

## **Bewertung der Qualität maschineller Übersetzung journalistischer Texte aus dem Deutschen ins Slowakische**

Oľga Wrede – Daša Munková – Renáta Pavlová

Katedra germanistiky, Filozofická fakulta, Univerzita Konštantína Filozofa v Nitre  
owrede@ukf.sk

Katedra translatológie, Filozofická fakulta, Univerzita Konštantína Filozofa v Nitre  
dmunkova@ukf.sk

Jazykové centrum, Filozofická fakulta, Univerzita Konštantína Filozofa v Nitre  
rpavlova@ukf.sk

**Schlüsselwörter:** maschinelle Übersetzung, journalistische Texte, Fehlertypologie, Übersetzungsqualität, Post-Editing, Evaluation, Deutsch, Slowakisch

**Key words:** machine translation, journalistic texts, error typology, quality of translation, postediting, evaluation, German language, Slovak language

### **Vorwort**

Auf dem Sprachdienstleistungsmarkt sind gegenwärtig etwa 330 000 Übersetzer und Dolmetscher<sup>1</sup> weltweit beschäftigt (Pym et al., 2012, S. 137). In den letzten Jahrzehnten stieg jedoch der Bedarf an Übersetzungen so massiv an, dass die Anfrage auf dem Übersetzungsmarkt das Angebot deutlich übersteigt. Dies ist in erster Linie als Folge vielfältiger Globalisierungsprozesse anzusehen, die inzwischen alle Lebensbereiche erfassen und sich auf allen Ebenen des menschlichen Handelns auswirken. Der rasch wachsende internationale Handel, die Intensivierung interkultureller und politischer Kontakte, der Zugang zum Internet und sonstigen Kommunikations- und Informationstechnologien einerseits, aber auch die Migrationswellen der letzten Jahre und die damit verbundenen Integrationsprozesse sowie die Sicherstellung von Menschen- und Grundrechten innerhalb und außerhalb der EU andererseits erfordern, dass Übersetzungen von Dokumenten diverser Art und aus diversen Bereichen in großem Umfang und einer oft kurzen Zeit angefertigt werden müssen.

Technologische Innovationen beschleunigten die Verbreitung von Informationen und veränderten grundsätzlich auch die Art und Weise der menschlichen Kommunikation. Dies betrifft auch Sprach- und Übersetzungstechnologien, die zwar eine relativ neue Erfindung sind, „Maschinen“ an sich aber bereits seit einigen Jahrzehnten auf unser Leben Einfluss nehmen. In Zeiten einer tiefgreifenden Umstrukturierung der Gesellschaft zur Informationsgesellschaft kann sodann angenommen werden, dass Maschinen und die mit ihnen zusammenhängenden Technologien zu einem nicht mehr wegzudenkenden und unabdingbaren Bestandteil der Menschheit geworden sind.

Maschinelle Übersetzung wird angesichts ihrer technischen Realisierung als ein Werkzeug betrachtet, das mittels komplizierter Befehle und Algorithmen jeglichen Text in kleinere Einheiten zerlegen kann. Diese können anschließend mittels einer Statistik (d. h. aufgrund der Wahrscheinlichkeit und Regelmäßigkeit des Auftretens) übersetzt werden, wobei umfangreiche Textkorpora und Übersetzungsspeicher (engl. *translation memory*, kurz *TM*) verwendet werden, in denen die Maschine nach adäquaten Lösungen sucht (vgl. Munková, 2013, S. 37). Die Anfänge maschineller Übersetzung (engl. *Machine Translation*,

---

<sup>1</sup> Im gesamten Text wird das Maskulinum generisch für maskuline und feminine Bezeichnungen von Personen verwendet.

abgekürzt *MT*) gehen bis in die 1930er Jahre zurück. Burchardt – Porsiel (2017, S. 11) nehmen zur Entwicklung maschineller Übersetzung folgende Stellung ein:

Maschinelle Übersetzung erlebte seitdem viele Höhen und Tiefen. Nach einer langen Durststrecke ist maschinelle Übersetzung jetzt allerdings schon seit etlichen Jahren einmal mehr in aller Munde. Unter anderem auch deshalb, weil Rechen- und Prozessorleistung keine limitierenden Faktoren mehr sind, weder technisch noch finanziell, und weil inzwischen viele bilinguale Daten im notwendigen Umfang und gelegentlich auch in der benötigten Qualität zum Erzeugen der Systeme zur Verfügung stehen.

Maschinelle Übersetzung muss nicht vollkommen sein, um ihren Zweck gerecht zu werden. Dies bezieht sich insbesondere auf die Übersetzung von E-Mails, SMS, Diskussionsforen, Geschäftskorrespondenz u. ä. In den Jahren 2012 bis 2016 verzeichnete die Gesellschaft Google beinahe eine doppelt so häufige Nutzung des Google Übersetzers. Während 2012 mittels Google Übersetzers 75 Milliarden Wörter täglich übersetzt wurden, stieg 2016 die Zahl der täglich übersetzten Wörter auf 143 Milliarden in 100 Sprachkombinationen. Ähnliche Zahlen veröffentlichte auch die Gesellschaft Facebook. Im Jahr 2016 wurden etwa zwei Milliarden Wörter von einer Milliarde Benutzer täglich übersetzt (vgl. Way, 2018, S. 161).

Maschinelle Übersetzung wird darüber hinaus auch von Gesellschaften in unterschiedlichem Umfang und in unterschiedlichen Bereichen eingesetzt (Recht, Produktion, Informationstechnologien, Marketing, Finanzen u. a.). Als Motivation für die Nutzung maschineller Übersetzung sind für die Gesellschaften der Preis sowie die Schnelligkeit der Distribution von Übersetzungen zu nennen, unter der Voraussetzung, dass die Qualität der Übersetzung nicht zu den Prioritäten der Gesellschaft gehört. Hier ist jedoch fairerweise anzumerken, dass es eher unwahrscheinlich ist, dass eine Maschine eine qualitativ gleichwertige Übersetzung wie ein Übersetzer anfertigen kann. Ungeachtet dessen verwenden auch große Organisationen und Gemeinschaften wie beispielsweise die Europäische Union maschinelle Übersetzung vor allem dann, wenn große Mengen an Dokumenten schnell übersetzt werden müssen, um zu entscheiden, welche von ihnen wichtig sind und welche noch „warten können“.

Manche Gesellschaften bieten Übersetzungslösungen mit mehreren Domänen an wie beispielsweise SYSTRAN (fünf Domänen) oder KantanMT (acht Domänen), andere Gesellschaften wiederum nur eine spezielle Domäne wie beispielsweise SDL Government, Canopy Innovations für Gesundheitswesen oder Lingua Custodia für Finanzen.

Auch wenn die Übersetzungslösungen durch die Nutzung von CAT-Werkzeugen oder MT größtenteils automatisiert sind, sind sie nach wie vor vom Übersetzer abhängig. Dieser partizipiert an dem Übersetzungsprozess entweder in der Phase des sog. *Pre-Editings* (Vorbereitung für die MT) oder in der Phase des sog. *Post-Editings* (Nachbearbeitung der MT) (vgl. Hutchins, 2005; Quah, 2006).

Der Fortschritt im Bereich der maschinellen Übersetzung baut auf der Bewertung der Qualität des neu entwickelten Systems maschineller Übersetzung auf, und dies mittels Evaluation mit dem Ziel, seine höhere Effektivität im Vergleich zu bereits existierenden Systemen nachzuweisen. Wenn auch die Qualität maschineller Übersetzung nicht die Qualität menschlicher Übersetzung erreicht, sollten sich die Kriterien ihrer Bewertung voneinander grundsätzlich nicht unterscheiden. Beide Arten der Übersetzung sollten dasselbe Ziel verfolgen, d. h., eine maschinell generierte Übersetzung sollte genauso vertrauensvoll, akzeptabel (verständlich) und stilistisch „geschliffen“ sein wie eine vom Übersetzer angefertigte Übersetzung.

Die Evaluation der maschinellen Übersetzung stützt sich auf zwei grundlegende Ansätze. Es ist die sog. *Glass-Box-Evaluation* und die sog. *Black-Box-Evaluation*.<sup>2</sup> In dem hier vorliegenden Beitrag wird auf die Black-Box-Evaluation Bezug genommen, die mittels interner und externer Methoden die Genauigkeit und Nutzbarkeit maschineller Übersetzung bewertet.

## 1 Bewertung der Übersetzungsqualität

Äquivalenz gehört unmissverständlich zu den zentralen Begriffen der linguistisch orientierten Übersetzungstheorien (vgl. Prunč, 2002, S. 33). Der Begriff der Äquivalenz wurde in den 1950er Jahren aus den technischen Disziplinen bzw. aus der formalen Logik in die (deutsche) Translationswissenschaft übernommen, als man in der Euphorie der ersten Erfolge der maschinellen Übersetzung noch hoffte, in jeder Sprache Elemente finden zu können, die mit Hilfe eines einfachen Regelsystems durch äquivalente Elemente einer anderen Sprache ersetzt werden könnten (ibid.).

Die Äquivalenz wird jedoch in den Übersetzungstheorien unterschiedlich wahrgenommen und gewichtet. Manche Autoren halten die Äquivalenz für äußerst wichtig (Jakobson, Nida, Catford, House u. a.), von anderen wird das Thema der Äquivalenz kontrovers diskutiert (Reiß, Vermeer, Snell-Hornby u. a.). Unterschiedliche Wahrnehmung der Äquivalenz führt sodann zu verschiedenen Konzeptionen der Übersetzungsqualität sowie zu voneinander abweichenden Arten ihrer Bewertung.

House (2015, S. 12) betrachtet die Übersetzung einerseits als kognitiven Prozess, der im Kopf des Übersetzers vorgeht, andererseits aber auch als soziale, interlinguale und interkulturelle Erfahrung des Übersetzers. House hält den Äquivalenzbegriff in der Übersetzungstheorie für unverzichtbar. Im Vergleich zu anderen Äquivalenztheorien definiert House den Äquivalenzbegriff aber etwas anders, um Übersetzungen von Bearbeitungen zu unterscheiden. House (2009, S. 33) stellt die Funktionsäquivalenz von Ausgangs- und Zieltext in den Mittelpunkt und gelangt schließlich zu zwei Arten der Übersetzung, der *offenen* und der *verdeckten* Übersetzung (engl. *overt and covert translation*). Die erstere Art der Übersetzung deckt historische Reden und literarische Texte ab, die in einer Kultur und in einer Sprache stark verankert sind, und somit in der Zielkultur wie ein Original wirken. Zur zweiten Art der Übersetzung zählt House (2001, S. 263) beispielsweise journalistische Texte, die in multinationalen Texten erscheinen, Werbeschriften für Produkte, die in verschiedenen Sprachgemeinschaften verbreitet sind, Informationsbroschüren für Touristen verschiedener Nationalitäten und schließlich auch Texte, die in globalisierten Firmen verwendet werden.

Die folgenden Ausführungen lehnen sich an die Theorie von House (2015, S. 14) an, die unter Übersetzung den Ersatz eines in der Ausgangssprache abgefassten Textes durch einen in der Zielsprache abgefassten Text versteht. Múglová (2009, S. 111) erfasst in diesem Kontext die Äquivalenz als „Relation zwischen den einzelnen Sprachzeichen des Ausgangs- und Zieltextes (Mikroebene des Textes) und die Relation zwischen dem Ausgangs- und Zieltext als Ganzem (Makroebene des Textes)“. In diesem Sinne stellt die Übersetzung nicht nur eine sprachliche Handlung dar, sondern auch eine sekundäre Handlung der Kommunikation zwischen Kulturen. Der Übersetzer, metaphorisch ausgedrückt, schlägt Brücken zu anderen Sprachen und Kulturen und erfüllt in der Kommunikation die Funktion eines Mediators.

<sup>2</sup> Die Glass-Box-Evaluation bewertet die Qualität des Systems anhand interner Merkmale. Diese Methode ist auf die linguistische Reichweite des Systems wie auch auf Theorien, die bei der Bearbeitung natürlicher Sprachen angewendet werden, gerichtet. Die Black-Box-Evaluation misst dagegen die Qualität des Systems aufgrund der eigentlichen (maschineller) Übersetzung (vgl. Munková – Wrede – Absolon, 2019, S. 239).

In den 1990er Jahren sind auch andere Aspekte der Übersetzung und der Bewertung von Übersetzungsqualität in den Fokus des Interesses gerückt. Diese Aspekte ergaben sich aus den zweisprachigen Korpora, die auf TM gebaut haben, und zum Gegenstand der empirischen Forschungsansätze und der Korporaforschung wurden. Die Übersetzungsqualität wurde dann korpusbasiert bewertet, um Subjektivität und Nichteindeutigkeit der Übersetzungen zu eliminieren. Ergebnisse der korpusbasierten Forschung lieferten wichtige und nützliche Formeln in der Übersetzung wie beispielsweise lexikalische Vereinfachungen, Explizitierungen oder Verwendung von standardisierten Sprachformen und Einfluss der Sprachstruktur des Ausgangstextes auf Übersetzungslösungen.

Doherty (2016) vertritt die Meinung, dass zweisprachige Korpora zu einer komparativen mehrdimensionalen Evaluation der Übersetzungsqualität führen. Die Mehrdimensionalität der Evaluation bezieht sich so dann auf Lesbarkeit und Verständlichkeit (Doherty, 2012), diagnostische Bewertung (Gaspari et al., 2014), Verwendungsmaß (Doherty – O’Brien, 2014) und kognitive Ansprüchigkeit (Doherty – O’Brien – Carl, 2010).

Die Bewertung der Übersetzungsqualität ist äußerst kompliziert. Es gibt verschiedene Qualitätskonzeptionen und Meinungen, was in der Übersetzung akzeptabel ist und was nicht mehr. All dies verursacht, dass die Bewertungskriterien der Übersetzung nicht einheitlich sind. So hat Newmark (1991) nach Kriterien einer „guten Übersetzung“ gesucht, House (1997), Nord (1991) u. a. unternahmen den Versuch, Übersetzungsfehler im Vergleich zu sprachlichen Fehlern zu identifizieren. Gouadec (1989) stellte eine Liste von möglichen Übersetzungsfehlern auf, aufgrund deren relative Kriterien der Übersetzungsqualität bestimmt wurden (siehe auch Williams, 1989). Einige Translationswissenschaftler berücksichtigten bei der Suche nach objektiven Kriterien der Übersetzungsqualität die Beschaffenheit von Übersetzungsfehlern sowohl aus linguistischer als auch pragmatischer Sicht (siehe Williams, 1989; Hewson, 1995; Kussmaul, 1995; Nord, 1991 u. a.). House (1981) ging bei der Festlegung von Kriterien der Übersetzungsqualität aus linguistischer Textanalyse aus, wodurch die Analyse mit der eigentlichen Übersetzungsbewertung verbunden wurde. Larose (1989) und Dancette (1989) bauten auf den Ausführungen von House auf, indem sie bei der Festlegung von Kriterien der Übersetzungsqualität besonderen Wert auf die einzelnen Textebenen gelegt haben und Zusammenhänge zwischen den Textebenen und Übersetzungsfehlern suchten. Dancette (1992) und Snell-Hornby (1995) gingen bei der Festlegung von Kriterien der Übersetzungsqualität noch weiter, da sie Bewertung der Übersetzungsqualität mit der psycholinguistischen Theorie von „scenes and frames“ verbanden. Alle diese Versuche waren aber theoretischer Natur, daher sind die daraus resultierenden Kriterien für die Bewertung der Übersetzungsqualität in der Praxis weniger anwendbar bzw. „greifbar“.

Maschinelle Übersetzung bedarf trotz ihrer Entwicklung und Erforschung nach wie vor eines Eingriffs des Übersetzers. Dies geschieht entweder in Form des Post-Editings oder durch die Überprüfung der Qualität der Übersetzung vor ihrer Veröffentlichung. Bei der Bewertung der Qualität maschineller Übersetzung (insbesondere bezüglich effektiver Methoden und Werkzeuge) gehen jedoch die Ansichten der Fachleute aus der Übersetzungsindustrie und der Translationswissenschaft teilweise auseinander. Die Gründe für die Suche nach anderen Werkzeugen für die Bewertung der Übersetzungsqualität waren u. a. der Zeitaufwand sowie die Nichteindeutigkeit von Kriterien der Angemessenheit und Flüssigkeit maschineller Übersetzung (Sprachfluss). Papineni et al. (2002) merken in diesem Zusammenhang an, dass Methoden und Werkzeuge für die Entwicklung von Systemen maschineller Übersetzung zu langsam und kostspielig sind, für die bezüglich der Übersetzungsqualität eine schnelle Rückmeldung wichtig ist. Vilar et al. (2006) konstatieren wiederum, dass die Subjektivität, die für eine „manuelle“ (menschliche) Bewertung charakteristisch ist, gewisse Voreingenommenheit der Bewertenden gegenüber maschineller



Übersetzung, sowie eine nicht klar definierte numerische Skala der Adäquatheit und Flüssigkeit maschineller Übersetzung in sich birgt. Snover et al. (2006) behaupten weiter, dass sich die Nichtkonsistenz der Meinungen von Bewertenden maschineller Übersetzung in niedriger Korrelation zwischen den Metriken menschlicher (manueller) Evaluation widerspiegelt. Thurmair (2005) sieht gleich mehrere Gründe für die Nichtkonsistenz der Meinungen von Bewertenden maschineller Übersetzung wie beispielsweise unterschiedliche Sprachkenntnisse, Übersetzungskompetenzen, Fachkenntnisse bezüglich der gegenständlichen Problematik u. a.

Obgleich beide Kriterien (Angemessenheit und Sprachfluss) für die Bewertung maschineller Übersetzung weiterhin verwendet werden, werden heute auch andere Methoden (Algorithmen) automatischer Bewertung entwickelt. Diese basieren meistens auf Messung der Ähnlichkeit zwischen maschineller Übersetzung, die bewertet wird, und menschlicher Übersetzung, der sog. *Referenzübersetzung*. Doherty (2016, S. 959) vertritt die Meinung, dass solche Werkzeuge für die Bewertung maschineller Übersetzung zwar weit von menschlichem „ausgereiften“ Urteil sind, dafür aber „schnelle wenn auch nicht saubere“ Lösung bieten, die für die Forschung und Entwicklung von besonderer Bedeutung ist.

## 2 Frameworks zur Qualitätsbewertung maschineller Übersetzung

Bis vor Kurzem gab es zwei Frameworks zur Qualitätsbewertung maschineller Übersetzung. Das erste ist das *dynamische Framework* (Dynamic Quality Framework, DQF), der von *Translation Automation User Society* (TAUS) entwickelt wurde. Das DQF bietet industriell entwickelte Werkzeuge für die Bewertung der Übersetzungsqualität an. Eine der DQF-Methoden geht von der Fehlertypologie und Fehlerfrequenz aus. Die DQF-Fehlertypologie wurde später zur Grundlage von mehrdimensionalen Metriken der Qualität (Multidimensional Quality Metrics, MQM). Das zweite Framework zur Qualitätsbewertung maschineller Übersetzung, das MQM-Framework, benennt und definiert Probleme, Beziehungen zwischen ihnen sowie Mechanismen zwecks Bestimmung des Qualitätscores. Das MQM eignet sich für die Bewertung von Eigenschaften der Übersetzungsqualität, die überprüfbar und nicht subjektiv sind. Das MQM-Framework wurde 2015 entwickelt und kann nicht nur bei Übersetzungen, sondern auch Ausgangstexten angewandt werden.<sup>3</sup> Der Zugang zur Qualitätsbewertung ist in diesem Fall eher analytischer als holistischer Natur.

Beide Frameworks definieren sechs identische Bereiche der Fehlerhaftigkeit bezüglich der Übersetzungsqualität:

- DQF Accuracy (formerly Adequacy) = MQM Accuracy (accuracy)
- DQF Design (formerly Layout) = MQM Design (design)
- DQF Fluency (formerly Language) = MQM Fluency (fluency)
- DQF Locale convention (formerly Country standards) = MQM Locale convention (locale-convention)
- DQF Style = MQM Style (style)
- DQF Terminology = MQM Terminology (terminology)

In dem vorliegenden Beitrag werden lediglich zwei dieser Bereiche besprochen, und dies die *Genauigkeit des Ausdrucks* (accuracy) sowie seine *grammatische Korrektheit* (fluency). Genauigkeit des Ausdrucks besagt, inwieweit der Zieltext den Inhalt (Informationen) des Ausgangstextes widerspiegelt. Die grammatische Korrektheit/Richtigkeit des Zieltextes wird aus Sicht des Endrezipienten bewertet.

Das kategoriale Framework zur Analyse von Fehlerhaftigkeit der Übersetzung ins Slowakische korrespondiert mit dem MQM-Framework sowie mit der DQF-Fehlertypologie.

<sup>3</sup> Vgl. < <http://www.qt21.eu/mqm-definition/definition-2015-06-16.html> > [Cit. 2020-12-01.].

Vaňko (2017, S. 96) definiert die Beziehung zwischen dem kategorialen Framework und der Fehlertypologie folgendermaßen:

1. Language („Sprache“), d. h. Grammatik, in unserer Auffassung Morphologie und Syntax bzw. Morphosyntax. Im präsentierten Modell korrespondieren mit ihr die Kategorien wie Prädikation, syntaktisch-semantische Korrelation (Kongruenz), Syntax zusammengesetzter Sätze und teilweise auch Modalität (z. B. Negation); 2. Accuracy („Genauigkeit“): es handelt sich insbesondere um nicht korrekte Bedeutungen im Text der Zielsprache, um Auslassung eines Lexems u. ä. Im Modell sind diese Elemente eines nicht korrekten Transfers im Block „Lexik“ vertreten; 3. Terminology („Terminologie“): hier geht es um einen negativen Transfer eines Terminus der Ausgangssprache; in unserem Modell tritt die Problematik des Transfers von Termini als Unterkategorie der Lexik auf. Bei der letzten Kategorie Style („Stil“) geht es um eine Unstimmigkeit zwischen dem Stil des Ausgangstextes und dem Stil des Zieltexes. (e. Ü.)

### 3 Quantitative Fehleranalyse maschineller Übersetzung

Im Folgenden wird der Frage nachgegangen, wie die einzelnen Kategorien Vaňkos Modells der Fehlertypologie (2017), welches für Slowakisch entworfen wurde, mit dem dynamischen Framework DQF der Fehlertypologie zusammenhängen. Analysiert wird das Maß der Abhängigkeit zwischen ausgewählten Parametern (Angemessenheit und Sprachfluss) beider Frameworks bei der maschinellen Übersetzung von journalistischen Texten. Überprüft wird weiter, ob es bei der maschinellen Übersetzung aus dem Deutschen ins Slowakische bedeutende Unterschiede im Fehlervorkommen nach der Fehlerkategorisierung von Vaňko (2017) gibt. Zum Schluss soll aufgezeichnet werden, welche Fehler am häufigsten bzw. am wenigsten vorkommen und welche in gleichem Ausmaß vertreten sind.

#### 3.1 Hypothesen

Bei den hier aufgestellten Hypothesen geht es um statistische Nullhypothesen:

H01: *Das Sprachniveau ist vom Fehlervorkommen in den untersuchten Kategorien nicht abhängig. (Fehlerkategorien: 1, 2, ..., 12)*

H02: *Das Niveau der Sprachgenauigkeit ist vom Fehlervorkommen in den untersuchten Kategorien nicht abhängig. (Fehlerkategorien: 1, 2, ..., 12)*

#### 3.2 Das Sprachkorpus

Untersucht wurden deutschsprachige Nachrichtentexte, welche der Webseite [www.faz.net](http://www.faz.net) im Zeitraum von Oktober 2016 bis Mai 2017 entnommen wurden. Das Sprachkorpus bilden insgesamt 150 Normseiten. Publizistische Texte wurden u. a. auch deswegen gewählt, weil ihre Sprache relativ einfach ist, sie über aktuelles Geschehen berichten und Antworten auf Fragen *wer? was? wann? wo? wie? warum?* liefern. Die Übersetzung von journalistischen Texten zählt zugleich zu den häufigsten im Rahmen der maschinellen Übersetzung.

#### 3.3 Die Methodologie

Deutschsprachige Ausgangstexte wurden ins Slowakische mittels *Google Translate* übersetzt. Die Wahl des Translators wurde dadurch bedingt, dass es sich um eine frei zugängliche Software handelt, welche die Sprachkombination Slowakisch – Deutsch unterstützt. Mit der Beurteilung der Übersetzungsqualität wurden professionelle Übersetzer beauftragt.

Für die Qualitätsanalyse der maschinellen Übersetzung wurde das OSTPERE-Online System für Übersetzung, Post-Editing, Revision und Evaluation verwendet. Im OSTPERE-System wurde nicht nur das Post-Editing, sondern auch die Klassifikation der Grundfehler maschineller Übersetzung (Sprache, Genauigkeit, Terminologie und Stil) von fünf professionellen Übersetzern durchgeführt. Jedes Segment wurde zuerst den Anweisungen

entsprechend korrigiert (post-editiert). Dabei war zu beachten, dass die maschinelle Übersetzung durch marginale Eingriffe korrigiert werden sollte, so dass der Text verständlich ist. Der Stil spielte dabei keine besonders wichtige Rolle. Anschließend wurden Fehler in puncto Sprache, Genauigkeit, Terminologie und Stil identifiziert. Anhand dessen wurde beurteilt, ob diese Fehler schwerwiegend, groß, klein oder neutral sind und ob sie das Textverständnis behindern. Bei der Identifikation der Fehler wurde die sog. *Drei-Sekunden-Regel* angewandt, d. h., dass innerhalb von drei Sekunden bestimmt werden sollte, um was für einen Fehler es geht. Zwecks Qualitätssicherung der Übersetzung wurden alle post-editierten Texte von einem Editor (einem slowakischen Muttersprachler) überprüft und für die Identifikation und Klassifikation der Fehler maschineller Übersetzung vorbereitet. Die Identifikation und Klassifikation wurde durch sechs Slowakisten nach Vaňkos kategorialem Modell (2017) durchgeführt.

### 3.4 Methoden

#### 3.4.1 Parameter der Fehlerhäufigkeit maschineller Übersetzung – translatologischer Aspekt

- Der Fehler „Sprache“ (Language) umfasst grammatische, syntaktische und Interpunktionsfehler.
- Der Fehler „Genauigkeit“ (Accuracy) bezieht sich auf Ungenauigkeit der Bedeutung bzw. Auslassen oder Hinzufügen von Bedeutungen in der Übersetzung.
- Der Fehler „Terminologie“ (Terminology) bedeutet die Nichteinhaltung konsistenter Terminologie wie auch der Terminologie allgemein.
- Der Fehler „Stil“ (Style) kann subjektiv sein, er sollte jedoch nach den für die Schriftform der jeweiligen Sprache geltenden Standards beurteilt werden.

#### 3.4.2 Das kategoriale Modell für die Fehleranalyse maschineller Übersetzung (Vaňko, 2017) – linguistischer Aspekt

I. Prädikation	II. Das modale und kommunikative Modell des Satzes	III. Syntaktisch-semantische Korrelation (Kongruenz)	IV. Syntax zusammengesetzter Sätze	V. Lexikale Semantik
<b>1. Prädikationskategorien</b> i. Tempus ii. Modus  <b>2. Kongruente Kategorien</b> i. Kongruenz in der Person ii. Kongruenz im Numerus iii. Kongruenz im Genus	<b>1. Modalität</b> i. Notwendigkeit ii. Pflicht iii. Möglichkeit iv. Intention v. Epistemic (gewisse Modalität) vi. Evaluation (Bewertung) vii. Emotionalität viii. Andere... ix. Negation  <b>2. Kommunikative Funktionen</b> i. Behauptung ii. Interrogativ iii. Imperativ iv. Wunsch	<b>1. Nominale Morphosyntax</b> i. Kongruenz in dem determinativen Syntagma ii. Rektion in dem determinativen Syntagma  <b>2. Pronominale Morphosyntax</b>  <b>3. Numerale Morphosyntax</b>  <b>4. Verbale Morphosyntax</b> i. Rektion ohne Präposition ii. Rektion mit der Präposition <b>5. Wortstellung</b>	<b>1. Identifikation der Satzanzahl</b>  <b>2. Identifikation der semantischen Beziehungen zwischen den Sätzen</b>  <b>3. Konnektivität zwischen den Sätzen (Elidierung der konnektiven Ausdrücke)</b>  <b>4. Temporale Verschiebungen des Verbs im Nebensatz</b>  <b>5. Andere Erscheinungen</b>	<b>1. Die adäquate Übertragung der Wortbedeutung</b>  <b>2. Polysemie</b>  <b>3. Das Erhalten der semantischen Kompatibilität</b>  <b>4. Stilistische Kompatibilität</b>  <b>5. Homonymie</b>  <b>6. Termini</b>  <b>7. Derivationserrscheinungen</b> <b>8. Auslassen der Lexeme</b>

		<b>6. Andere morphosyntaktische Erscheinungen</b>		<b>9. Andere Erscheinungen</b>
--	--	---	--	--------------------------------

**Tabelle 1** Das kategoriale Modell für die Fehleranalyse maschineller Übersetzung nach Vaňko (2017)

### 3.4.3 Parameterfreie Korrelation

Zum Testverfahren der Abhängigkeit und Maßbestimmung der Abhängigkeit wurde die parameterfreie Korrelation verwendet. Konkret wurde der Kendall'sche Koeffizient Tau gewählt, da die untersuchten Variablen aus einer unbekannten Verteilung kommen. Das Maß der Abhängigkeit wurde in Anlehnung an Cohen (1988) interpretiert: sgn. < 0.1 triviale Abhängigkeit, 0.1-0.3 kleine Abhängigkeit, 0.3-0.5 mittlere Abhängigkeit, 0.5-0.7 große Abhängigkeit, > 0.7 sehr große Abhängigkeit.

### 3.5 Ergebnisse

**H01:** *Das Sprachniveau ist vom Fehlervorkommen in den untersuchten Kategorien nicht abhängig (Fehlerkategorien: 1, 2, ..., 12).*

Sprache & Kategorien von Fehlern	Valid N	Kendall Tau	Z	p-value
sprache & 4#gruppe_kommunikation	2891	0.03	2.41	0.016032
sprache & 7# gruppe_numerale_morphosyntax	2891	0.06	4.68	0.000003
sprache & 1# gruppe_prädikation_kategorien	2891	0.12	9.64	0.000000
sprache & 3# gruppe_modalität	2891	0.13	10.32	0.000000
sprache & 6# gruppe_pronominale_morphosyntax	2891	0.14	11.22	0.000000
sprache & 8# gruppe_verbale_morphosyntax	2891	0.14	11.26	0.000000
sprache & 9# gruppe_wortstellung	2891	0.20	15.93	0.000000
sprache & 10# gruppe_andere_morphosyntax	2891	0.24	19.03	0.000000
sprache & 5# gruppe_nominale_morphosyntax	2891	0.24	19.31	0.000000
sprache & 11#gruppe_syntax_zusammengesetzte_sätze	2891	0.25	20.00	0.000000
sprache & 2# gruppe_kongruente_kategorien	2891	0.28	22.92	0.000000
sprache & 12# gruppe_lexikale_semantik	2891	0.45	36.52	0.000000

**Tabelle 2** Parameterfreie Korrelation: Sprache & Fehlerkategorien (Bánik et al., 2019, S. 160)

Bei der Kategorie 4 (Tabelle 2) wird die Nullhypothese einer Zuverlässigkeit von 95 % abgelehnt. Bei weiteren Kategorien (Tabelle 2) wird die Nullhypothese einer Zuverlässigkeit von 99.9 % abgelehnt.

Für alle Fehlerkategorien (Tabelle 2) wurde eine direkt proportionale Abhängigkeit festgestellt, d. h., die Werte ändern sich gemeinsam in die gleiche Richtung (Kendall Tau > 0). Je schlechter das Sprachniveau, desto höher ist das Fehlervorkommen. Markante Unterschiede sind allerdings bezüglich des Abhängigkeitsmaßes zu verzeichnen. Während das Maß der Abhängigkeit zwischen dem Sprachniveau und Fehlervorkommen bei den Fehlerkategorien 4 und 7 (Tabelle 2) nur trivial ist (Kendall Tau < 0.1), wurde bei der Fehlerkategorie 12 (Tabelle 2) ein mittleres Abhängigkeitsmaß (Kendall Tau = 0.45) bestimmt. Bei den restlichen Kategorien (Tabelle 2) wurde ein niedriges Abhängigkeitsmaß (Kendall Tau: 0.1-0.3) zwischen dem Sprachniveau und Fehlervorkommen identifiziert.

Die Fehlerkategorie „Sprache“ (Language) umfasst, wie bereits angegeben (siehe Punkt 3.4.1.), grammatische und syntaktische Fehler sowie Interpunktionsfehler. Beim ersten



Kriterium der Bewertung „Sprache“ ist nicht vorauszusetzen, dass die lexikalische Semantik (Wortsemantik) eine Erscheinung ist, die am bedeutendsten vertreten ist. Die Lexik ist vor allem für die Terminologie und Genauigkeit von Bedeutung. Die lexikalische Semantik steht jedoch an der ersten Stelle der Wichtigkeit bei allen Kriterien, unabhängig davon, ob es um den Stil, die Angemessenheit oder eine andere Bewertung geht. Es ist im Allgemeinen der meist vorkommende Fehlerbereich bei einer niedrigeren Bewertung der Übersetzung. Bei der Kategorie „Sprache“ zeigt sich jedoch eine relativ niedriger Abhängigkeitsindex (0.45). Bei anderen Kriterien werden auch die Werte über 0.50 nachgewiesen. Die zweite bedeutendste Fehlerkategorie stellen die Kongruenzkategorien dar (siehe Beispiel). Hier spielt die Grammatik eine wichtige Rolle, da in den flektierenden Sprachen die Kongruenz des Subjekts und des Prädikats ein wichtiges Element der sprachlichen Korrektheit ist. In den analysierten journalistischen Texten kamen diese Fehler häufig vor (vgl. Bánik et al., 2019, 159–162).

*Beispiel* (vgl. Bánik et al., 2019, S. 161):

AT: *Sie mussten den Tod eigener Verwandter mitansehen und selbst dem IS entkommen – viele Kinder in den Flüchtlingslagern im Nordirak sind schwer traumatisiert.*

MÜ: *Museli svedkom smrti svojich vlastných príbuzných a uniknúť je samo osebe – mnoho detí v utečeneckých táboroch v severnom Iraku boli ťažko traumatizovaná.*

PEMÜ: *Museli byť svedkami smrti svojich vlastných príbuzných a uniknúť Islamskému štátu – mnoho detí v utečeneckých táboroch v severnom Iraku je ťažko traumatizovaných.*

**H02:** *Das Niveau der Sprachgenauigkeit ist vom Fehlervorkommen in den untersuchten Kategorien nicht abhängig. (Fehlerkategorien: 1, 2, ..., 12)*

Genauigkeit & Kategorien von Fehlern	Valid N	Kendall Tau	Z	p-value
genauigkeit & 4# gruppe_kommunikation	2891	0.04	2.84	0.004543
genauigkeit & 7# gruppe_numerale_morphosyntax	2891	0.04	3.48	0.000494
genauigkeit&6#gruppe_pronominale_morphosyntax	2891	0.09	7.57	0.000000
genauigkeit & 1#gruppe_prädikation_kategorien	2891	0.11	8.82	0.000000
genauigkeit & 9# gruppe_wortstellung	2891	0.11	9.19	0.000000
genauigkeit & 8# gruppe_verbale_morphosyntax	2891	0.12	9.76	0.000000
genauigkeit & 3# gruppe_modalität	2891	0.13	10.20	0.000000
genauigkeit & 10# gruppe_andere_morphosyntax	2891	0.13	10.67	0.000000
genauigkeit & 5# gruppe_nominale_morphosyntax	2891	0.15	12.27	0.000000
genauigkeit & 2# gruppe_kongruente_kategorien	2891	0.21	17.01	0.000000
genauigkeit&11#gruppe_syntax_zusammengesetzte_sätze	2891	0.23	18.81	0.000000
genauigkeit & 12# gruppe_lexikale_semantik	2891	0.50	40.41	0.000000

**Tabelle 3 Parameterfreie Korrelation: Sprachgenauigkeit & Fehlerkategorien (Bánik et al., 2019, S. 163)**

Bei der Kategorie 4 (Tabelle 3) wird die Nullhypothese einer Zuverlässigkeit von 99 % festgestellt, d. h., das Niveau der Sprachgenauigkeit hängt vom Fehlervorkommen in der Kategorie 4 ab. Bei weiteren Kategorien (Tabelle 3) wird die Nullhypothese einer Zuverlässigkeit von 99.9 % abgelehnt.

Für sämtliche Kategorien (Tabelle 3) wurde eine direkt proportionale Abhängigkeit bestimmt, die Werte ändern sich gemeinsam in die gleiche Richtung (Kendall Tau > 0). Je schlechter das Niveau der Sprachgenauigkeit, desto höher ist das Fehlervorkommen.

Markante Unterschiede bestehen allerdings im Abhängigkeitsmaß. Während bei den Fehlerkategorien 4, 7 und 6 (Tabelle 3) das Abhängigkeitsmaß zwischen der

Sprachgenauigkeit und dem Fehlervorkommen nur trivial ist (Kendall Tau < 0.1), wurde bei der Fehlerkategorie 12 (Tabelle 3) ein hohes Abhängigkeitsmaß (Kendall Tau = 0.50) bestimmt. Bei restlichen Kategorien (Tabelle 3) wurde ein niedriges Abhängigkeitsmaß (Kendall Tau: 0.1 - 0.3) zwischen der Sprachgenauigkeit und dem Fehlervorkommen identifiziert.

Die Kategorie *Genauigkeit* (Accuracy) umfasst folgende Subkategorien: Hinzufügen (eines Elements/Ausdrucks) – Addition, falsche Übersetzung (Mistranslation), Auslassen (eines Elements/Ausdrucks) – Omission, ein nicht übersetztes Element/Ausdruck (Untranslated) (vgl. Munková & Vaňko et al., 2017, S. 69).

Bei der Bewertung der Genauigkeit steht neben der üblichen lexikalischen Semantik das Fehlervorkommen im Bereich der Syntax zusammengesetzter Sätze im Vordergrund. Die Abhängigkeit ist zwar gering, aber größer als bei der Bewertung der Sprache und Korrektheit.

Interessant ist jedoch, dass die Fehler im Bereich der Syntax zusammengesetzter Sätze relativ selten vorkamen. In der Statistik nehmen sie die fünfte Position mit einer Häufigkeit von 26,12 % ein (siehe Tabelle 3). Sie bilden eine homogene Gruppe zusammen mit der Wortstellung. Die Fehler im Bereich der lexikalischen Semantik stellen dagegen 75,75 % aller Fehler dar. Trotz der niedrigen Häufigkeit haben die Fehler einen relativ hohen prozentualen Anteil an der niedrigen Qualität der Übersetzung. Dies ergab sich aus der Bewertung der Post-Editoren. Unter den vier Kategorien der Bewertung – Genauigkeit, Terminologie, Stil und Angemessenheit – haben sie die zweite Stelle und in zwei Kategorien – Sprache und Korrektheit – haben sie die dritte Stelle besetzt. Anders gesagt: Fehler im Bereich der Syntax zusammengesetzter Sätze sind zwar nicht zahlreich vertreten, wenn sie aber vorkommen, wird dadurch die Richtigkeit der Übersetzung erheblich beeinträchtigt.

Die Syntax zusammengesetzter Sätze stellt im Allgemeinen einen der problematischsten Bereiche der maschinellen Übersetzung dar. Wie die vorherige Untersuchung bereits gezeigt hat, identifiziert das System komplizierte Satzkonstruktionen nicht immer richtig. Im Deutschen ist die Struktur eines einfachen Satzes ähnlich wie im Slowakischen, was aber die Beziehungen zwischen den Sätzen in einem Satzgefüge angeht, gibt es im Deutschen bestimmte Spezifika (siehe Beispiel). Ein größeres Fehlervorkommen in diesem Bereich hängt auch mit dem journalistischen Stil zusammen (vgl. Bánik et al., 2019, S. 162–166).

*Beispiel* (vgl. Bánik et al., 2019, S. 164):

AT: „Dann sitzt an einer Stelle, an der es ein Höchstmaß an Verantwortungsbewusstsein braucht, ein offenbar verantwortungsloser Mann,“ sagte Schulz in einem Interview.

MÜ: „Potom sediaci na mieste, kde to potrebuje väčší pocit zodpovednosti, čo je zrejme nezodpovedného človeka,“ uviedol Schulz v rozhovore.

PEMÜ: „Potom bude na mieste, ktoré si vyžaduje najvyššiu mieru zodpovednosti, sedieť nezodpovedný človek,“ uviedol Schulz v rozhovore.

## Schlussfolgerungen

Die hier präsentierten Forschungsergebnisse weisen auf tiefere Gründe bezüglich der Qualität bzw. Nicht-Qualität maschineller Übersetzung hin. Analysiert und besprochen wurden die Rahmenbewertungen in zwei (Sprache und Genauigkeit) von insgesamt sechs Qualitätskriterien (Sprache, Genauigkeit, Terminologie, Stil, Sprachfluss, Angemessenheit). Gezeigt wird, welche sprachlichen Fehler sich auf die Qualität der Übersetzung negativ auswirken. Als Textkorpus dienten deutschsprachige journalistische Texte.

Bei allen Bewertungen maschineller Übersetzung in den zwei vorgenannten Fehlerkategorien (Sprache und Genauigkeit) wurden die bedeutendsten Fehler im Bereich der lexikalischen Semantik bestimmt, gefolgt von Syntax zusammengesetzter Sätze und Grammatik. Fehler in der Syntax zusammengesetzter Sätze dominieren bei verringerter

Bewertung von Genauigkeit und Angemessenheit. Ein weiterer bedeutender Faktor verringerter Qualität der Übersetzung sind grammatische Fehler bezüglich der Satzglieder. Gemeint sind insbesondere Fehler im Bereich der Prädikation (Kongruenz, Verbum finitum, Subjekt-Prädikat-Beziehung). In zwei Kriterien der Bewertung belegen diese Fehler die 2. Position des Vorkommens (Sprache und Sprachfluss), in weiteren drei Kategorien die 3. Position (Genauigkeit, Terminologie, Stil) und nur bei der Bewertung der Angemessenheit haben diese Fehler eine niedrigere Position.

Als nächste problematische Erscheinung zeigt sich die nominale Morphosyntax, was jedoch bei der Übersetzung in eine flektierende Sprache (Slowakisch) zu erwarten ist. Dieser Fehlerbereich ist bei allen Bewertungen relativ ausgewogen vertreten. Ein bedeutender Unterschied zwischen der Fehlerhäufigkeit in den Übersetzungen und ihrem Vorkommen bei der geringen Bewertung der Übersetzung weist die Kategorie der Syntax zusammengesetzter Sätze auf. Diese Fehler kommen selten vor, wenn diese jedoch auftreten, beteiligen sie sich an der geringen Bewertung der Übersetzung markant (vgl. Bánik 2019 et al., 191–192).

Wie erlernen nun Maschinen das Übersetzen? Genauso wie ein Mensch. Auch eine Maschine muss sowohl der Ausgangs- als auch der Zielsprache mächtig sein, sie muss sich Wörter, Wortverbindungen, Sätze und Texte aneignen und vor allem merken. Das Gedächtnis einer Maschine ist im Unterschied zum Menschen eigentlich unbeirrbar. Warum irrt die Maschine dann mehr als der Mensch? Die Maschinen können den komplizierten Prozess des Erlernens einer Sprache (Mutter- oder Fremdsprache) nicht nachahmen. Auch wenn heute eine Maschine mittels Methoden maschinellen Lernens erlernen kann, wie man eine Katze von einem Pferd unterscheidet, indem man der Maschine das jeweilige Tier durch typische Merkmale definiert, ist es mit der Sprache nicht so einfach. Die Sprache ist ein vielschichtiges Phänomen, welches nicht nur geradlinig aufgebaut ist (Katze gleich vier Beine, Augen, spitze Ohren, Fell und Schnurrhaare). Durch die Sprache wird nicht nur die denotative Bedeutung ausgedrückt, sondern auch viele Konnotationen und Beziehungen zwischen den einzelnen Elementen der Sprache, die eine Maschine (noch) nicht erkennen bzw. wiedergeben kann. Dies trifft schließlich auch auf die maschinelle Übersetzung zu (Prunč, 2002, S. 33):

Zwischen den einzelnen Sprachen, so nahm man zunächst an, gebe es stets eine Art Symmetrie, aufgrund welcher auch ihre jeweiligen Elemente einander zuordenbar wären. Man bräuchte, so glaubte man, „nur“ noch das Problem der Mehrdeutigkeit in realen Texten zu lösen und die einzelnen Transferelemente zu isolieren. Die Ernüchterung folgte sehr bald. Man musste erkennen, dass zwischen den einzelnen Sprachen eben keine Symmetrie herrscht, sondern dass die Asymmetrie der Sprachen ein Faktum ist, dessen man sich zuwenig bewusst war.

Die Sprache und die damit verbundene Übersetzung sind zu „unregelmäßig“, als dass ein absolut verlässlicher Algorithmus für eine Maschine programmiert werden kann. Eine Maschine müsste demnach allzu viele Variablen erlernen, um sprachliche Nuancen zu erkennen, was letztendlich auch statistisch nicht möglich ist. Eine Maschine verlangt genau definierte Begriffe, die jedoch wissenschaftlich nur schwer qualifizierbar sind. Obgleich sich der Mensch immer mehr auf Technologien verlässt, darf nicht vergessen werden, dass Maschinen nach wie vor spezifische vom Menschen erteilte Anweisungen erfordern. Da Maschinen zu Fehlern neigen und sich durchaus irren können, darf ihnen nicht nur „kopflös“ vertraut werden.

### **Abkürzungen:**

e. Ü.	eigene Übersetzung
AT	Ausgangstext
MÜ	Maschinelle Übersetzung

PEMÜ post-editierte maschinelle Übersetzung

### Literaturverzeichnis:

- BÁNIK, T. et al. (2019): *Wie irrt die Maschine? Probleme maschineller Übersetzung*. Hamburg: Verlag Dr. Kovač.
- BURCHARDT, A. – PORSIEL, J. (2017): Was kann maschinelle Übersetzung und was nicht? In: A. Burchardt (ed.): *Maschinelle Übersetzung. Grundlagen für den professionellen Einsatz*. Berlin: BDÜ Weiterbildungs- und Fachverlagsgesellschaft mbH, s. 11–17.
- DANCETTE, J. (1992): Modèles sémantique et propositionnel de l'analyse de la fidélité en traduction. In: *Meta*, 37/3, s. 440–449.
- DOHERTY, S. (2012): *Investigating the Effects of Controlled Language on the Reading and Comprehension of Machine Translation Texts: A Mixed-Methods Approach*. PhD Thesis. Dublin, Ireland: Dublin City University.
- DOHERTY, S. (2016): The impact of translation technologies on the process and product of translation. In: *International Journal of Communication*, 10, s. 947–969. [Cit. 2019-20-02.] Dostupné na internete: <<https://ijoc.org/index.php/ijoc/article/view/3499/0>>
- DOHERTY, S. – O'BRIEN, Sh. (2014): Assessing the usability of raw machine translated output: a user-centered study using eye tracking. In: *International Journal of Human-Computer Interaction*, 30/1, s. 40–51.
- DOHERTY, S. – O'BRIEN, Sh. – CARL, M. (2010): Eye tracking as an MT evaluation technique. In: *Machine Translation*, 24/1, s. 1–13.
- GASPARI, F. – ALMAGHOUT, H. – DOHERTY, S. (2014): A survey of machine translation competences: Insights for translation technology educators and practitioners. In: *Perspectives: Studies in Translatology*, 22/4, s. 1–26.
- GOUADAC, D. (1989): *Le traducteur, la traduction et l'entreprise*. Paris: Afnor.
- HEWSON, L. (1995): Detecting cultural shifts: some notes on translation assessment. In: I. Mason (ed.): *Cross-Words: Issues and Debates in Literary and Non-Literary Translating (L3 – Liège language and literature)*, s. 101–108.
- HOUSE, J. (1981): *A Model for Translation Quality Assessment*. Tübingen: Narr.
- HOUSE, J. (1997): *Translation Quality Assessment: a Model Revisited*. Tübingen: Narr.
- HOUSE, J. (2001): Übersetzen und Deutschunterricht. In: G. Helbig – L. Götze – G. Henrici – H. J. Krumm (eds.): *Deutsch als Fremdsprache. Ein internationales Handbuch*. Berlin: de Gruyter, S. 258–268.
- HOUSE, J. (2009): *Translation*. Oxford: Oxford University Press.
- HOUSE, J. (2015): *Translation Quality Assessment: Past and Present*. London, NY: Routledge.
- HUTCHINS, J. (2005): Current commercial machine translation systems and computer-based translation tools. [Cit. 2019-01-09.] Dostupné na internete: <<http://www.hutchinsweb.me.uk/IJT-2005.pdf>>
- KUSSMAUL, P. (1995): *Training the Translator*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- LAROSE, R. (1989): *Théories contemporaines de la traduction*. Sillery (Québec): Presses de l'Université du Québec.
- MUNKOVÁ, D. (2013): *Prístupy k strojovému prekladu: modely, metódy a problémy strojového prekladu*. Nitra: UKF Nitra.
- MUNKOVÁ, D. – VAŇKO, J. et al. (2017): *Mýliť sa je ľudské (ale aj strojové). Analýza chýb strojového prekladu do slovenčiny*. Nitra: Univerzita Konštantína Filozofa v Nitre.
- MUNKOVÁ, D. – WREDE, O. – ABSOLON, J. (2019): Vergleich der menschlichen, maschinellen und Post-Editing-Übersetzung aus dem Slowakischen ins Deutsche mittels automatischer Evaluation. In: *Zeitschrift für Slawistik*, 64/2, s. 231–261.
- MÜGLOVÁ, D. (2009): *Komunikácia. Preklad. Tlmočenie. Alebo prečo spadla Babylonská veža?* Nitra: Engima.
- NEWMARK, P. (1991): *About Translation: Multilingual Matters*. Philadelphia, Adelaide: Multilingual Matters Ltd.



- NORD, Ch. (1991): *Text Analysis in Translation*. Amsterdam: Rodopi.
- PAPINENI, K. – ROUKOS, S. – WARD, T. – ZHU, W. (2002): BLEU: a method for automatic evaluation of machine translation. In: *ACL*, 2, s. 311–318.
- PRUNČ, E. (2002): *Einführung in die Translationswissenschaft. Band 1. Orientierungsrahmen*. Graz: Universität Graz, Institut für Theoretische und Angewandte Translationswissenschaft.
- PYM, A. et al. (2012): *The Status of the Translation Profession in the European Union*. London, New York: Anthem Press.
- QUAH, Ch. K. (2006): *Translation and Technology*. New York: Macmillan.
- SNELL-HORNBY, M. (1995): *Translation Studies: An Integrated Approach*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- SNOVER, M. et al. (2006): A study of translation edit rate with targeted human annotation. In: *Proceedings of Association for Statistical Machine Translation in the Americas 2006*. [Cit. 2020-25-06.] Dostupné na internete: <[https://www.cs.umd.edu/~snover/pub/amta06/ter\\_amta.pdf](https://www.cs.umd.edu/~snover/pub/amta06/ter_amta.pdf)>
- THURMAIR, G. (2005): Automatic means of MT evaluation. In: *ELRA-HLT Evaluation Workshop 2005*. [Cit. 2020-25-06.] Dostupné na internete: <<http://www.mt-archive.info/HLT-2005-Thurmair.pdf>>
- VANĀKO, J. (2017): Kategoriálny rámec pre analýzu chýb strojového prekladu. In: D. Munková – J. Vaňko et al. (eds.): *Mýliť sa je ľudské (ale aj strojové). Analýza chýb strojového prekladu do slovenčiny*. Nitra: Univerzita Konštantína Filozofa v Nitre, s. 83–100.
- VILAR, D. – XU, J. – D'HARO, L. F. – NEY, H. (2006): Error analysis of statistical machine translation output. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-06)*, s. 697–702.
- WAY, A. (2018): Quality expectations of machine translation. In: J. Moorkens – Sh. Castilho – F. Gaspari – S. Doherty (eds.): *Translation Quality Assessment: From Principles to Practice*. Cham: Springer, s. 159–178.
- WILLIAMS, M. (1989): Creating credibility out of chaos: the Assessment of translation Quality. In: *TTR*, 282, s. 13–33.

## Summary

### Assessing machine translation quality in translating journalistic texts from German into Slovak

This study focuses on the identification and classification of machine translation errors in translating journalistic texts from German into the highly inflectional Slovak language. The aim of the research is to evaluate the quality of machine translation via product analysis, namely using error analysis. We used the framework for the identification and classification of machine translation errors devised by the Vaňko and MQM-DQF error typology. We analyse two issues of error defined in the MQM-DQF error typology: accuracy, how precisely the target text represents the content of the source text, and fluency, linguistic (grammatical) correctness of the target text. We discovered that the most significant errors are in the area of lexical semantics, followed by complex sentence syntax and grammar.

*Die im vorliegenden Beitrag diskutierten Ergebnisse wurden im Rahmen eines Forschungsprojektes gewonnen, das von den Agenturen APVV (Vertrag Nr. APVV-18-0473) und VEGA (Vertrag Nr. VEGA-1/0809/18) unterstützt wird.*