorpus mit einer einzigen Textsorte war, kann man in Anlehnung an die Fachliteratur (z. B. Bowker/Pearson 2002: 45–48, Cheng 2012: 32) davon ausgehen, dass diese Größe ausreicht, um generelle Tendenzen zu beobachten. Die gesammelten Texte wurden aligniert, um die korrekte Untersuchung der Übersetzungen zu ermöglichen, und mittels Sketch Engine² mit Part-of-Speech-Tags annotiert. Andere benutzte Tools waren AnalyzeMyWriting,³ Ratte (Regensburger Analysetool für Texte)⁴ und DyLan TextTools.⁵ Für einige Analyseteile wurde das ganze Korpus berücksichtigt, während in anderen Analyseschritten, bei denen eine allgemeine Untersuchung nicht möglich war, nur eine kleinere Auswahl an Texten (30, d. h. 10 pro Sprache) betrachtet wurde.

Bei der Untersuchung der Texte lag der Fokus auf den Aspekten, die einen Einfluss auf die Verständlichkeit für das Laienpublikum haben. Dabei wurden die unterschiedlichen Sprachebenen berücksichtigt: die lexikoterminologische Ebene, die morphosyntaktische Ebene, die textuelle Ebene, aber auch das Layout und das Design (inklusive der Multikodalität) und die genderinklusive Sprache. Lesbarkeitsindizes und weitere Werte (wie lexikalische Dichte und

² <https://www.sketchengine.eu/>

³ <https://www.analyzemywriting.com/>

⁴ <https://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/downloads/ratte/index.html>

⁵ https://www.ilc.cnr.it/dylanlab/apps/texttools/?tt_user=guest

TTR, d. h. Type-Token-Relation) wurden ebenfalls herangezogen, um die Verständlichkeit der Texte zu bewerten. Wichtige Bezugswerke für die Textanalyse waren Handbücher und Leitfäden zum Thema Einfacher und Leichter Sprache sowie die Leitfäden über die Zusammenfassungen.

4 Befunde (Auswahl)

Auf *lexikoterminologischer Ebene* wurden zuerst die häufigsten Lemmata (aus den offenen PoS-Kategorien: Substantive, Adjektive, Verben und Adverbien) mithilfe von Sketch Engine extrahiert, um deren Vorkommen bzw. Rang in Frequenzwörterbüchern (Davies/Gardner 2010, De Mauro 2016, Tschirner/Möhring 2020) zu überprüfen. Dann wurden die Fachtermini mit Morphemen griechischer und lateinischer Herkunft (die bekannterweise schwieriger zu verstehen sind als ihre Äquivalente germanischer Herkunft im Deutschen und Englischen) anhand einer dreisprachigen Liste (Magris 1992: 31–50) extrahiert. Schließlich sollten die Sätze mit Definitionen, Erklärungen sowie Synonymen und Oberbegriffen sowohl in den Ausgangstexten als auch in den Zieltexten berücksichtigt werden. Zu diesem Zweck wurden die häufigsten Verben und Konstruktionen extrahiert, die diese Kategorien einleiten (z. B. *mean* und *be known as* für Englisch). Wie erwartet ist die Mehrheit der am häufigsten vorkommenden Lemmata auch in den Frequenzwörterbüchern verzeichnet. Die Anzahl der altsprachlichen Morpheme im italienischen Subkorpus ist viel höher als im Ausgangskorpus, während sie im deutschen Subkorpus niedriger als in den Ausgangstexten ist. Das Übergewicht an griechisch-lateinischen Morphemen im italienischen Subkorpus ist naturgemäß auch auf die Sprachfamilie, zu der das Italienische gehört, zurückzuführen. Die geringere Zahl entsprechender Morpheme im Deutschen im Vergleich zum Englischen ist dagegen auffälliger, wenn man bedenkt, dass beide germanische Sprachen sind, also Ausgangs- wie Zieltexte zur selben Sprachfamilie gehören. Der Unterschied zwischen den Korpora ist daher auf eine bewusste Wahl seitens der Übersetzer:innen statt auf sprachliche Unterschiede zurückführbar. Dies wird durch den Befund gestützt, dass Übersetzer:innen deutscher Texte deutlich häufiger dazu tendieren, den Fachtermini Erklärungen hinzuzufügen, als die italienischen Kolleg:innen. Sie scheinen daher die Bedürfnisse der Leser:innen stärker zu beachten, möglicherweise weil sie sich des in den germanischen Sprachen besonders gravierenden Problems der terminologischen Registerunterschiede bei altsprachlichen Termini und ihren Lehnübersetzungen bewusst sind. Die italienischen Übersetzer:innen achten hingegen weniger auf die Verständlichkeit der medizinischen Begriffe und benutzen mitunter Termini, die der fachinternen Kommunikation vorbehalten sein sollten.

Auf *morphosyntaktischer Ebene* wurden hauptsächlich drei Merkmale untersucht: das Genus Verbi, die Nominalisierung und die Tendenz zur Parataxe oder Hypotaxe. Die Anzahl der Passivformen nimmt im deutschen Subkorpus ab und ist ein wenig höher im italienischen Subkorpus. Interessanterweise wurden in Bezug auf dieses Merkmal erhebliche Unterschiede zwischen Texten von bestimmten Pharmakonzernen festgestellt, was darauf hinweist, dass nur einige Übersetzer:innen diesem Aspekt in ihren Texten Aufmerksamkeit schenken. Bezüglich der Nominalisierungstendenz wurden unter anderem Substantive mit nominalisierenden Suffixen aus den drei Subkorpora extrahiert; weitere Substantive, die nicht automatisch extrahiert werden konnten, wurden nur anhand einer kleineren Auswahl an Texten untersucht. Der Hauptbefund ist, dass die Nominalphrasen in den Übersetzungen (insbesondere in den deutschen Texten) zunehmen. Diese Tendenz deutet darauf hin, dass Übersetzer:innen die

Besonderheiten der Einfachen Sprache nicht besonders beachten, sondern auf Automatismen zurückgreifen, die sie in der traditionellen Fachübersetzung entwickelt haben.

In Bezug auf Parataxe und Hypotaxe wurden unter anderem Subordinatoren und Koordinatoren mittels Sketch Engine extrahiert: Das Ausmaß der Subordination und Koordination nimmt in den Übersetzungen ab, vor allem im Deutschen. Diese Ergebnisse könnten darauf hinweisen, dass Übersetzer:innen der italienischen und deutschen Texte versuchen, die Komplexität der Zusammenfassungen zu vermindern. Allerdings entspricht diese Abnahme an Subordination nicht einer Zunahme an Hauptsätzen, sondern der vorher beschriebenen Tendenz, verbale Formen zu nominalisieren, was einen höheren Komplexitätsgrad mit sich bringt.

Auf *textueller Ebene* wurden Informationshierarchie und Struktur der Zusammenfassungen, metakommunikative Elemente, Allokutiva sowie die Textkohäsion berücksichtigt. Die im Rahmen der Analyse untersuchten Texte weisen generell eine klare Informationshierarchie auf. Außerdem sind sie durch nicht-exzessive intertextuelle und intratextuelle Verweise gekennzeichnet, was mit einer gewissen Informationsredundanz einhergeht, da diese Angaben mehrmals im Text präsentiert werden. Auch metakommunikative Elemente spielen bei diesen Texten eine sehr wichtige Rolle. Textfunktion und Zielgruppe werden (meistens schon am Anfang der Texte) sehr deutlich gemacht. Die Textkohäsion wurde anhand der Extraktion einer Liste von meist adverbialen Konnektoren (wie *allerdings*, *ansonsten*, *deshalb* und *auch*) durch Sketch Engine untersucht. Die übersetzten Subkorpora sind durch eine höhere Anzahl an Konnektoren als die Ausgangstexte gekennzeichnet. Diese Tendenz ist auch diesmal im deutschen Subkorpus viel stärker ausgeprägt.

Die Texte der einzelnen Pharmakonzerne zeigen jeweils meistens ein äußerst ähnliches Layout und Design. Analysiert wurden Aspekte wie Seitenlayout, Schriftart, -größe und -farbe, Hervorhebungsarten (Fettung, Kursivierung usw.), Abbildungen und Grafiken, die auch in den Richtlinien genau thematisiert werden. Es hat sich herausgestellt, dass dies die Textebene ist, auf der die Richtlinien am meisten eingehalten wurden: In der Regel werden serifenlose Schriftarten mit einer Größe von ungefähr 11-13 verwendet. Die Texte sind linksbündig, mit klaren Unterschriften und Absätzen. Die Wahrnehmbarkeit ist aber nicht so weitgehend optimiert wie bei Leichte-Sprache-Texten. Die Zusammenfassungen enthalten beispielsweise viele Tabellen und Abbildungen, z. B. um die Herkunftsländer der Teilnehmer:innen zu zeigen oder unerwünschte Ereignisse und deren Häufigkeit aufzulisten. Bilder sind generell konvergent, d. h., sie passen genau zu dem entsprechenden Textteil und sollten daher die Leser:innen nicht verwirren.

Die Verständlichkeit der Texte wurde dann auch durch Lesbarkeitsindizes überprüft, um die bisherigen Analysen zu vervollständigen. Obwohl diese Indizes ein umstrittenes Thema sind, da sie in der Regel rein statistische, nicht semantische und stilistische Werte angeben, sind sie in der Gesundheitskommunikation sehr weit verbreitet. Diese Analysephase wurde erst nach den anderen durchgeführt, damit es zu keinen Verzerrungen bei der Auslegung der Daten kommen konnte. Für diesen Teil der Untersuchung wurden die Tools *Analyze My Writing*, *Ratte* und *DyLan TextTools* verwendet, die unterschiedliche Indizes berechnen. Diese sind dann anhand von Skalen auszuwerten, die die Richtwerte für ein bestimmtes Publikum (hier ein Laienpublikum für Einfache Sprache) angeben. Aufgrund der Ergebnisse kann man sagen, dass die Zusammenfassungen viel schwieriger sind, als sie sein sollten, um von einem allgemeinen Publikum leicht verstanden zu werden.

5 Fazit und Ausblick

Zum Schluss lässt sich sagen, dass die Zusammenfassungen anhand ihrer Merkmale in einer Varietät von Einfacher Sprache geschrieben sind, die eher der Standardsprache als der Leichten Sprache nahe ist. Mit Blick auf die Ergebnisse der Lesbarkeitsindizes (sowie von denen der anderen Analysen) können sie im überwiegenden Fall nur schwer als Beispiel von Einfacher Sprache betrachtet werden. Sie könnten daher noch deutlich verbessert werden, insbesondere auf morphosyntaktischer Ebene. Auch bei den aktuellen Leitfäden für Zusammenfassungen besteht viel Raum für Verbesserungen. Sie sind in ihrer aktuellen Fassung sehr unstrukturiert, mit Informationen, die nicht einfach zu finden sind; sie sind teilweise lückenhaft, teilweise redundant. Terminologische Hinweise werden häufig wiederholt, während Aspekte wie die Nominalisierung nie erwähnt werden (was, wie bewiesen, schlechte Auswirkungen auf die Texte hatte). Die Empfehlungen sind zudem auch nicht nach Sprachebenen geordnet, sondern es wirkt, als wären die verschiedenen Aspekte nach dem Zufallsprinzip aufgelistet. Die Verbesserungsvorschläge für aktuelle Empfehlungen in den Leitfäden wurden deswegen in der Dissertation nach Ebenen (lexikoterminologisch, morphosyntaktisch, textuell, graphisch und allgemein) geordnet und durch weitere Empfehlungen aus den Handbüchern und Leitfäden zu Einfacher und Leichter Sprache vervollständigt.

Als weitere Forschungsdesiderata kann man Studien zum tatsächlichen Verständnis der Texte bei Laien sowie zur Rezeption und Anwendung der vorgeschlagenen erweiterten Leitfäden durch Übersetzer:innen nennen.

Literatur

- Ahrens, Sarah / Schulz, Rebecca / Kröger, Janina / Hernández Garrido, Sergio / Keller, Loraine / Rink, Isabel (Hrsg.) (2022): *Accessibility – Health Literacy – Health Information. Interdisciplinary Approaches to an Emerging Field of Communication*. Berlin: Frank & Timme.
- Bowker, Lynne / Pearson, Jennifer (2002): *Working with Specialized Language. A Practical Guide to Using Corpora*. London / New York: Routledge.
- Cheng, Winnie (2012): *Exploring Corpus Linguistics. Language in Action*. London / New York: Routledge.
- Davies, Mark / Gardner, Dee (2010): *A Frequency Dictionary of Contemporary American English Word Sketches, Collocates and Thematic Lists*. London: Routledge.
- De Mauro, Tullio (2016): „Il Nuovo vocabolario di base della lingua italiana.“ <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana/> (31.05.2023).
- GLSP (2021): „Good Lay Summary Practice.“ https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-10_en (03.02.2023).
- Hansen-Schirra, Silvia / Maaß, Christiane (2020): „Easy Language, Plain Language, Easy Language Plus: Perspectives on Comprehensibility and Stigmatisation.“ *Easy Language Research: Text and User Perspectives*. Hrsg. Silvia Hansen-Schirra / Christiane Maaß. Berlin: Frank & Timme. 17–40.
- Maaß, Christiane / Rink, Isabel (2017): „Leichte Sprache: Verständlichkeit ermöglicht Gesundheitskompetenz.“ *Public Health Forum* 25.1: 50–53.
- Maaß, Christiane / Rink, Isabel (2018): „Das nennt Ihr Arzt: Rigor.‘ Medizinische Fachtexte in Leichter Sprache und der Ansatz der Situated Translation.“ *Barrieren abbauen, Sprache gestalten*. Hrsg. Susanne Jekat / Martin Kappus / Klaus Schubert. Winterthur: ZHAW. 24–38.

- Magris, Marella (1992): „La traduzione del linguaggio medico: analisi contrastiva di testi in lingua italiana, inglese e tedesca.“ *Traduzione, società e cultura*. Bd. 2. Hrsg. Gabriella Di Mauro / Federica Scarpa. Trieste: EUT. 3–82.
- Pedrini, Giulia (2022): „Plain and Easy Language as a Means to Increase Health Literacy on COVID-19: A Contrastive Analysis of English and German Texts.“ *trans-kom* 15.1: 142–155.
- Pedrini, Giulia (2024): *Medical Communication between Plain Language and Einfache Sprache: A Corpus Analysis of Layperson Summaries of Clinical Trials in English, German, and Italian*. Berlin: Frank & Timme.
- Recommendations of the Expert Group on Clinical Trials (2018): „Summaries of Clinical Trial Results for Laypersons. Recommendations of the Expert Group on Clinical Trials for the Implementation of Regulation (EU) No 536/2014 on Clinical Trials on Medicinal Products for Human Use.“ https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-10_en (14.01.2024).
- Ross, Dolores / Magris, Marella (2022): „Semplificazione linguistica nei Paesi Bassi e in Germania: il caso della comunicazione istituzionale sul COVID-19.“ *Rivista Internazionale di Tecnica della Traduzione* 24:105–131.
- Tschirner, Erwin / Möhring, Jupp (2020): *A Frequency Dictionary of German Core Vocabulary for Learners*. Oxon / New York: Routledge.