



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue Moderne per la Comunicazione e la Cooperazione Internazionale
Classe LM-38

Tesi di Laurea

A Study on Automatic Machine Translation Tools: A Comparative Error Analysis Between DeepL and Yandex for Russian-Italian Medical Translation

Relatore

Prof. Giorgio Maria Di Nunzio

Correlatrice

Prof.ssa Viviana Nosilia

Laureanda

Giulia Cambedda

n° matr.1205696 / LMLCC

Anno Accademico 2020 / 2021

TABLE OF CONTENTS

Introduction	1
1. THE EVOLUTION OF MACHINE TRANSLATION: FROM THE ORIGINS TO NEURAL MACHINE TRANSLATION	3
1.1. The pioneers of Machine Translation	3
1.2. A focus on Natural Language Processing	7
1.3. The ALPAC report	10
1.4. Rule-Based Machine Translation	12
1.5. Example-Based Machine Translation	14
1.6. Statistical Machine Translation	16
1.7. Neural Machine Translation	17
2. A COMPARISON BETWEEN DEEPL AND YANDEX: PURPOSES AND SCOPE OF THE RESEARCH	21
2.1. Translating the Russian language of medicine	21
2.2. Machine Translation Evaluation: an open question	26
2.3. Performing a comparative error analysis	37
2.4. An overview of DeepL Translate.....	39
2.5. An overview of Yandex Translate	41
3. A COMPARATIVE ERROR ANALYSYS BETWEEN DEEPL AND YANDEX	44
3.1. The biological therapy in the COVID-19 era.....	44
3.2. Clinical management of children with a disease caused by the new coronavirus infection (SARS-CoV-2).....	54
3.3. Coronavirus SARS-Cov 2: complexities of the pathogenesis, search for the vaccines, and future pandemics	68
3.4. Coronavirus was imported into Russian at least 67 times	89
3.5. Unknown lethality - Why we do not know the real extent of COVID-19.....	103
3.6. COVID-19 mortality rate - A demographer's perspective on the statistics of causes of death in Russia and worldwide	116
4. QUANTITATIVE ANALYSIS OF THE RESULTS	132
4.1. DeepL's translation performance	133
4.2. Yandex's translation performance.....	140
4.3. A comparison between DeepL's and Yandex's translation performances.....	150

4.4. BLEU metric's evaluation	153
Conclusion	157
Резюме	160
REFERENCES	176

Abstract

L'obiettivo della tesi qui proposta è quello di condurre un'analisi delle attuali capacità traduttive di due sistemi di traduzione automatica neurale al centro del panorama traduttivo degli ultimi anni, DeepL e Yandex, per quanto riguarda la traduzione medica dal russo all'italiano. Ai fini della nostra ricerca, sono stati selezionati tre articoli medici specialistici e tre articoli medici di carattere divulgativo in lingua russa sul tema della pandemia da coronavirus. La scelta di tale tema è stata orientata dalla volontà non solo di prendere in considerazione articoli recentemente pubblicati, ma anche di analizzare il comportamento dei due sistemi di traduzione automatica di fronte al fenomeno linguistico causato dalla pandemia da coronavirus, che ha comportato l'introduzione, tanto nella terminologia medica, quanto nel linguaggio quotidiano, di una grande quantità di termini precedentemente inesistenti o il cui utilizzo era limitato alla comunicazione scientifica professionale.

Il primo capitolo è dedicato ad una breve panoramica sullo sviluppo storico della traduzione automatica, a partire dalle sue origini, fino al recente sviluppo della traduzione automatica neurale. Particolare attenzione viene rivolta alla descrizione dei diversi approcci alla traduzione automatica proposti nel corso degli anni, al fine di delineare le principali caratteristiche e funzionalità dei due sistemi di traduzione automatica analizzati.

Nel corso del secondo capitolo vengono poste le basi teoriche della nostra ricerca. Inizialmente, viene fornita una descrizione dei principali aspetti del linguaggio medico russo e delle sue correlazioni storiche con il mondo della traduzione. Successivamente, vengono illustrati i principali metodi di valutazione dei sistemi di traduzione automatica, sottolineando la distinzione, non sempre ben definita, tra metodi di valutazioni automatici e non automatici. Particolare attenzione viene dedicata alla descrizione dell'analisi comparativa degli errori, il metodo di valutazione utilizzato ai fini della nostra ricerca. Ne vengono quindi stabilite la procedura e le specifiche categorie di errori, scelte sulla base delle caratteristiche linguistiche dei testi di partenza. Infine, le principali caratteristiche e funzionalità di DeepL e Yandex vengono descritte.

Nel terzo capitolo, l'analisi comparativa degli errori viene condotta nella seguente modalità: i titoli e gli abstract dei testi specialistici, così come i titoli e i primi paragrafi dei testi divulgativi, vengono divisi in frammenti di uguale lunghezza, e successivamente tradotti utilizzando DeepL e Yandex. Le traduzioni ottenute vengono inserite in una tabella, insieme al corrispondente frammento del testo originale. Gli errori contenuti in ogni frammento di traduzione vengono quindi individuati, brevemente discussi, e associati ad una o più delle categorie traduttive precedentemente selezionate.

Nel quarto capitolo, i frammenti contenenti gli errori individuati durante l'analisi comparativa vengono mostrati all'interno di una tabella, nella quale vengono presentati accanto al corrispondente frammento originale e a quello proveniente da una traduzione umana appositamente realizzata e utilizzata come traduzione di riferimento. Successivamente, al fine di garantire una chiara e immediata interpretazione dei dati, i risultati ottenuti a seguito dell'analisi comparativa degli errori vengono mostrati attraverso una serie di grafici. I grafici mostrano una valutazione delle capacità traduttive di DeepL e Yandex, in relazione alle due differenti tipologie testuali analizzate. I dati ottenuti mostrano che entrambi i sistemi di traduzione automatica commettono una quantità significativa di errori nella resa della struttura sintattica e del lessico dei testi di partenza, così come nell'uso degli articoli. Tuttavia, DeepL risulta ottenere risultati migliori rispetto a Yandex, sia nella traduzione dei testi specialistici sia in quella dei testi divulgativi. Infine, le traduzioni realizzate dai due sistemi di traduzione automatica vengono valutate utilizzando il sistema di valutazione automatica BLEU (Bilingual Evaluation Understudy), che calcola la percentuale di somiglianza testuale tra le traduzioni realizzate da DeepL e Yandex e le traduzioni umane di riferimento. La comparazione dei risultati ottenuti dall'analisi comparativa degli errori e dal sistema di valutazione BLEU rivela una generale corrispondenza tra i due differenti sistemi di valutazione, con alcune eccezioni. Inoltre, fornendo una valutazione specifica di ogni articolo, il sistema BLEU ci permette di analizzare le capacità traduttive dei due sistemi di traduzione automatica in relazione ad ognuno dei testi analizzati.

La ricerca condotta ci ha permesso di confermare il significativo sviluppo, in termini di qualità e efficienza, che ha caratterizzato i sistemi di traduzione automatica neurale negli ultimi anni. Sono stati tuttavia rilevati una serie di punti deboli nelle capacità traduttive di entrambi i sistemi di traduzione automatica qui analizzati. Costante ricerca

si rende infatti necessaria al fine di raggiungere standard traduttivi che possano avvicinare la traduzione automatica a quella umana. Una possibile ricerca futura può essere orientata, partendo dai risultati qui ottenuti, ad un'analisi linguistica delle categorie di errore analizzate, al fine di attribuire loro un grado di rilevanza nell'ambito della traduzione medica dal russo all'italiano. In questo modo, sarà possibile individuare gli errori che maggiormente impediscono ai sistemi di traduzione automatica analizzati nel corso della nostra ricerca di raggiungere gli standard qualitativi desiderati e agire di conseguenza per apportare i dovuti miglioramenti. Senza dubbio, questa ricerca mette in luce la necessità di una ridefinizione del ruolo del linguista/traduttore, che, con le sue conoscenze linguistiche, prenderà parte in misura sempre maggiore all'attuale e futuro sviluppo della traduzione automatica.

Introduction

The terrific and continuous development that Machine Learning and Natural Language Processing have been experienced over the last years undoubtedly is before our very eyes in the daily life. Suffice is to think about Machine Translation, speech-recognition tools, customer-support chatbots, and virtual assistants, just to mention a few, which we end up using at least once a day in order to perform a set of diverse ordinary tasks. Among the numerous Natural Language Processing applications, Machine Translation not only constitutes, having been investigated and developed for a relatively long time, a forerunner in the field but remains also of considerable relevance to this day, continuously posing new challenges that constantly involve the most varied sorts of professionals, from linguists to engineers, from IT experts to translators. Machine Translation, indeed, has played and still plays a crucial role in the transmission of knowledge of various kind, which has triggered and has been triggered by the last-decades globalized world. This sounds particularly true for medicine, as the sharing of knowledge undoubtedly constitutes the most effective way to spread the latest research results and conduct collaborative studies despite the geographical distance and the language barriers dividing one country from another. Although, as mentioned above, the enormous growth of Natural Language Processing applications, including Machine Translation, has become increasingly visible, one may wonder which level of development has been reached so far. We can easily say this represents a pointed question, as evaluation constitutes the starting point for a successful attempt of enhancement. For the purposes of the study that we are presenting in the following pages, we have therefore chosen to analyse and evaluate the translation performances of two prominent Neural Machine Translation tools, namely DeepL and Yandex, with regard to Italian-Russian medical translation. More specifically, we have selected three highly specialized and three popular-science medical Russian articles concerning coronavirus pandemic. Such a choice is justified by the willingness not only to analyse recent documents but also to investigate the particular linguistic implications of the occurrence of an unexpected and dramatic global event that has been totally monopolizing the political, social, and scientific discourse. 2020's coronavirus pandemic outbreak has indeed introduced in every-day

communication a whole set of terms whose use was previously limited to the language of science, as well as coined a group of new terms, which all of a sudden entered the boundaries of scientific terminology. We have considered this existing linguistic phenomenon as a proper condition to test Machine Translation Tools behaviour and performances. Our analysis, which will be presented in more detail throughout the following chapters, shed light on the strengths as well as the weaknesses of present-day Machine Translation tools with regard to a language pair, namely Russian-Italian, which, because of its structural discrepancies at all linguistic levels, can be considered sufficiently challenging. Apart from that, it is however intended to highlight two additional concepts, namely the importance of careful evaluation and the ever increasingly close collaboration between professionals coming from different fields of study as basis of Machine Translation performances improvement. On the one hand, Machine Translation Evaluation, although highly controversial in its methods and criteria, may pave the path for effective and necessary adjustments of the programs under examination. On the other hand, only a balanced synergy between linguists and software developers, representing the two main souls of Machine Translation tools, may provide functioning and up-to-date translation systems. This last aspect directly leads to the burning question of the current and future role of the human translator against the background of Artificial Intelligence's ever increasing development. Over the last years, the human translator, as professional figure, has indeed been often considered bound to disappear under the threat of scientific evolution. However, the radical changes affecting Machine Translation research and development may lead to a complete redefinition of his tasks and required skills, on which the translators of tomorrow need to build their professional training.

1. THE EVOLUTION OF MACHINE TRANSLATION: FROM THE ORIGINS TO NEURAL MACHINE TRANSLATION

Over the years, Machine Translation (MT) has undergone rapid growth, mainly due to extensive research in the field and an increasing worldwide interest in computer science. As illustrated in Figure 1, its history dates back over ninety years and involves experts in different disciplines, including linguistics, engineering, mathematics, and computer science. In the present chapter, we will briefly outline the stages of development of Machine Translation, from its origins to the latest achievements in the field of Neural Machine Translation.

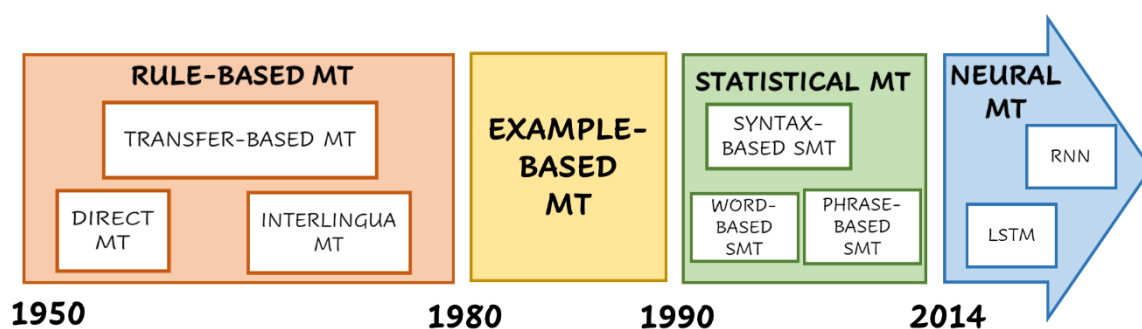


Figure 1 Timeline of Machine Translation development

1.1. The pioneers of Machine Translation

In 1933, the French engineer Georges Artsrouni and the Soviet scientist Petr Trojanskij were granted a patent for the invention of two mechanical devices that can be considered, at different extents, the first precursors of Machine Translation systems. Artsrouni designed the so-called *mechanical brain*, a multipurpose machine that, among its several applications, could also be used for translation. The device had four constituent parts, namely a memory, containing a set of words in four different languages, a keyboard to insert the input word, a search mechanism, aimed at detecting the corresponding word in the memory, and an output mechanism, which displayed the input word, its approximate word-to-word translation into the needed language and some other linguistic information. Although Artsrouni's proposals aroused considerable interest at the Paris Universal Exhibition in 1937, they were not seriously considered until 1950, when MT

research concerning the possibility to use computers for translation purposes began in France. Unlike the *mechanical* brain, which had a number of different applications and was comparable to a mere mechanized dictionary, Trojanskij's *machine for selecting and typing words when translating from one language into another or several others simultaneously*¹ was specifically designed for translation. Indeed, the Soviet scientist made a considerable step forward and paved the path for further studies in the field of Machine Translation. Firstly, as mentioned above, he designed a device whose only aim was the translation. His translation machine consisted of a desk with a large, perforated belt, which was positioned on several apertures. It worked as follows: by moving the belt, the source text word was located in the aperture that showed its corresponding word in the target language. Afterward, the source word was typed using a code indicating its main linguistic features, which was initially based on Esperanto and named by Trojanskij *signs for logical parsing*², and its combination with the target word was photographed in a tape. As final steps, a provisional translation was made by analyzing the tape and then reviewed by a bilingual editor, who provided the ultimate version. Secondly, unlike Artsrouni, Trojanskij theorized a real translation process, divided into three stages, and, by proposing the use of *signs for logical parsing*, introduced the concept of interlingua and its possible role as translation intermediary between the source and the target language. Finally, the Soviet scientist anticipated the idea of post-editing and stressed the need for a bilingual editor, who was in charge of reviewing the translation performed by the machine. Nonetheless, Trojanski's patent was rejected by the USSR Academy of Sciences in 1939 and the scientist died leaving a set of inspirational, although ignored ideas (Hutchins, 2004). Over the following decades, significant development of computer science and the appearance of the first computers triggered an increase of interest in the field of Machine Translation and the awareness of the concrete possibility of using them in the translation process.

“I have wondered if it were unthinkable to design a computer which would translate” (Weaver, 1947). In his letter to Professor Norbert Wiener of Massachusetts Institute of Technology on March 4, 1947, Warren Weaver, at the time responsible at Rockefeller Foundation, mentioned the possibility of making use of a computing device

¹ Машина для подбора и печатания слов при переводе с одного языка на другой или на несколько других одновременно (Trojanskij, 1935)

² Условные обозначения для логического разбора (Trojanskij, 1935)

for translating from one natural language to another. Despite the discouraging response from Wiener, who considered the idea of a mechanized translation premature and incapable to properly convey all the precise nuances of meaning, Weaver conducted pioneering research in the field, acquiring the worldwide reputation as one of the precursors of Machine Translation. Indeed, the mathematician was not only concerned about the detrimental effects of linguistic barriers on international relationships and, consequently, the importance of translation, but also aware of the fact that a correct transmission of multiple meanings represented a major issue in translation. Starting from the assumption that, since invented and used through the same human organs, all the languages but few exceptions share some basic features, Weaver introduced the concept of the logical structure of languages and remarked its relevance as a starting point of the research in the field of mechanized translation. Hence, he suggested to focus first on the traits that the two languages under consideration have in common, and then, having obtained a more general view, try to find the most effective way to properly transfer the intended meaning from the source to the target language. Together with Weaver, this was of great concern to many other scholars of the same time, including Dr. Andrew D. Booth, who, throughout their career, investigated the possible practical applications of computers to the translation process, limiting the scope of their research to the mechanization of dictionaries. Unlike his contemporaries, Warren Weaver felt the need to design, in cooperation with experts in computer science and linguistics, a computer able to translate different kinds of texts (Weaver, 1949). One of the most significant contributions that the scientist made to the further studies concerning Machine Translation doubtless is his strong will to overcome the word-to-word translation, which seemed to be, at that time, the only possible option with regards to mechanized translation, although it was hardly capable to translate the most technical and repetitive texts and to provide satisfactory results in terms of quality and reliability. Weaver stressed the relevance of context and explored the possibility of taking into account not only the word that has to be translated but also the words or the group of words on either side. According to his studies, the proper number of side words to be considered depends on the distinctive features of each language and on the degree of ambiguity that usually affects the different elements of a sentence, namely nouns, adjectives, adverbs, articles, and prepositions (Weaver, 1949). Warren Weaver's considerations laid the theoretical foundations for some of the major

issues in the field of Machine Translation, allowing further scholars to conduct in-depth research, within a constantly evolving international context. Moreover, the mathematician contributed to “A Mathematical Theory of Communication”, published by Claude E. Shannon in the *Bell System Technical Journal* in 1948 and considered a turning point for Machine Translation studies.

In “A Mathematical Theory of Communication”, the communication process is extensively analyzed, and particular attention is devoted to defining the concept of information and the different stages of its transmission from the sender to the receiver. According to Claude E. Shannon, information can be defined as any kind of message that is intended to be delivered from a sender to a receiver, regardless of whether it has a complete meaning or not (Shannon, 1948).

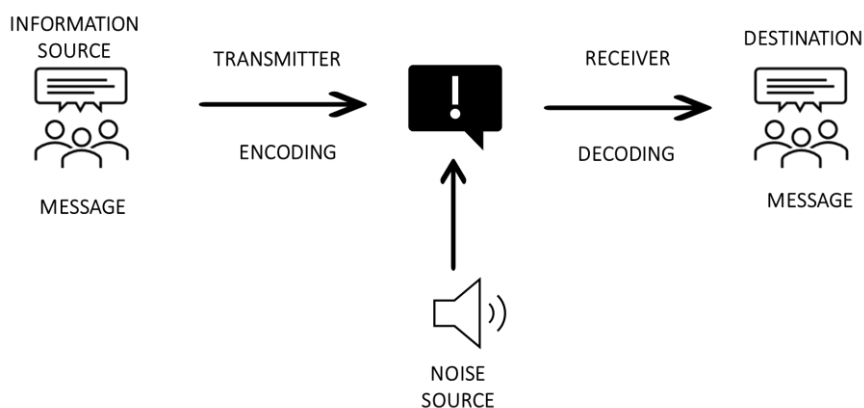


Figure 2 The communication process

Figure 2 illustrates the communication process described by Shannon: the information is generated from an information source and then passed to a transmitter, which is responsible for encoding the message and sending it to the channel. Once arrived at the receiver, the message is decoded and then sent to the destination. As mentioned before, in the article, any type of information is considered, be it continuous, such as music and an oral speech, or not continuous, i.e. consisting of a set of discrete symbols, like in written language. Moreover, the mathematician strongly remarks on its logarithmic constitution as a major feature in the communication process. By its very nature, when transmitted, information can indeed be expressed in a set of new symbols, which have to

be chosen according to the channel of communication. The concept of information coding formulated in this article forms the very core of information technology and constitutes one of the first attempts to illustrate how information can be processed by a computing device. In addition, another major issue regarding information processing was investigated, namely its statistical nature. Given a text, Shannon studied a sequence of letters, a word, or even a sentence, and aimed to estimate the likelihood of the adjacent letters, words, or sentences. According to the results, since the generation of language is ruled by probability, an effective solution to the problem can be found in the probability distribution, a mathematical function able to describe the likelihood of occurrence of different outputs in an experiment (Shannon, 1948). As mentioned above, Claude E. Shannon provided a significant contribution to further studies concerning computational linguistic, as well as the mathematical fundamental principles on which the entire field of machine learning is based. His speculations can be indeed easily considered as the foundations of natural language processing, among whose main applications there are machine translation and machine learning.

1.2. A focus on Natural Language Processing

Natural Language Processing (NLP) can be defined as a branch of computer science devoted to training computers to understand, process, and generate natural language. In other words, it concerns the implementation of diverse systems aimed at allowing computers to deal with natural languages as humans do, to perform a variety of tasks (Reshamwala et al., 2013). Given the enormous complexity of natural language, it does not come as a surprise that NLP constitutes even nowadays one of the most challenging areas of research in the field of computer science. We can easily say that what is extremely difficult when dealing with natural language processing is, broadly speaking, variety. Variety regards the three core elements of a general NLP system, namely the language itself, the available input, and the desired and usually demanded by the market output. Starting from the most general level, natural language is highly diverse. Indeed, since it is created and acquired by humans in their natural environment by the time they are children, it develops a set of characteristics that are not easy to be processed by computers.

Not only natural languages are organized in a number of levels, which continuously interact with each other, but they also evolve and change over time. Besides, one natural language may include several different varieties, which differ according to the specific place where the same language has developed (Bates, 1995).

As mentioned above, NLP's main aim is to train computers to understand, process, and generate natural language as they were humans. Since humans, in their every-day interactions, use every level in which their natural language is organized, the same an optimal NLP system is expected to do. By making a considerable simplification, a general NLP system works as follows: given an input, i.e a natural language document, be it spoken or written, the system has to analyze and process it in order to provide the desired output. Analyzing the input means extracting its deep sense, which depends on the interactions between all the levels of the natural language at hand. Therefore, a general NLP system processes the input according to the so-called *levels of language approach*. It may start from phonology, concerned with the interpretation of speech sounds, then move to morphology, which analyses a word's constituent parts in order to define its linguistic features and its interactions with the other words of the sentence. Syntax aims at establishing the role of each word in the sentence. A syntactic analysis can be conducted either by focusing on the grammar rules of a particular language or by means of a procedure called *parsing* and consisting of representing the syntactical structure of a sentence as a tree, which branches according to the syntactical interactions among its components. Syntax is generally followed by semantics, which investigates the meaning of each word that constitutes the sentence and, consequently the meaning of the sentence itself. Semantics significantly differs from pragmatics, the next level. Pragmatics indeed tries to extract the actual meaning of a word or a sentence within the overall context. Especially with regard to long written documents, it is rather difficult for an NLP system to detect the general context of an input. This is mainly due to the fact that the context elements are scattered throughout the whole document, at a great distance from the fragments of texts under analysis (Khurana, Koli, Khatter and Singh, 2017). An incorrect understanding of the context leads to ambiguity, not only one of the most difficult barriers in human communication but also a major impediment that prevents computers from correctly understanding natural language. Ambiguity may especially affect syntax, semantics, and pragmatics and most of the current research in the field is devoted to

implementing disambiguation techniques in order to enable NLP systems to detect the context and to resolve the input ambiguity. Generally speaking, how many and which language levels are processed by an NLP system highly depends on the input and the type of output it is intended to provide. Since both NLP systems' input and output are extremely diverse, not always all the language levels are considered, but just the ones that are specifically suitable for the initial piece of information, i.e. the input and for the task or the tasks to be accomplished, i.e. the desired output (Bates, 1995).

An NLP system may process an incredible variety of language information, which ranges from speeches to complete sentences or fragments of sentences. In addition, its input can be perfectly grammatical or not, contain capitalization and punctuation, or lack these elements. The same applies to output. NLP systems cover indeed a wide range of different applications, including speech recognition, information retrieval, chatbox, virtual assistants, machine translation, text extraction, text classification, and text summarization, just to mention a few (Khurana et al., 2017). As a consequence, a great deal of diversely designed NLP systems exist and, at present, there is not a common training method applicable to all of them. Therefore, we will briefly illustrate the most widespread training methods, with the knowledge that this is an extremely diverse and ever-evolving field of research.

NLP systems are usually trained on the basis of great amounts of natural language data, from which grammatical rules are extracted and used to develop the corresponding algorithms. In order to accomplish this not easy task, several methods can be adopted, such as sentence segmentation, which divides the sentences of a text usually according to the punctuation marks; tokenization, which splits each sentence into smaller parts carrying a syntactical meaning, called tokens; and Part-of-Speech Tagging, which assigns a word-class to each token (Bates, 1995). In this respect, an innovative approach called Distributed Representation can represent many linguistic features concerning words or phrases of a sentence through vectors, which can be thereafter composed in order to represent the meaning of the whole sentence. As for the NLP applications that involve the generation of natural language, increasing importance is being drawn to the concept of the statistical nature of language introduced by Claude E. Shannon in "A mathematical Theory of communication". In particular, starting from the assumption that the generation

of natural language is ruled by probability, N-grams models are implemented to estimate the likelihood of n-consecutive words (Mikolov et al., 2013).

In the 1950s, initial studies in the field of NLP started, and Machine Translation, which at the time stimulated considerable worldwide interest, was one of the first applications they were devoted to. Over the years, extensive research concerning NLP has provided Machine Translation with solid theoretical foundations on which it developed, achieving increasingly impressive results.

1.3. The ALPAC report

The Georgetown-IBM experiment, jointly conducted by IBM and Georgetown University in New York's IBM headquarters on 7 January 1954, and consisting of a completely automatic translation of more than sixty Russian sentences into English undoubtedly produced a considerable optimistic view concerning future perspectives and potentialities of Machine Translation. In the following days, it was indeed reported by the most popular American and European newspapers as the starting point of a constant and successful development of increasingly efficient Machine Translation systems able to translate great amounts of linguistic data in multiple languages (Hutchins, 2005). This led to an increase in research effort in the field and to the implementation of the first machine translation systems (Hutchins). Nonetheless, the submission of the report "Language and Machines", later known as ALPAC report, by the Automatic Language Processing Advisory Committee (ALPAC) in 1966, questioned the utility of Machine Translation in general, and consequently halted for some decades its development, especially in the US, URSS, and some European countries. In the report, the Automatic Language Processing Advisory Committee analyzed several areas concerning current Machine Translation research in the US and provided its own opinions and suggestions about possible paths to take in the future.

At the beginning of the report, the current need for translation in the US was investigated. ALPAC expressed considerable disagreement with whom, who overestimated the need for translation in the US on the basis of the large amount of scientific non-English written articles that were published at the time all around the world. According to the committee, translating all the foreign articles was indeed neither

essential nor worthwhile. On the contrary, providing good translations of the abstracts and then making the full translation of the most interesting articles available upon specific request could have been not only a good compromise but also money and time-saving practice. As for translators, the idea of a national shortage of translators was firmly denied. By analyzing the governmental data concerning the number of professional translators, ALPAC wanted to demonstrate that there was not a lack of translators, but a lack of authorized positions for translators instead. After having remarked the absence of any compelling need to reinforce the translation service in the US, particular attention was devoted to analyzing the existing state of development of Machine Translation.

The committee reflected on the current lack of machine translation systems able to follow the whole translation process. The post-editing phase was indeed left to human translators and it was significantly long and costly. Although at the time there were no reliable evaluation methods for translations, ALPAC subsequently compared machine translation with human translation and stated that not only human translation was able to provide higher quality translations but it was also cheaper than designing high-quality and reliable machine translation systems. As a consequence, the committee remarked the complete lack of sense in conducting further studies aimed at developing Machine Translation systems and proposed to focus on other fields of research, namely linguistics, computational linguistics, and computer-aided translation.

Although Machine Translation research conducted up to that time was judged completely pointless if applied to the design of Machine Translation systems, it permitted to develop several fields of study that deserved special attention. The idea of using computers to manage linguistic content led indeed to the emergence of computational linguistics and computer-aided translation. According to the committee, computational linguistics could have been beneficial for a comprehensive study of natural languages, whose deeper understanding might have significantly boosted the implementation of new techniques concerning language education, international communication, and information retrieval. As for computer-aided translation, several experiments had recently demonstrated that it was able to provide better results, with regard to quality, reliability, and speed, than human translation.

In conclusion, the committee indicated the direction to be taken in the future. It suggested that further research in the field of computer science should have been

conducted, with special attention to computational linguistics and computer-aided translation. Moreover, governmental expenditures should have been devoted not to the implementation of new Machine Translation systems, but to improving human translation, by adopting a common and efficient evaluation method, accelerating the translation process, enhancing the post-editing phase, and, finally, providing translators with adequate reference material, such as technical glossaries and comprehensive dictionaries (Automatic Language Processing Advisory Committee, 1966).

The harsh criticism by the Automatic Language Processing Advisory Committee can be undoubtedly justified on the grounds of the poor development that Machine Translation had experienced in the decades before the submission of the report. However, the controversial decision to consider the sole US situation concerning translation expenditures, problems, and growth, within the framework of an already globalized economy, has been the subject of intense debates and was accused of being a sign of the limitation of its scope (Hutchins, 1996). Despite the great impact of the report, Machine Translation research continued, although sharply reduced, both in the US and abroad, particularly in Japan, Canada, and some European countries, where new approaches were developed and tested, with the aim of designing increasingly reliable, fast and cheap products.

1.4. Rule-Based Machine Translation

Historically, Rule-Based Machine Translation (RBMT) was the technique applied to the first Machine Translation systems. It is based on linguistic information concerning the source and the target languages, including grammatical rules, semantic, syntactical, and morphological structures (Charoenpornasawat et al., 2002). This information can be retrieved from a number of different sources, such as grammars, dictionaries, glossaries, and other kinds of documents that come from the work of expert linguists (Lagarda et al., 2009). Over the course of its history, three different approaches have been developed within the scope of rule-based Machine Translation, namely direct Machine Translation, transfer-based Machine Translation, and interlingua Machine Translation.

Direct Machine Translation consists of a direct word-to-word translation of the source language document into the target language and is divided into several phases,

namely morphological analysis of the source language text, direct word-to-word translation using an automatic bilingual dictionary, generation of the target text, final review, and syntactical reorganization of the automatically generated text if needed. The morphological analysis is conducted by reducing the inflected forms of the words of the source document, which contain grammatical and syntactical information, to their basic form, called lemma. The translation is made by simply substituting the source document words with their counterparts in the target language, with no regard to the overall context or the linguistic relations between the words. Finally, the syntactical reorganization consists of an adaptation of the automatically generated text to the grammatical rules and syntactical patterns of the target language. Direct Machine Translation provides indeed word-to-word translations, on the basis of the linguistic information mainly contained in automatic bilingual dictionaries. As a consequence, this approach is suitable for just those kinds of text that do not require the analysis of the context, such as highly technical documents or lists of terms. Moreover, each direct translation system has to be designed for a specific language pair and is not able to provide multilingual translations.

A Transfer-based Machine Translation system is composed of three different modules. The first one, the analysis module, is devoted to analyzing the source text by parsing each sentence, in order to neatly establish its structure and the linguistic role of its components. The second one, the transfer module, contains the translation rules from the source language to the target language, specifically formulated for one language pair and one direction of translation. The third and final one, the generation module, concerns the generation of the target text, according to the grammatical rules of the target language. The addition of an intermediate module enables transfer-based Machine Translation systems not only to provide more accurate translations when compared to direct Machine Translation systems but also to consider the general context of the text. However, since transfer modules are suitable just for one direction of translation from one specific language to another, a great number of transfer modules are needed to implement a multilingual transfer-based Machine Translation system, with an exponential increase of design time, effort, and, consequently, cost.

Interlingua Machine Translation is based on the concept of Interlingua, a language created as an intermediate stage between the source and the target language. The source document is indeed translated into the Interlingua, and the newly generated text has to

carry the meaning of the initial document, being disconnected from the syntactical structure and the grammatical rules of the source language. The Interlingua text is a sort of neutral intermediary, which will be itself translated, according to the grammatical and syntactical rules of the target language. Unlike the transfer-based approach, the Interlingua approach does not imply the development of a great number of transfer modules but requires the existence of one Interlingua, common to all the language pairs. As a consequence, it might be suitable for multilingual translations. Nonetheless, the creation of an intermediate language able to properly convey the meaning of the source text, and simultaneously be neutral to the syntactical structures of all the languages involved, presents a number of difficulties, which so far prevented the purely Interlingua approach to be successfully adopted in translation (Naldi, 2014).

Generally speaking, despite its relatively low cost of implementation and the possibility of developing new techniques to extend the glossaries and the corpora containing grammatical rules, rule-based Machine Translation requires a great deal of linguistic knowledge, which is difficult to achieve and process. Moreover, since natural language is such a diverse and enormous field, it is currently impossible to properly formulate rules able to cover all a language (Charoenpornasawat et al., 2002).

1.5. Example-Based Machine Translation

Example-Based Machine Translation (EBMT) was theorized for the first time in a conference paper written by the Japanese computer scientist Makoto Nagao in 1981, however, in-depth research in the field only started in the late 1980s (Hutchins, 2005). Together with Statistical Machine Translation (SMT), which will be described later in this chapter, EBMT falls within the corpus-based Machine Translation approach, based on the analysis of bilingual text corpora. The main aim of the implementation of the first EBMT systems was undoubtedly the attempt to overcome the weaknesses of Rule-Based Machine Translation systems. RBMT systems were indeed found to be not only scarcely suitable to language pairs with significantly different syntactical structures and unable to convey the proper meaning of idiomatic expressions and collocations but also based on a complex and not easily updatable software architecture. By contrast, EBMT systems were built on a simpler structure, which provided better-quality outputs, also with regard to

languages with completely different structures. In the beginning, the academic community was divided between those scholars who considered EBMT as a complement to improve RBMT systems' performances and the ones that were convinced of the necessity to conduct extensive research in order to enable EBMT systems to deal with the whole translation process (Hutchins, 20051).

Despite the existence of a great variety of different techniques and approaches within the scope of Example-Based Machine Translation, we can briefly outline its main features and functionalities. Conceptually, an EBMT system follows the analogy process, the same used by the human brain in learning new languages or translating a text and based on the idea of transferring a certain meaning from one subject to another (Nagao, 1984). The core element of an EBMT system is a corpus containing several translation examples regarding a specific language pair. At the beginning of the translation process, the source text is subject to morphological analysis. The length of the strings considered by the software varies depending on the characteristics of each system and may range from fragments of text or whole sentences to, more likely, fragments of sentences or phrases. After the analysis, similar fragments are retrieved from the example corpus, with their corresponding translations. A number of likely translations are shown, together with their reliability factor, which depends on the distance between the source fragment and each example retrieved from the example database. The distance is calculated by a thesaurus, which contains linguistic information regarding the source document fragments. In case no similar examples are found in the corpus, the system informs the user about the impossibility to provide any translation. When a reliable translation is found, the adaptation module is in charge of adjusting it according to the grammatical rules and the syntactical structure of the target language (Sumita et al., 1991).

As mentioned before, the EBMT approach presents several features that make it preferable to the RBMT one. Unlike RBMT systems, which require the application of grammatical and syntactical rules formulated by linguists, EBMT systems are based on example corpora that can be easily updated by simply adding new examples of translation, which are reasonably easy to collect and do not change with time. In addition, being based on actual documents and translations, they are able to consider the overall context and provide more reliable translations (Hutchins, 2005). Nonetheless, the EBMT approach is far from being immune to complexities. Indeed, matching, extracting, adaptation, and

recombination algorithms able to ensure a reliable output are not easy to develop and need particular attention and continuous updating. Moreover, the mere addition of new examples to the example corpus does not necessarily produce a considerable improvement of the system's performances. The new examples have to be chosen carefully, in order not to burden the system with superfluous data (Hutchins, 2005).

1.6. Statistical Machine Translation

As mentioned above, both Statistical Machine Translation and Example-Based Machine Translation come under the scope of corpus-based machine translation. Nonetheless, the two methods differ widely, in terms of system architectures and performances. Statistical Machine Translation (SMT) theory was proposed for the first time by the IBM group in 1988. Its development doubtless owes a great deal to the formulation of the probability distribution function by Claude E. Shannon in "A Mathematical Theory of Communication". Given an input or a set of inputs, the probability distribution indicates the likelihood of a certain outcome (Shannon, 1948). Assuming the statistical nature of natural languages, when applying Statistical Machine Translation, a text is translated according to the probability distribution that a word, phrase, or group of words in the source language, corresponds to a word, phrase, or group of words in the target language.

SMT is based on two main statistical models, namely the translation model and the language model. The translation model is generated by analyzing a bilingual text corpus containing a number of documents in the source and target languages. In a preparatory phase, the documents from the bilingual corpus are divided into single words, and every single word in the source language is aligned to its corresponding word in the target language. The translation module constitutes the result of this alignment and indicates the statistical frequency regarding the translation from the source language into the target language. As for the language module, it indicates the most likely sequence of words in the target language and may be generated starting from the same bilingual corpus as the translation module (Hutchins, 2005). The translation process itself is based on the statistical frequency indicated by the translation module, while the generation of the target text is made according to the language module.

In the early stages of their development, SMT systems adopted a single-word based approach, which consisted, as described above, in a single word alignment between the source and the target texts contained in the bilingual corpus, and was afterward found insufficiently effective in considering the overall context and the linguistic relations between words (Zens et al., 2004). Hence, in order to include contextual information and to provide higher-quality outputs also with respect to languages with a significantly different word order, a phrase-based approach was implemented and entire groups of adjacent words, or phrases, started to be brought into correspondence (Och et al., 1999). Finally, to overcome the phrase-based approach's shortcomings in dealing with syntactically different languages, a significant step forward in the development of SMT was made with the implementation of the syntax-based approach. This new method, which requires the division of the texts into syntactical units, is able to incorporate information about the syntactical structure of the text, and consequently, to generate more reliable translations (Hadiwinoto, 2017).

Despite the leading role it has had over the decades within the framework of machine translation, and its better performances when compared to RBMT and EBMT, Statistical Machine Translation presents several downsides, both in terms of implementation costs and output quality. Therefore, it has been recently integrated with new techniques and methods, especially after the emergence of Neural Machine Translation.

1.7. Neural Machine Translation

Neural Machine Translation (NMT) started to emerge in 2014 as a supplement of SMT systems and subsequently developed its own techniques and systems. It falls within the scope of Machine Learning, which is devoted to programming computers to perform several human brain tasks by developing algorithms based on past experiences or example data. From the very beginning of its growth, NMT has proved itself better in performing translation tasks than the previously designed MT systems and still nowadays represents the state-of-the-art technology in the field of machine translation.

NMT systems' most innovative aspect, which undoubtedly departs them from the other MT systems, is the central role of word embeddings in the translation process. Broadly speaking, word embedding includes a set of techniques in the field of Natural

Language Processing (NLP), aimed at representing words using n-dimensional vectors. The idea of representing words using n-dimensional vectors comes from distributional semantics, a branch of computational linguistics, which deals with the semantic distribution of words in natural languages and is based on the distributional hypothesis. According to the distributional hypothesis, the semantic similarity of words can be quantified on the basis of the linguistic contexts in which they tend to occur together. In other terms, the more two words are used in the same or similar contexts, the more they are similar from a semantic point of view. The concept of semantic space is the common element of the highly diverse applications of the distributional hypothesis. Indeed, a semantic space is drawn by analogy with the geometric space, and each of its points, i.e. each word belonging to the semantic space, is represented by an n-dimensional vector. The vector itself does not have any semantic value, and its function is limited to indicating the position of a word in the semantic space and its distance from the other words. The Euclidean distance between the vector representations of two words corresponds to their semantic distance. Hence, words are located in the semantic space according to their semantic meaning. The most significant contribution of the distributional hypothesis to the linguistic theory in general and the research in the field of machine translation is the idea that the meaning of words needs to be investigated in a contextual framework (Lenci, 2010). Context has indeed to be included not only in linguistic analysis but also in all the applications of linguistic studies, including Machine Translation.

In order to create word embeddings, Neural Machine Translation systems use Artificial Neural Networks (ANN), sets of artificial neurons linked together on the model of the human brain's biological neural networks, and specifically trained to perform translation tasks. Generally speaking, exactly like a newborn baby's brain cells, once an artificial neuron has been created, it has to be trained in a specific domain. Nonetheless, being artificial, an ANN is able to ensure better performances, when compared to human neural networks. Indeed, not only they do not risk dying and lose the stored information, but they are also capable of a higher level of accuracy in repeating the same actions an indefinite number of times, and a greater speed of processing (Wołk et al., 2015). There are several categories of artificial neural networks, including Multilayer Perception Neural Network and Repetitive Neural Network (RNN). The former consists of layers of artificial neurons so that the output of a layer is the input of the next layer. The latter is

formed by a set of looped artificial neurons, where each artificial neuron is alimented by its own output. Because of their structure, Repetitive Neural Networks are able to maintain the stored information over time and are consequently particularly effective in performing translation tasks.

One of the most popular NMT architecture consists of three core elements, namely an encoder, a decoder, and an attention model. The encoder is a Bidirectional Recurrent Neural Network (BRNN) (Wang et al., 2017), which, using an n-gram model, analyzes the input sentence from right to left and vice versa and extracts a fixed-length vector representation of the source sentence. Afterward, the decoder, starting from the vector representation, creates a variable-length sequence, i.e. the target sentence (Van Merriënboer, 2014). The attention model is a Multilayer Perception Neural Network and is in charge of aligning the source sentence words with the corresponding target sentence words.

As mentioned above, an RNN may be particularly effective in performing translation tasks. Nonetheless, this kind of neural network is not able to link pieces of information that are distributed over long distances. Therefore, when contextual information is scattered over a long source text, NMT systems based on RNN cannot by their nature provide reliable translations. In order to overcome RNN's limitations, the more sophisticated Long Short Term Memory (LSTM) neural networks have been implemented. Unlike RNN, which consists of one only layer, LSTM neural networks have four layers and can solve long-distance problems and properly convey contextual meaning (Wu et al., 2016).

Despite Statistical Machine Translation has been the most widely adopted approach for decades, Neural Machine Translation is proven to provide more reliable translations. This is mainly due to the fact that, by using word embeddings, NMT systems succeed in including contextual information in the translation process. For this very reason, a number of originally Statistical Machine Translation systems have recently evolved by adopting the Neural Machine Translation approach. Moreover, Neural Machine Translation is considered the direction in which Machine Translation research has to move in the near future.

2. A COMPARISON BETWEEN DEEPL AND YANDEX: PURPOSES AND SCOPE OF THE RESEARCH

As mentioned in the introduction, the thesis that we are presenting in these pages is devoted to comparing two machine translation tools that feature prominently in today's neural Machine Translation framework, namely DeepL and Yandex. The research will be conducted by observing how the two translation programs perform the translation of a number of Russian highly specialized and popular-science medical texts concerning the Coronavirus pandemic outbreak. Before delving into the actual investigation, we believe it is worth clarifying the purposes and scope of our research, which will be discussed in the present chapter. The first section deals with the historical correlations of medical language within the field of Machine Translation, its main linguistic features, and the major aspects regarding the translation of the Russian language of medicine. The second and the third sections are aimed at giving a relatively brief insight into the controversial issue regarding Machine Translation Evaluation (MTE), by introducing some of the evaluation methods that have been implemented over time, with a focus on the comparative error analysis approach, chosen for the purposes of our study. Finally, in the last section, DeepL's and Yandex's respectively architectures are presented.

2.1. Translating the Russian language of medicine

Over the last decades, increasingly comprehensive linguistic studies have been devoted to the language of science. This is undoubtedly due to the terrific development of scientific knowledge as well as the crucial role that international communication has assumed in scientific research. Generally speaking, the language of science is quite heterogeneous, especially with regard to lexis. It is indeed formed by three different broad vocabularies, namely a general vocabulary, consisting of terms borrowed from the common language, a scientific vocabulary, made of general scientific words also that are used in every-day communication, and a terminological vocabulary, containing scientific terms that specifically refer to scientific concepts and entities (Guščina, 2005). Moreover,

it is characterized by rapid and constant evolution, as new terms, with different extents of specificity, need to be continuously coined to frame new concepts, as well as obsolete words are expected to either slowly disappear from the linguistic landscape of a certain country or be replaced by newly formed ones. Within the vast and diverse scope of the language of science, the language of medicine certainly not only constitutes one of its most ever-developing and vivid sectors but also is of major interest for the purposes of our research, as it has always been strictly related to translation studies.

Medicine and translation share a long and rich history. In fact, medicine has undoubtedly been one of the first fields of knowledge that has constantly required, since the beginning of its evolution, the support of translation (Karwacka, 2014). This is mainly due to its historical development, which began in Greece, moved to the Roman and Arabian empires to finally settle in the new-born eastern and western European countries, with the emergence of the first national medical languages. Over the decades, each transfer of medical knowledge, which was accomplished through the conquests of foreign lands, as well as the peregrinations of scientists and doctors invited to spread their medical expertise, resulted in a primary phase of preservation of books, treatises, and other documents in the original language, followed by the second phase of adjustment of the medical terminology to the semantic and syntactical rules of the receiving-country language, and a final one of actual translation of the medical material, and the emergence of a completely new medical terminology (Fischbach, 1986). Still nowadays, translation plays a key role in sharing medical knowledge across the world, particularly with respect to the publication of the latest research results by the international scientific community, the marketing campaigns of the new pharmaceutical products, and the linguistic support provided to foreign patients in the communication with physicians. Since in most cases the success of a certain medical procedure or the global spread of accurate medical data depends on medical translation, its quality is a matter of considerable relevance. A translation error may indeed provoke serious or even life-threatening consequences (Karwacka, 2014).

Generally speaking, medical language presents several features that may facilitate its translation from one language to another, namely the universality of the topic, the relatively easy availability of reference material, and a certain lexical equivalence. The human body and its peculiarities are indeed the very same regardless of the country under

consideration. Hence, although the language used to describe them may considerably differ, the subject remains universally known and most of the time the conveyed message does not require to be adjusted according to the receiving culture. Moreover, medicine and health care have always been of great concern for human beings and the research in the field is not only extensive but also highly documented. As a consequence, there is a considerable availability of medical documents that can serve as reference material for medical translation. Finally, since medical terminology has its roots in Greek and Latin and consequently tend to contain Greek or Latin prefixes and suffixes, a certain degree of lexical equivalence is universally widespread. However, medical translation implies a set of challenges related to some of the features of medical language as well as to the incredibly rapid development of medicine and science technology in general (Fischbach, 1986).

As mentioned above, medicine not only has a long and rich history but also experiences a continuous and rapid development, which necessarily affects medical language and terminology. This has required the creation of a whole new set of medical terms in order to describe newly discovered or recently spread illnesses, the latest medical treatments, as well as state-of-the-art medical technology, and the new medical branches. Moreover, nowadays medicine appears to be ever increasingly compartmentalized in highly specific branches, whose borders are undoubtedly more defined than in the past (Mičić, 2013). Therefore, in order to perform a high-quality translation, a medical translator is expected not only to fully master the source and the target language but also to acquire a certain familiarity with the very specific subject of the translation under examination (Fischbach, 1962). Finally, being a technical language, medical language displays a number of distinctive features that require particular attention when it comes to translation.

Medical language can be defined as the “occupational register of physicians and it is largely opaque outside the medical community” (Mičić, 2013). It is one of the so-called Languages for Specific Purposes (LSP), i.e. those specific registers adopted by professionals to exchange information and knowledge in professional contexts. The medical language shares some features with the other LSP and has developed over time a set of specific characteristics. A high degree of impersonality aimed at maintaining distance and objectivity undoubtedly stands out in medical texts and treatises. It usually

comes along with a marked tendency towards the nominalization of verbs and adjectives, which leads to extensive use of extended nominal groups. Moreover, medical language is prone to passivize active verbs and includes highly technical phrases, which constitute the medical jargon. As for lexis, the historical origins of medical language are responsible for the persistence, within the medical vocabulary, of Greek and Latin terms in their original form, as well as the formation of new words starting from Greek and Latin suffixes and prefixes (Guščina, 2005). Moreover, eponymy has always represented a highly widespread phenomenon in medical language. With eponymy the tendency to name diseases after the scientists who first discovered and studied them is meant. At the first stages of medical research, being a certain illness's pathogenesis still almost completely unknown, the researcher's name is frequently used to denote it. Afterward, once the illness's main features have been uncovered, a term describing them is chosen. Among the medical science's above-mentioned characteristics, a rich and ever-increasing terminology undoubtedly stands out. In fact, due to medicine's rapid and continuous development, whole new sets of medical terms, which directly reflect state-of-the-art technology applied to medicine, as well as new treatments and illnesses (Mičić, 2013) has emerged, so that medical terminology has become, over the last twenty years, one of the vastest terminological systems and frequently borrows terms from a great deal of other scientific fields (Guščina, 2005). Generally speaking, terms are words or groups of words defining a specific concept or phenomenon, usually referring to a particular field of knowledge. They are frequently used in professional communication and play a crucial role in information interchange. From a linguistical point of view, not only terms can be analyzed according to the same linguistic categories as general-vocabulary words, but they are also involved in the same lexical relations, first and foremost, synonymity (Fedina, 2017). In linguistics, synonymity indicates the existence of phonetically different words or groups of words conveying the same meaning and is owed to the coexistence of terms originally coming from different terminological systems that, for a variety of reasons, become part of the same terminological vocabulary. In the language of medicine, synonymic pairs may be formed by a highly specialistic and a more general term, a term directly borrowed from Latin and Greek and its equivalent in a specific country's modern language, an eponym and a term accurately describing the pathogenesis of a certain illness, as well as an obsolete term and its contemporary substitute (Fedina, 2017). While

a great deal of significantly different cases of synonymy may occur, its origins have to be searched in the historical development of languages. In fact, only some sort of linguistic contact between two distinct terminological systems can result in the overlapping of the mentioned systems and the emergence of pairs of different terms conveying the same meaning. Among plenty of practical possibilities, the geographical proximity of different cultures, the sharing of knowledge between people speaking different languages, and, with specific regard to eponymy, several scientists who simultaneously discover the same diseases in different countries, seem to present the greatest likelihood (Fedina, 2011). By complicating the translation process, synonymy may represent a barrier to international communication and the sharing of medical knowledge and consequently deserves special attention when translating medical texts as well as with regard to the creation of new domestic medical terms. Nowadays, English constitutes the lingua franca of medicine, as Greek and Latin did in the past, as well as the model for other countries in creating their own language of medicine (Mičić, 2013).

Russian language of medicine started to emerge, along with professional medicine itself, in the 17th century, which undoubtedly constituted a turning point in the development of Russian medical terminology. In fact, before the 17th century, popular words were used to indicate diseases and symptoms, whereas afterward the first contact between Russian and international medical terminology occurred and new medical terms, as well as new highly differentiated medical professions and disciplines emerged. The development of Russian medical terminology owes indeed a great deal to the European professionals who brought to Russia their medical knowledge and expertise and contributed to the emergence of new national medical terms. Cyrillic alphabet undoubtedly constituted a barrier to communication, and, in the first medical textbooks and treaties, European terms derived from Latin or Greek were displayed together with their Russian translation (Olekhovich and Olshvang, 2017). Nowadays, the Russian medical language shares some features with the other technical languages and developed its own ones. Objectivity and impersonality, represented by impersonal phrases and constructions, and reflexive verbs, constitute the two main features of the Russian medical language. Moreover, nouns are more widely used than verbs and a marked tendency to formulate grammatically relatively easy sentences has been observed. However, the existence of concatenations of genitives along with widespread use of gerunds and

participles certainly contributes to complicate the overall grammatical structure. As mentioned above concerning medical language in general, the Russian language of medicine is likewise characterized by a widespread presence of synonyms, which attests that close contact between Russian and international terminology took place. Although in medical terminology, international synonyms are usually preferred for reasons of systematization, translatability, ease of international communication, and spread of medical knowledge, the Russian medical language displays the coexistence of international terms and their Russian equivalents (Polackova, 2001).

2.2. Machine Translation Evaluation: an open question

Translation Quality Assessment (TQA) has always constituted a key issue in Translation Studies and its relevance has significantly increased with the emergence of Machine Translation and the growth of the translation industry. Reaching an adequate understanding of how to properly evaluate Machine Translation systems is vital to the development of Machine Translation research. A good evaluation method can indeed shed light on a specific MT system's strengths and weaknesses and consequently suggest the necessary modifications and the appropriate line to be taken in the future. Moreover, it represents an essential tool for Machine Translation professionals to monitor the increasingly rapid progress of their systems and for users to sensibly choose the MT programs that best suit their needs (Papineni et al., 2002). Nonetheless, since the beginning of Machine Translation growth, its evaluation has always been an extremely controversial issue, and still nowadays represents an open question.

Before moving to the enumeration and description of the most widely used Machine Translation evaluation techniques that have been developed so far, we will start by clarifying what is commonly meant by Machine Translation quality. Generally speaking, when assessing a translation, be it performed by humans or computer programs, worth underling undoubtedly is the assumption that, given a source sentence, or text, there is not only one "perfect" translation, as well as there may be several "acceptable" translations (Papineni et al., 2002). Moreover, with regard to Machine Translation evaluation, a core concept is introduced in the following sentence by Papineni et al.

(2002): “The closer a machine translation is to a professional human translation, the better it is”. Hence, in order to undertake quality assessment for MT systems’ output, a machine translation is usually judged on the basis of a numerical metric that measures its closeness to a set of reference professional human translations. Having said this, fluency and adequacy represent two major evaluation criteria when it comes to evaluating the output of a Machine Translation system. According to the Linguistic Data Consortium (2005), a fluent translation is “one that is well-formed grammatically, contains correct spellings, adheres to the common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker”, whereas adequacy, also called accuracy or fidelity, is defined as “How much of the meaning expressed in the gold standard translation or the source is also expressed in the target translation” (Monti & Montella, 2015). A whole set of additional evaluation parameters can be considered, including readability, comprehensibility, acceptability, and usability. Readability represents the degree of ease or complexity with which a written text can be read by a reader or a group of readers. It is related to linguistic features such as word frequency and sentence length, as well as extralinguistic features, concerning, among others, text formatting. While a universal definition of readability seems to be commonly accepted, the same does not hold for comprehensibility. In fact, different definitions of comprehensibility have been provided over time, and the nature of its relationship with readability still is disputed. Broadly speaking, comprehensibility can be defined as the extent to which a given text can be understood by a reader or a group of readers. However, whether it mainly depends on the type of text or the type of reader, whether it is part of the scope of readability or constitutes an independent evaluation parameter and other related issues have not received an official explanation yet. Despite the diverse definitions given by the academic community, acceptability deals with how a text, and, in this case, a translation, meets the reader’s needs and expectations, and consequently, it is accepted by its intended audience. Finally, usability concerns the degree of possibility for an MT output to be useful to its users in order to accomplish a set of intended tasks within a certain context of use (Castilho et al., 2018). These last two criteria undoubtedly constitute core concepts of translation evaluation as the target audience and the purposes of the translation service, agreed in advance between the requester and the provided, carry considerable weight in translation quality assessment. Finally, the degree of statistical closeness between

machine translation and human translation represents the ultimate and controversial evaluation parameter to be considered, as it constitutes the basis for the other criteria, as well as the most subjected to criticism.

Machine translation evaluation can be performed by means of automated, semi-automated, and human techniques. It must be said that a clear distinction between the mentioned categories cannot be considered obvious, as their boundaries are quite blurry, and a juxtaposition is not difficult to occur. Moreover, since they all present a set of positive and negative sides, determining the most effective approach still constitutes a matter of dispute among scholars. Starting from automated Machine Translation evaluation, it refers to those assessing methods that do not involve human judgment or rather in which human intellect is confined to a number of side activities, such as data collection, preparation of the reference translations, or annotation. It is divided into three broad categories, namely reference translation-based metrics, confidence or quality estimation metrics, and diagnostic evaluation based on checkpoints. The first technique is based on the comparison between the machine translation to be assessed with a set of reference professional human translations, also known as golden translations, which constitute the benchmark to judge its quality (Chatzikoumi, 2020). A score is indeed assigned on the basis of their closeness, and, generally speaking, the higher their values, the higher the quality of the machine translation under examination (Munkova et al., 2020). Several different metrics are used to calculate the closeness between machine translations and reference translations. It can be defined according to the edit distance, or Levenshtein³ distance between the gold and the machine translation, i.e. the number of editing steps necessary to transform the latter into the former. Word Error Rate (WER) metric undoubtedly constitutes the most notorious representative of this category. It is designed to evaluate a candidate translation according to the number of words that need to be inserted, deleted, or substituted in the candidate translation to make it identical to the reference one. More precisely, the ratio of the necessary edit operations to the total number of the words constituting the reference translation is calculated. A number of

³ In information theory, computer science and computational linguistics, the Levenshtein distance is a metric used for determining the difference between two strings. More specifically, it measures the amount of edit operations needed to transform one string into another. Levenshtein distance was named after Vladimir Iosifovič Levenštejn (Владимир Иосифович Левенштейн), a Soviet scientist, who investigated it in 1965 (Levenštejn, 1965).

other implemented metrics, which we will briefly enumerate in the following lines, represent greater or lesser significant variations of WER (Chatzikoumi, 2020). Position independent Error Rate (PER) constitutes a variant of WER in which the word order of the candidate sentence is not taken into account and, consequently, an excessively optimistic evaluation of the candidate translation risks being provided (Munkova et al., 2020). On the other hand, Cover Disjoint Error Rate (CDER) considers the word order by adding to the evaluation process a further operation, namely the block movement. CDER's design is indeed based on the assumption that there may be several correct translation variants for the same source sentence. In most cases, these variants differ from each other in the order in which the blocks of words are collocated throughout the sentences. Hence, along with the classical operations constituting the Levenshtein distance, long jump movements, i.e. the relocation of whole blocks of words, are included in the assessment of the candidate translation. Moreover, in order to narrow the gap between human judgment and automatic evaluation of MT systems', the proposal to combine two opposed evaluation metrics like PER and CDER has been put forward by several scholars. Despite minor differences in terms of implementation and reliability, WER, PER, and CDER metrics tend not to consider the degree of similarity between the candidate translation's words and their counterparts in the reference translations. In other terms, when two different words occupy two corresponding places in the candidate and the reference translations, an editing step is counted, irrespective of whether they have slightly or completely different meanings. In the reality, the two cases are profoundly distinct, and would doubtless receive different assessments by a human evaluator (Leusch et al., 2006). Over the decades, an increasingly large amount of metrics of error rate has been designed. Broadly speaking, they roughly share the same basic principles, with slight architectural differences, and tend to mostly differ in the number of the reference human translations used as the basis of judgment and the inclusion of the displacement of words and phrases between the group of editing steps to be considered (Chatzikoumi, 2020).

On the other hand, precision and recall metrics evaluate the quality of a Machine Translation output according to the measure of textual similarities between an MT system's output and the reference human translations. In particular, precision refers to the

ratio of n-grams⁴ in the translation under examination that occur in any of the reference translations to the total number of n-grams contained in the translation under examination. The recall is defined by calculating the ratio of n-grams in the translation under examination that occur in any of the reference translations to the total number of n-grams of the reference translations (Chatzikoumi, 2020). The Bilingual Evaluation Understudy (BLEU) metric undoubtedly represents one of the most popular precision and recall metrics as well as the benchmark to judge other automatic Machine Translation evaluation systems. The algorithm is composed of two main constituent parts, a numerical metric designed to measure the closeness between the candidate machine translation and the reference translations, and a corpus containing a number of reliable human reference translations. In evaluating the candidate translation, the BLEU metric considers three core factors, namely word choice, word order, and length. A score is indeed assigned to each n-gram of the candidate translation on the basis of its similarity with the reference translations. Afterward, all the scores are averaged over the entire corpus. In this way, a comprehensive corpus-based evaluation indicating the candidate translation's overall quality is provided. The score given to each n-gram is calculated according to a precision measure: the words contained in the n-gram under examination that occur also in any reference translation are counted and their total number is then divided by the number of words contained in the n-gram (Chatzikoumi, 2020). However, since MT systems tend to generate a greater amount of words when compared to any reference translation, in order to achieve a reliable evaluation, a modified unigram precision measure is adopted. The modified precision measure takes into account not only whether a candidate translation n-gram appears in any reference translation, but also its maximum number of times of occurrence in any of the reference translations, which is afterward divided by the number of times the same n-gram appears in the candidate translation itself. The size of the n-grams used as units of comparison has a strong effect on the overall evaluation. Shorter n-grams, such as unigrams, i.e. consisting of a single word, are proven to be useful to assess the adequacy, whereas longer n-grams are better in determining the fluency of the candidate translation (Papineni et al., 2002). As mentioned above, for the candidate translation to be given a high score, it should correlate well with the reference translations

⁴ In computational linguistics, sequences of n consecutive words, usually retrieved from a written or speech corpus.

also in terms of length. Moreover, the candidate translation is expected to contain just one of the synonymic words that appear in the different reference translations as variants of the same potential source word. In this case, a candidate sentence could indeed be given a high score as it contains many words that occur also in the reference translations, without being a reliable translation, nor having, in many cases, full meaning. In order to include the length factor in the evaluation process and try to overcome the recall problem, a brevity penalty is introduced. The reference translation that better reflects the candidate translation's length is selected and the same procedure is performed at the sentence level. Afterward, all the length best matches are summed and compared to the total length of the machine translation output under examination. BLEU metric's scores range from a minimum of 0 and a maximum of 1, which would be assigned to a candidate translation proven to be identical to one of the reference translations contained in the corpus. Moving to the last two categories of automated machine translation evaluation, namely confidence or quality estimation metrics, and diagnostic evaluation based on checkpoints, we will just say that they cannot be considered actual evaluation metrics, as they rather represent proxies for them, which can be useful in those cases when there is no availability of reference human translations or an evaluation at the segment level is needed (Chatzikoumi, E. 2020). Nonetheless, they fall outside the scope of our research.

As for human evaluation methods, they include all those approaches that directly or not involve human judgment. More specifically, they can be divided into two main groups, namely DEJ-based and non-DEJ-based according to whether Directly Expressed Judgment (DEJ) is performed or not. When DEJ-based techniques are adopted, the human judge, better known as an annotator, directly expresses an evaluation of the candidate machine translation, by comparing it with the source text or specially made by him reference translation (Chatzikoumi, 2020). Given a machine translation system's output, human evaluators can be asked to assess its quality according to several different criteria and highly diverse procedures, which we will briefly describe in the following lines. Adequacy and fluency annotation tasks involve an evaluation of the candidate translation's adequacy and fluency, which are assessed according to a scale of values. As an example, the Dynamic Quality Framework (DQF), developed by the Amsterdam-based Translation Automation User Society (TAUS) in 2011, provides the users with a platform where they can upload the candidate machine translation in a spreadsheet format

and choose the criteria according to which it will be evaluated by human annotators, between fluency, adequacy or a combination of them. After the translation under examination has been assessed, the final document containing the evaluators' annotations can be downloaded. The quality-checking annotation tasks and the Human UCCA-Based MT Evaluation represent slightly different variations of DQF. The former marks indeed an MT system's output as acceptable, can easily be fixed and none of them, whereas the latter provides an analysis of how the machine translation at stakes semantically reflects the source document. Finally, a ranking and a direct assessment can be conducted. When ranking is performed, human evaluators are asked to compare usually three and better no more than five translations produced by different MT systems, and either choose the best option or rank them according to their overall quality. On the contrary, Direct Assessment (DA) consists of a general evaluation of an MT system's output, which is collocated on a continuous scale.

Non DEJ-based evaluation metrics, as deductible by their name, correspond to the evaluation methods when human judgment is indirectly expressed. Semi-automated metrics, which include the use of automated metrics along with human annotations undoubtedly fall within this category. The task-based evaluation also can be performed. In this case, humans, not necessarily professional translators, are intended to check whether the provided machine translations actually fulfill the purposes for which they were originally created. As an example, they can be asked to read the candidate translations and answer several questions about its content, in order to determine, among the others, the clarity of its linguistic structure and the degree of compliance with the source text. Aside from the two mentioned techniques, a third one has been implemented, namely the error classification and analysis approach (Chatzikoumi, 2020). Generally speaking, error analysis consists of assessing machine translation quality by annotating each error and marking it with an error-tag. It can be performed manually, automatically, or semi-automatically. As illustrated in Figure 3, which describes a general manual error analysis process, annotators are usually provided with additional reference material, such as the source text or a set of reference human translations, or both, and asked to classify the errors encountered throughout the translation according to a number of previously agreed error categories.

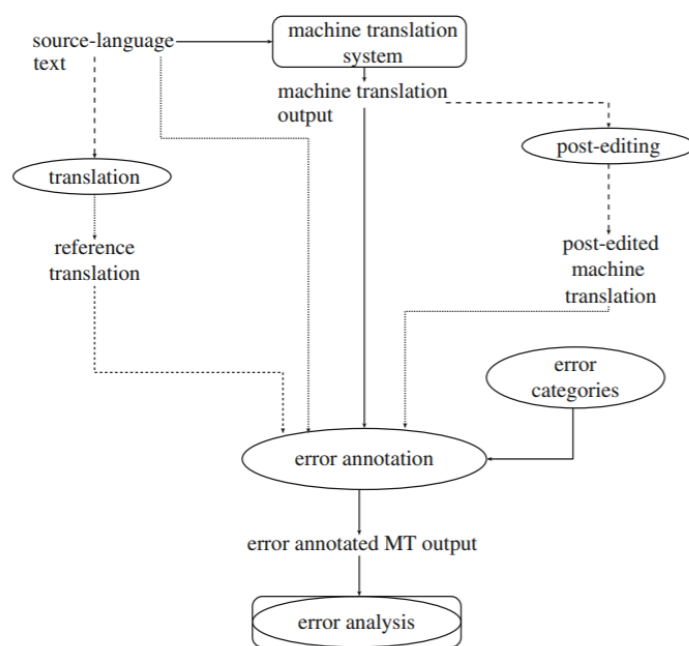


Figure 3 General manual error analysis procedure (Popovic', 2018)

In order to be comprehensive, error analysis needs careful planning. Firstly, it is necessary to consider annotators' knowledge and background and, in case the evaluation is performed by a group of evaluators, to increase inter-annotator agreement by organizing a specific training devoted to clarifying the scope and purposes of the analysis. Secondly, the choice of error classes requires special attention. In fact, a great amount of error categories undoubtedly contributes to providing an exhaustive error analysis. Nonetheless, in this case, since the boundaries between the different categories are relatively blurry, it may be difficult to properly assign each error to its corresponding category. Moreover, error classes are to be carefully chosen according to the type of analysis that has to be conducted, on which depends the importance given to each error, and the relevant linguistic features of the text under examination. Finally, not only the number and type of error classes have to be considered, but also an exact and unambiguous definition of each error class needs to be provided. In this way, correct classification is easier to perform by the evaluator or the group of evaluators.

A universal approach to the problem of properly setting error analysis has not been implemented yet and different scholars proposed their own error typology. However, the adoption of a common method suitable for any kind of text and language pair would

undoubtedly increase the consistency in Machine Translation Quality assessment with a positive impact on Machine Translation development and overall quality (Popovic', 2018). A significant step towards error analysis standardization has been carried out with the implementation of the Multidimensional Quality Metrics (MQM) by the German Research Centre for Artificial Intelligence (DFKI) from 2012 to 2014. MQM aims at providing a comprehensive list of error categories sorted by their specificity. In other words, each primary error category, or dimension, branches into one or more subcategories, which constitute more specific types of issues that fall within a certain error category. The primary categories and their corresponding subcategories displayed by the MQM metric are not expected to be used in their entirety and simultaneously. On the contrary, only the ones of them that suit a specific analysis requirement are intended to be chosen and used by the annotators to undertake that error analysis. MQM includes the following primary categories:

- Accuracy
- Design
- Fluency
- Internationalization
- Local convention
- Style
- Terminology
- Verity

As mentioned above, each primary category subsequently branches into a number of subcategories, gradually more and more specific. Not only MQM provides a set of error categories but also defines four different levels of severity, namely critical, major, minor, and null. Critical refers to errors that prevent a certain target text from fulfilling its intended purpose, risking causing damage with legal or economic implications to its final user. Major errors have a serious effect on the comprehensibility of the target text, whereas minor ones are expected to be noticed by target-language speaking users, but do not prevent them from properly understanding the meaning conveyed by the translated text. Finally, some words contained in the translated text may be marked as null errors, although they are not exactly errors, in case they have been simply changed by a reviewer after the translation had been delivered. Along with severity, which regards the impact

of each error on the overall translation, also the importance of each error category requires consideration. In fact, the error classes that are used to perform a specific error analysis are given importance according to the purposes of the analysis and the type of texts under examination. Hence, a certain weight is assigned to each error class (Lommel, 2018).

Finally, the post-editing process analysis can be included in human evaluation techniques. With postediting is meant the whole set of operations necessary for a specific MT system's output to be published or delivered to the intended purchaser. In this specific case, in order to evaluate the candidate translation under examination, the needed intellectual and temporal effort for the evaluators to accomplish the post-editing operations is calculated. The less the number and extent of post-editing operations, the higher the quality of the candidate translation at stake (Chatzikoumi, 2020).

Having described the most widely used non-automatic and automatic methods for assessing Machine Translation, we can easily say that both evaluation approaches suffer from substantial shortcomings. On the one hand, human evaluation of Machine Translation has been strongly criticized for depending on a specific evaluator's or group of evaluators' linguistic knowledge and subjective opinion, and it is undoubtedly slower and more costly when compared to automatic evaluation metrics. Moreover, human Machine Translation evaluation implies not only a significantly costly and time-consuming implementation (Papineni et al., 2002) but also the impossibility of reproducing the same patterns an indefinite number of times and the risk of being too subjective, as human beings may be biased by a set of different factors, related to the external environment, as well as the evaluator's previous knowledge and physical or psychological conditions (Munkova et al., 2020). However, the scholars who support the human evaluation of Machine Translation pointedly remark the idea that a human assessment of Machine Translation is needed firstly because the output of an MT system is meant to be received, understood, and used by human beings. Secondly, only human perception of the world can efficiently detect eventual errors made by MT systems and assess their severity (Guzmán et al., 2015). Thirdly, professional human evaluators master the linguistic knowledge necessary to deeply analyze a translation, and integrating such a piece of knowledge into an automatic evaluation system undoubtedly would not only be difficult but also costly and suitable for just a limited number of language pairs.

Extensive research has been conducted over the decades to measure and improve the effectiveness of human assessment as an evaluation method applicable to the output produced by Machine Translation systems. In doing so, special attention has been devoted not only to estimating human beings' ease, speed, and consistency in performing the evaluation tasks but also to determining the most favorable conditions under which a proper human assessment can be performed. On the other hand, automatic evaluation metrics tend to equally weigh all the words, or n-grams contained in the translation, regardless of the degree of informativeness of their content. Hence, the substitution of articles or interjections risks being equalized by the system to one of highly informative parts of speech, such as nouns and verbs. This kind of evaluation is indeed responsible for penalizing the candidate translations that, although correct, present a low degree of similarity with the golden translations and being unable to properly detect long-distance linguistic relationships to provide a corpus-level quality assessment (Chatzikoumi, 2020). In addition, by comparing MT systems' output against a corpus of human reference translations, automatic Machine Translation evaluation metrics necessarily include in the evaluation process a subjective human element and limit the scope and accuracy of the assessment itself. Indeed, for a given source sentence, there is a great number of reliable translation variants, undoubtedly more than the ones that can be contained in a limited, although exhaustive corpus. However, not only automated translation evaluation certainly constitutes a fast and low-cost evaluation method but is also able to provide a degree of consistency and that cannot be reached by human beings alone. In addition, the recent increasing development of Neural Machine Translation has posed new technical challenges to machine translation evaluation metrics, be them human or automated, as a higher degree of precision is demanded. In fact, for the evaluation techniques to comply with the newly implemented Neural Machine Translation systems and constitute a useful tool for their development, they need to present a deeper sensitivity to linguistic nuances, be able to perform extensive analysis at the document level, and be specifically designed to focus on specific linguistic features (Chatzikoumi, 2020). As a consequence, a universal and commonly accepted Machine Translation evaluation method has not been implemented yet, and MTE remains an open question (Castilho et al., 2018).

2.3. Performing a comparative error analysis

Having provided a general overview of Machine Translation Evaluation, in the present section we will illustrate in more detail the methods of our research, which will be presented in the following chapter. As previously mentioned, we will translate the titles and the abstracts of three Russian highly specialized medical articles and the titles and the first paragraphs of three Russian popular-science medical texts, by using DeepL and Yandex. For the mentioned analysis to be easily readable and understandable by the intended readers, each article will be divided into several fragments according to its total length. Afterward, in order to compare the two popular Machine Translation systems' effectiveness and reliability, we will perform a manual comparative error analysis of their translations. We will indeed detect the errors committed by the two translation tools and provide a brief linguistic analysis. Moreover, each error will be marked as belonging to one or more of the following error categories, specifically selected according to the linguistic features of the texts and the language pair under examination:

- Syntax
 - Has the syntactical structure of the original document been respected?
 - How have the syntactical relations of the source text been correctly rendered?
 - Are the provided translations conform to the syntactical rules of the target language?
- Grammar
 - Does the translated text observe the target language grammatical rules?
 - Is the translated text easily readable and understandable by a native speaker?
- Lexis
 - Has the original meaning been adequately rendered in the target language?

- Use of the articles
 - Is the use of the articles consistent with the target language grammatical rules?
 - Does the use of the articles render the intended meaning?
- Acronyms
 - Have the acronyms of the source document been properly rendered?
- Terminology
 - Has terminology been used consistently and adequately?
 - Have domain-specific terms been properly conveyed?
- Culture-specific references
 - Have culture-specific references been adequately translated with their equivalent in the target culture?
 - How have culture-specific references been rendered in case no equivalents are found in the target culture?
- Theme-rheme pattern
 - Has the original theme-rheme pattern been correctly rendered?
- Omissions
 - Have any omissions been made?
 - Does the omission prevent the intended readers from a complete understanding of the text?
- Untranslated elements
 - Has any element of the original document been maintained in the source language?
 - Has the source text been entirely and adequately translated?
- Consistency
 - Does the translated text sound natural?
 - Is the translated text consistent?
- Orthography
 - Have words been correctly spelled?

- Have capitalization rules been respected?
- Transliteration
 - Have proper nouns been correctly transliterated?
- Format
 - Has the original document format been properly reflected?
 - Have the format conventions of the source language been respected?

Once marked all the errors, two tables will be compiled to provide a visual overview of the quality of the translations provided by the two translation tools. Each table will show how the erroneous fragments are displayed in the source document, in the provided translations, and a specifically made human translation, used as reference translation. Each erroneous fragment will be moreover associated with one or more of the above-mentioned error categories and then summed according to the error category within it falls. Afterward, we will analyze the obtained results, with regard to specialized texts, popular science texts as well as overall translation quality. Finally, several graphs will be drawn showing the distribution patterns of the translation errors.

2.4. An overview of DeepL Translate

DeepL translator was launched in 2017 by DeepL GmbH, known since 2009 as Linguee and based in Germany. The free-of-charge translation tool currently supports 11 languages, including English, German, French, Portuguese, Italian, Dutch, Polish, Russian, and the newly implemented Chinese and Japanese, added in March 2020. Moreover, DeepL translator allows the user to directly upload an up-to-5000-characters source document and supports three formats, namely .docx, .pptx, and .txt. In addition to automatically detecting the source language, it offers the user a set of functionalities with regard to the post-editing phase, such as a list of translation options, and the possibility of copying, saving, and sharing the translation or downloading it as a text file. Since September 2019 it is possible to integrate DeepL into Windows and Mac operating systems and made the translation tool available for all the computer programs and accessible via a keyboard shortcut. Along with a free-of-charge version, DeepL Pro

subscription is available and provides the users with additional functions, including unlimited text translation, the retention of the original text formatting, wider customization options, the possibility of ensuring the security of confidential documents, editing the translations and enabling their integration into CAT tools (www.deepl.com).

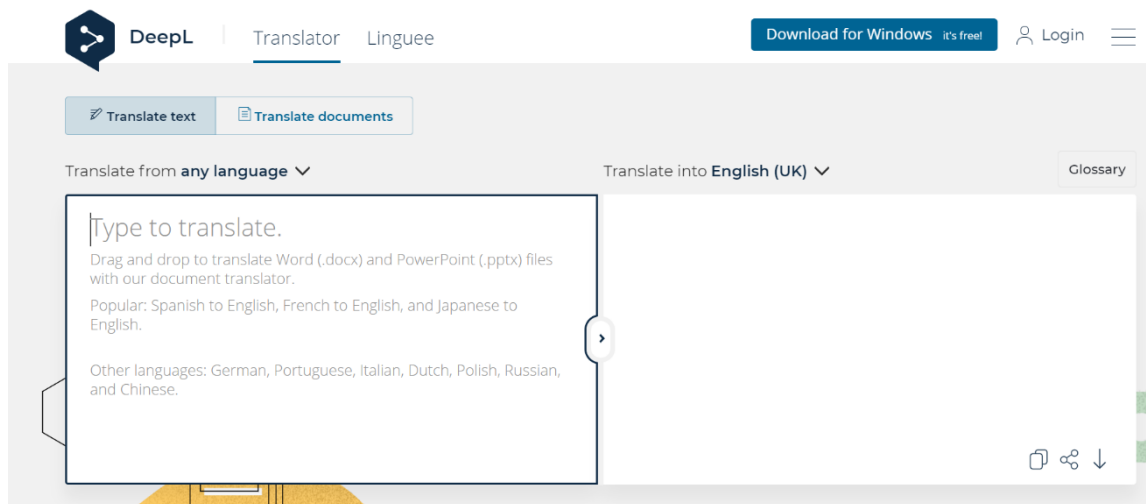


Figure 4 DeepL's translator user interface (www.deepl.com)

Although the company decided not to release a great deal of information concerning DeepL's architecture, merely mentioning the key role of artificial neural networks, several websites attribute its accuracy to the use of a particular type of artificial neural network, namely Convolutional Neural Networks (CNNs). By simultaneously detecting all the features of a source sentence, CNNs are indeed able to take into account the relationships between words within a sentence more extensively, when compared to Recurrent Neural Networks (RNNs). As a consequence, a machine translation system using CNNs may turn out to provide more natural and reliable output translations (Gehring et al., 2017).

2.5. An overview of Yandex Translate

Yandex Translate was launched in 2011 by the Russian Internet company Yandex as a Statistical Machine Translation tool mainly focusing on Russian, and more generally, Slavic languages. It consisted of three main components, namely a translation model, a language model, and a decoder. Given a specific language pair, after analyzing a great amount of source texts and their translations into the target language, the system created a comprehensive matrix containing source segments and their possible translations with an indication of their probability. The language model was based only on the target language and was aimed at analyzing the statistical likelihood of occurrence of words and word combinations, which were provided with an ID indicating their statistical frequency in that specific language. The actual translation process was carried out by the decoder, which combined the information provided by the translation model and the language model, analyzed it, and selected the most statistically relevant options. Therefore, texts were translated by retrieving from the matrix several translation options, which were evaluated according to the language model, and selecting the one that best complied with the target language at stake. Moreover, Yandex dictionary was in charge of the translation of single words and word sequences (www.Yandex.com).

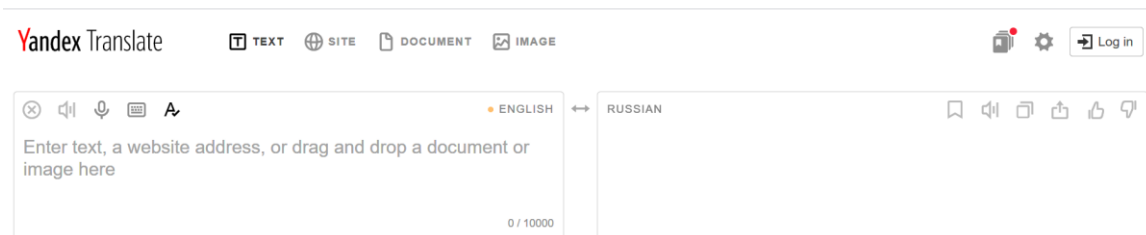


Figure 5 Yandex's user interface (www.Yandex.com)

Over the years, Yandex Translate has experienced continuous development, both in terms of architecture, and functionalities. In 2017 it shifted from a purely statistical translation system to a hybrid translation system, consisting of both a statistical and a neural approach. The text is consequently translated by using both the approaches and the provided translations are evaluated by Yandex gradient boosting, which selects the best option. The hybrid translation system permits to benefit from the quality potential of

Neural Machine Translation, which by analyzing the source sentences as a whole, and considering the overall context and consequently, is able to provide human-like and reliable translations, and simultaneously not sacrificing the benefits given by the statistical approach, which is better at translating words and word combinations that rarely occur in the training data. The hybrid translation system was initially launched for the English-Russian language pair and afterward implemented for several other language pairs over time (www.Yandex.com).

Yandex Translate currently supports 90 languages and automatically detects the source language. It provides the users with a great deal of functions, including the possibility to translate not only texts but also sites as well as images. Moreover, vocal synthesis, a dictionary, and a list of related words are available. Once the translation is completed, the user can add it to a collection, read, copy, share it, or give an assessment, by liking or disliking the provided translation. Yandex Translator is accessible via web as well as by downloading the mobile application, which additionally enables the users to consult an offline dictionary (Van Hees et al., 2015).

3. A COMPARATIVE ERROR ANALYSYS BETWEEN DEEPL AND YANDEX

In order to compare the effectiveness and reliability of DeepL and Yandex, we will perform a manual comparative error analysis of the translations provided by the two Machine Translation tools according to the criteria established in the previous chapter. For the purposes of our study, we have selected three Russian highly specialized medical texts and three Russian popular-science medical texts concerning the Coronavirus pandemic outbreak. The analysis and translation of the whole documents lay outside the scope of the present research. Nonetheless, several fragments of each document will be translated into Italian and analysed. More specifically, special attention will be devoted to the specialized texts' titles and abstracts, and the popular-science texts' titles and first paragraphs. In the following paragraphs, we will start analysing the three specialized texts and then move to the three popular-science ones. A table consisting of three columns will be used to simultaneously display the source text and the translations provided by DeepL and Yandex. For the texts' fragments to be exhaustively and clearly readable and understandable to the reader, we will divide the documents in different parts, according to their total lengths.

3.1. The biological therapy in the COVID-19 era

The first text, *Биологическая терапия в эру COVID-19* concerns the effectiveness and risks of the biological therapy as medical treatment against the background of Coronavirus SARS-CoV-2 pandemic outbreak. More specifically, the present research regards the likelihood that patients treated with the biological therapy contract the new Coronavirus SARS-CoV-2 and the implications that such a therapy has in patients that already contracted the virus.

SOURCE TEXT	DEEPL	YANDEX
Биологическая терапия в эру COVID-19	La terapia biologica nell'era COVID-19	Terapia biologica NELL'era COVID-19

We can easily notice that DeepL and Yandex correctly convey the general meaning of the title, with a difference in the use of the article. In fact, while in DeepL's translation the Italian singular feminine determined article *La* is used, in Yandex's one no articles introduce the subject of the sentence, namely *Terapia biologica*. According to Italian language grammatical rules, the determined article is used to introduce a noun representing an entity that has previously been mentioned in a given text, universally notorious or known by the intended reader (Lepschy & Lepschy, 1993). In this case, the document under examination concerns the effectiveness and risks of the biological therapy during the Coronavirus SARS-CoV-2 pandemic outbreak. In other words, the core issue is represented by the discussion regarding the biological therapy against the background of the Coronavirus pandemic outbreak and not the biological therapy itself. The biological therapy can be therefore considered as already well known by the intended reader, which in all probability is familiar with the medical environment. As a result, the translation version provided by DeepL is more exhaustive and grammatically correct than Yandex's one. Moreover, Yandex's translation presents a format error, as part of the title is written in capital letters, whereas in the original text lower-case letters are used.

In both translation variants, no prepositions introduce the term *COVID 19*. This may be due to the fact that it is interpreted as the name defining the Italian noun *era*, which constitutes the correct translation of the Russian term *эра* (Kovalev, 2020). Since the provided translations are able not only to convey the source document meaning but also to perfectly suit this specific context and text genre, they can be considered correct.

SOURCE TEXT	DEEPL	YANDEX
В связи с широким применением генно-инженерных биологических препаратов (ГИБП) в лечении иммуноопосредованных воспалительных заболеваний остро встает вопрос о дальнейшей терапевтической тактике ведения таких пациентов	In relazione all'ampio uso di preparati biologici geneticamente modificati (GIBP) nel trattamento delle malattie infiammatorie immuno-mediate, solleva la questione di ulteriori tattiche terapeutiche per tali pazienti, data la grave situazione epidemiologica causata dalla pandemia del	In connessione con l'ampia applicazione di farmaci biologici geneticamente ingegnerizzati (GIBP) nel trattamento delle malattie infiammatorie immuno-mediate, sorge acutamente la questione di ulteriori tattiche terapeutiche di riferimento tali pazienti, tenendo conto della grave situazione epidemiologica

с учетом тяжелой эпидемиологической обстановки, вызванной пандемией нового коронавируса SARS-CoV-2.	nuovo coronavirus SARS-CoV-2.	causata dalla pandemia del nuovo coronavirus SARS-CoV-2.
-----------------------------------------------------------------------------------------------------	-------------------------------	----------------------------------------------------------

In the first part of the abstract, significant lexical discrepancies between the two translation variants are observed. The Russian expression *в связи* is indeed translated by DeepL as *In relazione*, whereas the expression *In connessione* occurs in Yandex' s translation. The Italian term *connessione* conveys the figurative meaning of interdependence relationship between concrete or abstract entities (Zingarelli, 2020), which not only makes it a synonym of *relazione* but also a potentially correct translation for the Russian term *связь* (Kovalev, 2020). However, the Russian expression *в связи* is specifically translated into Italian as *in rapporto* or *in relazione* (Kovalev, 2020). Moreover, the Italian term *relazione* along with the simple preposition *in* constitute the collocation *in relazione*, which perfectly suits the context described by the document, whereas the same does not hold for the term *connessione* (Tiberii, 2018).

The Russian term *применение* can be translated into Italian as both *uso* and *applicazione*, as suggested by the two translation tools (Kovalev, 2020). Nonetheless, in the document it is used in association with the Russian term *препарат*, which indicates a laboratory product with pharmaceutical or cosmetic properties and can be properly translated into Italian with the general term *preparato* rather than the more specific *farmaco* (Kovalev, 2020). Since in Italian *uso* and *preparato* are often used in combination (Tiberii, 2018), the translation provided by DeepL undoubtedly sounds more natural to a native speaker.

Генно-инженерный is assigned different translations by the two translation tools, namely *geneticamente modificati* and *geneticamente ingegnerizzati*. Both adjectives refer to organisms whose genetic material has been modified by means of diverse genetic engineering techniques (Treccani, 2020). In Italian language, while the word sequence *geneticamente modificato* is widely used in highly specialized texts as well as popular-science ones, *geneticamente ingegnerizzato*, which directly derives from the Italian term *ingegneria*, is not only less common but also more frequently found in highly specialized

texts (Zingarelli, 2020). Since the document under examination is a medical highly specialized text, both variants are accepted. Furthermore, the Russian acronym ГИБП, which in the source document stands for *генно-инженерные биологические препараты*, is simply transliterated by both translation tools as *GIBP*. Although the transliteration is correct, the provided acronym does not convey any official meaning neither in English nor in Italian. Moreover, in the Italian and English language no existing equivalents of the acronym under examination have been found. As a result, since the mere transliteration cannot be considered in this case an acceptable translation variant, the omission of the acronyms would not represent a translation error.

In the translations, a complex preposition is used to express the specification relation between the Russian terms *лечение* and *заболевания*, correctly translated into Italian by both programs as *trattamento* and *malattie* respectively. When translating from Russian into a language that does not have cases, like Italian, the selection of the articles and prepositions conveying the syntactical structure of the sentences represents a core issue. In Italian, complex prepositions are created by directly adding an article to simple prepositions. In this case, the article is correctly inserted as contributes to conveying specificity and expressing the familiarity of the intended readers with the term it introduces.

The Russian adverb *оспо* is rendered by Yandex with the Italian adverb *acutamente*, whereas it is completely omitted by DeepL. Since the adverb under examination constitutes an important element of the original sentence as well as an integral part of the source document's meaning, DeepL's translation variant cannot be considered neither complete nor acceptable. Moreover, although the Italian adverb *acutamente*, used by Yandex, may represent a correct rendering of the Russian adverb *оспо* (Kovalev, 2020), it does not suit this specific context, as it is not adequate to accompany the Italian noun *questione*, which constitutes the proper translation for the Russian term *вопрос* (Kovalev, 2020). On the contrary, the Italian expression *con urgenza* may be a possible translation option, as it not only properly conveys the original meaning but also frequently comes in association with the Italian noun *questione*, so that the two terms constitute a collocation (Tiberii, 2018).

The main verb of the sentence under examination *встаем* is translated as *solleva* by DeepL Translate and *sorge* by Yandex Translate. In the source text, the verb is

preceded by the adverb *очпо*, which is, as mentioned above, ignored by DeepL and translated by Yandex as *acutamente* (Kovalev, 2020), and followed by the noun *вопрос*, properly translated by the two translation tools as *questione* (Kovalev, 2020) and representing its subject. The verbs selected by DeepL and Yandex are not only different in meaning but also in their syntactical relation with the Italian term *questione*. In DeepL's translation *questione* constitutes indeed the object of the verb, whereas in Yandex's one it assumes the role as subject. Moreover, in DeepL's translation the main verb does not have an actual subject and consequently cannot be considered grammatically correct. Being, in the original document, *вопрос* the subject of the verb *встаем*, Yandex's translation, which in addition displays a complete and correct grammatical structure, constitutes the only proper translation variant of this section of the sentence, as it conveys both the meaning and the structure of the source text.

The Russian term *тактике* is translated by both translation tools with the Italian noun *tattiche*. Starting from saying that, from a lexical point of view, it can be considered an acceptable translation variant (Kovalev, 2020), being the original term displayed in its singular form, the plural form used by DeepL and Yandex to render the Russian noun under examination and its corresponding adjectives cannot be accepted.

While the Russian term *ведение* is not translated by DeepL, it is rendered by Yandex with the Italian noun *riferimento*. Starting from underling that a more correct translation for *ведение* in this context would be the Italian term *gestione* (Kovalev, 2020), although DeepL does not translate the term, by adding the simple preposition *per*, it is able to accurately convey the meaning of the source sentence. On the other hand, Yandex Translate does translate, although inaccurately, the Russian term *ведение*, but it does not add any prepositions to link it with the Russian term *пациенты*, failing to render the specification relation that the two terms have in the source sentence, and consequently to provide a completely correct translation.

Finally, the Russian word sequence *с учетом* is rendered differently by the two translation tools. On the one hand, DeepL translates it with the Italian past participle *data*, correctly linked with the Italian noun *situazione*, which constitutes the correct rendering of the Russian term *обстановки* (Kovalev, 2020). On the other hand, in Yandex's translation, the Italian expression *tenendo conto della* is displayed. Being both translation

variants able to properly convey the original meaning (Kovalev, 2020), they can be considered equally acceptable.

SOURCE TEXT	DEEPL	YANDEX
В обзоре собраны актуальные данные о патогенезе COVID-19 с развитием острого респираторного дистресс-синдрома, обусловленного синдромом высвобождения цитокинов («цитокиновый шторм»). Рассматриваются влияние ГИБП на патогенез COVID-19 и их роль в лечении тяжелых форм COVID-19.	La revisione contiene dati aggiornati sulla patogenesi di COVID-19 con lo sviluppo della sindrome da distress respiratorio acuto causato dalla sindrome da rilascio di citochine ("tempesta di citochine"). L'influenza dell'GIBP sulla patogenesi del COVID-19 e il loro ruolo nel trattamento delle forme gravi di COVID-19 sono considerati.	La revisione raccoglie dati attuali sulla patogenesi di COVID-19 con lo sviluppo della sindrome da distress respiratorio acuto causata dalla sindrome da rilascio di citochine («tempesta di citochine»). Vengono considerati gli effetti della GIBP sulla patogenesi di COVID-19 e il loro ruolo nel trattamento delle forme gravi di COVID-19.

In the second part of the abstract, the Russian term *обзор* is translated by both translation tools with the Italian term *revisione*. Since in Italian language the term *revisione* can be defined as a review aimed at correcting or modifying errors and imperfections, a periodic check of machines and devices, or the philological adaptation of ancient texts (Zingarelli, 2020), it does not perfectly suit this context. A more appropriate translation, instead, would be the Italian terms *rassegna* and *panoramica*, which not only properly convey the meaning of the original sentence but also comply with its specific context (Kovalev, 2020).

The first verb *собрать*, which is used in the original text in the passive form, is correctly rendered by Yandex and DeepL in its active form and translated with the two different Italian verbs *contenere* and *raccogliere* respectively. Although the Russian verb *собрать* carries a meaning that is undoubtedly closer to the verb *raccogliere* than *contenere* (Kovalev, 2020), in this specific case both translations can be considered correct as they reach by different means the same result, namely properly rendering the

intended meaning and being understandable by the intended readers. By using the Italian verb *contenere*, DeepL indeed describes the effect of the action contained in the original text, i.e. the data are contained in the article because they have been previously collected. On the contrary, Yandex considers the action itself and consequently translates *собрать* as *raccogliere*. We can easily say that Yandex's translation is more accurate than DeepL's one, as perfectly interprets the meaning and the sense of the original sentence (Kovalev, 2020). Nonetheless, to an Italian speaker both translations not only sound easily understandable and natural but also seem to accurately reflect the original document's intents.

In the source text the noun *данные*, correctly translated by DeepL and Yandex as *dati* comes along with the Russian adjective *актуальные*, which is rendered by the two translation tools with the Italian adjectives *aggiornati* and *attuali*. These terms can be considered as synonyms in some contexts, and, by indicating entities that relate to the present time, they both potentially constitute a translation of the Russian adjective *актуальный* (Kovalev, 2020). However, they do not occur with the same frequency with the Italian noun *dati*. Indeed, the adjective *aggiornato* very frequently refers to *dati*, so that the two terms constitute an actual collocation. As a consequence, DeepL's translation not only accurately reflects contemporary Italian language but also provides a more natural translation of the sentence, when compared to Yandex's one.

The term *COVID-19* is introduced by DeepL by the complex preposition *del*, whereas in Yandex's translation variant the simple preposition *di* is displayed. Starting from saying that the mentioned prepositions properly render the specification relation expressed in the source document, defining the term under examination a specific and notorious disease, a definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). Moreover, the gender of the term *COVID-19* still constitute an open question of contemporary Italian linguistics. In fact, denoting the proper name given to a syndrome, which in the Italian language is a feminine noun, it should be feminine itself. However, since it appears in highly-specialized and popular-science articles as a masculine as well as feminine term, both a masculine and a feminine definite article is accepted. As a consequence, DeepL's version can be undoubtedly considered correct, whereas the same does not apply to Yandex's one.

The translation of the Russian past participle *обусловленного*, derived from the Russian verb *обусловят/обусловить* undoubtedly constitutes an interesting basis for discussion. Both translation tools correctly render it with the Italian verb *causare* (Kovalev, 2020). On the one hand, DeepL translates *обусловленного* with the singular masculine past participle *causato*, which reasonably refers to the Italian term *sviluppo*, i.e. the translation of the Russian *развитие*. On the other hand, Yandex as well translate it with the past participle of the Italian verb *causare*, but in its singular feminine form, *causata*, which is linked to *sindrome*, i.e. the translation of the Russian term *дисстресс-синдром*. In the original document, being *обусловленного* in the genitive case, it clearly refers to the term *острого респираторного дисстресс-синдрома*. As a consequence, in this case only Yandex is able to provide a correct translation and accurately reflect the grammatical structure of the source sentence.

Although the Russian word sequence *цитокиновый шторм* is correctly translated by both translation tools as *tempesta di citochine* (Kovalev, 2020), Yandex's translation contains a format error, as it is inserted into a hybrid pair of inverted commas, consisting of Russian inverted commas and Italian inverted commas. On the contrary, DeepL remains consistent with the Italian format and uses a pair of Italian inverted commas.

The last sentence of this section deserves special attention. The original document displays the Russian reflexive verb *рассматриваются*, which refers to the Russian terms *влияние* and *роль*. DeepL and Yandex properly render it with the passive form of the Italian verb *considerare* (Kovalev, 2020), as, according to Russian grammatical rules, the reflexive form of the verbs is employed, among its various uses, to express verb passivity. Nonetheless, a difference is observed between the two translations, namely the position of the Italian verb. DeepL indeed collocates it at the end of the sentence, whereas in Yandex's translation it is placed at the very beginning, before the terms to which it refers. Although both translation variants can be considered grammatically correct, Yandex's one seems to sound more natural to an Italian speaker as the main verb introduces the elements to which it refers as well as the whole sentence. Moreover, by collocating the verb under examination at the beginning of the sentence, Yandex perfectly reflects the theme-rheme pattern of the source document and the intents of its writer. In other words, in the source text, the final elements of the section under examination are

put into sharp focus as they constitute the new piece of information to which the author aims at drawing the reader’s attention. The same applies for Yandex’s translation variant, where the syntactical structure is organized for the focal points to be located at the end of the sentence.

The Russian term *влияние* is translated by DeepL and Yandex with two different Italian terms, namely *influenza* and *effetti* respectively. From a lexical point of view, the two terms can be considered in this context as synonyms and consequently both translation tools provide an appropriate translation (Zingarelli, 2020). However, the same does not hold for grammar. In the source document, the Russian term *влияние* is indeed used in its singular form and translated by DeepL with the singular noun *influenza* and by Yandex with the plural one *effetti*. From a grammatical point of view, Yandex’s translation is not accurate as it does not respect the number of the original Russian term.

The Russian acronym *ГИБП* stands for *генно-инженерные биологические препараты*, which refers to genetically engineered biological products (Korotaeva, 2015) and is translated by Yandex and DeepL with two different acronyms, namely *HIBP* and *GIBP*. Starting from saying that both acronyms do not carry an actual meaning neither in Italian nor in English, Yandex provides a correct transliteration of the original acronyms whereas DeepL does not. In this case, a possible correct translation option could be a direct translation of all the words contained in the acronym.

SOURCE TEXT	DEEPL	YANDEX
В обзоре отражены последние рекомендации международных ассоциаций/консенсусов и наблюдения врачей различных специальностей по вопросу прерывания/продолжения терапии ГИБП с оценкой последствий в случае прерывания биологической терапии.	La revisione riflette le ultime raccomandazioni delle associazioni/consensuali internazionali e le osservazioni dei medici di varie specialità sulla questione dell'interruzione/continuazione del trattamento HSI con valutazione delle conseguenze in caso di interruzione della terapia biologica.	La revisione riflette le ultime raccomandazioni delle associazioni internazionali di consenso e le osservazioni dei medici di varie specialità sulla questione dell'interruzione / continuazione della terapia GIBP con una valutazione delle conseguenze in caso di interruzione della terapia biologica.

In the last part of the abstract, the Russian past participle *отражены*, which is used as part of a passive construction and refers to the word sequence *последние рекомендации* is correctly translated by both translation tools with the Italian verb *riflettere* (Kovalev, 2020) in its active form. As a consequence, the whole sentence undergoes a profound transformation. Firstly, the Russian local adjunct introduced by the preposition *в* becomes the subject of the translated sentences and is translated by DeepL and Yandex with the Italian term *revisione*. As mentioned with regard to the previous section of the article, since in Italian *revisione* usually refers to a review aimed at correcting or modifying errors and imperfections, a periodic check of machines and devices, or the philological adaptation of ancient texts (Zingarelli, 2020), it does not represent a correct translation for this specific context. On the contrary, other terms such as *rassegna* and *panoramica* would be able to reflect the meaning and the context of the source document (Kovalev, 2020). Secondly, the subject of the original sentence, i.e. the Russian word sequence *последние рекомендации*, is correctly rendered by both translation tools as *ultime raccomandazioni* (Kovalev, 2020), and constitutes the object of the Italian verb *riflette*. Finally, the main verb switches from the passive form to the active one, and both translation tools succeed in maintaining the present tense used in the original sentence.

The Russian term *консенсус* is translated by DeepL and Yandex with two different Italian terms, namely the adjective *consensuali* and the noun *consenso*, linked by Yandex to the term *associazioni* with a specification relation. In this case, the translations provided by both translation tools cannot be considered correct, as the Russian term *консенсус* is a plural noun and it is used in association with the Russian adjective *международных*, correctly translated by both translation tools with the Italian adjective *internazionali*. For the target text to be consistent and complaint with the source document, the Italian terms *accordi* and *vertici* would be possible translation options as they frequently occur with the adjective *internazionali* (Tiberii, 2018), and properly suit the context within the Russian term *консенсус* falls in the original document.

The Russian term *терапия* is translated as *trattamento* by DeepL and *terapia* by Yandex. Starting from clarifying that both translations are acceptable (Kovalev, 2020), they differ in specificity. The term *trattamento* carries indeed a great deal of different meanings, which fall within a number of diverse fields, such as, among others, medicine,

economics, technology, cinematography (Zingarelli, 2020). On the contrary, in contemporary Italian language, the noun *terapia* is strictly related to the medical sphere (Zingarelli, 2020) and consequently perfectly suits the context of the original document.

As mentioned above, the Russian acronym *ГИБИ* stands for *генно-инженерные биологические препараты* (Korotaeva, 2015), and can be translated into English as genetically engineered biological products. Both DeepL and Yandex translate it with another acronym. Although the former does not adhere to the transliteration rules whereas the latter does, the acronyms provided by both translation tools do not exist neither in Italian nor in English. As a consequence, both translations cannot be considered correct as they fail in exactly conveying the meaning of the original text. A good variant could be instead the translation into Italian of all the words that constitute the Russian acronym.

The Russian term *оценка* is correctly translated by both translation tools with the Italian term *valutazione* (Kovalev, 2020). However, while in Yandex’s translation it is introduced by means of an indetermined article, in DeepL’s one no articles are used. In this case, the Russian term *оценка* reflects an entity that has not been already mentioned in the text as well as completely unknown by the intended reader. The indetermined article is indeed necessary to convey these nuances (Lepschy & Lepschy, 1993), and render Yandex’s translation more accurate when compared to DeepL’s one.

3.2. Clinical management of children with a disease caused by the new coronavirus infection (SARS-CoV-2)

The second specialised text, *Ведение детей с заболеванием, вызванным новой коронавирусной инфекцией (SARS-CoV-2)*, regards the medical treatment of children who contracted serious diseases triggered by Coronavirus SARS-CoV-2 infection. In particular, special attention is devoted to underline the prevention strategies, diagnosis, and treatment of the above-mentioned pediatric diseases and their potential complications.

SOURCE TEXT	DEEPL	YANDEX
Ведение детей с заболеванием, вызванным новой	Conduzione di bambini con una malattia causata da una nuova infezione da	Gestione dei bambini con una malattia causata da una nuova infezione da

коронавирусной инфекцией (SARS-CoV-2)	coronavirus (SARS-CoV-2)	coronavirus (SARS-CoV-2)
---------------------------------------	--------------------------	--------------------------

The translations of the title of the article undoubtedly deserve careful analysis. The Russian term *ведение* is translated by DeepL and Yandex with two different Italian terms, namely *conduzione* and *gestione*. Although they can be considered as synonyms in some contexts (Zingarelli, 2020), *gestione* more frequently comes in association with the Italian term *bambini* (Tiberii, 2018), i.e., the proper translation of the Russian term *детей* (Kovalev, 2020), and perfectly suits the specific context of the source text. As a consequence, Yandex's translation better reflects the meaning and structure of the original title. Moreover, the Russian terms under examination, namely *Ведение* and *детей* are linked in the source document with a specification relation, which is properly rendered in the provided translation variants with the Italian preposition *di*. Nonetheless, while in DeepL's version the simple preposition *di* is displayed alone, in Yandex's one it comes in association with the Italian plural masculine definite article *i*, so that the complex preposition *dei* is formed. In this specific case, the Italian term *bambini* given a certain degree of specificity by the rest of the title, namely *con una malattia causata da una nuova infezione da coronavirus (SARS-CoV-2)*. In other words, the article does not regard the children in general, but specifically the ones who contracted an illness related to the coronavirus infection. As a consequence, being *bambini* a specific term, the defined article is necessary for the title under examination to comply with Italian grammatical rules (Lepschy & Lepschy, 1993).

Aside from their first part, when comparing the translations provided by the two translation tools, we can easily notice that they are identical. DeepL and Yandex indeed correctly translate the Russian noun *заболеванием* as *malattia* (Kovalev, 2020) and succeed in properly conveying the syntactical relations of the original text, in particular with regard to the Russian past participle *вызванным*, rendered with the Italian feminine past participle *causata*, which clearly refers to the term *malattia*. Moreover, particularly accurate is the translation of the Russian adjective *коронавирусной*, which refers to *инфекцией* and is translated by both translation tools with the Italian word sequence *da Coronavirus*. This translation variant perfectly respects Italian grammatical and lexical rules, as the simple preposition *da* is often used to express the relation cause within

medical contexts (Treccani, 2020). Finally, in both translations, the Russian word sequence *новой коронавирусной инфекцией*, which assumes the role of agent, is introduced by the indefinite article *una*. As clarified in the previous chapter, in the Italian language, the indefinite article is frequently used to introduce a non-specific or still unknown entity, as well as an element that has never been mentioned in a given text (Lepschy & Lepschy, 1993). In this case, being the text under examination about the treatment of pediatric diseases caused by Coronavirus infection, we can reasonably assume that the intended reader not only is familiar with the medical environment in general but also has vast knowledge concerning Coronavirus infection. As a consequence, the indefinite article is unable to fully convey the sense of the source document and a definite article would perfectly suit the context and the meaning of the original title.

SOURCE TEXT	DEEPL	YANDEX
С целью обеспечения детского населения эффективной медицинской помощью в условиях пандемии новой коронавирусной инфекции Минздравом России совместно с профессиональными ассоциациями и экспертами в области педиатрии, инфекционных болезней и реанимации разработаны методические рекомендации «Особенности клинических проявлений и лечения заболевания, вызванного новой коронавирусной инфекцией (COVID-19), у детей».	Al fine di fornire alla popolazione infantile un'efficace assistenza medica nelle condizioni di una nuova pandemia di infezione da coronavirus, il Ministero della Salute della Russia, insieme ad associazioni professionali ed esperti nel campo della pediatria, delle malattie infettive e della rianimazione, ha sviluppato raccomandazioni metodologiche "Caratteristiche delle manifestazioni cliniche e trattamento della malattia causata dalla nuova infezione da coronavirus (COVID-19) nei bambini".	Con l'obiettivo di garantire alla popolazione pediatrica efficace aiuto medico in condizioni di pandemia di nuova coronavirusной infezione dal ministero della salute Russia in collaborazione con le associazioni professionali e di esperti nel campo della pediatria, malattie infettive e rianimazione sviluppato linee guida «Caratteristiche delle manifestazioni cliniche e trattamento della malattia, causata da un nuovo coronavirusной infezione (COVID-19), nei bambini».

When observing the translations provided by the two translation tools concerning the first part of the abstract, several lexical and grammatical discrepancies are observed. Starting from the very first sentence, the Russian multiword expression *С целью* is translated by DeepL with the Italian word sequence *al fine di*, whereas it is rendered by Yandex as *Con l'obiettivo di*. Both Italian terms *fine* and *obbiettivo* constitute accurate translation options of the Russian noun *цель* (Kovalev, 2020) and the expressions used by DeepL and Yandex properly suit the original text genre and specific context. As a consequence, the two translations can be considered equally correct.

The Russian noun *обеспечения* is translated by both translation tools with two Italian verbs. This translation option is particularly accurate, as the Italian expressions *al fine di* and *Con l'obiettivo di* are usually followed by verbs, which assume the role of object (Tiberii, 2018). Nonetheless, DeepL and Yandex chose two different verbs, namely *fornire* and *garantire*. Generally speaking, their meaning differs to various extents according to the contexts they fall within (Zingarelli, 2020). In this specific case, although they have a slightly different meaning, both *fornire* and *garantire* can be accepted as translation variants of the original sentence, as they are able to transmit to an Italian speaker the intents of the original document. Moreover, the lexical similarity between the two verbs is strengthened by the association with their objects (Tiberii, 2018), which will be discussed in the following lines.

The Russian word sequence *детского населения* is translated by DeepL as *popolazione infantile* and by Yandex as *popolazione pediatrica*. The Russian adjective *детский* can be translated into Italian as both *infantile* and *pediatrico* (Kovalev, 2020). The two Italian terms undoubtedly differ in their context of use (Zingarelli, 2020). *Infantile* is indeed used in non-specialized as well as highly specialized texts, whereas *pediatrico* frequently appears in specialized texts, and its meaning is strictly linked to the medical environment (Tiberii, 2018). Although both translations can be accepted, being the original document a highly specialized text, Yandex's translation variant better suits the specific features and the lexical nuances of the source text.

The translation of the Russian word sequence *Медицинской помощью* deserves careful analysis. It is indeed translated by DeepL and Yandex with two lexically similar Italian nouns, namely *assistenza* and *aiuto*. Moreover, in DeepL's translation variant, the singular feminine indefinite article *un'* is used to introduce the Italian noun *assistenza*,

whereas in Yandex's one, no articles precede the term *aiuto*. From a lexical point of view, although both terms can be considered correct translation options for the Russian noun *помощью* (Kovalev, 2020), the Italian noun *assistenza* frequently comes in association with the Italian adjective *medica*, so that they constitute an actual collocation (Tiberii, 2018). As a consequence, DeepL's translation not only properly conveys the meaning of the original document but also sounds more natural to an Italian speaker when compared to Yandex's one. As for grammar, in this specific case, the presence or absence of an article introducing the two Italian terms do not alter neither the meaning nor the syntactical structure of the whole sentences.

Both translation tools correctly translate the Russian term *условиях* with the Italian term *condizioni* (Kovalev, 2020) and succeed in rendering the temporal adjunct introduced by the Russian preposition *в*. Nonetheless, in DeepL's translation, the Italian noun *condizioni* is introduced by the complex preposition *nelle*, whereas in Yandex's one a simple preposition is used. As mentioned in the previous paragraph, complex prepositions are formed by adding defined articles to simple prepositions. In this specific case, a definite article better suits the context and accurately describes the coronavirus pandemic as an existing condition that is affecting our lives at a global level (Lepschy & Lepschy, 1993). Therefore, DeepL's translation can be considered an appropriate translation variant of this section of the document.

A similar situation arises in the following section. The Russian term *пандемии*, correctly rendered as *pandemia* by both translation tools (Kovalev, 2020), is indeed introduced by the simple preposition *di* in Yandex's translation, whereas in DeepL's one *in* is accompanied by the feminine indefinite article *una*. As mentioned above, being the Coronavirus pandemic universally known, both translation variants cannot be considered correct, as they fail in conveying the precise meaning of the original document. The addition of a definite article to the simple preposition *di* would instead perfectly suit this specific context. Moreover, DeepL inappropriately links the Italian adjective *nuova*, which constitutes the translation of the Russian adjective *новой* (Kovalev, 2020) and in the source text undoubtedly refers to *инфекции*, to the Italian noun *pandemia*, completely distorting the sense of the original document. On the other hand, in Yandex's translation variant, the Italian adjective *nuova* properly refers to the Italian term *infezione*, respecting the structure of the original document. Nonetheless, the Russian adjective

коронавирусной is not translated and remains in the source language. The same does not hold for DeepL, which properly renders the Russian adjective with the Italian expression *da Coronavirus*, complying with the target language grammatical rules (Treccani, 2020).

The translation of the Russian multiword expression *Минздравом России* represents a crucial point in our analysis. It is indeed rendered by DeepL and Yandex with two different Italian word sequences, namely Ministero della Salute della Russia and minister della salute Russia. Although they can be considered equally correct from a lexical point of view (Kovalev, 2020) the same does not hold for syntax and spelling. In fact, Yandex's translation variant does not use any prepositions to introduce the Italian noun Russia and consequently fails in properly rendering the specification relation expressed in the source document. As for spelling, denoting the word sequence under examination the official designation given to a governmental institution, the Italian terms minister and salute should begin with an upper-case letter (Treccani, 2020), which is not displayed in Yandex's translation. Moreover, generally speaking, for the translation to be more natural and readable by an Italian speaker, the term *Россия* could simply be translated with the Italian adjective *russo*. As for syntax, DeepL succeeds in conveying the structure of the source text, whereas Yandex does not. In the original document, *Минздравом России* plays the role of agent and is linked to the past participle *разработаны*. In DeepL's translation, *Ministero della salute della Russia* indeed correctly assumes the role of subject and is properly linked to its corresponding active verb, namely *ha sviluppato*, which constitutes the translation of *разработаны* (Kovalev, 2020). Moreover, DeepL not only maintains the tense of the original verb but also correctly transforms the original subject *методические рекомендации* into the object of this new sentence. On the contrary, in Yandex's translation the structure of the original sentence is maintained and the Italian multiword expression *Ministero della salute Russia*, introduced by the complex preposition *dal*, is given the role of agent. This translation variant per se could be correct if Yandex did not fail to be consistent with the syntactical structure of the sentence. The Russian verb *разработаны* is indeed rendered with the Italian past participle *sviluppato*, which not only does not meet the tense of the source document but is also completely unlinked to its subject, namely *linee guida*, which represents the translation of the Russian word sequence *методические рекомендации*.

The Russian adverb *совместно с* is properly rendered as *insieme ad* by DeepL and in *collaborazione con* by Yandex (Kovalev, 2020). Although the two Italian word sequences may carry slightly different meanings (Zingarelli, 2020), in this specific context they can equally convey the sense of the original document. As a consequence, both translation variants can be considered acceptable.

A major discrepancy between the two provided translation variants is observed in the second part of the section under examination. Although all the nouns are correctly rendered in the same way by the two translation tools (Kovalev, 2020), the same does not hold for the syntactical structure of the sentence. Firstly, in DeepL's translation variant, no articles are used to introduce the Italian terms *associazioni professionali* and *esperti*, whereas in Yandex's one the word sequence *associazioni professionali* is preceded by a plural feminine definite article. In this case, being the above-mentioned terms neither universally known nor already mentioned in the text, the definite article does not suit the context (Lepschy & Lepschy, 1993) and a simple preposition is sufficient to link the terms. Secondly, in Yandex's translation, the syntactical structure of the original document is not respected, as the Italian term *esperti* is introduced by the simple preposition *di* and seems to be directly linked to the term *associazioni* with a specification relation. In actual fact, in the original text, the two Russian terms *ассоциациями* and *экспертами* are collocated at the same syntactical level and linked by a coordinating conjunction. As a consequence, only DeepL's translation, in which these conditions are respected, can be considered acceptable. Finally, in DeepL's translation the Italian terms *pediatria*, *malattie infettive*, and *rianimazione* are introduced by three complex prepositions, whereas in Yandex's one only a complex preposition is used. Starting from underlining the correctness of using a complex preposition as an introduction of specific and universally known entities (Lepschy & Lepschy, 1993), DeepL's translation is more consistent when compared to Yandex's one, as the three terms, originally combined in a coordinating relationship, are equally preceded by three complex prepositions.

The Russian multiword expression *методические рекомендации* is rendered differently by the two translation tools. Although DeepL's variant *raccomandazioni metodologiche* can be considered acceptable (Kovalev, 2020), Yandex's one, namely *linee guida* undoubtedly better reflects contemporary Italian language. The two Italian

terms constitute indeed an actual collocation (Tiberii, 2018) and consequently sound more natural to an Italian speaker. As mentioned above, in Yandex’s translation the Italian word sequence is not correctly linked to its verb, whereas in DeepL’s one the original syntactical structure is fully respected. Therefore, in this case, Yandex’s translation option is preferable from a mere lexical point of view, whereas DeepL’s variant can be considered, on the whole, more accurate.

The Russian term *лечения*, properly rendered by DeepL and Yandex with the Italian noun *trattamento*, is not introduced by any prepositions in both translation variants. Being in the source document the term under examination linked by a coordinating conjunction with the Russian noun *проявлений*, the two terms are located at the same syntactical level and consequently linked with a specification relation to the Russian noun *обстановки*, which is properly rendered by both translation tools as *caratteristiche* (Kovalev, 2020). Nonetheless, since, as mentioned before, in the provided translations no prepositions precede the term *trattamento*, it seems to be located at the same syntactical level of the Italian noun *caratteristiche*. This undoubtedly contributes to misinterpreting the meaning of the original document and distorting its syntactical structure. The Italian simple preposition *di* is indeed necessary to provide an acceptable translation version.

As mentioned with regard to the previous lines, DeepL properly renders the Russian multiword expression *коронавирусной инфекцией* as *infezione da coronavirus* (Kovalev, 2020). On the contrary, in Yandex’s translation, the Russian adjective *коронавирусной* is erroneously maintained in the source language, as happened in the first part of the section under examination.

SOURCE TEXT	DEEPL	YANDEX
При разработке документа был учтен практический опыт специалистов не только нашей страны, но и зарубежных коллег. Особое внимание уделено доказательной базе представленных	Nello sviluppo del documento si è tenuto conto dell'esperienza pratica di specialisti non solo del nostro Paese, ma anche di colleghi stranieri. Particolare attenzione è stata prestata alla base di prova dei dati presentati,	Nello sviluppo del documento è stata presa in considerazione l'esperienza pratica di specialisti non solo del nostro paese, ma anche colleghi stranieri. Particolare attenzione è rivolta alla base di prove

<p>данных, а также вопросам эффективности и безопасности лекарственных препаратов, применяемых при лечении данной инфекции и ее осложнений.</p>	<p>nonché all'efficacia e alla sicurezza dei farmaci utilizzati nel trattamento di questa infezione e delle sue complicazioni.</p>	<p>dei dati presentati, nonché all'efficacia e alla sicurezza dei farmaci utilizzati nel trattamento di questa infezione e delle sue complicanze.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------

When comparing the two translation variants of the second part of the abstract, several discrepancies are observed. The Russian term *разработка* is translated by both translation tools with the Italian term *sviluppo* and preceded by the complex preposition *nello*, which correctly renders the Russian preposition *при* followed by the prepositional case (Glazunova, 2016). While the original syntactical structure is indeed respected, from a lexical point of view, a better option would be the Italian term *elaborazione*, as it more accurately conveys the original document's meaning (Kovalev, 2020) and frequently refers to the Italian term *documento*, so that the two nouns constitute a collocation in Italian (Tiberii, 2018).

The main verb of the first sentence, namely the past passive form of the Russian verb *учесть* is rendered differently by the two translation tools. On the one hand, in DeepL's translation, it is indeed rendered with the past impersonal form of the Italian verb *tenere conto*, afterward correctly linked to its original subject with a specification relation. On the other hand, Yandex uses the past passive form of the Italian verb *prendere in considerazione*, fully reflecting the syntactical structure of the original document. Even though the same clearly does not hold for Yandex's translation variant, it does not represent a translation mistake. Therefore, since the original meaning is accurately conveyed from a lexical point of view (Kovalev, 2020), and both the syntactical structure and the grammatical rules of the target language are respected, both translation variants can be considered fully acceptable.

The Russian word sequence *зарубежных коллег* is correctly translated by DeepL and Yandex as *colleghi stranieri* (Kovalev, 2020). However, while the two translation variants are equally accurate from a lexical point of view, the same does not hold for syntax. Indeed, while DeepL properly links the word sequence *colleghi stranieri* to the

preceding words and respects the specification relationship of the original document, in Yandex's translation the terms are not preceded by the simple preposition *di* and thus completely disconnected from the rest of the sentence. As a consequence, only DeepL's variant of the section under examination can be considered acceptable.

A major dissimilarity is observed in the rendering of the main verb of the second sentence of the present section. In fact, the passive form of the Russian verb *уделить* is rendered differently by DeepL and Yandex. DeepL renders it with the past passive form of the Italian verb *prestare*, whereas in Yandex's variant, the present passive form of the Italian verb *rivolgere* is used. Although the two Italian verbs may significantly differ in meaning, in this specific context they can be considered synonyms (Zingarelli, 2020). Moreover, in contemporary Italian language, they both frequently occur in association with the Italian noun *attenzione*, forming a natural-sounding collocation (Tiberii, 2018). Nonetheless, the provided translation variants do not share a key feature, namely the verb tense. Starting from saying that the Russian passive verb under examination can be rendered into Italian with a verb at past tense as well as the present tense (Glazunova, 2016), both translations can be considered correct as they succeed in conveying the meaning of the source text and their syntactical difference does not affect the rendering of the whole sentence.

The Russian word sequence *доказательной базе* is given two slightly different translation variants, namely *base di prova* and *base di prove*. Since the two provided word sequences not only do not reflect the meaning of the original document but also do not present an accepted sense in the Italian language (Zingarelli, 2020), they cannot be considered correct. The Italian expression *all'evidenza scientifica dei dati presentati* may constitute a possible translation option as it is able to describe the idea described in the source text, namely the special attention that is given to the scientific evidence of the presented data (Kovalev, 2020).

The Russian term *вопросам* is completely ignored by both translation tools and its rendering is consequently omitted. Despite the omission, the two translation variants do not fail in fully conveying the meaning of the source sentence and they can therefore be considered acceptable. Nonetheless, the integration of the term under examination, which can be rendered with the Italian noun *questione* (Kovalev, 2020) within the provided translations, would make them undoubtedly more accurate and complete.

A final discrepancy between the provided translation variants regards the Russian noun *осложнений*, which is rendered as *complicazioni* by DeepL and *complicanze* by Yandex. In this case, since the Italian terms suggested by the two translation tools can be considered synonyms (Zingarelli, 2020) and equally properly convey the source document’s meaning (Kovalev, 2020), both translation variants are accepted.

SOURCE TEXT	DEEPL	YANDEX
В статье на основании указанных методических рекомендации авторы освещают вопросы профилактики, диагностики, лечения патологических состояний, обусловленных COVID 19. Тактика ведения пациента представлена в зависимости от возраста и степени тяжести течения болезни, терапия рассмотрена с позиции этиологической, патогенетической и симптоматической направленности.	Nell'articolo sulla base delle linee guida citate gli autori evidenziano la prevenzione, la diagnosi e il trattamento delle condizioni patologiche causate da COVID 19. La tattica di gestione del paziente viene presentata a seconda dell'età e del grado di gravità della malattia; la terapia viene considerata a partire dall'orientamento eziologico, patogenetico e sintomatico.	Nell'articolo, sulla base di queste raccomandazioni metodologiche, gli autori coprono le questioni di prevenzione, diagnosi, trattamento di condizioni patologiche dovute a COVID 19. Le tattiche di gestione del paziente sono presentate in base all'età e alla gravità del decorso della malattia, la terapia è considerata da una posizione di orientamento eziologico, patogenetico e sintomatico.

The third part of the abstract is characterized by several dissimilarities between the two provided translation variants. Starting from the beginning of the first sentence, we can easily notice that in Yandex’s translation a comma is inserted between the Italian term *articolo* and the following words, whereas in DeepL’s one, no punctuation marks are used in the given section. According to Italian grammatical rules, in this case, the comma is not mandatory (Treccani, 2020). Nonetheless, it undoubtedly contributes to enhancing the clarity of the whole sentence and, more generally, conveying the meaning of the original document. As a consequence, Yandex’s version is preferable.

The Russian participle *указанных* is rendered by DeepL and Yandex with two different adjectives, namely *cite* and *queste*, respectively. *Указанный* constitutes the

past participle form of the Russian verb *указать* (Kovalev, 2020) and is used in the original document as an attribute referring to the terms *методических рекомендаций*. Being *указать* usually translated into Italian with the verbs *spiegare*, *indicare*, *comunicare* (Kovalev, 2020), DeepL's translation variant can be considered acceptable, as it is able to convey the meaning displayed in the source document, where the Russian participle is used to express the idea that the guidelines are cited in the article. On the contrary, since, as just said, the source sentence refers to the guidelines discussed throughout the article and not previously mentioned in the abstract, the Italian adjective *queste*, displayed in Yandex's translation cannot be considered correct.

The Russian word combination *методических рекомендаций*, contained in the first part of the abstract, occurs here a second time. While in the first rendering it is translated as *raccomandazioni metodologiche* by DeepL and *linee guida* by Yandex, in the second one, the same Italian terms are used, but oppositely. In other words, DeepL renders it as *linee guida*, whereas in Yandex's translation the Italian word combination *raccomandazioni metodologiche* is used. As mentioned with regard to the first part of the abstract, although from a lexical point of view both variants can be considered acceptable (Kovalev, 2020), the Italian word sequence *linee guida*, displayed in this case in DeepL's translation, more frequently occurs in contemporary Italian, so that the two constituting terms became a collocation (Tiberii, 2018). Therefore, since DeepL's variant undoubtedly sounds more natural to an Italian speaker and properly conveys the meaning and the sense of the original document, it is considered more accurate when compared to Yandex's one.

The Russian verb *освещают*, along with its object *вопросы* are rendered differently by the two translation tools. On the one hand, in DeepL's variant *освещают* is translated with the Italian verb *evidenziare*, whereas *вопросы* is not assigned an equivalent in the target language. On the other hand, Yandex translates the Russian terms under examination as *coprono le questioni*. When comparing the two verbs suggested by the translation tools, only the Italian verb *evidenziare* can be considered an accurate translation option. The Italian verb *coprire* is indeed not only unable to properly convey the meaning of the original verb but also does not match with the Italian term *questione* (Tiberii, 2018). Moreover, although, as mentioned before, DeepL erroneously omits the Russian term *вопросы*, this is not considered a grave translation mistake, as the original meaning is preserved and the terms constituting the sentence are syntactically correctly

linked to one another. Hence, DeepL's variant is preferable, even though not completely correct.

Both translation tools accurately translate the Russian terms *патологических состояний* with the Italian word sequence *condizioni patologiche* (Kovalev, 2020) and comply with the syntactical structure of the original document. Both DeepL and Yandex are indeed able to render the specification relation that links the terms under examination with *trattamento*, which constitutes a correct translation option of the Russian noun *лечения* (Kovalev, 2020). Nonetheless, a discrepancy can be observed in the preposition used to link the above-mentioned terms. In fact, while DeepL uses the complex preposition *delle*, consisting of the simple preposition *di* and the plural feminine article *le*, in Yandex's version, no articles are added to the simple preposition *di*. In this case, being the terms *condizioni patologiche* given a certain degree of specificity by its following words, namely *causate da COVID 19*, a definite article is necessary to introduce them and convey the meaning of the original text (Lepschy & Lepschy, 1993). As a consequence, DeepL's variant is considered more accurate when compared to Yandex's one.

The Russian past participle *обусловленных* is rendered by DeepL and Yandex as *causate da* and *dovute a*, respectively. Not only both translation options are correct from a lexical point of view (Kovalev, 2020) but they are also compliant with the gender and the number of the word sequence to which they are linked, namely *condizioni patologiche*.

The Russian noun *тактика* is rendered by the two translation tools with the same Italian term, namely *tattica*, but in different number forms. DeepL indeed uses the singular form, whereas in Yandex's translation the plural form is observed. Since in the original document the term at stake is used in its singular form, DeepL's variant is undoubtedly more accurate. However, from a lexical point of view, both translation options cannot be considered completely correct, as a more precise translation would be the Italian term *piano*, which better conveys the meaning of the source text and complies with this specific context, especially when considering the rest of the sentence (Kovalev, 2020).

A major dissimilarity is observed when analyzing the rendering of the Russian noun *течения*. While DeepL does not translate it, in Yandex's variant it is rendered with

the Italian term *decorso*. In this case, the omission of the noun under examination significantly distorts the sense of the translation and certainly affects its accuracy. Therefore, DeepL translation is not able to properly convey the original meaning and consequently cannot be considered acceptable.

In the source document, the Russian term *тяжести*, correctly rendered by both translation tools as *gravità* (Kovalev, 2020), comes in association with the Russian noun *степеню*, which is correctly translated by Yandex with the Italian term *grado*, and omitted in Yandex's translation variant. Although the omission of the term under examination does not lead to a serious distortion of the original meaning, its rendering clearly contributes to making Yandex's version more accurate, when compared to DeepL's one.

The last section constitutes an interesting basis for discussion. DeepL and Yandex provide two different translation versions, which deserve special attention. Despite the use of two different auxiliary verbs, the Russian past participle *рассмотрена*, along with its subject *терапия* are equally correctly rendered by DeepL and Yandex (Kovalev, 2020). Nonetheless, the second part of the sentence presents some issues in both variants, especially concerning the rendering of the Russian word sequence *с позиции направленности*. While DeepL translates it with the Italian expression *a partire dall'orientamento*, in Yandex's version a more literal translation is provided, and the Italian word sequence *da una posizione di orientamento* is used. Starting from saying that the rest of the sentence is accurately rendered by both translation tools (Kovalev, 2020), a more appropriate translation option for the Russian word sequence *с позиции направленности* could be the Italian multiword expression *a partire dal punto di vista*, which not only properly conveys the sense of the original document (Kovalev, 2020) but is also easily readable and natural sounding for an Italian speaker (Tiberii, 2018). The choice of using a definite article in association with the Italian preposition *da* is justified by the assumption that being the document under examination of highly specialized character, as mentioned before, the intended reader is significantly familiar with the medical environment and consequently owns extensive knowledge concerning the covered topics (Lepschy & Lepschy, 1993). In this case, DeepL's variant should be considered more accurate than Yandex's one in which an indefinite article is displayed.

3.3. Coronavirus SARS-Cov 2: complexities of the pathogenesis, search for the vaccines, and future pandemics

The third specialized article, *Коронавирус SARS-Cov 2: сложности патогенеза, поиски вакцин и будущие пандемии* regards the usefulness of applying the concept of peptide-protein relatedness continuum as a possible approach to better understand the complexities of the coronavirus pathogenesis and consequently develop an effective and safe vaccine against covid-19. Moreover, the likelihood of future coronavirus pandemics is discussed.

SOURCE TEXT	DEEPL	YANDEX
Коронавирус SARS-Cov 2: сложности патогенеза, поиски вакцин и будущие пандемии	SARS-Cov 2 coronavirus: patogenesi complessa, ricerche di vaccini e future pandemie	Coronavirus SARS-Cov 2: difficoltà di patogenesi, ricerca di vaccini e pandemie future

The renderings of the title of the article undoubtedly present several dissimilarities. Starting from its very beginning, we can easily notice that the two translation variants display different word orders. More precisely, while Yandex fully reflects the pattern of the source document and locates the term *coronavirus* before its official scientific designation *SARS-Cov 2*, DeepL reverses their original collocation. Although from a syntactical point of view both translations can be considered acceptable (Kovalev, 2020), Yandex’s version is clearly preferable, as it not only reflects the original title’s structure but also contemporary Italian lexis. Indeed, being coronavirus a common noun denoting a group of viruses rather than a specific virus, it comes in this case in association with the scientific designation given to the specific coronavirus responsible for covid 19, namely *SARS-Cov 2*. As a result, the word sequence *Coronavirus SARS-Cov 2* specifically constitutes the name indicating the virus that became extremely notorious with 2020’s pandemic outbreak (Giovine, 2020), and therefore represents the only acceptable translation option.

The second part of the title is doubtless not immune from translation issues. The rendering of the Russian word sequence *сложности патогенеза* is indeed assigned by DeepL and Yandex two different Italian equivalents, namely *patogenesi complessa* and

difficoltà di patogenesi, respectively. Starting from DeepL's translation variant, although the Italian adjective *complessa* and the Russian noun *сложности* convey similar meanings (Kovalev, 2020), they do not share the same syntactical role and hence cannot be considered proper translation equivalents. On the other hand, Yandex not only is unable to properly convey the intended meaning but it also makes a grave mistake, as it provides a nonsensical translation version. Indeed, denoting the Italian noun *patogenesi* the complex of modes of onset of a certain disease (Zingarelli, 2020), Yandex's translation variant, namely *difficoltà di patogenesi*, seems to convey the idea that the illness at stake, i.e. covid-19 encounters difficulties in arising, which undoubtedly does not correspond to the one expressed in the source document as well as it does not make any sense. In particular, the Russian term *сложности* is here considered as synonym of *трудности*, as frequently happens with regard to colloquial language. However, this correlation does not apply to the specific context at stake, where the Russian word sequence *сложности патогенеза* is used to indicate the complexities of COVID-19 pathogenesis (Kovalev, 2020), which includes a great deal of diverse elements. Indeed, the Italian expression *complessità della patogenesi* may constitute a possible translation option (Kovalev, 2020).

The Russian noun *поиски* is rendered by DeepL and Yandex with the same Italian term, displayed in two different grammatical numbers. Indeed, while DeepL renders it with the Italian noun *ricerche*, *ricerca* is suggested by Yandex. In this case, being the original Russian term in its plural form, DeepL's translation variant can be considered perfectly correct, as it properly conveys not only the meaning but also the grammatical features of the original Russian term (Kovalev, 2020), whereas the same does not apply to Yandex's one.

Both translation tools equally correctly (Kovalev, 2020) render the remaining part of the title. We can easily notice that the Italian nouns used by DeepL and Yandex are not linked by any kind of articles. In this specific case, being the terms under examination not specific and mentioned for the very first time, an article is not necessary to properly convey the intended meaning and suit the particular context of the title under examination (Lepschy & Lepschy, 1993).

SOURCE TEXT	DEEPL	YANDEX
<p>Актуальность. Вакцина против коронавируса SARS-Cov-2 рассматривается как наиболее перспективное средство для укрощения вызванной им нынешней пандемии и воспрепятствования возникновению новой. В числе трудностей создания вакцин выбор иммунодоминантных антигенов, обеспечивающих их эффективность и безвредность. Цель исследования -показать полезность применения концепции пептидного континуума родства белков (ПКРБ) для понимания сложности патогенеза Covid-19, поиска вакцин против Covid-19 и обсудить возможную природу будущих пандемий.</p>	<p>Pertinenza. Un vaccino contro il coronavirus della SARS-Cov-2 è considerato lo strumento più promettente per domare l'attuale pandemia e prevenire la comparsa di una nuova pandemia. Le sfide nello sviluppo del vaccino includono la selezione di antigeni immunodominanti per garantire l'efficacia e l'innocuità. Lo scopo dello studio è quello di dimostrare l'utilità del concetto di continuum della proteina peptidica (PPCR) per comprendere la complessità della patogenesi di Covid-19, trovare vaccini contro Covid-19 e discutere la possibile natura di future pandemie.</p>	<p>Attualità. Il vaccino contro il coronavirus SARS-Cov - 2 è considerato il mezzo più promettente per domare l'attuale pandemia che ha causato e impedire l'emergere di una nuova. Tra le difficoltà della creazione di vaccini, la scelta di antigeni immunodominanti che garantiscono la loro efficacia e innocuità. Lo scopo della ricerca - mostrare l'utilità dell'applicazione del concetto di пептидного continuum di parentela proteine (ПКРБ) per comprendere la complessità della patogenesi Covid-19, ricerca di vaccini contro Covid-19 e discutere la possibile natura delle pandemie future.</p>

In the first part of the abstract, the Russian term *Актуальность* is rendered by DeepL and Yandex with two different Italian terms, namely *pertinenza* and *attualità*, respectively. *Pertinenza* undoubtedly better conveys the original meaning (Kovalev, 2020) and contributes to making DeepL's translation version preferable when compared to Yandex's one.

The second sentence of the section under examination presents a number of translation issues. Starting from its very beginning, the Italian noun *vaccino*, which constitutes the proper translation of the Russian term *Вакцина* (Kovalev, 2020) is introduced by the singular masculine indefinite article *un* in DeepL's version, whereas in

Yandex's one the singular masculine definite article *il* is displayed. In this specific case, the Russian term *Вакцина* and consequently its Italian equivalent denote a non-specific and still theoretical entity. In other words, since a vaccine against coronavirus had not been developed yet when the article under examination was written, the name describing it refers to a general concept, i.e. the fact that only a vaccine may restrain the existing pandemic to exacerbate and prevent the emergence of a new one. As a result, an indefinite article would more appropriately contribute to conveying the original meaning (Lepschy & Lepschy, 1993).

Proceeding in order of appearance within the source text, in DeepL's translation variant, the term *SARS-Cov-2* is preceded by the complex preposition *della*. In this case, as mentioned before with regard to the title of the article, the word sequence *Coronavirus SARS-Cov-2* represents the name assigned to the virus responsible for covid 19 and, forming a compact ensemble, its constituting parts cannot be divided from one another. As a consequence, only Yandex's translation version can be considered acceptable.

DeepL and Yandex render the Russian term *средство* with two different Italian nouns, namely *strumento* and *mezzo*, respectively. Although the two terms may convey a great deal of different meanings according to their context of use (Zanichelli, 2020), they can be considered, in this specific case, as synonyms and represent, in actual fact, equally accurate translation variants of the Russian term under examination (Kovalev, 2020).

The Russian term *укрошения* is translated, by both translation tools, with the Italian verb *domare*. Starting from saying that, generally speaking, rendering the term under examination with a verb is a good translation option that perfectly reflects contemporary Italian language, from a lexical point of view, the verb chosen by DeepL and Yandex does not properly suit this specific context. Indeed, although the Italian verb *domare* may get lexically close to the Russian noun *укрошения* (Kovalev, 2020), it cannot be considered appropriate to the topic covered by the article and the term to which it refers, namely *pandemia*. In fact, although the association of the Russian term under examination and the medical context still persists in the colloquial language, it undoubtedly does not fit a highly specialized medical article. Instead, the Italian term *frenare* may represent a possible translation option, as it is not only able to convey the original meaning but also suits the medical context and frequently comes in association with the Italian term *pandemia* (Tiberii, 2018).

The Russian past participle *вызванной*, followed by the personal pronoun *им* is properly rendered by Yandex with the Italian relative clause *che ha causato* (Kovalev, 2020), whose subject is constituted by the term *coronavirus*, whereas it is completely omitted in DeepL's translation version. Although Yandex's variant can be considered correct, it may lead to a certain degree of ambiguity. In fact, by the sentence, it is not entirely clear whether the relative clause under examination refers to the term *coronavirus* or to *vaccino*. Logically, there is no doubt that 2020's pandemic has been caused by a virus and not by a vaccine. However, for the sake of completeness, the indication of the relative clause's referent may be useful for Yandex to provide a remarkably accurate as well as absolutely consistent translation. However, since the origin of the pandemic is implicit and notorious, the Russian past participle under examination does not constitute a fundamental piece of information and can be omitted without affecting the understandability of the whole section. As a consequence, also DeepL's version can be considered acceptable.

DeepL and Yandex render the Russian noun *воспрепятствования* with two different Italian verbs, namely *prevenire* and *impedire*, respectively. In this case, both translation variants may be considered acceptable from a syntactical as well as from a lexical point of view. The translation choice of rendering the Russian term under examination with a verb reflects indeed the grammatical rules of the target language and perfectly suits the syntactical structure of the provided versions. As for lexis, although the verbs suggested by the two translation tools may significantly differ in meaning (Zingarelli, 2020) according to their context of use, in this specific case, they are both able to convey the original meaning (Kovalev, 2020) and adequate to the particular topic discussed in the abstract under examination. In other words, since the core concept developed in this section of the abstract, namely the idea that a vaccine represents the most effective method to prevent a new pandemic from emerging, is properly rendered in meaning (Kovalev, 2020) and expressed with a grammatically correct structure in both translation variants, they can be considered equally accurate.

The Russian term *возникновению* is given two different translation options. Indeed, it is translated by DeepL with the Italian noun *comparsa*, whereas the verb *emergere*, used as a noun, is displayed in Yandex's translation. Being the two Italian terms not only synonyms in this specific context (Zingarelli, 2020) but also exact Italian

equivalents for the Russian term under examination (Kovalev, 2020), the two translation versions can be considered equally acceptable.

Particularly interesting is the rendering of the Russian adjective *новой*, which, in the source document, clearly refers to the Russian term *пандемии* and is rendered by DeepL as *una nuova pandemia*, and by Yandex simply as *una nuova*. Being *новой*, a nominal phrase with an implied head, the noun to which the Russian adjective *новой* originally refers is implied. As a consequence, the repetition of the Italian term *pandemia*, already presented as the translation of the Russian noun *пандемии*, is not downright necessary to properly convey the original meaning. Therefore, although DeepL's translation version cannot be considered erroneous, it may result redundant. On the contrary, Yandex one not only perfectly reflects the meaning (Kovalev, 2020) and syntactical structure of the source document but also turns out to be more natural and easily readable by an Italian speaker.

The Russian word sequence *В числе трудностей* is rendered differently by the two translation tools. On the one hand, DeepL entirely reorganizes the original sentence as follows. The Russian term *трудностей* is correctly translated with the Italian noun *sfide* (Kovalev, 2020), which is assigned the role of subject of this new sentence and linked to the Italian noun *sviluppo*, constituting the proper translation of the Russian term *создания* (Kovalev, 2020) with the complex preposition *nello*. Moreover, the Italian verb *includono* is inserted. On the other hand, Yandex fully respects the original structure and correctly renders the Russian expression *В числе трудностей* with the Italian word sequence *Tra le difficoltà*, which is linked to the Italian term *creazione*, i.e. the equivalent of the Russian term *создания*, with a specification relation, which exactly reflects the genitive case used in the source document. In addition, the translation provided by Yandex is a nominal sentence, as no verbs are added. Both translation versions present a number of issues, which we will enumerate in order of appearance within the text. Firstly, although the Italian term *creazione* may represent a possible translation equivalent for the Russian noun *создания* in some contexts (Kovalev, 2020), it does not usually refer to the Italian noun *vaccino* (Tiberii, 2020), whereas the term suggested by DeepL, namely *sviluppo* is preferable, as it suits this particular context (Tiberii, 2018). Secondly, when translating the Russian plural genitive *вакцин* as *del vaccino*, DeepL properly renders the specification relationship expressed in the source document but does not respect the

original grammatical number of the Russian term. On the contrary, Yandex is able to provide a grammatically correct translation version. Thirdly, the Italian term *vaccino* is introduced by the complex preposition *del* by DeepL, whereas in Yandex's translation version the simple preposition *di* is displayed. In this case, having the Italian term under examination already been mentioned in the present abstract, a definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). Finally, according to Russian grammatical rules (Glazunova, 2016), in the original sentence, the verb *есть*, i.e. the third singular person of the Russian verb *быть*, which corresponds to the Italian verb *essere* (Kovalev), is omitted, and both translation tools provide an acceptable version of the section under examination. Indeed, the Italian verb *includono*, added by DeepL, contributes to properly conveying the original meaning and, as mentioned above, is inserted into a well-formed syntactical structure. On the other hand, Yandex suggests a nominal sentence, as no verbs are added. Although nominal sentences may result to be misleading for the readers, in this case, the one provided by Yandex fully reflects the meaning (Kovalev, 2020) and the syntactical structure of the source document and can be therefore considered acceptable.

The Russian noun *выбор* is translated by DeepL and Yandex with two different Italian nouns, namely *selezione* and *scelta*, respectively. Although, generally speaking, the two terms convey a similar meaning (Zingarelli, 2020), they present a different degree of specificity. Indeed, while in the Italian language, the noun *scelta* is intended to define a choice that is made between an indefinite and potentially infinite number of possibilities (Zingarelli, 2020), *selezione* refers to a choice between a restricted and well-defined number of possibilities. In this particular case, being the number of available antigens restricted and, in any case, not unlimited, the term *selezione* undoubtedly better conveys the intended meaning (Kovalev, 2020) and suits this specific context.

The Russian present participle *обеспечивающих* surely constitutes an interesting basis for discussion. It is translated by DeepL with the Italian infinitive verb *garantire*, preceded by the simple preposition *per*, and by Yandex with the relative clause *che garantiscono*. Starting from saying that the present participle under examination derives from the Russian verb *обеспечивать*, whose correct translation into Italian is represented by the verb used by both translation tools, namely *garantire* (Kovalev, 2020), Yandex's translation version better reflects the syntactical structure of the source document and

constitutes the most proper rendering of the Russian present participle, which is usually translated into Italian with a relative clause (Glazunova, 2016).

The rendering of the possessive adjective *их*, which in the source document refers to the Russian terms *эффективность* and *безвредность*, correctly translated by both translation tools as *efficacia* and *innocuità* (Kovalev, 2020) constitutes another discrepancy between the two provided translation versions. Yandex indeed properly renders it with the Italian possessive adjective *loro*, whereas in DeepL's version it is completely ignored. In this specific case, the omission of the possessive adjective may cause ambiguity and prevent the intended readers from fully understanding the provided translation variant. As a consequence, only Yandex's translation can be considered acceptable.

Starting from saying that the Russian term *исследования* is translated by DeepL and Yandex with two different Italian terms, namely *studio* and *ricerca*, which can be equally considered proper translation options for the Russian term under examination (Kovalev, 2020), the rendering of the remaining part of the sentence undoubtedly presents some issues. In Yandex's translation version, indeed, the structure of the original sentence is not adjusted according to Italian grammatical rules and the hyphen, used in the Russian language as a substitute of the verb to be, is maintained. Since it is not an element of the Italian contemporary language, it cannot be considered acceptable and needs to be given an equivalent that is compliant with the target language's grammatical rules. In this specific case, DeepL's translation, which displays the Italian verb *è* followed by the Italian pronoun *quello*, may represent a proper translation variant of the sentence under examination, as it accurately conveys the intended meaning (Kovalev, 2020) and is able, at one time, to properly express it in the target language. Moreover, the Russian verb *показать* is translated by DeepL and Yandex with two different Italian verbs, namely *dimostrare* and *mostrare*, respectively. Although they both can be considered, generally speaking, possible translation equivalents of the verb *показать* (Kovalev, 2020), they convey a quite different meaning. Indeed, while the Italian verb *mostrare* refers to the action of simply showing something to a person or a group of people (Zingarelli, 2020), *dimostrare* may denote in some cases the same action, but aimed at convincing those to which the action is addressed of its usefulness. When considering the specific context at stake, we can easily say that DeepL's version is more adequate than Yandex's one.

Nonetheless, DeepL's translation version is not immune to issues. The Russian term *применения* is indeed completely omitted by the translation tool, whereas in Yandex translation it is correctly rendered with the Italian term *applicazione* (Kovalev, 2020). Representing the term under examination a meaningful element of the sentence, its omission prevents DeepL's translation version from properly conveying the intended meaning and consequently being considered acceptable.

The Russian multiword expression *концепции пептидного континуума родства белков (ПКРБ)* undoubtedly represent an interesting basis for our analysis. Both translation variants provided by DeepL and Yandex, namely *concetto di continuum della proteina peptidica (PPCR)* and *concetto di пептидного continuum di parentela proteine (ПКРБ)* cannot be considered acceptable. It is worth noting that, although, as mentioned above, neither of the versions is correct, in Yandex's one the Russian adjective *пептидного* is maintained in the source language and the same applies to the Russian acronym (*ПКРБ*), which is not even transliterated. Constituting the Russian expression under examination a specific term which is used by the author of the article we are analysing but do not present any close correspondence in the Russian scientific literature, it undoubtedly represents an interesting translation issue. In this case, we decided to try to explain the scientific phenomenon described by the Russian term *концепции пептидного континуума родства белков (ПКРБ)*, taking as a reference the existing Russian scientific literature regarding the structure of proteins. As a consequence, the Italian expression *concetto di molteplicità della relazioni delle strutture proteiche* may represent a possible translation option, as it seems to express the intended meaning of the source text (Volkenštein, 1975). By doing so, we focus on making the Italian translation of the document under examination as understandable as possible to the intended readers. In addition, since no acronyms for this technical term are found in the Italian scientific literature, the Russian acronym (*ПКРБ*) can be omitted.

The Russian term *патоге́неза*, properly rendered by both translation tools as *potogenesi* (Kovalev, 2020) is linked with its related term *Covid 19* by the simple preposition *di* in DeepL's translation version, whereas in Yandex's one no prepositions are used to connect them. Defining in the Italian language the term *patogenesi* the typology of onset and development of a certain illness (Zingarelli, 2020), it needs to be linked with the name of the specific illness to which it refers by a specification relation.

As a consequence, Yandex's translation cannot be considered acceptable. On the other hand, although DeepL's version expresses the above mentioned and necessary specification relation, being the term *covid 19* the specific and extremely notorious name of an illness, a definite article has to come in association with the simple preposition *di*. The gender of the term *covid 19* is still a controversial issue, as, since it refers to the name of a syndrome, which in the Italian language is feminine, the term should be logically considered feminine itself. However, it is circulating in both specialized and non-specialized literature with the feminine as well as masculine gender (Giovine, 2020). As a consequence, both a definite masculine and a definite feminine article would be accepted.

The same applies when the term *covid 19* occurs again, later in the sentence, after the Russian preposition *пpомus*, which is correctly rendered by DeepL and Yandex as *contro* (Kovalev, 2020). In both translation variants, no articles indeed introduce the term. Describing *covid 19*, as mentioned above, a specific and well-known syndrome, a definite article, either masculine or feminine (Giovine, 2020), is necessary to properly express the intended meaning.

Particularly worth mentioning is also the fact that the two terms *пoнимания* and *пoуска* are rendered in DeepL's translation variant by using two Italian verbs. Indeed, in this case, being the two terms under examination along with the Russian verb *oбсyдить* linked by coordinating conjunctions and consequently located at the same syntactical level, this translation choice undoubtedly contributes to creating a certain constituency throughout the sentence. However, the Russian noun *пoуска* is translated with the Italian verb *trovare*, which does not perfectly correspond to the meaning expressed in the source document. In fact, since the Russian term *пoуска* is usually translated into Italian with the noun *ricerca*, the Italian verbs *cercare* and *ricercare* may represent possible and certainly more accurate translation option. As for Yandex's translation variant, it is equal to DeepL's one, except for the rendering of the Russian term *пoуска*, which is translated with the Italian noun *ricerca*. Although it may represent a proper Italian equivalent of the Russian term under examination (Kovalev, 2020), in this specific case, as mentioned above, a verb would more adequate and consistent with the rest of the sentence.

Finally, the Italian word sequence *future pandemie*, which constitutes the correct equivalent given by both translation tools to the Russian terms *бyдyщих пандемий*

(Kovalev, 2020) is introduced by the simple preposition *di* in DeepL’s translation variant, whereas in Yandex’s one the complex preposition *delle* is displayed. In this case, denoting the expression under examination not an existing entity, but an event that is likely to happen in a hypothetical future, there is no need to insert a definite article in order to convey the intended meaning (Lepschy & Lepschy, 1993).

SOURCE TEXT	DEEPL	YANDEX
<p>Материалы и методы. Для выявления компьютерным анализом пептидного (иммуноэпитопного) родства S, M и N белков SARS-Cov-2 с белками человека и других вирусов был выполнен поиск гомологичных последовательностей. Источниками первичных последовательностей белков служили доступные в Интернете базы данных. Результаты. S-белку свойственно пептидное (иммуноэпитопное) родство со многими белками человека, локализующимися на поверхности клеток или циркулирующими в крови, и вирусов. Образование антител к SARS-Cov-2, перекрестно реагирующих с гомологичными последовательностями в белках человека, может</p>	<p>Materiali e metodi. È stata effettuata una ricerca di sequenze omologhe per identificare la correlazione peptidica (immunoepitopo) delle proteine S, M e N della SARS-Cov-2 con quelle dei virus umani e di altri virus mediante analisi computerizzate. Le fonti delle sequenze proteiche primarie erano le banche dati disponibili su Internet. Risultati. La parentela peptidica (immunoepitopo) della S-proteina con molte proteine umane, localizzate sulla superficie cellulare o in circolazione nel sangue, e i virus è caratteristica. La formazione di anticorpi contro la SARS-Cov-2 che reagiscono in modo incrociato con sequenze omologhe nelle proteine umane può aggravare il decorso di Covid-19.</p>	<p>Materiali e tecniche. Per identificare l'analisi computerizzata della parentela peptidica (immunoepitopica) delle proteine S, M E N SARS-Cov-2 con proteine umane e altri virus, è stata eseguita una ricerca di sequenze omologhe. Le fonti di sequenze primarie di proteine erano database disponibili su internet. Risultati. La proteina S è caratterizzata da una relazione peptidica (immunoepitopica) con molte proteine umane localizzate sulla superficie delle cellule o circolanti nel sangue e virus. La formazione di anticorpi contro SARS-Cov-2, che reagiscono trasversalmente a sequenze omologhe nelle proteine umane, può appesantire il corso di Covid-19.</p>

отягощать течение Covid-19.		
--------------------------------	--	--

The second part of the abstract undoubtedly presents several translation issues. Starting from its very beginning, the Russian noun *методы* is translated by DeepL and Yandex with two different Italian terms, namely *metodi* and *tecniche*, respectively. Although the two nouns may convey a similar meaning (Zingarelli, 2020), *metodi* undoubtedly more frequently appears in scientific papers and often describes the particular process with which a specific scientific research is conducted. As a consequence, DeepL's translation variant better suits this context.

The second sentence of the section under examination certainly constitutes a core element of our analysis as the two translation tools provide significantly different translation variants. By taking a brief look, we can easily notice that Yandex's translation fully reflects the syntactical structure and word order of the source document. The same does not apply to DeepL's version, where the original sentence's order is reversed, and its syntactical structure readjusted accordingly. Proceeding in order of appearance in the source document, the Russian noun *выявления* is properly rendered by DeepL and Yandex with two different Italian verbs, namely *è stata effettuata* and *è stata eseguita*. Since these renderings perfectly suit the original sentence's context, meaning (Kovalev, 2020) and tense and the two suggested verbs may be considered, in this case, as synonyms (Zingarelli, 2020), both translation variants are equally acceptable.

The Russian word sequence *компьютерным анализом*, expressing, in the source text, the method used to conduct the research discussed throughout the article, is given completely different equivalents by the two translation tools. On the one hand, DeepL collocates the terms at the very end of the sentence under examination and, although it correctly renders their original syntactical role as a complement of means, as well as their meaning (Kovalev, 2020) it does not fully reflect their grammatical number, as translates them with the plural noun *analisi* and its corresponding plural adjective *computerizzate*, even though they are originally displayed in their singular form. On the other hand, in Yandex's translation version, the rendering of the terms at stake is erroneously given the role of the direct object, depending on the Italian verb *identificare*, completely distorting

the original meaning. As a consequence, Yandex's version of this particular section cannot be considered acceptable.

The Russian adjective *иммуноэпитопное* is incorrectly translated by DeepL as *immunoepitopo*, whereas the Italian adjective *immunoepitopica* is suggested by Yandex. Since, unlike *immunoepitopico*, *immunoepitopo* does not have any correspondence in Italian scientific literature (Treccani, 2020), only Yandex's version can be considered acceptable.

The term *SARS-Cov-2* is introduced, in DeepL's translation variant, by the complex preposition *della*, whereas in Yandex's one no linking elements are displayed. Being in the source document the term under examination clearly linked to the preceding Russian noun *белков*, which is correctly translated as *proteine* (Kovalev, 2020) by both translation tools, with a specification relation, the simple preposition *di* is necessary to properly render the syntactical structure of the original sentence. Moreover, since the term *SARS-Cov-2* constitutes the specific and extremely notorious scientific designation assigned to the virus responsible for 2020's pandemic, a masculine definite article has to come in association with the before mentioned simple preposition. As a result, the complex preposition *del* may constitute a possible translation option, as it is able to convey the intended meaning and properly reflect the source text's sentence.

The Russian term *родства* is rendered differently by the two translation tools. Indeed, DeepL translates it with the Italian noun *correlazione*, whereas in Yandex's translation variant the Italian term *parentela* is displayed. Since *parentela* not only constitutes a more accurate Italian equivalent for the Russian term at stake (Kovalev, 2020) but is also frequently used in association with the term *proteine* in Italian scientific literature (Treccani, 2020) only Yandex's version can be considered acceptable.

Yandex's translation variant displays a format error, as the coordinating conjunction *e*, located between the two proteins' proper nouns *M* and *N*, is capitalized. In this case, being *e* a conjunction collocated in the middle of a sentence, according to Italian grammatical rules, a lower-case letter is needed (Accademia della Crusca, 2020).

In DeepL's translation variant, the Russian word sequence *с белками человека* is rendered as *con quelle dei virus umani*. Starting from saying that, since the Italian term *proteine* is already mentioned shortly before in the sentence, the Italian pronoun *quelle* represents a correct translation option for the Russian term *белками*, the translation tool

fails in rendering the remaining part of the section under examination. Indeed, the Italian noun *virus* is inserted, even though it does not have an equivalent in the source text. On the other hand, although Yandex adequately renders the Russian genitive *человека* with the Italian adjective *umane*, properly conveying the intended meaning (Kovalev, 2020) and providing a natural-sounding translation, it fails in properly reflecting the specification relation expressed by the Russian genitive *других вирусов*, as no prepositions are used to introduce the Italian terms *altri virus*, which, from a mere lexical point of view, would constitute a correct translation variant (Kovalev, 2020).

Moving to the following sentence, further translation issues deserve special attention. Firstly, the Russian word sequence *первичных последовательностей белков* is rendered as *sequenze proteiche primarie* by DeepL and *sequenze primarie di proteine* by Yandex. Starting from underling that both versions can be considered acceptable (Kovalev, 2020), DeepL's rendering of the Russian genitive *белков* with the Italian adjective *proteica*, correctly linked with the Italian noun *sequenza*, not only properly conveys the intended meaning but also better reflects Italian medical language, where the two terms constitute an actual collocation (Treccani, 2020). Moreover, the rendering of the Russian word sequence under examination is introduced by the complex preposition *delle* in DeepL's translation version, whereas the simple preposition *di* is used in Yandex's one. Although they are both intended to express the specification relation displayed in the source document, denoting *первичных последовательностей белков* a well-defined and finite number of entities, a definite article has to come in association with the simple preposition *di*, for the provided translations to properly convey the original meaning (Lepschy & Lepschy, 1993).

Secondly, both translation tools render the Russian verb *служили* as *erano*. Since the provided translations reflect the meaning (Kovalev, 2020), the tense, the imperfective aspect, and the other grammatical features of the verb displayed in the source document, they can be considered acceptable.

Thirdly, the Russian multiword expression *базы данных* is translated by DeepL and Yandex with two different terms, namely *banche dati* and *database*, respectively. Both nouns can properly convey the source sentence's meaning (Kovalev, 2020) and perfectly suit this specific context. In particular, although *database* officially constitutes an English term, it has become so frequently used at an international level, so that it also

became part of the Italian contemporary language (Zingarelli, 2020) and it is nowadays as widespread in the country as its original Italian equivalent. However, while, as said before, the provided translation can be considered equally accurate from a lexical point of view, the same does not hold for syntax. Indeed, in DeepL's translation variant the Italian term *banche dati* is introduced by the plural feminine definite article *le*, whereas in Yandex's one no articles precede database. In this case, indicating the original term under examination a non-specific number of databases, chosen between the ones available on the Internet, a definite article is not adequate to suit this particular context. In fact, the insertion of a definite article may convey the erroneous idea that all the databases available on the Internet have been used as sources for the mentioned research. As a consequence, Yandex's translation version is overall preferable when compared to DeepL's one.

Finally, although the Russian word *Интернете* is rendered by the two translation tools with the same Italian term *internet*, a spelling difference can be observed. In DeepL's translation version, it indeed begins with the capital letter, whereas in Yandex's one, a lower-case initial letter is used. Since, according to Italian grammatical rules, the term under examination can be written with an upper-case, as well as a lower-case first letter (Treccani, 2020), both translation variants can be equally accepted.

The following sentence is given totally different equivalents by the two translation tools. Indeed, while Yandex properly renders it by respecting the original meaning and the syntactical structure of the source document, except for some inaccuracies, which we will enumerate in the following lines, DeepL seems to completely distort the sense of the original text and provide a translation variant that cannot be considered compliant with Italian grammatical rules.

Starting from Yandex's translation version, the Russian genitive *вирусов* is simply rendered with the Italian term *virus*, which is linked to the preceding part of the sentence by the Italian coordinating conjunction *e*. Although it may represent a correct translation equivalent of the Russian term under examination from a lexical point of view (Kovalev, 2020), the translation tool fails in properly rendering the specification relation that links the Russian terms *вирусов* and *белками* in the original document. In this case, the complex preposition *dei* may be a correct introduction for the Italian noun *virus*. Moreover, depending the Russian genitives *человека* and *вирусов* on the same Russian

term *белками* and being consequently located at the same syntactical level, for the translation variant to remain consistent, the Italian expression *dell'essere umano* may constitute a good translation option for *человека* (Kovalev, 2020). As mentioned above, the remaining part of Yandex's translation variant can be undoubtedly considered acceptable.

As for DeepL's translation variant, it completely fails in rendering the syntactical relationships of the source document. In particular, the rendering of the Russian term *свойственно*, namely the Italian expression *è caratteristica*, although lexically correct (Kovalev, 2020), seems to be completely disconnected from the rest of the sentence, whereas in the original text it refers to the Russian term *S-белку*, which is moreover erroneously translated as *S-proteina*. Even though the remaining part of the translation variant may be considered acceptable, the mentioned translation mistakes clearly prevent it from properly conveying the original meaning and consequently being considered acceptable.

The last sentence of the present section doubtless deserves our attention. The term *SARS-Cov-2* is introduced by the singular feminine definite article *la* in DeepL's translation version, whereas in Yandex's one no articles precede the term under examination. In this case, as mentioned above, constituting *SARS-Cov-2* the scientific designation given to a specific and notorious virus, the definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). Moreover, defining the term under examination a virus, which in the Italian language is masculine, its scientific name, according to Italian grammatical rules, should be masculine itself (Treccani, 2020). As a consequence, the Italian singular masculine article *il* would perfectly fit this specific context.

The Russian adverb *перекрестно* is rendered by DeepL and Yandex with two different Italian expressions, namely in *modo incrociato* and *trasversalmente*, respectively. In this case, the Italian adverb *trasversamente* is able to better convey the original meaning of the source document, when compared to the one suggested by DeepL (Kovalev, 2020).

The Russian verb *отягощать* is translated differently by the two translation tools. In fact, while DeepL renders it with the Italian verb *aggravare*, in Yandex's translation variant the Italian verb *appesantire* is displayed. In this case, the verb suggested by

Yandex not only better conveys the original meaning (Kovalev, 2020) but also perfectly suits this specific context, as it is frequently used in the medical field to describe the worsening of an illness (Zingarelli, 2020). As a consequence, Yandex’s version is undoubtedly preferable.

Finally, the two translation tools render the Russian noun *течение* with two different Italian terms, namely *decorso* and *corso*, which may constitute a possible correct translation option for the Russian noun under examination (Kovalev, 2020), but in this specific context cannot be considered equally adequate. In fact, the Italian noun *corso*, suggested by DeepL, conveys a general meaning and, as mentioned above, although it can be used as Italian equivalent for the Russian term *течение*, it does not perfectly fit this specific context as *decorso*, displayed in DeepL’s translation variant, does. Indeed, not only the term conveys a significantly close meaning to the one expressed in the source document (Kovalev, 2020) but it also frequently refers, in Italian contemporary language, to illness-related subjects (Tiberii, 2018). Moreover, both translation tools link the above mentioned Italian terms with the noun *covid 19* by using the Italian simple preposition *di*. As discussed above, defining *covid 19* the specific illness caused by the infamous *SARS COV 2*, a definite article should come in association with the simple preposition *di*. As for the gender of the needed article, being the gender of *covid 19* still an open issue in Italian linguistics, both masculine and feminine articles would be accepted (Giovine, 2020).

SOURCE TEXT	DEEPL	YANDEX
Присутствие таких гомологичных последовательностей в вакцине против Covid-19 связано с риском развития аутоиммунных осложнений и гетерологичного иммунитета. Вывод. Концепция пептидного континуума родства белков (ПКРБ) представляется полезной	La presenza di tali sequenze omologhe nel vaccino Covid-19 è associata al rischio di complicazioni autoimmuni e di immunità eterologa. Conclusione. Il concetto di continuità della proteina peptidica (PKRB) appare utile nella ricerca di epitopi immunitari per i vaccini Covid-19 e permette di prevedere i	La presenza di tali sequenze omologhe nel vaccino Covid-19 è associata al rischio di complicanze autoimmuni e immunità eterologica. Output. Il concetto di peptide continuum of protein Parenthood (PCRB) sembra essere utile nella ricerca di epitopi immunitari per i vaccini Covid-19 e

<p>в поисках иммунных эпитопов для вакцин против Covid-19 и позволяет спрогнозировать возможные риски, связанные с их применением. По-видимому, в будущем коронавирусные вспышки и пандемии будут чаще, чем пандемии гриппа.</p>	<p>possibili rischi associati al loro utilizzo. Le epidemie di coronavirus e le pandemie sembrano essere più frequenti delle pandemie influenzali in futuro.</p>	<p>consente di prevedere i possibili rischi associati alla loro applicazione. Apparentemente, in futuro, le epidemie di coronavirus e le pandemie saranno più comuni delle pandemie influenzali.</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The third and last part of the abstract undoubtedly constitutes an interesting basis for discussion. The Russian preposition *против*, which in the source document links the two Russian terms *вакцине* and *Covid-19*, properly rendered as *vaccino* and *Covid-19* (Kovalev, 2020), is completely ignored by both translation tools. In fact, in the two provided translation versions no prepositions precede the term *Covid-19*. In this case, being the rendering of the preposition under examination essential to convey the original meaning, its omission cannot be considered acceptable and the same applies to DeepL's and Yandex's translation variants. Moreover, as mentioned with regard to the previous section, denoting Covid-19 a specific and notorious disease, a definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). Since its gender is still controversial (Giovine, 2020), a feminine as well as a masculine definite article would be equally accepted.

The Russian term *осложнений* is translated by DeepL and Yandex with two different Italian nouns, namely *complicazioni* and *complicanze*, respectively. Constituting both Italian nouns exact Italian equivalents for the Russian term at stake and constituting in this context a synonymic pair (Zingarelli, 2020), they can be considered equally acceptable.

The Russian term *развития*, which in the original text is linked by a specification relation with the two Russian word sequences *аутоиммунных осложнений* and *гетерологического иммунитета*, lexically correctly rendered by DeepL and Yandex as *complicanze autoimmuni* and *immunità eterologa* (Kovalev, 2020), is completely ignored by both translation tools. In this case, the omitted Russian term's meaning is implied in

the general sense of the sentence. In other words, by mentioning the risks of the vaccine's complications, the fact that those risks arise with the development of the complications appears obvious. As a consequence, the Russian term under examination can be omitted without distorting the overall meaning of the sentence and the two provided translation variants can be therefore considered acceptable.

We can notice that, in DeepL's translation version, the Italian word sequence *immunità eterologica*, is introduced by the simple preposition *di*, whereas in Yandex's one it is not. As previously mentioned, the word sequence under examination, along with *complicanze autoimmuni*, should be linked by a specification relation with the rendering of the Russian term *развития*. Being the Russian term *развития* omitted in both translation variants, the two words sequences seem to depend on the Italian term *rischio*, i.e. the correct translation of the Russian term *риск*, to which they are linked by a specification relation. As a consequence, the simple preposition *di* is necessary to reflect the original syntactical structure. Nonetheless, since they are linked to each other by the Italian coordinating conjunction *e* and consequently collocated at the same syntactical level, the simple preposition introducing the first word sequence, according to Italian grammatical rules, performs its function with regard to the second as well. As a result, both versions can be considered acceptable.

The Russian term *Вывод* is rendered differently by the two translation tools. Indeed, it is translated with the Italian term *conclusione* in DeepL's translation variant, whereas in Yandex's one, the English noun *output* is displayed. Starting from DeepL's rendering, the noun *conclusione* not only constitutes a proper Italian equivalent for the Russian term under examination but also suits this specific context as in the Italian contemporary language it frequently appears in correlation with the subject covered by the abstract (Tiberii, 2018). Moreover, it seems to be consistent with the Italian terms used to render the introductions of the three abstract's sections. As for Yandex's version, although the English term *output* is nowadays extremely widespread in the Italian contemporary language and highly frequently appears in Italian scientific paper to indicate the results of a certain scientific research (Zingarelli, 2020), it does not perfectly suit this specific context and, for a matter of consistency throughout the whole abstract, an Italian term would be certainly preferable.

The Russian multiword expression *Концепция пептидного континуума родства белков (ПКРБ)* is erroneously rendered by both translation tools. On the one hand, DeepL completely ignores a core element of the Russian word sequence under examination, namely *родства*, distorting its overall meaning and displays the original acronym (*ПКРБ*) in Cyrillic letters. On the other hand, Yandex renders most of the Russian multiword expression in English, which could theoretically be a possible translation option in case the expression at stake was well known and widespread in Italian scientific literature. Since no correspondences have been observed, it cannot be considered acceptable. Moreover, in Yandex's translation version, the Russian acronym (*ПКРБ*) is simply transliterated even though it does not present a direct equivalent either in Italian or in English. As mentioned above with regard to the first section of the abstract, the Italian multiword expression *concetto di molteplicità della relazioni delle strutture proteiche* may constitute a possible translation option (Volkenštein, 1975). In this case, since no Italian or English equivalent have been found for the Russian acronym (*ПКРБ*), it can be exceptionally omitted.

The rendering of the Russian verb *представляется* constitutes a discrepancy between the two provided translation variants. It is indeed translated as *sembra essere* by DeepL and *appare essere* by Yandex. Since the two Italian verbs are not only synonyms in the Italian language (Zingarelli, 2020) but also able to properly convey the intended meaning (Kovalev, 2020), they can both be considered correct translation options.

The Russian word sequence *в поучках*, usually followed, in the Russian language, by the genitive case, is rendered by both translation tools with the Italian expression *alla ricerca di*. In this case, since the provided translation versions perfectly the intended meaning remaining compliant with Italian grammar and lexis, they can be considered

In both translation variants, the Russian preposition *против* is completely omitted and no prepositions are used to introduce the term *Covid 19*. As mentioned above with regard to the previous lines of this section, being the Russian preposition under examination essential to render the syntactical structure as well as the meaning of the source document, its omission cannot be considered acceptable. The Italian preposition *contro*, displayed in association with the singular masculine defined article *il*, may represent a possible translation option, as it is able to reflect the original meaning (Kovalev, 2020) and the grammatical features of the term it is intended to introduce.

The Russian verb *позволяет* is rendered differently by DeepL and Yandex with two different Italian verbs, namely *permette* and *consente*, respectively. In this case, since the two suggested verbs convey the intended meaning (Kovalev, 2020) and perfectly suit this specific context, they can be considered equally correct.

The same does not apply to the Russian term *применением*. It is indeed translated as *utilizzo* by DeepL and *applicazione* by Yandex. In this case, although the two Italian terms may convey a similar meaning (Zingarelli, 2020), they do not equally suit this specific context. Indeed, the Italian noun *applicazione* more frequently refers to the field of medicine and comes in association with the proper names of medical products (Tiberii, 2018). As a result, Yandex's translation version results preferable.

The last sentence of the section under examination undoubtedly deserves special attention. Firstly, the Russian expression *По-видимому* is translated by Yandex with the Italian adverb *apparentemente* and completely ignored in DeepL's translation variant. Being *По-видимому* a meaningful element of the sentence, its omission cannot be considered acceptable. On the other hand, DeepL's version, although preferable when compared to Yandex's one, does not perfectly reflect the meaning of the source document (Kovalev, 2020). The Italian adverbs *evidentemente* and *probabilmente* may instead constitute better translation options (Kovalev, 2020).

The Russian expression *в будущем* is correctly rendered by both translation tools as *in futuro* (Kovalev, 2020). However, since only Yandex perfectly respects the original sentence's word order, its translation version results preferable when compared to DeepL's one.

The Russian verb *будут* is correctly rendered by Yandex with the Italian verb *saranno*, which perfectly reflect the grammatical features of the original verb (Kovalev, 2020). On the contrary, in DeepL's translation variant, the Italian verb *sembrano essere* is displayed and, since it is not able to convey the source document's meaning (Kovalev, 2020), it cannot be considered acceptable.

The rendering of the Russian adjective *коронавирусные* undoubtedly represents a translation issue. In fact, although it is correctly translated by DeepL and Yandex with the Italian expression *di coronavirus*, it is erroneously collocated, in both translation variants, immediately after the Italian term *pandemie*, so that it seems to be solely connected to it. In this case, referring the adjective under examination, in the source

document, to the Russian terms *вспышки* and *пандемии*, correctly rendered by both translation tools as *epidemie* and *pandemie* (Kovalev, 2020), it needs to be located after both of them in order to properly convey the original meaning.

The Russian adjective *чаще*, which appears, in the source document, in its comparative form, is rendered differently by the two translation tools. Indeed, it is translated by DeepL with the Italian adjective *più frequenti*, whereas in Yandex’s translation version the Italian adjective *più comuni* is displayed. In this case, the adjective suggested by DeepL undoubtedly better conveys the meaning of the original verb (Kovalev, 2020) and has to be consequently considered preferable.

Finally, more correct usage of Italian punctuation can be observed in Yandex’s translation, when compared to DeepL’s one. Indeed, while in DeepL’s version no commas are displayed, in Yandex’s one they are used to properly convey the original text’s meaning and comply with Italian grammatical rules (Treccani, 2020).

3.4. Coronavirus was imported into Russian at least 67 times

The first popular-science text, *Коронавирус завозили в Россию не менее 67 раз*, concerns a study conducted by specialists from several Russian research institutes, aimed at discovering when exactly the Coronavirus SARS CoV-2 entered the Russian Federation borders for the first time, triggering the outbreak of the pandemic in the country.

SOURCE TEXT	DEEPL	YANDEX
Коронавирус завозили в Россию не менее 67 раз	Il Coronavirus è entrato in Russia almeno 67 volte	Il coronavirus è stato importato in Russia almeno 67 volte

The first discrepancies between the two provided translation variants are observed in the rendering of the title of the article. The Russian noun *Коронавирус*, which plays the role of subject of the sentence, is correctly rendered by both translation tools as *coronavirus*, with a spelling difference. Indeed, in DeepL’s variant, the term is written

with a capital letter, whereas in Yandex’s one a lower-case initial letter is used. The term coronavirus is a loanword from English, originally used to refer to a group of viruses characterized by a particular crown-shape form, but currently indicating a specific type of coronavirus, namely SARS CoV-2, become enormously notorious with the outbreak of 2020’s pandemic. Being *coronavirus* a common noun (Accademia della Crusca, 2020), denoting a type of virus and not a specific virus, Yandex’s version is undoubtedly more compliant with Italian grammatical rules.

The second part of the title presents a number of issues. The Russian verb *завозили* is indeed rendered differently by the two translation tools. On the one hand, DeepL translates it with the Italian present perfect *è entrato* and correctly matches it with its subject, namely the noun *Coronavirus*. On the other hand, in Yandex’s version, the Italian impersonal past perfect *è stato importato* is displayed and the noun *coronavirus* plays the role of object. The Italian verb used by Yandex not only gets closer to the meaning of the original Russian verb (Kovalev, 2020) but also fully reflects the syntactical structure of the source document. However, although DeepL’s translation variant does not perfectly comply with the original title neither from a lexical (Kovalev, 2020) nor from a syntactical point of view, it succeeds in properly conveying the intended meaning, naming the effect of the action described by the source document, namely the import of the virus to Russian Federation. In other words, the virus has entered the Russian Federation borders because it has been somehow and somewhen imported. As a consequence, Yandex’translation is clearly more accurate, but also DeepL’s one can be considered acceptable, especially as the title of a popular-science article.

SOURCE TEXT	DEEPL	YANDEX
<p>Сначала коротко</p> <p>Проблема: Многие в России считают, что переболели COVID-19 ещё в декабре 2019 года или в январе 2020.</p> <p>Можно ли узнать, когда действительно в России началась эпидемия коронавируса и откуда</p>	<p>Prima, brevemente</p> <p>Problema: molte persone in Russia credono di essere state infettate da COVID-19 nel dicembre 2019 o nel gennaio 2020. È possibile scoprire quando l'epidemia di coronavirus è iniziata davvero in Russia e da dove è stata portata a</p>	<p>Prima breve</p> <p>Problema: molti in Russia credono che COVID-19 sia stato malato a dicembre 2019 o a gennaio 2020. È possibile scoprire quando l'epidemia di coronavirus è iniziata in Russia e da dove è stata portata a noi?</p>

его к нам завезли? Ответ дала биоинформатика.	noi? La bioinformatica ha dato la risposta.	La risposta è stata data dalla bioinformatica.
-----------------------------------------------	---------------------------------------------	------------------------------------------------

In the first part of the article, an interesting basis for discussion can be observed. Starting from the very beginning, the title of the first paragraph is rendered differently by the two translation tools. In particular, the Russian adverb *коротко* is translated with the Italian adverb *brevemente* by DeepL, whereas the adjective *breve* is used in Yandex’s translation variant. In this case, only DeepL’s translation is acceptable. The Italian adverb *brevemente* indeed not only correctly conveys the source sentence meaning (Kovalev, 2020) and syntactical structure but is also compliant with the target language grammatical rules and syntax. Nonetheless, a better translation variant could be the Italian expression *in breve*, which better suits this specific context and reflects contemporary Italian language (Zingarelli, 2020).

The Russian pronoun *многие*, playing, in this case, the role of subject of the sentence, is translated by DeepL with the word sequence *molte persone* and by Yandex with the Italian pronoun *molti*. Starting from saying that both translations can be considered correct (Kovalev, 2020), Yandex’s one better suits this specific context, as it perfectly conveys the source text meaning (Kovalev, 2020) avoiding any possible redundancy and constituting an easily readable and natural translation option. In fact, in this specific case, the Italian pronoun *molti* sufficiently clearly transmits the intended meaning and does not necessarily need the noun *persone*, as there are no doubts that the people are actually the subject of the sentence under examination.

The Russian verb *переболели* undoubtedly represents a core issue in the translation process. It is rendered by DeepL with the Italian passive past infinite verb *essere state infettate*. Although it is correctly linked to its subject and fully reflects the tense and the aspect of *переболели*, it is lexically distant from the original verb (Kovalev, 2020). However, the verb chosen by DeepL represents the cause of the condition described in the source document and consequently may be able to convey, as a matter of fact, the intended meaning. Closer from a lexical point of view undoubtedly is the verb displayed in Yandex’s translation variant, namely the Italian past perfect subjunctive *sia stato malato* (Kovalev, 2020). Moreover, the term COVID 19 is introduced, in DeepL’s translation version, by the simple preposition *da*. Starting from saying that the suggested

preposition correctly renders the syntactical structure of the source document, constituting COVID 19 the definition of a specific and well-known disease, a definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). Since the gender of the term under examination still represents an open question of Italian linguistics (Giovine, 2020), both a masculine and a feminine definite article would be accepted. Nonetheless, although Yandex's verb is compliant with the original verb's meaning (Kovalev, 2020), tense, and aspect, not only it is erroneously linked to the intended subject but is also inserted within an unclear and incorrect syntactical structure. Indeed, being the term *COVID-19* located immediately after the relative conjunction *che* and containing the verb used in Yandex's version a masculine past participle, it seems like *COVID-19* plays the role of subject of the sentence. Therefore, since Yandex's translation variant completely distorts the original meaning, only DeepL's translation can be considered acceptable.

The Italian nouns indicating the months when the coronavirus possibly entered Russian Federation borders are introduced by the complex preposition *nel* in DeepL's translation, whereas in Yandex's one the simple preposition *a* is used. Starting from saying that both prepositions constitute a proper translation option of the original sentence, the correctness of the addition of the definite article depends on the type of preposition. In this case, a definite article is necessary in association with the simple preposition *in*, whereas would be erroneous if added to the simple preposition *a*. As a consequence, both versions are equally acceptable.

The Russian adverb *действительно* may be given a number of different Italian equivalents (Kovalev, 2020). In this specific case, DeepL's translation can be considered correct and the Italian term *davvero* is moreover properly located within the sentence. Its position indeed makes DeepL's translation variant readable and easily understandable by an Italian reader. On the contrary, Yandex does not translate the term, which is completely omitted.

The Russian singular masculine personal pronoun *его*, which in the source document clearly refers to the Russian term *коронавирус*, is rendered by both translation tools with the feminine past participle *portata*, i.e. a part of the main verb of the last part of this section. Being feminine, the past participle undoubtedly seems like to be directly linked to the Italian feminine term *epidemia*, which constitutes the proper translation of

the Russian noun *эпидемия* (Kovalev, 2020). In actual fact, both translation variants fail in accurately rendering the syntactical structure of the original sentence and cannot therefore be considered acceptable.

Both translation tools translate the Russian expression *к нам*, which in the original document comes in association with the Russian verb *завезли*, as *a noi*. Being the translations on the whole correct (Kovalev, 2020), this cannot be considered a translation mistake. However, the Italian pronoun *ci*, collocated immediately before the Italian verb *portare* may constitute a better translation option, as it is perfectly able to convey the original meaning and underline the spatial opposition between Russia and Europe that is expressed in the source document.

Finally, a discrepancy can be observed in the last sentence of the section under examination. DeepL indeed fully reflects the syntactical structure and the word order of the source document. In fact, its version consists of the Italian active present perfect verb *ha dato*, the subject, namely the Italian term *bioinformatica*, located at the beginning of the sentence, and the object *risposta* at its end. On the contrary, DeepL distorts the original syntactical structure by displaying the passive present perfect verb *è stata data* and consequently giving *risposta* the role of subject and *bioinformatica* the role of agent. Starting from saying that both translation tools provide lexically correct translation variants (Kovalev, 2020) and respect the tense of the original verb, they cannot be considered equally accurate. In fact, although DeepL fully observes the syntactical structure of the source document, Yandex succeeds in reflecting the original theme-rheme pattern and perfectly complies with the intents of the writer of the article. In the original document, denoting the Russian term *биоинформатика* a new piece of information, it is located at the end of the last sentence and consequently brought into sharper focus than the other terms. The same applies to Yandex’s variant, where the original syntactical structure has been adjusted for the Italian noun *bioinformatica* to be displayed as the last term of the sentence.

SOURCE TEXT	DEEPL	YANDEX
Решение: Учёные сравнили мутации 211 геномов коронавируса от пациентов из 25	Soluzione: Gli scienziati hanno confrontato le mutazioni di 211 genomi di coronavirus di pazienti	Soluzione: gli scienziati hanno confrontato le mutazioni di 211 genomi di coronavirus da pazienti

<p>регионов России. На основе этих данных они построили эволюционные деревья вируса. Оказалось, что SARS CoV-2 завезли к нам из Европы в промежутке между концом февраля и началом марта. А первый случай его внутрироссийской передачи состоялся не раньше 11 марта 2020 года.</p>	<p>in 25 regioni della Russia. Sulla base di questi dati, hanno costruito alberi evolutivi del virus. Si è scoperto che la SARS CoV-2 è stata introdotta dall'Europa tra la fine di febbraio e l'inizio di marzo. E il primo caso di trasmissione intra-russo è avvenuto non prima dell'11 marzo 2020.</p>	<p>provenienti da 25 regioni della Russia. Sulla base di questi dati, hanno costruito alberi evolutivi del virus. Si è scoperto che SARS CoV- 2 è stato portato a noi Dall'Europa tra la fine di febbraio e l'inizio di marzo. E il primo caso del suo trasferimento intra-russo ha avuto luogo non prima dell'11 marzo 2020.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The translations of the second part of the article present several dissimilarities, primarily concerning lexis and syntax. Proceeding in order of appearance in the text, the Russian preposition *от*, linking the terms *коронавирус* and *пациентов*, is rendered differently by the two translation tools. DeepL translates it with the Italian simple preposition *di*, which connects the two Italian terms *coronavirus* and *pazienti* with a specification relationship. On the contrary, in Yandex’s translation variant, the Italian noun *pazienti* is assigned the role of origin of the coronavirus and consequently linked by the simple preposition *da*. Starting from saying that, indicating the term *pazienti* a non-specific, indefinite entity, the addition of an article to the above-mentioned simple prepositions would be incorrect, the two translation versions cannot be considered equally adequate. Indeed, according to Russian grammatical rules, the preposition *от* precisely introduces the complement of origin, in case it is performed by human entities (Glazunova, 2016). As a consequence, Yandex’s translation is considered more accurate when compared to DeepL’s one, as it perfectly reflects the syntactical structure of the source document. Nonetheless, although DeepL erroneously renders the source text syntax, it is able to convey the original meaning, namely the fact that the coronavirus genomes have been extracted from the genetic material of the mentioned patients. Therefore, it can be considered acceptable, even though Yandex’s translation variant is undoubtedly preferable.

The Russian preposition *из*, which connects in the original document the two terms *пациентов* and *регионов*, represents in the Russian language the spatial origin when performed by, among others, indoor places, countries, cities, villages, and public institutions (Glazunova, 2016) is displayed in the original text. In DeepL's translation variant, it is rendered with the Italian simple preposition *in*, which in the Italian language is usually used to express the place where a certain action is performed. Hence, DeepL not only does not properly reflect the source document's syntactical structure but also distorts the original meaning. The same does not hold for Yandex, which translates the Russian preposition using the Italian word sequence *provenienti da*, perfectly respecting the original meaning and syntax. For this reason, only Yandex's translation version can be considered acceptable.

The rendering of the Russian multiword expression *эволюционные деревья* certainly represents a point of resemblance between the two translation variants. Both translation tools indeed literally translate it as *alberi evolutivi*. In Russian, *эволюционные деревья* refers to a diagram representing the evolutionary relations among living organisms according to their physical or genetic features (Naumov, 2006). However, the same concept is expressed in the Italian language with the multiword expression *albero filogenetico* (Treccani, 2020). Both translation tools are hence not able to accurately convey the meaning of the source document and consequently cannot be considered acceptable.

In DeepL's translation, the term *SARS CoV-2* is introduced by the Italian singular feminine definite article *la*, whereas in Yandex's one no articles precede it. The international term *SARS CoV-2* is an acronym that stands for Severe Acute Respiratory Syndrome Coronavirus 2 and constitutes the technical name given to the virus responsible for coronavirus disease 19. Describing the name of a virus, which in the Italian language is masculine, also *SARS CoV-2* is masculine (Accademia della Crusca, 2020). Hence, the feminine article displayed in DeepL's version is not compliant with Italian grammatical rules. As mentioned before, Yandex does not insert articles to introduce the term *SARS CoV-2*. In the Italian language, the terms used to name illnesses are considered common nouns and written with the lower-case letter (Giovine, 2020), and in this specific case, an Italian singular masculine definite article would consequently be appropriate. Therefore, both translation variants cannot be considered totally acceptable.

The active Russian verb *завезли* is rendered by DeepL and Yandex with the passive form of two different Italian verbs, namely *introdurre* and *portare*. Their meaning may vary to various extents according to the context of use (Zingarelli, 2020), and in this specific case, we can say it is slightly different. The verbs indeed express a similar idea, namely moving something from one place to another, which is located close to the speaker, but with a different degree of specificity. In fact, *introdurre* describes the precise action of introducing something to a place for the very first time (Zingarelli, 2020), whereas *portare* presents the more general meaning of transferring something towards the speaker (Zingarelli, 2020). As a consequence, from a lexical point of view, DeepL's variant is preferable, as it better conveys the original meaning, i.e. the first introduction of coronavirus within the Russian Federation borders (Kovalev, 2020). As mentioned above, from a syntactical point of view, both translation tools correctly render the active Russian verb with a passive form and properly accordingly reorganize the syntactical structure of the sentence, by matching the past participles *introdotta* and *portata* in gender and number with the term that is assigned the role of subject, namely *SARS CoV-2*. Moreover, as mentioned before with regard to the first part of the article, Yandex's rendering of the Russian expression *к нам* as *a noi* can be acceptable, even though the Italian pronoun *ci* may constitute a better and more natural-sounding translation option (Zingarelli, 2020). Nonetheless, DeepL's translation variant undoubtedly results to be more easily readable and understandable to an Italian speaker when compared to Yandex's one.

In Yandex's translation, a format error is displayed, as the Italian complex preposition *dall'*, preceding the Italian term *Europa*, is written with a capital letter. According to Italian grammatical rules, being *dall'* a preposition collocated not at the beginning of a new sentence, a lower-case letter has to be used (Accademia della Crusca, 2020).

The Russian noun *непедачи* is translated by DeepL with the Italian term *trasmissione*, whereas in Yandex's translation variant the Italian term *trasferimento* is used. Starting from saying that both Italian nouns may convey a great number of different meanings according to the context of use (Zingarelli, 2020), in this specific case they can be considered synonyms (Zingarelli, 2020). However, in the contemporary Italian language, the term *trasmissione* frequently refers to a virus (Tiberii, 2018), so that it

specifically describes the process through which a virus moves from one organism to another (Zingarelli, 2020). As a result, DeepL’s translation variant is preferable to Yandex’s one. However, DeepL’s does not translate the Russian possessive adjective *его*, which in the original text refers to the term *SARS CoV-2*. On the contrary, Yandex properly renders it with the Italian singular masculine possessive adjective *suo* preceded by the complex preposition *del*. Although DeepL’s omission does not alter the meaning of the source text, the rendering of the possessive adjective would have made its translation variant undoubtedly more accurate and complete.

In the source document, the Russian adjective *внутрироссийской* clearly refers to the Russian noun *передачи*, which is rendered, as mentioned before, as *trasmissione* by DeepL and *trasferimento* by Yandex. While in Yandex’s translation variant, the adjective under examination is correctly linked to its corresponding term *trasferimento* and is consequently displayed in its masculine form *intra-russo*, in DeepL’s one, being in its masculine form, it seems like to refer to the Italian term *caso*, rather than to *trasmissione*. As a result, DeepL fails in reflecting the syntactical structure of the source document and thus properly conveying the original meaning.

Finally, the Russian verb *состоялся* is translated by DeepL and Yandex with the present perfect form of two different Italian verbs, namely *avvenire* and *avere luogo*. Since both translation variants respect the tense and the meaning of the original verb (Kovalev, 2020), they can be considered equally acceptable.

SOURCE TEXT	DEEPL	YANDEX
Теперь подробнее Исследовательская группа из Высшей школы экономики и Сколтеха, совместно со специалистами НИИ гриппа им. А.А. Смородинцева в Санкт Петербурге и ИППИ им. А.А. Харкевича РАН установили, что коронавирус SARS-CoV 2 независимо проникал	Ora per maggiori dettagli Un team di ricerca della Scuola Superiore di Economia e Skoltech, insieme a specialisti dell'A.A. Smorodintsev Influenza Research Institute di San Pietroburgo e dell'A.A. Kharkevich IPPI RAS, ha scoperto che il coronavirus SARS-CoV 2 è entrato in Russia almeno 67 volte,	Ora leggi di più Un gruppo di ricerca della Graduate School of Economics e Skoltha, in collaborazione con gli specialisti dell'Istituto di ricerca per l'influenza. A. A. Riborodintseva a San Pietroburgo e Ippi loro. A. A. HARKEVICH Ran ha stabilito che il coronavirus SARS-CoV 2 è penetrato indipendentemente nel

на территорию России не менее 67 раз, главным образом в конце февраля и начале марта 2020 года.	principalmente a fine febbraio e inizio marzo 2020.	territorio della Russia almeno 67 volte, principalmente alla fine di febbraio e all'inizio di marzo 2020.
-------------------------------------------------------------------------------------------------	-----------------------------------------------------	-----------------------------------------------------------------------------------------------------------

Starting from the title of this last section, which is clearly connected to the one introducing the first part of the article, we can easily notice several dissimilarities between the versions provided by the two translation tools. In particular, the Russian adverb *подробно*, in its comparative form *подробнее*, is rendered by DeepL and Yandex with two different Italian expressions, namely *per maggiori dettagli* and *leggi di più*, respectively. Although both translation variants do not strictly stick to the original title, they cannot be considered equally inaccurate. In fact, starting from saying that a proper translation option for the Russian adverb *подробно* may be the Italian expression *in dettaglio*, which becomes *più in dettaglio* in its comparative form (Koavlev, 2020), DeepL's variant seems to convey the original meaning (Kovalev, 2020) and be particularly suitable as the title of a popular-science article's section. Moreover, it provides the whole translation with a certain degree of consistency, as it constitutes an undoubtedly acceptable continuation of the title displayed in the first part of the article. The same does not apply for Yandex's title, as, although it may, in broad terms, convey the original meaning, it not only does not constitute a proper title for a popular-science article's section but is also inconsistent with the title of the first section of the article under examination.

The Russian noun *группа* is translated by DeepL with the English term *team*, whereas in Yandex's translation version, the Italian noun *gruppo* is used. Starting from underling that the two terms convey the same meaning (Collins English Dictionary, 2020), they can be both considered acceptable, as nowadays the English term *team* occurs in the Italian language so frequently that it is perfectly understandable by an Italian-speaking reader.

The Russian preposition *из*, indicating in the Russian language the spatial origin (Glazunova, 2016), is rendered by both translation tools with the complex preposition *della*, consisting of the simple preposition *di*, which in the Italian language expresses the specification relation, and the singular feminine definite article *la*. The two translation

variants fail indeed in accurately reflecting the syntactical structure of the original document. However, this is not considered a grave mistake, as they are able to convey, although not perfectly, the meaning of the source document, namely the fact that the researchers who are members of the research group come from the Higher School of Economics and Skoltech.

The proper nouns *Высшей школы экономики* and *Сколтеха*, respectively naming the famous Higher School of Economics of Moscow and the Skolkovo Institute of Science and Technology are rendered differently by the two translation tools. On the one hand, DeepL indeed correctly translates the former with its Italian equivalent *Scuola Superiore di Economia* and the latter with the highly widespread acronym of Skolkovo Institute of Science and Technology, namely *Skoltech*, also used in the website of the Institute (www.skoltech.ru/en/). Nonetheless, DeepL partially fails in conveying the specification relation expressed in the original document. In its translation version, the complex preposition *della* indeed correctly precedes the noun *Scuola Superiore di Economia*, whereas no prepositions are used to introduce *Skoltech* and the term seems to be completely disconnected from the rest of the sentence. On the other hand, in Yandex's translation variant, not only the proper nouns indicating the two Russian scientific institutes are given non-existent equivalents but also, as mentioned concerning DeepL's version, no prepositions precede the acronym indicating the Skolkovo Institute of Science and Technology.

The Russian adverb *совместно со* is translated as *insieme a* by DeepL and *in collaborazione con* by Yandex. The two Italian expressions, although different, not only properly reflect the meaning of the original document (Kovalev, 2020) but also suit this specific context. As a result, both translation variants can be considered equally acceptable.

The prepositions introducing the Italian term *specialisti*, which constitutes the correct translation of the Russian term *специалистами*, undoubtedly represents a good point for our comparative analysis. In DeepL's translation variant, the term is indeed introduced by the simple preposition *a*, whereas in Yandex's one the plural masculine definite article *gli* comes in association with the simple preposition *con*. In this specific case, being the term *specialisti* given a certain degree of specificity by the addition of the following words indicating the research institute from which they come from, the

insertion of the definite article is doubtless compliant with Italian grammatical rules (Lepschy & Lepschy, 1993). Therefore, Yandex's version of this specific section is preferable.

The Russian acronym *НИИ*, which stands for *Научно-исследовательский институт* and comes in association with the word sequence *группа имени А.А. Смородинцева* is correctly translated by DeepL with the proper noun that appears in the research institute's website, namely *A.A. Smorodintsev Influenza Research Institute* (www.influenza.spb.ru/en/). Even though the term used by the translation tool is in English, we can assume that, if not already known by the intended readers, it can be in all probabilities easily accessible with a quick online search. By contrast, in Yandex's translation version, the Italian word *Istituto di ricerca per l'influenza A. A. Riborodintseva* is used. An Italian rendering could be supposedly correct, as provides the potential readers with sufficient clarity with regard to what sphere of research the institute at stake performs, without the need for an online search. Nonetheless, Yandex fails in properly transliterating the proper noun *Смородинцева*, distorting the actual name of the research institute and consequently the message conveyed by the source document. In this case, DeepL's translation variant can be therefore considered preferable.

The Russian proper name *Санкт Петербурге* is preceded, in the source document by the Russian preposition *в*, indicating, in the Russian language, the spatial complement (Glazunova, 2016). It is rendered by DeepL and Yandex with two different Italian word sequences, namely *di San Pietroburgo* and *a San Pietroburgo*, respectively. Starting from saying that the Russian noun under examination is equally correctly rendered by both translation tools from a lexical point of view (Kovalev, 2020), the same does hold for syntax. In fact, while Yandex fully reflects the source document's syntax and perfectly renders the spatial complement displayed in the original text. On the other hand, DeepL, by translating the original Russian preposition with the Italian preposition *di*, usually indicating, in the Italian language, the specification complement (Treccani, 2020), does not totally respect the original syntactical structure. However, conveying both Italian expressions the intended meaning, namely that the mentioned research institutes are located in Saint Petersburg and consequently figuratively belong to the city. As a consequence, both translation variants can be considered correct.

Both translation tools totally fail in properly rendering the Russian acronyms *ИППИ* and *РАН*, which stand for *Институт Проблем Передачи Информации* and *Российская Академия Наук*. The English noun that appears on the institute's website as well as a translation into Italian of each word constituting the acronyms may represent good translation options. Moreover, in both translation versions, the Russian proper noun *Харкевича* is erroneously transliterated.

The Russian verb *установить* is rendered by DeepL and Yandex with two different Italian verbs, namely *ha scoperto* and *ha stabilito*, respectively. We will start from saying that both versions do not respect all the grammatical features of the original verb. Indeed, while in the source document the Russian verb refers to the group of mentioned researchers and is consequently displayed in its plural form, in the provided translations, it is rendered in its singular form and it seems like to be just linked to the target terms *gruppo* and *team*. From a mere lexical point of view, Yandex's translation version not only fully reflects the original meaning (Kovalev, 2020) but also seems to perfectly suit this specific context. Indeed, although both verbs can be used to express the results of research, a profound lexical difference is observed (Zingarelli, 2020). The Italian verb *scoprire* usually refers to a finding that may be due to a sudden realization, a stroke of luck as well as extensive research in a certain field. On the contrary, *stabilire* implies that comprehensive studies have been conducted, empirical data have been collected and analyzed, and the corresponding results have been obtained. Hence, the verb used in Yandex's translation more specifically reflects the meaning of the original document and can be consequently considered preferable.

Two different Italian verbs are used to render the Russian verb *проникал*. DeepL indeed translates it as *è entrato*, whereas in Yandex's translation version the Italian verb *è penetrato* is displayed. Although both translation tools respect the tense of the original verb, from a lexical point of view, they cannot be considered equally accurate. In this specific context, the two Italian verbs express indeed a similar idea, namely the entering of an entity into a certain place (Zingarelli, 2020), but with a different degree of specificity. In fact, while *entrare* carries a more neutral meaning, *penetrare* presents a greater number of lexical nuances and may refer to something sneaky crossing the threshold of somewhere and afterward spreading out of control (Zingarelli, 2020). Therefore, Yandex's verb seems not only to better convey the meaning of the original

verb (Kovalev, 2020) but also to perfectly suit this specific context as it properly expresses the wide and rapid spread of a virus. We can therefore say that Yandex's translation version of this section is preferable.

The Russian adverb *независимо* is omitted by DeepL and rendered by Yandex with the Italian adverb *indipendentemente*. Based on the premise that the rendering of the Russian adverb under examination is essential for the translation variant to convey the source text meaning, DeepL's translation variant cannot be accepted. Nonetheless, Yandex's version undoubtedly presents some issues. In this case, both the form and the location of the Italian adverb can indeed be misleading for the reader. The Russian adverb expresses indeed the idea that the coronavirus entered the Russian Federation borders 67 times that do not present any direct correlations with one another. Hence, since *indipendentemente* displays several different meanings (Zingarelli, 2020), linking the adverb to the Italian noun *volte*, which constitutes the correct translation of the Russian term *раз* (Kovalev, 2020), and using the Italian word sequence *indipendenti l'una dall'altra*, may represent a possible and more clear translation variant of this section of the article as well as a strategy to better convey the original meaning.

Yandex undoubtedly more accurately translates the Russian word sequence *меппуморуо Рoccus* as *territorio russo* and properly links it to the rest of the sentence, expressing the location complement displayed in the source document. By contrast, in DeepL's translation variant, it is roughly rendered as *Russia* and the Russian term *меппуморуо* is completely ignored. We can easily say that Yandex's translation version not only better conveys the original meaning but also shows a higher degree of consistency with the source document. Nonetheless, although it is more complete and clearly preferable when compared to DeepL's one, also this last one can be considered acceptable, as it does not fail in conveying the source document meaning and its version is consequently correctly understandable by an Italian speaker.

Finally, the Italian terms *fine* and *inizio*, constituting respectively the proper translations of the Russian nouns *конец* and *начале* are introduced in Yandex's translation variant by the complex prepositions *all* and *all'*, whereas in DeepL's one only *fine* is preceded by the simple preposition *a*. In this specific case, both translations can be considered equally acceptable, as there are no significant differences between using complex or simple prepositions when introducing the above-mentioned terms. Moreover,

being the two Italian nouns linked by the coordinating conjunction *e*, the preposition introducing the first one performs the same action with regard to the second.

3.5. Unknown lethality - Why we do not know the real extent of COVID-19

The second popular-science text, *Неизвестная летальность - Почему мы не знаем истинных масштабов COVID-19*, focuses on the key issues that have been discussed at the webinar of the International Laboratory for Population and Health Research held by the Higher School of Economics. In particular, special attention is devoted to underline the complexities of acquisition and analysis of numerical data on the coronavirus pandemic, which prevents governmental institutions from providing a comprehensive framework of the current health situation at a global level.

SOURCE TEXT	DEEPL	YANDEX
Неизвестная летальность - Почему мы не знаем истинных масштабов COVID-19	Letalità sconosciuta - Perché non conosciamo la vera portata di COVID-19	Letalità sconosciuta- perché non conosciamo la vera scala di COVID-19

When observing the two provided translation variants with regard to the title of the article, we can easily notice three main translation issues. Firstly, the Russian noun *масштабов* is rendered by DeepL and Yandex with two different Italian terms, namely *portata* and *scale*. Although both Italian nouns may constitute proper Italian equivalents for the Russian term under examination (Kovalev, 2020), *portata* undoubtedly better conveys the source document's sense (Kovalev, 2020) and perfectly suits this specific context.

Secondly, the term *COVID-19* is introduced, in both translation versions, by the simple preposition *di*. In this case, constituting *COVID-19* the name defining the specific and notorious syndrome caused by the new coronavirus SARS-COV 2, a definite article is necessary to properly convey the intended meaning (Lepschy & Lepschy, 1993). The gender of the Italian term *COVID-19* still represents an open question of Italian

linguistics. In fact, denoting the name of a specific syndrome, which in the Italian language is a feminine term, it should be feminine itself. However, since, from the beginning of its circulation, it has been frequently mistaken for the name of the virus responsible for the mentioned syndrome, which, in the Italian language, is masculine, it currently appears in highly specialized and popular-science texts as a feminine as well as a masculine term (Giovine, 2020). As a result, both a feminine and a masculine definite article would be considered acceptable.

Finally, Yandex’s translation variant presents an orthography error, as the Italian adverb *perché* begins with a lower-case letter. In this case, being the title formed by two independent sentences, an upper-case initial letter should be displayed (Treccani, 2020).

SOURCE TEXT	DEEPL	YANDEX
Пандемия коронавируса поставила демографов на передовую, но пока без оружия: достоверно оценить ситуацию они не могут из-за отсутствия данных. И это при том, что интернет, казалось бы, переполнен статистикой. О парадоксах количественных подходов к COVID-19 говорили на вебинаре Международной лаборатории исследований населения и здоровья ВШЭ.	La pandemia di coronavirus ha messo i demografi in prima linea, ma finora senza armi: non possono valutare la situazione in modo affidabile a causa della mancanza di dati. E questo nonostante il fatto che Internet sembra traboccare di statistiche. I paradossi degli approcci quantitativi al COVID-19 sono stati discussi al webinar del Laboratorio Internazionale per la Ricerca sulla Popolazione e la Salute della Scuola Superiore di Economia.	La pandemia del coronavirus ha messo i demografi in prima linea, ma finora senza armi: valutare in modo affidabile la situazione, non possono a causa della mancanza di dati. E questo nonostante il fatto che Internet sembra essere pieno di statistiche. I paradossi degli approcci quantitativi a COVID-19 hanno parlato al webinar del Laboratorio Internazionale di ricerca sulla popolazione e sulla salute HSE.

In the first part of the article, the Russian term *коронавируса*, correctly rendered by both translation tools with the Italian noun *coronavirus* (Kovalev, 2020), is introduced by the simple preposition *di* in DeepL’s translation variant, whereas in Yandex’s one the complex preposition *del* is displayed. Starting from saying that both translation versions render the specification relation expressed in the source document, in this case, the Italian

term *pandemia*, which constitutes the proper translation of the Russian noun *пандемия* (Kovalev, 2020), in the Italian language, is frequently linked to the name of the illness to which it refers by means of the simple preposition *di* (Treccani, 2020). As a result, DeepL's translation's variant is preferable.

The final part of the first sentence undoubtedly constitutes an interesting basis for discussion. Indeed, the two translation tools provide lexically correct translation variants (Kovalev, 2020), which, however, display a partly different word order with regard, in particular, to the rendering of the main verb. On the one hand, Yandex perfectly reflects the structure of the main Russian verb, and correctly renders its two constituting elements, namely the infinitive *оценить* and the Russian modal verb *не могут*, in the same order with which they appear in the source document. Moreover, the two verbs are originally separated by the Russian term *ситуацию*, whose Italian equivalent, i.e. the Italian noun *situazione* (Kovalev, 2020), is inserted between the Italian verbs *valutare* and *non possono*. Although Yandex's translation variant undoubtedly remains consistent with the original text, since, in the Italian language, the modal verbs usually directly precede the verbs to which they refer, it does not sound absolutely natural to an Italian native speaker. On the other hand, DeepL remains compliant with Italian grammatical rules and collocates the whole verb, followed by its direct object, at the very beginning of the section under examination. Constituting DeepL's translation a lexically correct (Kovalev, 2020), syntactically well-structured, and Italian sounding sentence, it can be undoubtedly considered preferable when compared to Yandex's one.

The Russian adjective *переполнен* is rendered by DeepL and Yandex with two different Italian verbs, namely *traboccare* and *essere pieno*, respectively. Starting from saying that rendering the Russian adjective under examination with a verb may, in this case, constitute a proper translation option, since the verbs suggested by the two translation tools properly convey the intended meaning (Kovalev, 2020) and constitute a synonymic pair (Zingarelli, 2020), they can be considered equally correct.

Both translation tools render the Russian term *статистикой* as *di statistiche*, perfectly conveying the meaning (Kovalev, 2020) and the syntactical structure of the source document. The same does not hold for the grammatical number of the original Russian term, which is displayed in its singular form and translated with a plural Italian noun. Nonetheless, since, according to Italian grammatical rules, the verbs and adjectives

denoting the quality or state of being full are usually followed by the singular form of uncountable nouns and the plural form of countable ones (Treccani, 2020), in this case, being *statistica* a countable noun, its plural form undoubtedly better conveys the intended meaning and constitutes a more natural-sounding translation option.

The last sentence of the section under examination presents some major translation issues. Firstly, the term *COVID-19* is introduced by the complex preposition *al* in DeepL's translation variant, whereas in Yandex's one the simple preposition *a* is displayed. Starting from saying that both versions seem to correctly render the syntactical structure of the source document, being *COVID 19* the specific name of a notorious syndrome, a definite article should be inserted in association with the simple preposition *a*. Since, as mentioned above, the gender of the term under examination is still controversial (Giovine, 2020), both a feminine and a masculine definite article can be accepted. As a result, DeepL translation variant is undoubtedly preferable when compared to Yandex's one.

Secondly, the rendering of the Russian verb *говорили*, displayed in the source document in its impersonal active form (Glazunova, 2016) and referring to a non-specific plural subject, undoubtedly constitutes a serious discrepancy between the provided translation variants. It is indeed translated by DeepL and Yandex with two different Italian verbs, namely *sono stati discussi* and *hanno parlato*, respectively. Although both suggested verbs may represent, from a lexical point of view, correct Italian equivalents for the original verb (KOvaley, 2020), they do not equally reflect the syntactical structure of the source sentence. On the one hand, as mentioned above, DeepL renders the Russian verb under examination with the passive Italian verb *sono stati discussi* and assigns the Italian word sequence *I paradossi degli approcci quantitativi al COVID-19*, i.e. the Italian equivalent of the original Russian topic complement *О парадоксах количественных подходов к COVID-19*, the role of subject of the new sentence. On the other hand, Yandex suggests the Italian verb *hanno parlato*, which perfectly reflects the grammatical form of the source verb. However, the translation tool directly links the verb to the rendering of the original topic complement, namely the Italian expression *I paradossi degli approcci quantitativi a COVID-19*, so that it seems like to constitute the subject of the sentence. In so doing, not only in the provided translation version the meaning and syntactical structure of the source document are completely distorted but also a

nonsensical expression is displayed, as a verb usually defining, in the Italian language, a human action (Zingarelli, 2020) is connected to an inanimate subject.

Finally, the Russian multiword expression *Международной лаборатории исследований населения и здоровья ВШЭ* is rendered by DeepL and Yandex with two different Italian word sequences, namely *Laboratorio Internazionale per la Ricerca sulla Popolazione e la Salute della Scuola Superiore di Economia* and *Laboratorio Internazionale di ricerca sulla popolazione e sulla salute HSE*, respectively. Since no official Italian equivalents of the multiword expression under examination have been found, both translation variants can be considered acceptable as they properly express the intended meaning in a well-formed syntactical structure. However, two factors make DeepL's version preferable when compared to Yandex's one. Firstly, it is closer to the official English equivalent displayed on the Higher School of Economics official website (www.demogr.hse.ru/en/). Secondly, although Yandex renders the final acronym *ВШЭ*, which stands for *Высшая школа экономики*, with its correct international equivalent *HSE* (www.hse.ru/en/), the Italian translation suggested by DeepL seems to contribute to maintaining a certain consistency within the text.

SOURCE TEXT	DEEPL	YANDEX
<p>Доступно и несопоставимо</p> <p>По счетчикам в Сети видно, что страны очень отличаются как по заболеваемости, так и по числу смертей.</p> <p>Показатель Case Fatality Ratio (CFR), то есть соотношение умерших к заболевшим, разный даже в государствах с соразмерным уровнем экономического развития, в которых эпидемия началась приблизительно в одно время. Например, в</p>	<p>Accessibile e non comparabile</p> <p>I contatori sul Web mostrano che i paesi variano notevolmente sia nell'incidenza delle malattie che nel tasso di mortalità. Il Case Fatality Ratio (CFR), ovvero il rapporto tra decessi e casi, è diverso anche nei paesi con un livello di sviluppo economico commisurato, dove l'epidemia è iniziata più o meno nello stesso periodo. Ad esempio, in Germania il CFR è di circa</p>	<p>Disponibile e disparabile</p> <p>Secondo i contatori della rete, è chiaro che i paesi sono molto diversi sia per incidenza che per numero di morti. Il Case Fatality Ratio (CFR), cioè il rapporto tra i morti e i malati, è diverso anche negli Stati con un livello proporzionato di sviluppo economico, in cui l'epidemia è iniziata approssimativamente nello stesso momento. Ad esempio, in Germania</p>

Германии CFR около 2%, во Франции — более 10%.	il 2%, in Francia è superiore al 10%.	CFR circa il 2%, in Francia — più del 10%.
------------------------------------------------	---------------------------------------	--------------------------------------------

The second part of the article under examination is undoubtedly not immune to translation issues. Starting from the title of the first section, the Russian term *Доступно* is rendered by DeepL and Yandex with two different Italian adjectives, namely *accessibile* and *disponibile*, respectively. Although both Italian terms may constitute correct Italian equivalents for the Russian adjective under examination (Kovalev, 2020), they do not equally suit this specific context. Indeed, in the Italian language, *accessibile* is more frequently used to express the concept of availability of data (Tiberii, 2018) and consequently better conveys the meaning of the source document. As a result, DeepL's translation variant is preferable when compared to Yandex's one.

The same does not hold for the second adjective of the title under examination, namely *несопоставимо*. It is indeed rendered by DeepL with the Italian adjective *non comparabile*, whereas in Yandex's translation version the Italian adjective *disparabile* is displayed. Starting from saying that *disparabile* does not have any correspondence in the Italian language (Zingarelli, 2020) and cannot, therefore, be considered acceptable, the Italian adjective suggested by DeepL perfectly suits this specific context and properly conveys the intended meaning (Kovalev, 2020).

Particularly interesting is the rendering of the Russian expression *По счетчикам в Сети видно*, as the two translation tools provide significantly different translation variants. On the one hand, Yandex remains perfectly consistent with the original syntactical structure and organizes the sentence as follows. The first part of the section under examination, namely *По счетчикам* is correctly rendered with the Italian expression *Secondo i contatori*, the Russian term *Сети* (Kovalev, 2020) is properly translated with the Italian noun *rete* and the Russian predicative *видно* is given the Italian equivalent *è chiaro che* (Kovalev, 2020). Yandex's translation variant is overall accurate, as it perfectly reflects the syntactical structure of the source document simultaneously complying with Italian grammatical rules. On the other hand, DeepL completely readjusts the original sentence and provides a natural-sounding translation variant. The Italian expression *I contatori*, which constitutes the rendering of the Russian terms *По*

счетчикам, is assigned the role of subject of the sentence and is directly linked to the Italian verb *mostrano*, used as the Italian equivalent for the Russian adverb *видно*. Finally, the Russian term *Сету* is correctly translated with the English noun *Web*, which, although originally foreigner, has become so sufficiently widespread in Italy that is perfectly understandable by the intended readers. As a result, although remarkably distinct, both translation variants can be considered equally acceptable.

The Russian verb *отличаются* is translated by DeepL and Yandex with two different Italian verbs, namely *variano* and *sono diversi*. In this case, since both suggested verbs not only correctly convey the intended meaning (Kovalev, 2020) but also suit this specific context, they are equally acceptable.

By observing the remaining part of the section under examination, we can easily notice a number of translation issues. The Russian expression *числу смертей* is rendered as *tasso di mortalità* by DeepL and *numero di morti* by Yandex. In this case, the suggested Italian equivalents, although significantly different, can be considered mutually interchangeable, as they equally properly convey the intended meaning and suit the particular context they fall within. The same undoubtedly applies to several other Russian terms, namely *то есть*, translated as *ovvero* and *cioè, приблизительно*, which is given the Italian equivalents *più o meno* and *approssimativamente, в которых*, associated with the Italian adverbs *in cui* and *dove*, and, finally, *время*, rendered with the Italian nouns *periodo* and *momento*.

The rendering of the Russian past participles *умерших* and *заболевшим* undoubtedly constitutes an interesting basis for discussion. The former is indeed translated as *decessi* by DeepL and *morti* by Yandex. In this case, although the noun *decessi*, in the Italian language, defines an inanimate entity and is used as equivalent for an animate one, can be considered a correct translation variant, as it properly conveys the intended meaning (Kovalev, 2020) and suits this specific context. On the contrary, the Italian term *morti* perfectly reflects the meaning (Kovalev, 2020) and the grammatical features of the original Russian term. As for the latter, it is given by DeepL and Yandex two completely different Italian equivalents, namely *casi* and *malati*, respectively. On the one hand, the Italian noun suggested by Yandex perfectly renders the meaning and the grammatical features of the source document. On the other hand, the one displayed in DeepL's translation variant, namely *casi*, may convey a great deal of different meanings

according to its contexts of use (Zingarelli, 2020). However, since, in the Italian language, it frequently denotes the people who have contracted a certain illness, it not only conveys the intended meaning but also suits this specific context and can therefore be considered acceptable.

The Russian adjective *соразмерным* is translated by DeepL and Yandex with two different Italian adjectives, namely *commisurato* and *proporzionato*, respectively. In this case, since the adjective suggested by DeepL more accurately conveys the meaning of the original Russian term and better suits this particular context, has to be considered preferable. (Kovalev, 2020). Nonetheless, the Italian adjective *comparabile* would undoubtedly be a more accurate translation variant (Kovalev, 2020).

The Russian verb *быть*, which implicitly occurs twice in the last sentence of the section under examination, the first one when it is omitted, according to Russian grammatical rules, immediately after the acronym *CFR* and the second one when it is replaced by an hyphen, is correctly rendered by DeepL with the Italian verb *è*, and completely ignored in Yandex’s translation version. In this case, being the verb under examination a meaningful element for the translation versions to properly convey the intended meaning (Kovalev, 2020) and remain compliant with Italian grammatical rules, according to which it cannot absolutely be omitted, only DeepL’s translation version can be considered acceptable.

The international acronym *CFR* is introduced, in DeepL’s translation variant, by the definite article *il*, whereas in Yandex’s one no articles precede the term underexamination. In this case, constituting it a notorious and already-mentioned in the text entity, a definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993).

Finally, the Russian adjective *более*, which is displayed in the source document in its comparative form, is rendered as *superiore al* by DeepL and *più del* by Yandex. In this case, since the suggested equivalents correctly convey the original meaning and suit this specific context, they can both be considered equally correct.

SOURCE TEXT	DEEPL	YANDEX
Кроме того, отчетность по CFR меняется и не	Inoltre, il reporting CFR sta cambiando e non è	Inoltre, la segnalazione di CFR sta cambiando e non

<p>факт, что причина динамики в росте смертности. Так, во Франции в начале апреля CFR резко прибавил. Оказалось, там просто начали учитывать смертность в домах престарелых, до этого собирались только сведения из медучреждений. Инна Данилова: «Несмотря на доступность данных, пока не очень понятно, что стоит за ними и насколько они сопоставимы между странами».</p>	<p>certo che la causa della dinamica sia un aumento della mortalità. In Francia, ad esempio, il QCR è aumentato notevolmente all'inizio di aprile. Come si è scoperto, hanno semplicemente iniziato a contare i decessi nelle case di cura, mentre in precedenza raccoglievano solo dati da istituti medici. Inna Danilova: "Nonostante la disponibilità di dati, non è ancora molto chiaro cosa ci sia dietro e quanto siano comparabili tra i vari Paesi.</p>	<p>è il fatto che la causa della dinamica è l'aumento della mortalità. Così, in Francia, ALL'inizio di aprile, il CFR ha bruscamente aggiunto. Si è scoperto che c'era appena iniziato a prendere in considerazione la mortalità nelle case di cura, prima che solo le informazioni raccolte dalle istituzioni mediche. Inna Danilova: "nonostante la disponibilità dei dati, non è ancora molto chiaro cosa c'è dietro di loro e quanto siano comparabili tra i paesi».</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Moving to the third part of the article, a number of discrepancies can be observed between the provided translation variants. The Russian term *отчетность* is rendered differently by the two translation tools. Indeed, it is translated by DeepL with the English noun *reporting*, whereas in Yandex's translation variant, the Italian term *segnalazione* is displayed. In this case, both suggested equivalents do not seem to properly convey the original meaning and cannot therefore be considered acceptable. The Italian terms *rappporto* and *rendiconto* may instead constitute possible translation options as they both reflect the source document's meaning (Kovalev, 2020) and suit this specific context.

The rendering of the Russian noun *факт* undoubtedly constitutes a major translation issue. It is indeed translated with the Italian adjective *certo* by DeepL, whereas in Yandex's translation version the Italian noun *fatto* is displayed. Although the two Italian terms may constitute possible Italian equivalents for the Russian noun under examination (Kovalev, 2020), they do not equally convey the intended meaning. Indeed, in this specific case, the Italian adjective *certo*, suggested by DeepL, is the only translation variant that perfectly conveys the source document's sense.

The Russian verb *быть*, implicitly present in the source document immediately after the Russian noun *динамики*, is rendered by the translation tools with the Italian verb

essere in two different verbal forms. Indeed, while in Yandex's translation version the mentioned verb appears in its simple present form, in DeepL's one the subjunctive is displayed. In this case, since, according to Italian grammatical rules, the subjunctive is usually used in the objective propositions that depend on verbs expressing uncertainty (Treccani, 2020), the verb suggested by DeepL undoubtedly constitutes an accurate translation equivalent, whereas Yandex's translation version cannot be considered acceptable.

The Russian noun *рочме*, correctly rendered by both translation tools as *aumento* (Kovalev, 2020) is introduced in DeepL's translation variant by the indefinite article *un*, whereas in Yandex's one the definite article *il* precedes the Italian noun under examination. In this case, since no relevant differences in meaning are observed when comparing the two translation variants, they can be both considered acceptable.

The second sentence of the section under examination clearly presents several translation issues. Firstly, DeepL and Yandex render the Russian adverb *Так* with two different Italian expressions, namely *ad esempio* and *così*, respectively. Although both suggested translation variants may constitute correct translation equivalents for the Russian adverb under examination (Kovalev, 2020), in this specific case, the Italian word sequence *ad esempio* is able to better convey the sense expressed in the source document when compared with the Italian adverb *così*, suggested by Yandex. As a consequence, DeepL's version can be undoubtedly considered preferable.

Secondly, the international acronym *CFR*, which does not theoretically need any translation, is correctly maintained by Yandex as originally written in the source document, and erroneously rendered by DeepL with the acronym *QCR*, which does not present any consistent correspondence either in English or in Italian language and cannot, therefore, be considered acceptable.

Thirdly, the Russian verb *прибавил*, which comes in association, in the original text, with the Russian adverb *резко*, is rendered by DeepL as *è aumentato notevolmente* and *ha bruscamente aggiunto* by Yandex. Starting from saying that both suggested Italian adverbs may represent correct Italian equivalents for the Russian adverb *резко* (Kovalev, 2020), the same does not apply to the renderings of the Russian verb under examination. Indeed, while the translation variant provided by DeepL perfectly conveys the source text's meaning (Kovalev, 2020), Yandex's one cannot be considered acceptable. Finally,

Yandex's translation version presents a format error, as the complex preposition *all'* is fully capitalized and, being a preposition located in the middle of a sentence, does not comply with Italian grammatical rules.

The following sentence constitutes a focal point of our analysis. In fact, a quick look is sufficient to notice that both translation variants present several issues. Proceeding in order of appearance in the source document, DeepL and Yandex render differently the Russian verb *Оказалось*. In fact, it is translated by DeepL as *Come si è scoperto*, whereas in Yandex's translation variant the Italian phrase *Si è scoperto che* is displayed. In this case, since both suggested equivalents succeed in properly conveying the intended meaning, they can be considered equally correct (Kovalev, 2020).

The Russian adverb *там* is completely ignored by both translation tools. In this specific case, constituting the term a meaningful element of the original sentence, its omission cannot be considered acceptable.

The Russian adverb *просто* is correctly rendered by DeepL with the Italian adverb *semplicemente*, which fully reflects the original meaning (Kovalev, 2020), whereas *appena* constitutes the erroneous translation equivalent provided by Yandex.

The Russian verb *начали* is assigned two different Italian equivalents by DeepL and Yandex, namely *hanno iniziato* and *c'era iniziato*, respectively. While the verb suggested by Yandex does not present any correspondence in the Italian language (Kovalev, 2020) DeepL's one perfectly conveys the intended meaning and the grammatical features of the original verb (Kovalev, 2020). As a consequence, only DeepL's translation version can be considered acceptable.

Although the translation variants provided by DeepL and Yandex for the Russian verb *учитывать*, namely *contare* and *prendere in considerazione* may significantly differ in meaning (Zingarelli, 2020) they can be both considered acceptable. Starting from saying the two Italian verbs may constitute correct Italian equivalents for the Russian term under examination (Kovalev, 2020), they seem to convey, in this specific case, a common message. Indeed, constituting the statistics on coronavirus consequences the main topic of the article at stake, an Italian reader would associate the Italian verb *prendere in considerazione*, among its diverse meanings (Zingarelli, 2020) to the idea of including the people coming from the nursing homes in the total number of coronavirus

cases. This undoubtedly makes the two suggested verbs lexically close and equally acceptable as equivalents of the Russian verb *учитывать*.

The same does not hold for the Russian noun *смертность*, which is rendered as *decessi* by DeepL and *mortalità* by Yandex. Indeed, although the two suggested Italian terms can be considered lexically related, do not constitute equally correct Italian equivalents for the Russian term under examination. In this specific case, *mortalità* clearly better conveys the intended meaning (Kovalev, 2020) and can be therefore considered preferable when compared to the Italian noun *decessi*.

The Italian equivalents suggested by DeepL and Yandex for the Russian expression *до этого*, namely *in precedenza* and *prima* can be considered equally acceptable (Kovalev, 2020). Nonetheless, DeepL's inclusion of the Italian adversative conjunction *mentre* in its translation variant contributes to underling the sense of disjunction expressed in the source document and, consequently, making DeepL's translation version preferable when compared to Yandex's one.

The Russian verb *собирались* is correctly rendered by DeepL with the Italian passive verb *venivano raccolti*, which perfectly reflects the meaning (Kovalev, 2020), the tense, the imperfect aspect, and the other grammatical features of the original Russian verb. On the contrary, the Italian past participle suggested by Yandex, namely *raccolte*, not only does not respect the grammatical features of the Russian verb under examination but also does not comply with the syntactical structure it is inserted within, so that is not able to assume the role of the main verb of the given sentence. As a result, it absolutely cannot be considered acceptable.

The Russian noun *сведения* is rendered by DeepL and Yandex with two different Italian terms, namely *dati* and *informazioni*, respectively, which may constitute, in theory, possible Italian equivalents for the Russian term under examination (Kovalev, 2020). However, they do not equally suit this specific context. In this case, indeed, referring the original Russian term *сведения* to the numerical data on which official statistics will be produced, the Italian noun *dati* is undoubtedly closer to the original meaning expressed in the source document. As a result, DeepL's translation variant is considered preferable when compared to Yandex's one. Nonetheless, in DeepL's version, no articles are displayed to introduce the Italian noun *dati*. Defining the mentioned term highly specific

and already well-known entities, a definite article is necessary to fully convey the intended meaning (Lepschy & Lepschy, 1993).

The Russian noun *медучреждений* is rendered differently by the two translation tools. Indeed, while it is translated by DeepL as *istituti medici*, in Yandex's translation variant, the Italian word sequence *istituzioni mediche* is displayed. Denoting the Russian term under examination the healthcare facilities that are intended to release the data to be included in the official statistics, DeepL's translation equivalent seems to properly convey the original meaning, whereas the same does not apply to Yandex's one (Zingarelli, 2020). However, in DeepL's translation variant, the Italian multiword expression *istituti medici* is introduced by the simple preposition *di*, which, although correctly renders the specification relation expressed in the source document, seems to simply refer to some non-specific healthcare facilities, failing in properly reflecting the original meaning. On the contrary, since all the French healthcare facilities are considered, a definite article has to come in association with the simple preposition *di* (Lepschy & Lepschy, 1993).

Yandex's translation variant displays an orthography error, as the rendering of direct speech, correctly reported between a couple of inverted commas, begins with a lower-case letter. Constituting it a completely-formed and independent sentence, an upper-case initial letter is undoubtedly needed to comply with Italian grammatical rules (Accademia della Crusca, 2020).

The Russian noun *данных* is introduced by the simple preposition *di* in DeepL's translation, whereas in Yandex's one the complex preposition *dei* is displayed. In this case, starting from saying that in both cases the specification relation expressed in the source document is properly rendered, defining the term under examination an already mentioned in the text and specific entity, the definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). Therefore, Yandex's translation version is undoubtedly preferable.

The rendering of the Russian dependent clause *что стоит за ними* clearly represents two main translation issues. In fact, it is rendered by DeepL as *cosa ci sia dietro*, whereas the Italian expression *cosa c'è dietro di loro* is suggested by Yandex. Firstly, depending on the clause under examination on the main verb expressing uncertainty, according to Italian grammatical rules, the subjunctive verbal form is preferable (Treccani, 2020). Secondly, the Russian expression *за ними*, as mentioned above, is

rendered differently by the two translation tools. It is indeed translated by DeepL with the Italian pronoun *ci*, and by Yandex with the Italian expression *dietro di loro*. In this case, although Yandex's version undoubtedly fully reflects the original syntactical structure, DeepL's one constitutes a more natural and Italian-sounding equivalent for the term under examination. All things considered, DeepL's translation variant has to be considered preferable when compared to Yandex's one.

Two translation issues also arise when considering how the two translation tools translate the Russian expression *между странами*, translated by DeepL as *tra i vari Paesi* and by Yandex as *tra i paesi*. We will start from saying that the Italian noun *paesi* represents a correct Italian equivalent for the Russian term under examination (Kovalev, 2020) and the simple preposition *tra* correctly renders the original syntactical structure. However, the two translation variants cannot be considered perfectly equal, as in DeepL's one the Italian term *paesi* begins with a capital letter, whereas in Yandex's one it is displayed in lower-case letters. In this case, since in the Italian language both forms are accepted, the two translation versions can be considered correct. Moreover, DeepL inserts the Italian adjective *vari*, which does not present any equivalent in the original text. Although this translation choice is not necessary to properly convey the original meaning, since the added adjective does not constitute a misleading element and seems to remain consistent with the overall sense and syntactical structure of the sentence, it cannot be considered a mistake.

Finally, both translation versions display a format error. On the one hand, DeepL fails in properly enclosing the quote in the last part of the section under examination between two pairs of quotation marks, as just the initial one is provided. On the other hand, Yandex's version, by displaying a pair of Italian quotation marks at the beginning and a pair of Russian quotation marks at the end of the mentioned quote, undoubtedly does not contribute to creating consistency within the text.

3.6. COVID-19 mortality rate - A demographer's perspective on the statistics of causes of death in Russia and worldwide

The third popular-science text, *Смертность от COVID 19 - Взгляд демографа на статистику причин смерти в России и мире*, regards the intense public debate

arose in Russia about the actual reliability of the accountability methods of the data concerning COVID-19 mortality within the country. More specifically, two experienced demographers, Sergej Timonin e Anatolij Višnevskij are asked to answer key questions relating to this burning topic.

SOURCE TEXT	DEEPL	YANDEX
Смертность от COVID 19 - Взгляд демографа на статистику причин смерти в России и мире	Mortalità da COVID 19 - Una prospettiva demografica sulle statistiche delle cause di morte in Russia e nel mondo	Mortalità da COVID 19 - Vista demografica sulle statistiche delle cause di morte in Russia e nel mondo

When observing the title of the article under examination, we can easily notice two interesting translation issues. Firstly, the Russian noun *Взгляд* is rendered by DeepL as *prospettiva* and by Yandex as *Vista*. In this case, starting from saying that the Italian noun *prospettiva* seems to better convey the original meaning, when compared to *vista* (Kovalev, 2020), the Italian terms *punto di vista*, *parere*, and *opinione* may undoubtedly constitute better translation equivalents for the Russian noun under examination (Kovalev, 2020).

Secondly, the Russian genitive *демографа* is rendered by both translation tools with the Italian adjective *demografica*. In this specific case, although the suggested adjective can be accepted, the Italian expression *di un demografo* would undoubtedly constitute not only a better equivalent from a lexical point of view (Kovalev, 2020) but also contribute to making the translated title more effective and consistent with the general topic discussed throughout the article, which explicitly mentions the two interviewed demographers.

SOURCE TEXT	DEEPL	YANDEX
Многочисленные статьи в зарубежных и русскоязычных СМИ о «русском чуде» и	Numerosi articoli sui media stranieri e in lingua russa sul "miracolo russo" e sul "mistero del	Numerosi articoli stranieri e russi dei MEDIA «russo miracolo» e «l'enigma di comportamento

<p>«загадке поведения коронавируса» в России, а также признания министра здравоохранения республики Дагестан о соотношении смертей от коронавируса и внебольничной пневмонии породили большую общественную дискуссию о методах учёта статистики причин смерти.</p>	<p>comportamento del coronavirus" in Russia, così come le ammissioni del Ministro della Salute della Repubblica del Daghestan sul rapporto tra i decessi da coronavirus e la polmonite fuori dall'ospedale hanno generato un importante dibattito pubblico sui metodi di contabilizzazione delle statistiche delle cause di morte.</p>	<p>coronavirus» in Russia, ma anche il riconoscimento del ministro della sanità della repubblica del Daghestan circa il rapporto di morti da coronavirus e polmonite acquisita in comunità generato un grande dibattito pubblico sui metodi di contabilità le statistiche sulle cause di morte.</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

A number of discrepancies between the two provided translation variants can be observed with regard to the first part of the article. Proceeding in order of appearance throughout the texts, the Russian compound adjective *русскоязычных* is translated by DeepL and Yandex with two different Italian expressions, namely *in lingua russa* and *russi*, respectively. Since the adjective under examination, as deductible from its constituting components, specifically describes the characteristic of being Russian-speaking and may refer to both animate and inanimate entities (Kovalev, 2020), the adjective suggested by DeepL undoubtedly better conveys the intended meaning and has to be therefore considered preferable.

The Russian acronym *СМИ*, which in the Russian language stands for *Средства массовой информации*, is correctly rendered by DeepL and Yandex with the Italian noun *media* (Kovalev, 2020). Nonetheless, the two translation tools do not equally reflect the syntactical structure and format of the source document. Indeed, the location complement expressed in the source document is perfectly reflected only by DeepL, which introduces *media* with the complex preposition *sui*. On the other hand, in Yandex's translation variant, the Italian noun under examination is preceded by the complex preposition *dei*, usually expressing, in the Italian language, the specification relation (Treccani, 2020). Moreover, since the rendering of the Russian adjectives *зарубежных* and *русскоязычных*, referring in the source document to the Russian term *СМИ*, are collocated immediately before the complex preposition *dei*, they seem to be related to the

Italian noun *articoli*, which constitutes the correct translation of the Russian term *статьи*. This translation mistake completely distorts the original meaning and prevents Yandex's translation from being acceptable. As for the format, in Yandex's translation variant the Italian noun *MEDIA* is completely capitalized. Constituting the term a common noun and not an acronym, like its Russian equivalent, there is no reason to write it in capital letters. All things considered, only DeepL's translation variant can be considered correct.

When comparing the two provided translation versions, it appears that Yandex almost completely fails in properly rendering the syntactical structure of a consistent part of the section under examination. The Russian preposition *о*, expressing, in the Russian language, the topic, and linking in the source document, the Russian acronym *СМИ* with the expressions «*русском чуде*» and «*загадке поведения коронавируса*», is indeed correctly rendered by DeepL with the Italian complex preposition *sul*. On the contrary, in Yandex's translation version it is completely ignored, so that the renderings of the term to which it originally refers, seem to be completely disconnected from the rest of the sentence. As a consequence, Yandex's translation cannot be accepted.

The Russian word sequence «*русском чуде*» is rendered differently by the two translation tools. Indeed, while DeepL correctly translates it as "*miracolo russo*", Yandex's translation variant displays the same Italian terms suggested by DeepL, but in the reverse order. In so doing, it fails in remaining compliant with Italian grammatical rules, according to which the adjectives usually follow the noun to which they refer (Treccani, 2020). Moreover, Yandex's version contains a format error, as a Russian pair of inverted commas is used to enclose the rendering of the expression under examination. As a result, only DeepL's translation can be considered acceptable.

The same happens shortly after in the sentence, where the expression *l'enigma di comportamento coronavirus*, suggested by Yandex as the Italian equivalent of the Russian word sequence *загадке поведения коронавируса* is inserted within a couple of Russian inverted commas. Besides its format, two other major translation issues may be observed in the renderings of the Russian expression under examination. Firstly, the Russian noun *загадке* is translated as *mistero* by DeepL and *enigma* by Yandex. Although both terms may be considered valid Italian equivalents (Kovalev, 2020), since *mistero* better conveys the intended meaning and suits this specific context, it can be

considered preferable (Zingarelli, 2020). Secondly, while in DeepL's translation version the specification relation expressed in the source document is rendered by introducing the term *coronavirus* with the complex *del*, in Yandex's one no prepositions precede the Italian term at stake, which seems to be completely disconnected by the rest of the sentence.

The Russian expression *а также* is rendered by DeepL and Yandex with two different Italian equivalents, namely *così come* and *ma anche*, respectively. In this case, both translation variants may be considered acceptable, as they are equally able to convey the intended meaning (Kovalev, 2020) and properly express the sense of continuity of the source sentence.

The two translation tools differently translate the Russian noun *признания*. It is indeed rendered by DeepL as *ammissioni*, whereas in Yandex's translation the Italian noun *riconoscimento* is displayed. Starting from saying that only the equivalent suggested by DeepL fully reflects the plural number of the original term, although, generally, speaking, from a lexical point of view, both Italian nouns may constitute correct translation option for the Russian term under examination (Kovalev, 2020), they undoubtedly do not equally suit this specific context. In fact, in this case, only *ammissioni* is able to convey the meaning of the source document and can be therefore be considered correct, whereas the same does not hold for *riconoscimento*.

Particularly interesting is the rendering of the Russian multiword expression *министра здравоохранения республики Дагестан*. It is indeed translated by DeepL as *del Ministro della Salute della Repubblica del Daghestan* and by Yandex as *del ministro della sanità della repubblica del Daghestan*. By comparing the provided translation variants, three discrepancies can be observed. Firstly, the Russian noun *здравоохранения* is translated by DeepL and Yandex with two distinct Italian terms, namely *Salute* and *sanità*, which moreover present a format difference, as the former begins with a capital letter, whereas the latter is written by using lower-case letters only. From a lexical point of view, the term suggested by Yandex is undoubtedly closer to the original Russian noun (Kovalev, 2020) and has to be therefore considered preferable. Nonetheless, according to Italian grammatical rules, constituting the expression under examination the official designation of a governmental institution, an initial upper-case letter is required

(Treccani, 2020). The same applies to the Italian terms *Ministro* and *Repubblica*, correctly displayed with an Italian capitalized letter only in DeepL's translation version.

The Russian preposition *о*, expressing in the source documents, as mentioned above, the topic and linking *Дагестан* and *соотношению*, correctly rendered by both translation tools as *Dagestan* and *rapporto* (Kovalev, 2020), is translated by DeepL with the complex preposition *sul*, whereas the preposition *circa* is displayed in Yandex's translation versions. Since both Italian equivalents correctly convey not only the intended meaning but also the original syntactical role of the Russian preposition under examination, they can be considered equally correct.

The Russian noun *смертей* is rendered by DeepL and Yandex with two different Italian terms, namely *decessi* and *morti*. Since the suggested equivalents equally accurately convey the original meaning and suit the specific context they fall between, they can undoubtedly be considered equally correct. Nonetheless, although lexically interchangeable, the two terms present a discrepancy. In fact, in DeepL's translation, the Italian term *decessi* is introduced by the complex preposition *tra i*, whereas in Yandex's one the simple preposition *di* precedes the Italian noun *morti*. Starting from saying that, defining the terms under examination not all the deaths, but just the specific ones caused by the new coronavirus, a definite article is necessary to convey the intended meaning, the complex preposition suggested by DeepL more frequently comes in association with the Italian term on which it depends, namely *rapporto* (Tiberii, 2018). As a consequence, DeepL's version is undoubtedly preferable.

The Russian medical term *внебольничной пневмонии* is translated by DeepL with the Italian expression *polmonite fuori dall'ospedale*, whereas in Yandex's translation version the Italian multiword expression *polmonite acquisita in comunità* is displayed. Since the original Russian word sequence refers to lung diseases contracted outside of the healthcare system (Mandell et al., 2007), only the Italian equivalent suggested by Yandex is able to properly convey the intended meaning. On the contrary, DeepL's translation variant cannot be accepted.

While DeepL accurately renders the Russian verb *породили* with the Italian verb *hanno generato*, which perfectly reflects the meaning (Kovalev, 2020), the tense and the grammatical features of the original verb, in Yandex's translation version the mere past participle *generato* is displayed. Since the verb suggested by DeepL cannot undoubtedly

assume the role of the main verb of the sentence under examination, and consequently be accepted, only DeepL's equivalent can be considered correct.

The Russian adjective *большую*, which, in the source document refers to the Russian noun *дискуссию*, correctly rendered as *dibattito* (Kovalev, 2020), is translated by DeepL and Yandex with two different Italian adjectives, namely *importante* and *grande*, respectively. The Italian equivalent suggested by Yandex is doubtless closer to the meaning of the source document (Kovalev, 2020). Nonetheless, although in the Italian language the two Italian adjectives may convey significantly different meanings according to their context of use (Zingarelli, 2020), when related to the Italian noun *dibattito*, they are both able to express the degree of the historical relevance of the debate to which it refers (Zingarelli, 2020). As a result, the two translation variants can be considered acceptable.

The Russian noun *учёта* is translated differently by the two translation tools. Indeed, it is rendered as *contabilizzazione* by DeepL and as *contabilità* by Yandex. Although both Italian terms are enumerated among the possible Italian equivalents for the Russian noun under examination (Kovalev, 2020), they do not equally suit this specific context. Indeed, denoting the Italian noun *contabilizzazione* the specific action of computing, counting, and recording in the appropriate accounting records (Zingarelli, 2020) it perfectly conveys the original meaning and should, therefore, to be considered preferable when compared to the one suggested by Yandex.

Finally, the rendering of the Russian noun *статистике*, correctly translated by both translation tools as *statistiche* (Kovalev, 2020) presents three translation issues. Firstly, since the Russian term under examination, displayed in the source document in its singular form, is used, in the Russian language, also to refer to statistical data, which usually constitute a plural noun in the Italian language, the suggested plural equivalent can be considered acceptable. Secondly, in DeepL's translation variant, the Italian term *statistiche* introduced by the complex preposition *delle*, whereas in Yandex's one the definite article *le* is displayed. Since only the complex preposition suggested by DeepL is able to properly convey the specification relation expressed in the original text, Yandex's translation version cannot be considered acceptable.

Finally, the Italian term *statistiche* is linked to its following noun *cause*, which constitutes the proper rendering of the Russian noun *причин* (Kovalev, 2020), by the

complex preposition *delle* in DeepL’s translation, whereas in Yandex’s one the complex preposition *sulle* appears. Although DeepL equivalent more directly reflects the original specification relation, since both translation options properly convey the intended meaning (Kovalev, 2020) and perfectly suit the specific context they fall within, they can be considered equally correct.

SOURCE TEXT	DEEPL	YANDEX
Насколько достоверны собираемые официальными ведомствами данные? Могут ли власти полагаться на них при принятии решений? Какие два типа статистических данных должны разрабатываться в условиях эпидемии — на эти и многие другие вопросы отвечают демографы Сергей Тимонин и Анатолий Вишнеvский.	Quanto sono affidabili i dati raccolti dalle agenzie ufficiali? Le autorità possono fare affidamento su di loro quando prendono decisioni? Quali due tipi di statistiche dovrebbero essere sviluppate nel contesto di un'epidemia - a queste e a molte altre domande rispondono i demografi Sergei Timonin e Anatoly Vishnevsky.	Quanto sono affidabili i dati raccolti dalle agenzie ufficiali? Le autorità possono fare affidamento su di loro quando prendono decisioni? Quali due tipi di dati statistici devono essere sviluppati in condizioni epidemiche- queste e molte altre domande rispondono demografi Sergey Tymonin e Anatoly Vishnevsky.

The second part of the article is given by DeepL and Yandex two quite similar translation versions, which on the whole present a small amount of translation issues. Proceeding in order of appearance in the source document, the Russian word sequence *статистических данных* is rendered as *statistiche* by DeepL, whereas the Italian multiword expression *dati statistici* is suggested by Yandex. Denoting the Italian noun *statistiche*, in the Italian language, the specific branch of science studying collective phenomena using mathematical methods as well as the actual masses of data collected to produce statistical reports of various nature, the suggested Italian equivalents may be considered, in this case, as synonyms conveying the meaning expressed in the source documents.

The Russian verb *должны разрабатываться* is rendered differently by the two translation tools. On the one hand, DeepL translates it with the Italian conditional verb

dovrebbero essere sviluppate. On the other hand, in Yandex's translation version the simple present verb *devono essere sviluppati* is displayed. In this case, both suggested equivalents seem to be equally correct from a lexical point of view (Kovalev, 2020). Moreover, although the two different verbal forms may render the two translation variants slightly different in style, they both properly suit this specific context and can be therefore considered equally acceptable.

The rendering of the Russian expression *в условиях эпидемии* undoubtedly constitutes a discrepancy between the two provided translation variants. In fact, while DeepL translates it as *nel contesto di un'epidemia*, the Italian multiword expression *in condizioni epidemiche* is suggested. On the one hand, the Italian noun *condizioni* is undoubtedly closer in meaning to the original Russian term than *contest* (Kovalev, 2020). On the other hand, the Italian equivalent *un'epidemia*, proposed by DeepL, better reflects the grammatical features of the Russian noun *эпидемии*. However, although significantly different, all things considered, the two provided Italian equivalents seem to properly convey the intended meaning (Kovalev, 2020) and constitute easily readable and completely understandable translation options. As a result, they can be considered equally acceptable.

The Italian terms *queste* and *molte altre*, which constitute, in both translation versions, the correct rendering of the original Russian terms *эти* and *многие другие* (Kovalev, 2020), are introduced in DeepL's translation, by the simple preposition *a*, whereas in Yandex's one, no prepositions precede the Italian terms under examination. In this case, being the simple preposition *a*, in the Italian language, usually inserted between the verb *rispondere*, i.e. the Italian equivalent of the Russian verb *отвечать*, linked, in the source document to the Russian terms *эти* and *многие другие*, and its object, the simple preposition *a* is necessary to convey the intended meaning and the syntactical structure of the source document. Moreover, since the Russian terms at stake, and consequently their Italian counterparts, are linked by a coordinating conjunction and collocated at the same syntactical level within the sentence, one simple preposition, preceding the first term would be sufficient, as it performs its intended action also on the second element of the couple. Nonetheless, the presence of two simple prepositions, located immediately before each term does not constitute a translation mistake.

The Italian noun *demografi*, which constitutes the Italian equivalent suggested by both translation tools for the Russian term *демографы*, is introduced in DeepL's translation by the definite article *i*, whereas in Yandex's one no articles precede the noun under examination, which seems to be completely disconnected from the rest of the sentence. In fact, in this case, since the proper names of the demographers whose point of view is discussed in the present article, is mentioned, *demografi* is characterized by a certain degree of specificity and a definite article is consequently necessary to convey the intended meaning (Lepschy & Lepschy, 1993).

Finally, several transliteration mistakes can be observed in both translation variants. Indeed, the second demographer's proper name *Анатолий Вишнеvский* is erroneously transliterated as *Anatoly Vishnevsky*, whereas, for the translations to remain compliant with the rules of scientific transliteration of Cyrillic alphabet, *Anatolij Višnevskij* should be displayed. The same holds for the Russian proper name *Сергей*, which is erroneously transliterated by DeepL as *Sergei* and as *Sergey* by Yandex. According to scientific transliteration rules, the Italian equivalent *Sergej* would represent a correct translation option.

SOURCE TEXT	DEEPL	YANDEX
Согласно отчёту ВОЗ, основанному на данных, полученных от официальных национальных источников, в России на 28 мая было зарегистрировано 379 тысяч подтвержденных случаев COVID-19. Из них 4,1 тысяча (1,09%) закончились летальным исходом. При сравнении этих данных со статистикой по другим странам обращает на себя внимание и вызывает неизбежные вопросы исключительно	Secondo il rapporto dell'OMS, sulla base di dati provenienti da fonti ufficiali nazionali, al 28 maggio sono stati registrati in Russia 379.000 casi confermati di COVID-19. Di questi, 4,1 migliaia (1,09%) si sono conclusi con la morte. Confrontando questi dati con le statistiche di altri paesi, il tasso di letalità estremamente basso causato dal coronavirus in Russia attira l'attenzione e solleva inevitabili domande.	Secondo il rapporto DELL'OMS, sulla base dei dati ottenuti da fonti nazionali ufficiali, in Russia il 28 maggio sono stati registrati 379 migliaia di casi confermati COVID-19. Di questi, 4,1 mille (1,09%) si è conclusa con un esito fatale. Quando si confrontano questi dati con le statistiche di altri paesi richiama l'attenzione e solleva questioni inevitabili eccezionalmente basso

низкий уровень летальности от коронавируса в России.		tasso di mortalità da coronavirus in Russia.
------------------------------------------------------	--	----------------------------------------------

Moving to the third part of the article, several interesting bases for discussion can be observed. Starting from the very beginning, a format error is displayed in Yandex’s translation variant. Indeed, the Italian acronym *OMS*, which stands for *Organizzazione Mondiale della Sanità* and constitutes the correct rendering of the Russian acronym *ВОЗ* (Kovalev, 2020), is properly introduced by the complex preposition *dell’*, which is fully capitalized. The same does not apply to DeepL’s translation, which consequently represents the only acceptable translation variant.

The Russian term *данных*, correctly rendered by both translation tools as *dati* (Kovalev, 2020) is introduced by the simple preposition *di* in DeepL’s translation variant, whereas in Yandex’s one the complex preposition *dei* is suggested. In this case, defining the Italian terms under examination specific entities, namely a certain type of data released by official agencies, a definite article is necessary to convey the intended meaning (Lepschy & Lepschy, 1993). As a result, Yandex’s translation version is undoubtedly preferable.

Undoubtedly interesting is the rendering of the Russian expression *официальных национальных источников*. The Russian word sequence is indeed translated by the two translation tools with the same correct Italian terms (Kovalev, 2020), displayed in different orders. In particular, DeepL suggests the Italian word sequence *fonti ufficiali nazionali*, whereas in Yandex’s translation variant the Italian expression *fonti nazionali ufficiali* appears. Since, according to Russian grammatical rules, the adjective usually precedes the term to which it is related (Glazunova, 2016), in this case, proceeding in order of appearance in the text, *официальных* refers to the Russian word sequence *национальных источников*. In addition, the second Russian adjective, namely *национальных* refers to the Russian noun *источников*. When evaluating the provided translation variants, we need to consider Italian grammatical rules, according to which, except for in particular cases, the adjective usually follows the noun to which it refers. Hence, the two provided translation variants do not convey the same meaning and cannot therefore be considered equal. In fact, only Yandex’s version perfectly refers to the original structure, as the adjective *ufficiali*, which constitutes the rendering of the Russian

adjective *официальных*, refers to both *fonti* and *nazionali*, whereas *nazionali*, i.e. the translation of the Russian adjective *национальных*, refers to *fonti* only, which represents the Italian equivalent of the Russian noun *источников*. Since the same does not apply to DeepL's translation variant, only Yandex's one can be considered acceptable.

The Russian date *на 28 мая* is rendered by DeepL and Yandex with two different Italian expressions, namely *al 28 maggio* and *il 28 maggio*, which, in Italian language, convey two significantly distinct meaning. Indeed, on the one hand, when the complex preposition *al* precedes a certain date, a period of time lasting from a starting date in the past to that exact date is considered (Zingarelli, 2020). On the other hand, the definite article *il* is used to introduce an exact date if only that specific date is taken into account. Since, in this specific case, DeepL's translation variant is able to perfectly reflects the grammatical structure of the source document and convey its intended meaning (Kovalev, 2020), namely the idea that the mentioned cases of death from the new coronavirus have been registered from the beginning of the pandemic to May 28, it can be considered correct. The same does not apply to Yandex's version, which cannot consequently be considered acceptable.

The Russian numerical expression *379 тысяч* is rendered differently by DeepL and Yandex. Indeed, while DeepL translates it with the number *379.000*, in Yandex's translation variant, the Italian expression *379 migliaia* is displayed. In this case, since the two translation versions not only constitute in this case a synonymic pair (Zingarelli, 2020) but are also able to convey the source document's meaning (Kovalev, 2020), they can be considered equally acceptable.

The same does not hold for the other numerical expression of the section under examination, namely *4,1 тысяча*, which is correctly rendered by DeepL as *4,1 migliaia* (Kovalev, 2020), whereas the expression *4,1 mille* is suggested by Yandex. Since no correspondences have been found in the Italian language for the expression suggested by Yandex (Treccani, 2020), only DeepL's translation version can be considered acceptable.

The term *COVID-19* is introduced, in DeepL's translation variant, by the simple preposition *di*, whereas, in Yandex's one, no prepositions precede the term under examination, which seems to be completely disconnected from the rest of the sentence. In this case, the simple preposition *di* is necessary to render the specification relation linking the Russian terms *случаев* and *COVID-19* in the source document. Therefore,

only DeepL's version can be considered acceptable. However, worth noting undoubtedly is the word sequence *casi COVID-19*, frequently used in the Italian language to define coronavirus cases (Treccani, 2020), in which the second term, namely *COVID-19*, assumes the role as the proper noun of *casi*. Hence, if the adjective *confermati* were not inserted between the two terms under examination, Yandex's translation variant would be considered acceptable as well.

The rendering of the Russian expression *летальным исходом* clearly constitutes a discrepancy between the two provided translation variants. It is indeed translated by DeepL as *con la morte*, whereas the Italian equivalent *con un esito fatale* is suggested by Yandex. Starting from saying that Yandex's translation doubtless more accurately reflects the structure of the original Russian expression, since the two versions are able to render the syntactical structure and the intended meaning of the source document (Kovalev, 2020), they can both be considered acceptable.

The last sentence of the section under examination provides an interesting basis for discussion. Indeed, aside from several translation issues, which we will enumerate later on in the paragraph, DeepL provides a syntactically correct translation variant that properly conveys the original meaning. The same does not apply to Yandex, whose version does not result compliant with Italian grammatical rules, especially with regard to its erroneous word order. Indeed, the main verb of the sentence and its subject are not only displayed in an order that does not respect Italian grammatical rules, according to which the verb, except for in some special cases, usually follows the corresponding subject, but they are also completely disconnected from one another. This undoubtedly contributes to distorting the meaning of the whole translation variant and preventing it to being acceptable.

The Russian expression *При сравнении* is translated by DeepL and Yandex with two different Italian expressions, namely *Confrontando* and *Quando si confrontano*, respectively. The two translation versions, although significantly different in structure, are able to convey the intended meaning (Kovalev, 2020) and can consequently be both considered acceptable. Nonetheless, the Italian gerund suggested by DeepL seems to be more easily readable and natural sounding to an Italian reader, when compared to Yandex's translation equivalent.

The Russian verb *обращает на себя*, along with its direct object *внимание*, is rendered differently by the two translation tools. Indeed, while DeepL translates it as *attira*, in Yandex's translation version the Italian verb *richiama* is displayed. In this case, not only both suggested verbs come frequently in association, in contemporary Italian language, with the Italian noun *attenzione* (Tiberii, 2018), which constitutes the correct rendering of the Russian term *внимание* (Kovalev, 2020) but also convey the same meaning, namely the idea of directing someone's attention towards someone or something (Zingarelli, 2020), which perfectly corresponds to the one expressed in the source document. As a consequence, the two translation variants can be considered equally correct.

The Russian noun *вопросы* is translated as *domande* by DeepL and *questioni* by Yandex. In this case, although both suggested Italian nouns may constitute correct Italian equivalents for the Russian term at stake, they do not equally suit this specific context. Indeed, since the demographers' answers to the people questions regarding the controversial accounting methods for covid deaths constitute the main topic covered by the article under examination, we can assume that the Russian noun *вопросы* refers to the above-mentioned people's questions. As a result, DeepL's version can be considered preferable.

Particularly interesting is the rendering of the Russian adverb *исключительно*, translated as *estremamente* by DeepL and *eccezionalmente* by Yandex. Although the two Italian adverbs are enumerated in the list of possible Italian equivalents for the Russian adverb under examination (Kovalev, 2020), they convey, in the Italian language, a distinct meaning (Zingarelli, 2020). However, in this specific case, they both seem to suit the entity to which they refer in the source document, namely the mortality rate of coronavirus in Russia. In fact, it is simultaneously extremely low as very few cases of death from coronavirus have been registered within the Russian Federation borders and exceptionally low, as, such a low mortality rate represents an exception when compared to the one registered in the other countries. As a consequence, both translation variants can be considered acceptable.

Finally, the Russian word sequence *уровень летальности* is rendered by DeepL and Yandex with two different Italian expressions, namely *tasso di letalità* and *tasso di mortalità*, respectively. In this case, although the Italian noun *letalità* undoubtedly

constitutes a more accurate Italian equivalent for the Russian term *летальность*, and the expression *tasso di letalità* does convey the intended meaning and is understandable by an Italian-speaking reader, Yandex's version can be considered preferable. Indeed, the expression *tasso di mortalità*, considered as a whole, is more widespread in the Italian language, so that its two components form an actual collocation (Tiberii, 2020), and consequently contributes to rendering DeepL's translation more natural sounding.

4. QUANTITATIVE ANALYSIS OF THE RESULTS

Having conducted, in the preceding chapters, a brief qualitative analysis of the translations provided by the two translation tools, we hereby mark each translation error as belonging to one or more of the error categories that we have previously selected according to the specific relevant linguistic features characterizing the texts under examination. As mentioned above, a table will be inserted, showing the erroneous fragments as displayed in the original documents, in the translations released by DeepL and Yandex, and in a specifically-made human translation, used as reference translation. Moreover, each erroneous fragment will be associated with one or more error categories, by flagging the corresponding table cell. For the mentioned table to be easily readable and understandable to the intended readers, it will be divided according to the translation tool and the text type that is displayed. Therefore, a total of four tables will be inserted.

Afterward, the total number of translation errors will be counted and shown by means of several graphs in order to provide a visual and direct understanding of the data displayed in the tables. More specifically, before delving into the actual comparison between DeepL's and Yandex's translation performances, careful attention will be devoted to analyzing the results obtained for each translation tool with regard to the two text types under examination, namely specialized and popular-science texts. The graphs will, in this case, display the translation errors occurring in each error category separately, in descending order of number of translation errors. Subsequently, the two translation tools' performances will be compared on the basis of the number of errors occurring in the translations of the three specialized texts, the three popular-science texts, and in all texts, considered as a whole. Two different colored columns will indeed be used to represent the two translation tools, whose translation errors will be shown according to the error category to which they belong.

Finally, in order to provide an overall assessment of each translation tool's performance, the percentage of closeness between the provided translations and a specifically made human translation, used as reference translation, will be displayed. For a matter of consistency, this human translation is the one that serves as a reference translation in the previously-mentioned tables.

4.1. DeepL's translation performance

In the present paragraph, DeepL's translation performance will be analyzed.

Table 1 Error analysis of DeepL's translation of specialized texts

Original	Automatic translation	Human translation	Syntax	Grammar	Lexis	Article	Acronym	Terminology	Culture-specific	Theme-rheme	Omission	Untranslated	Consistency	Orthography	Transliteration	Format
(ГИБП)	(GIBP)	/					X									
остро встает вопрос	solleva la questione	sorge con urgenza la questione	X								X					
о дальнейшей терапевтической тактике	di ulteriori tattiche terapeutiche	di un'ulteriore tattica terapeutica		X												
обзоре	revisione	rassegna			X											
с развитием острого респираторного дистресс-синдрома, обусловленного	con lo sviluppo della sindrome da distress respiratorio acuto causato	con lo sviluppo della sindrome da distress respiratorio acuto causata	X													
Рассматриваются влияние ГИБП на патогенез COVID-19 и их роль в лечении тяжелых форм COVID-19.	L'influenza dell'HIBP sulla patogenesi del COVID-19 e il loro ruolo nel trattamento delle forme gravi di COVID-19 sono considerati.	Vengono considerati l'effetto dei preparati biologici geneticamente modificati sulla patogenesi del COVID 19 e il loro ruolo nel trattamento delle forme gravi di COVID 19.								X						
ГИБП	HIBP	preparati biologici geneticamente modificati					X	X	X						X	
обзоре	revisione	rassegna			X										X	
консенсусов	consensuali	vertici			X											
ГИБП	HSI	della terapia dei preparati biologici geneticamente modificati					X	X	X							
с оценкой	con valutazione	con una valutazione				X										
Ведение детей	Conduzione di bambini	Gestione dei bambini			X	X										
в условиях пандемии	nelle condizioni di una nuova pandemia	nelle condizioni della pandemia	X			X										

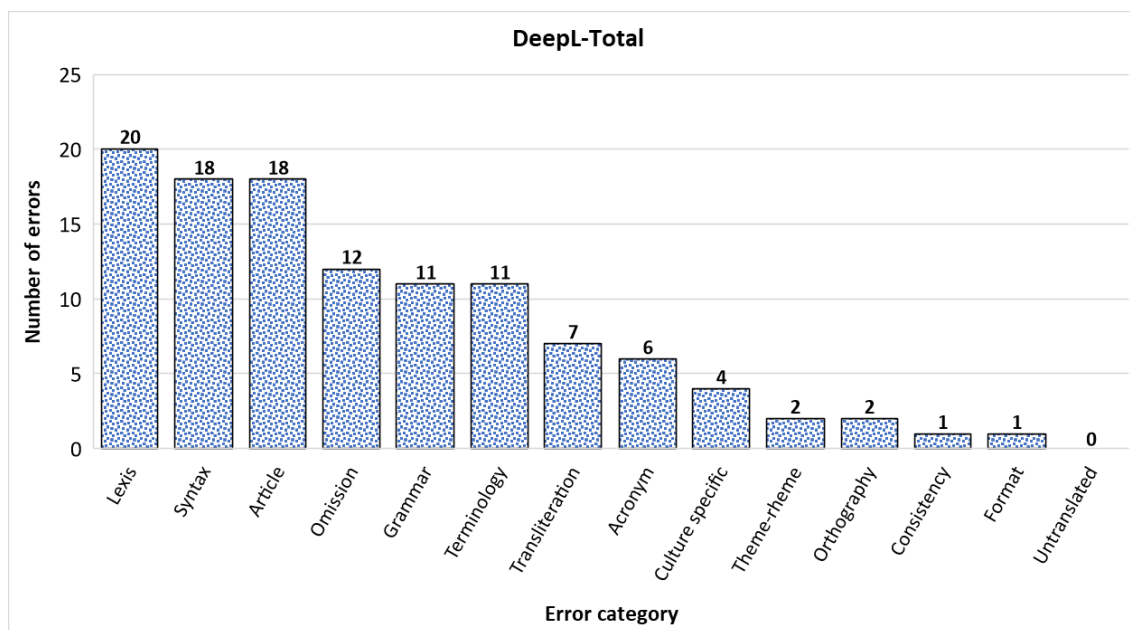
новой коронавирусной инфекции	di infezione da coronavirus	della nuova infezione da coronavirus				X											
Особенности клинических проявлений и лечения заболевания	Caratteristiche delle manifestazioni cliniche e trattamento della malattia	Caratteristiche delle manifestazioni cliniche e del trattamento della malattia	X														
При разработке	Nello sviluppo	Nell'elaborazione			X												
доказательной базе	alla base di prova	all'evidenza scientifica			X												
а также вопросам эффективности и безопасности	nonché all'efficacia e alla sicurezza	nonché alle questioni dell'efficacia e della sicurezza									X						
авторы освещают вопросы профилактики	gli autori evidenziano la prevenzione	gli autori evidenziano le questioni della prevenzione									X						
Тактика	La tattica	Il piano			X												
в зависимости от возраста и степени тяжести течения болезни	a seconda dell'età e del grado di gravità della malattia	in base all'età e al grado di gravità del decorso della malattia									X						
терапия рассмотрена с позиции этиологической, патогенетической и симптоматической направленности.	la terapia viene considerata a partire dall'orientamento eziologico, patogenetico e sintomatico.	la terapia viene considerata a partire dal punto di vista eziologico, patogenetico e sintomatico.			X												
Коронавирус SARS- Cov 2	SARS-Cov 2 coronavirus	Coronavirus SARS- Cov 2						X									
сложности патогенеза	patogenesi complessa	complessità della patogenesi	X														
коронавируса SARS-Cov-2	coronavirus della SARS-Cov-2	coronavirus SARS- Cov-2	X					X									
для укрощения	per domare	per fenare			X												
создания вакцин	sviluppo del vaccino	sviluppo dei vaccini		X													
обеспечивающих их эффективность и безвредность.	per garantire l'efficacia e l'innocuità.	che garantiscono la loro efficacia e innocuità.	X								X						
показать полезность применения концепции	dimostrare l'utilità del concetto	di dimostrare l'utilità dell'applicazione del concetto									X						
концепции пептидного континуума родства белков	concetto di continuum della proteina peptidica	concetto di molteplicità della relazioni delle strutture proteiche						X	X								
(ПКРБ)	(PPCR)	/				X										X	

патогенеза Covid-19	patogenesi di Covid-19	patogenesi del Covid-19				X											
поиска вакцин	trovare vaccini	ricercare vaccini			X												
вакцин против Covid-19	contro Covid-19	contro il Covid-19				X											
компьютерным анализом	mediante analisi computerizzate	attraverso un'analisi computerizzata		X													
иммуноэпитопного	immunoepitopo	immunoepitopica						X									
S, M и N белков SARS-Cov-2	delle proteine S, M e N della SARS-Cov-2	delle proteine S, M e N del SARS-Cov-2		X													
пептидного родства	correlazione peptidica	parentela peptidica						X									
с белками человека и других вирусов	con quelle dei virus umani e di altri virus	con quelle umani e di altri virus	X		X												
доступные в Интернете базы данных	le banche dati disponibili su Internet	banche dati disponibili su Internet				X											
S-белку свойственно пептидное (иммуноэпитопное) родство со многими белками человека, локализующимися на поверхности клеток или циркулирующими в крови, и вирусов.	La parentela peptidica (immunoepitopo) della S-proteina con molte proteine umane, localizzate sulla superficie cellulare o in circolazione nel sangue, e i virus è caratteristica.	La proteina S è caratterizzata da una relazione peptidica (immunoepitopico) con molte proteine dell'essere umano, localizzate sulla superficie delle cellule o circolanti nel sangue, e dei virus	X					X									
Образование антител к SARS-Cov-2	La formazione di anticorpi contro la SARS-Cov-2	La formazione di anticorpi contro il SARS-Cov-2		X													
течение Covid-19.	il decorso di Covid-19.	il decorso del Covid-19				X											
в вакцине против Covid-19	nel vaccino Covid-19	nel vaccino contro il Covid-19	X			X											
Концепция пептидного континуума родства белков (ПКРБ)	Il concetto di continuità della proteina peptidica (PKRB)	Il concetto di molteplicità della relazioni delle strutture proteiche	X				X	X				X					X
вакцины против Covid-19	i vaccini Covid-19	i vaccini contro il Covid-19	X			X											
применением	utilizzo	applicazione			X												
По-видимому	/	Probabilmente										X					
будут	sembrano essere	saranno		X	X												
коронавирусные вспышки и пандемии	Le epidemie di coronavirus e le pandemie	le epidemie e le pandemie di coronavirus	X														

Table 2 Error analysis of DeepL's translation of popular-science texts

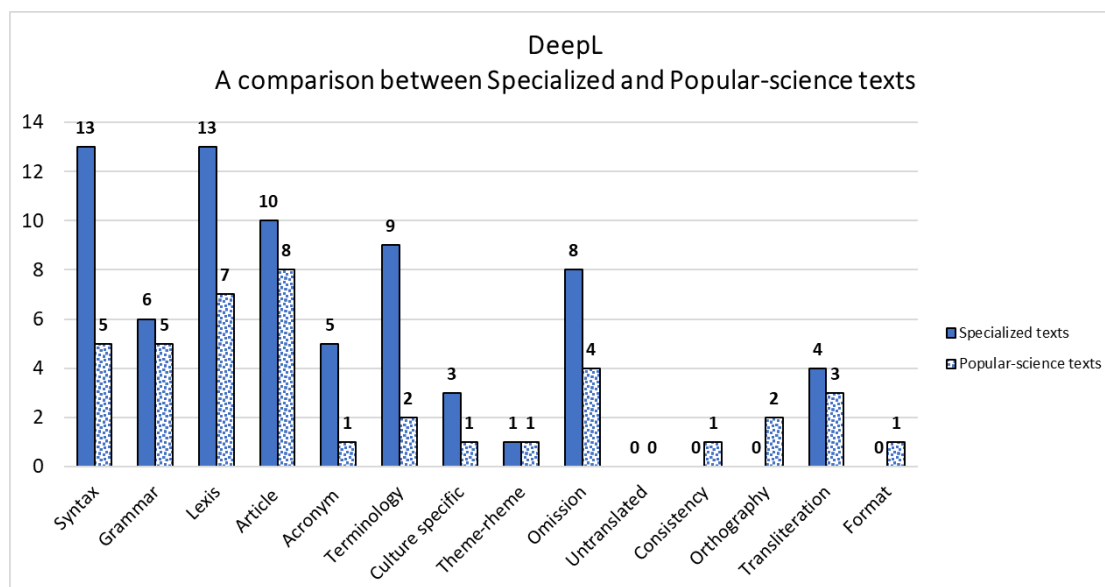
Original	Automatic translation	Human translation	Syntax	Grammar	Lexis	Article	Acronym	Terminology	Culture-specific	Theme-rheme	Omission	Untranslated	Consistency	Orthography	Transliteration	Format
Коронавирус	Il Coronavirus	Il coronavirus												X		
коротко	brevemente	In breve											X			
считают, что переболели COVID 19	credono di essere state infettate da COVID 19	pensano di aver avuto il COVID 19			X	X										
когда действительно в России началась эпидемия коронавируса и откуда его к нам завезли?	quando l'epidemia di coronavirus è iniziata davvero in Russia e da dove è stata portata a noi?	quando l'epidemia di coronavirus è iniziata esattamente in Russia e da dove il virus ci è stato portato?	X	X												
Ответ дала биоинформатика.	La bioinformatica ha dato la risposta.	La risposta è stata data dalla bioinformatica								X						
пациентов из 25 регионов России.	di pazienti in 25 regioni della Russia.	da pazienti provenienti da 25 regioni della Russia.	X													
эволюционные деревья	alberi evolutivi	alberi filogenetici						X								
Оказалось, что SARS CoV-2	Si è scoperto che la SARS CoV-2	È risultato che il SARS CoV-2		X												
первый случай его внутрироссийской передачи	il primo caso di trasmissione intra-russo	il primo caso della sua trasmissione intra-russa		X							X					
Исследовательская группа из Высшей школы экономики	Un team di ricerca della Scuola Superiore di Economia	Il gruppo di ricerca dalla Scuola Superiore di Economia	X			X										
и Сколтеха	e Skoltech	e di Skoltech	X													
совместно со специалистами	insieme a specialisti	in collaborazione con gli specialisti				X										
ИППИ им. А.А. Харкевича РАН	dell'A.A. Kharkevich IPPI RAS	dell'Istituto delle questioni di trasmissione dell'informazione A.A. Harkevič Rossiskja Akademija Nauk					X		X						X	
установили	ha scoperto	hanno stabilito		X	X											
проникал	è entrato	è penetrato			X											

что коронавирус SARS-CoV 2 независимо проникал на территорию России не менее 67 раз,	che il coronavirus SARS-CoV 2 è entrato in Russia almeno 67 volte	che il coronavirus SARS-CoV 2 è penetrato nel territorio russo almeno 67 volte indipendenti l'una dall'altra									X					
на территорию России	in Russia	nel territorio russo									X					
Почему мы не знаем истинных масштабов COVID-19	perché non conosciamo la vera portata di COVID-19	Perché non conosciamo la vera portata del COVID-19				X								X		
соразмерным уровнем экономического развития	con un livello di sviluppo economico commisurato	con un livello di sviluppo economico comparabile				X										
отчетность по CFR	il reporting CFR	il rapporto CFR				X										
CFR резко прибавил	il QCR è aumentato notevolmente	il CFR è aumentato notevolmente														X
там просто начали учитывать	hanno semplicemente iniziato a	allora si è semplicemente iniziato a									X					
до этого собирались только сведения	mentre in precedenza raccoglievano solo dati	in precedenza venivano raccolti solo i dati				X										
из медучреждений	da istituti medici	dagli istituti medici				X										
Несмотря на доступность данных	Nonostante la disponibilità di dati	Nonostante la disponibilità dei dati				X										
«Несмотря на доступность данных, пока не очень понятно, что стоит за ними и насколько они сопоставимы между странами».	"Nonostante la disponibilità di dati, non è ancora molto chiaro cosa ci sia dietro e quanto siano comparabili tra i vari Paesi.	"Nonostante la disponibilità dei dati, non è ancora molto chiaro cosa ci sia dietro e quanto siano comparabili tra i vari paesi".														X
Взгляд демографа	Una prospettiva demografica	L'opinione di un demografo	X			X										
внебольничной пневмонии	polmonite fuori dall'ospedale	polmonite acquisita in comunità								X						
демографы Сергей Тимонин и Анатолий Вишневецкий	i demografi Sergei Timonin e Anatoly Vishnevsky	i demografi Sergej Timonin e Anatolij Višnevskij													X	
основанному на данных	sulla base di dati	sulla base dei dati													X	
официальных национальных источников	fonti ufficiali nazionali	fonti nazionali ufficiali				X										
уровень летальности	tasso di letalità	tasso di mortalità				X										



Graph 1 DeepL's translation performance

With a total of 113 errors, DeepL displays an undoubtedly non-uniform error distribution. Indeed, starting from the left side of Graph 1, three error categories, namely Lexis, Syntax, and Article Usage clearly stand out for their high number of translation errors, followed by Omission, Grammar, Terminology, Transliteration, and Acronym. On the contrary, while Culture-Specific References, Theme-Rheme Pattern, Orthography, Consistency, and Format seem to display a minor, even negligible amount of translation errors, no translation errors are marked as belonging to Untranslated Elements category. When considering the language pair at stake, consisting of two highly different natural languages, both in terms of syntactical structure and article usage, it does not come as a surprise that the majority of the errors committed by DeepL fall in these two error categories. This may rather mean, generally speaking, that the translation tool under examination still encounters considerable difficulties in properly rendering two of the most challenging aspects of Russian-Italian translation. As for lexis, we will see later on in the present paragraph, when considering DeepL's translation performances with regard to specialized texts and popular-science ones separately, that the highly-specialized terms displayed in the specialized articles are proven to make a major contribution to these results.



Graph 2 A comparison between DeepL's translations of specialized and popular-science texts

When comparing DeepL's translation performances with regard to specialized and popular-science texts, with a total of 72 and 41 errors, respectively, we can easily notice that it performs significantly better in translating popular-science-texts. It indeed commits a higher number of translation errors with regard to specialized texts in nine out of fourteen selected error categories, except for Theme-Rheme Pattern, Untranslated Elements, Consistency, Orthography, and Format. However, those last categories display a small overall number of translation errors. Moreover, other error categories, namely Grammar, Article Use, Culture-Specific References, and Transliteration feature a narrow gap between the two text types under examination. On the contrary, Syntax, Lexis, Acronym, Terminology, Omission visibly stand out as a significant disparity can be observed in the renderings of specialized and popular-science texts. When analyzing Acronym and Terminology categories, we need to consider that, since the total number of occurrences of these elements throughout specialized and popular-science texts may significantly differ, the results are likely to be biased accordingly. As for Syntax, Lexis, and Omission, we may assume that the translation tool encounters considerable difficulties in rendering the significantly more elaborate syntactical structure and the specific lexis and terminology characterizing Russian specialized medical articles when compared to popular-science ones.

4.2. Yandex's translation performance

In the present paragraph, Yandex's translation performance will be analyzed.

Table 3 Error analysis of Yandex's translation of specialized texts

Original	Automatic translation	Human translation	Syntax	Grammar	Lexis	Article	Acronym	Terminology	Culture-specific	Theme-rheme	Omission	Untranslated	Consistency	Orthography	Transliteration	Format
Биологическая терапия в эру COVID-19	Terapia biologica NELL'era COVID-19	La terapia biologica nell'era COVID-19				X										X
В связи	In connessione	In relazione			X											
широким применением	con l'ampia applicazione	con l'ampio uso			X											
биологических препаратов	di farmaci biologici	di preparati biologici			X											
(ГИБП)	(GIBP)	/					X									
остро встает вопрос	sorge acutamente la questione	sorge con urgenza la questione			X											
о дальнейшей терапевтической тактике	di ulteriori tattiche terapeutiche	di un'ulteriore tattica terapeutica		X												
ведения таких пациентов	di riferimento tali pazienti	di gestione di tali pazienti	X		X											
обзоре	revisione	rassegna			X											
актуальные данные	dati attuali	dati aggiornati			X											
патогенезе COVID-19	sulla patogenesi di COVID-19	sulla patogenesi del COVID-19				X										
(«цитокиновый шторм»)	(»tempesta di citochine»)	(»tempesta di citochine»)														X
влияние	gli effetti	l'effetto		X												
ГИБП	GIBP	preparati biologici geneticamente modificati					X	X	X							
обзоре	revisione	rassegna			X											
консенсусов	di consenso	vertici	X		X											
ГИБП	GIBP	della terapia dei preparati biologici geneticamente modificati					X	X	X							

новой коронавирусной инфекции	da una nuova infezione da coronavirus	della nuova infezione da coronavirus				X											
медицинской помощью	aiuto medico	assistenza medica			X												
в условиях	in condizioni	nelle condizioni				X											
пандемии	di pandemia	della pandemia				X											
новой коронавирусной инфекции	di nuova coronavirus infectione	della nuova infezione da coronavirus				X						X					
Минздравом России	ministero della salute Russia	Ministero della Salute della Russia	X													X	
разработаны	sviluppato	ha sviluppato	X	X													
совместно с профессиональными ассоциациями и экспертами	in collaborazione con le associazioni professionali e di esperti	in collaborazione con associazioni professionali ed esperti	X			X											
в области педиатрии, инфекционных болезней и реанимации	nel campo della pediatria, malattie infettive e rianimazione	nel campo della pediatria, delle malattie infettive e della rianimazione	X										X				
Особенности клинических проявлений и лечения заболевания	Caratteristiche delle manifestazioni cliniche e trattamento della malattia	Caratteristiche delle manifestazioni cliniche e del trattamento della malattia	X														
вызванного новой коронавирусной инфекцией	causata da un nuovo coronavirus infectione	causata dalla nuova infezione da coronavirus				X						X					
При разработке	Nello sviluppo	Nell'elaborazione			X												
опыт специалистов не только нашей страны, но и зарубежных коллег	l'esperienza pratica di specialisti non solo del nostro paese, ma anche colleghi stranieri	l'esperienza pratica di specialisti non solo del nostro paese, ma anche di colleghi stranieri	X														
доказательной базе	alla base di prove	all'evidenza scientifica			X												
а также вопросам эффективности	nonché all'efficacia e alla sicurezza	nonché alle questioni dell'efficacia e della sicurezza									X						
на основании указанных методических рекомендации	sulla base di queste raccomandazioni metodologiche	sulla base delle linee guida citate			X												
авторы освещают вопросы	gli autori coprono le questioni	gli autori evidenziano le questioni			X												
вопросы профилактики, диагностики, лечения	le questioni di prevenzione, diagnosi, trattamento di condizioni patologiche	le questioni della prevenzione, della diagnosi, del trattamento delle				X							X				

патологических состояний		condizioni patologiche																
Тактика	Le tattiche	Il piano		X	X													
в зависимости от возраста и степени тяжести течения болезни	in base all'età e alla gravità del decorso della malattia	in base all'età e al grado di gravità del decorso della malattia										X						
терапия рассмотрена с позиции этиологической, патогенетической и симптоматической направленности.	la terapia è considerata da una posizione di orientamento eziologico, patogenetico e sintomatico.	la terapia viene considerata a partire dal punto di vista eziologico, patogenetico e sintomatico.			X	X												
сложности патогенеза	difficoltà di patogenesi	complessità della patogenesi			X	X												
поиски вакцин	ricerca di vaccini	ricerche di vaccini		X														
Актуальность	Attualità	Pertinenza			X													
Вакцина против коронавируса	Il vaccino contro il coronavirus	Un vaccino contro il coronavirus				X												
для укрощения	per domare	per fenare			X													
создания вакцин	creazione di vaccini	sviluppo dei vaccini			X	X												
выбор	scelta	selezione			X													
Цель исследования - показать	Lo scopo della ricerca-mostrare	Lo scopo della ricerca è quello di dimostrare			X							X	X					
концепции пептидного континуума родства белков	concetto di peptidico continuum di parentela proteine	concetto di molteplicità della relazioni delle strutture proteiche	X				X	X					X					
(ПКРБ)	(ПКРБ)	/					X						X					
патогенеза Covid-19	patogenesi Covid-19	patogenesi del Covid-19	X															
поиска вакцин	ricerca di vaccini	ricercare vaccini												X				
вакцин против Covid-19	contro Covid-19	contro il Covid-19				X												
возможную природу будущих пандемий	la possibile natura delle pandemie future	la possibile natura di future pandemie				X												
методы	tecniche	metodi			X													
Для выявления компьютерным анализом	Per identificare l'analisi computerizzata	attraverso un'analisi computerizzata	X															
S, M и N белков SARS-Cov-2	delle proteine S, M E N SARS-Cov-2	delle proteine S, M e N del SARS-Cov-2	X															X
с белками человека и других вирусов	con proteine umane e altri virus	con quelle umani e di altri virus	X															

Источниками первичных последовательностей белков	Le fonti di sequenze primarie di proteine	Le fonti delle sequenze proteiche primarie				X										
со многими белками человека, локализующимися на поверхности клеток или циркулирующими в крови, и вирусом.	con molte proteine umane localizzate sulla superficie delle cellule o circolanti nel sangue e virus	con molte proteine dell'essere umano, localizzate sulla superficie delle cellule o circolanti nel sangue, e dei virus	X										X			
Образование антител к SARS-Cov-2	La formazione di anticorpi contro SARS-Cov-2	La formazione di anticorpi contro il SARS-Cov-2				X										
перекрестно	in modo incrociato	trasversalmente			X											
может отягощать	può appesantire	può aggravare			X											
течение Covid-19.	il corso di Covid-19.	il decorso del Covid-19			X	X										
в вакцине против Covid-19	nel vaccino Covid-19	nel vaccino contro il Covid-19	X			X										
Вывод	Output	Conclusione			X								X			
Il concetto di continuo epitopico della parentela proteica (ПКРБ)	Il concetto di peptide continuum of protein Parenthood (PCRB)	Il concetto di molteplicità delle relazioni delle strutture proteiche					X	X							X	
вакцин против Covid-19	i vaccini Covid-19	i vaccini contro il Covid-19	X			X										
По-видимому	Apparentemente	Probabilmente			X											
коронавирусные вспышки и пандемии	Le epidemie di coronavirus e le pandemie	le epidemie e le pandemie di coronavirus	X													
чаще	più comuni	più frequenti			X											

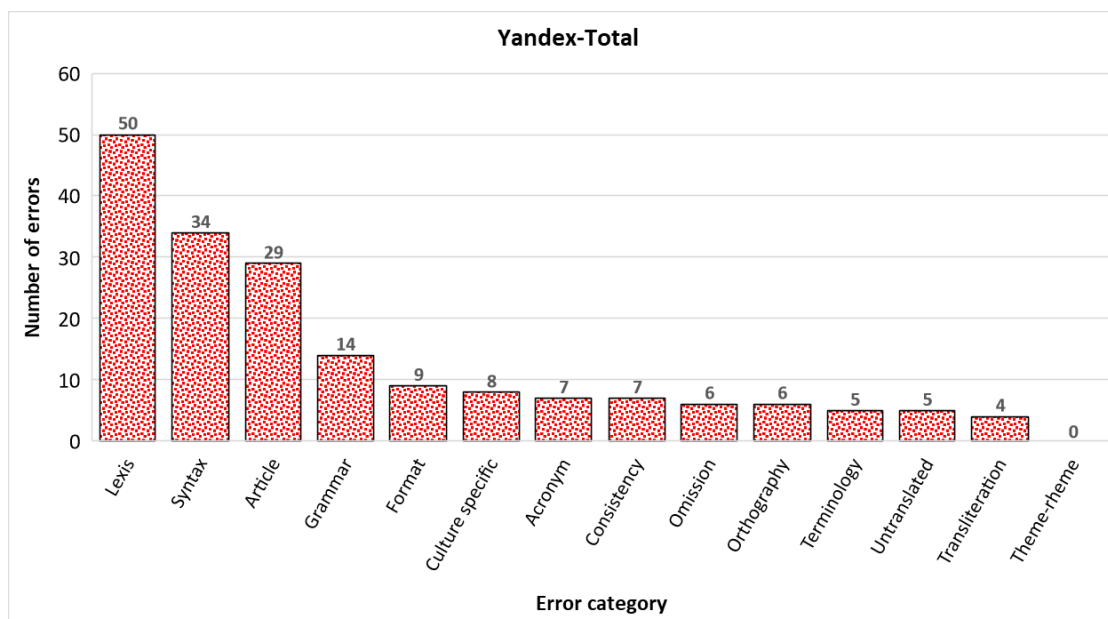
Table 4 Error analysis of Yandex's translation of popular-science texts

Original	Automatic translation	Human translation	Syntax	Grammar	Lexis	Article	Acronym	Terminology	Culture-specific	Theme-rheme	Omission	Untranslated	Consistency	Orthography	Transliteration	Format
коротко	breve	in breve		X												
считают, что переболели COVID-19	che COVID-19 sia stato malato	pensano di aver avuto il COVID 19	X			X										
когда действительно в России началась эпидемия коронавируса и откуда его к нам завезли?	quando l'epidemia di coronavirus è iniziata in Russia e da dove è stata portata a noi?	quando l'epidemia di coronavirus è iniziata esattamente in Russia e da dove il virus ci è stato portato?	X	X							X					
эволюционные деревья	alberi evolutivi	alberi filogenetici						X								
Оказалось, что SARS CoV-2	Si è scoperto che SARS CoV-2	È risultato che il SARS CoV-2				X										
из Европы	Dall'Europa	dall'Europa												X		
первый случай его внутрироссийской передачи	il primo caso del suo trasferimento intra-russo	il primo caso della sua trasmissione intra-russa			X											
Теперь подробнее	Ora leggi di più	Ora, più in dettaglio			X								X			
Исследовательская группа из Высшей школы экономики	Un gruppo di ricerca della Graduate School of Economics	Un gruppo di ricerca dalla Scuola Superiore di Economia							X							
и Сколтеха	e Skoltha	e di Skoltech	X						X							
НИИ гриппа им. А.А. Смородинцева	dell'Istituto di ricerca per l'influenza. A. A. Riborodintseva	dell'Istituto di ricerca per l'influenza A. A. Smorodintsev					X		X						X	
ИППИ им. А.А. Харкевича РАН	Ippi loro. A. A. HARKEVICH Ran	dell'Istituto delle questioni di trasmissione dell'informazione A.A. Harkevic Rossiskja Akademija Nauk					X		X						X	
установили	ha stabilito	hanno stabilito		X												
что коронавирус SARS-CoV 2 независимо проник на территорию России не менее 67 раз	che il coronavirus SARS-CoV 2 è penetrato indipendentemente nel territorio della Russia almeno 67 volte	che il coronavirus SARS-CoV 2 è penetrato nel territorio russo almeno 67 volte indipendenti l'una dall'altra			X											
Почему мы не знаем истинных масштабов COVID-19	perché non conosciamo la vera scala di COVID-19	Perché non conosciamo la vera portata del COVID-19			X	X								X		
Пандемия коронавируса	La pandemia del coronavirus	La pandemia di coronavirus				X										

достоверно оценить ситуацию они не могут	valutare in modo affidabile la situazione, non possono	non possono valutare la situazione in modo affidabile	X															
О парадоксах количественных подходов к COVID-19 говорили на вебинаре	I paradossi degli approcci quantitativi a COVID-19 hanno parlato al webinar	I paradossi degli approcci quantitativi al COVID-19 sono stati discussi al webinar	X			X												
на вебинаре Международной лаборатории исследований населения и здоровья ВШЭ	al webinar del Laboratorio Internazionale di ricerca sulla popolazione e sulla salute HSE	al webinar del Laboratorio Internazionale per la Ricerca sulla Popolazione e la Salute della Scuola Superiore di Economia							X					X				
Доступно	Disponibile	Accessibile				X												
несопоставимо	disparabile	non comparabile				X												
соразмерным уровнем экономического развития	con un livello proporzionato di sviluppo economico	con un livello di sviluppo economico comparabile				X												
Например, в Германии CFR около 2%, во Франции — более 10%.	Ad esempio, in Germania CFR circa il 2%, in Francia — più del 10%.	Ad esempio, in Germania il CFR è di circa il 2%, in Francia è superiore al 10%	X			X							X					
отчетность по CFR	la segnalazione di CFR	il rapporto CFR				X												
и не факт, что причина динамики в росте смертности	e non è il fatto che la causa della dinamica è l'aumento della mortalità	e non è certo che la causa della dinamica sia un aumento della mortalità			X	X												
Так	Così	Ad esempio				X												
в начале апреля	ALL'inizio di aprile	all'inizio di aprile																X
CFR резко прибавил	il CFR ha bruscamente aggiunto	il CFR è aumentato notevolmente				X												
там просто начали учитывать	c'era appena iniziato a	allora si è semplicemente iniziato a			X	X							X					
смертность	decessi	mortalità				X												
до этого собирались только сведения	prima che solo le informazioni raccolte	in precedenza venivano raccolti solo i dati	X	X	X													
из медучреждений	dalle istituzioni mediche	dagli istituti medici				X												
Несмотря на доступность данных	nonostante la disponibilità dei dati	Nonostante la disponibilità dei dati														X		
что стоит за ними	cosa c'è dietro di loro	cosa ci sia dietro			X													

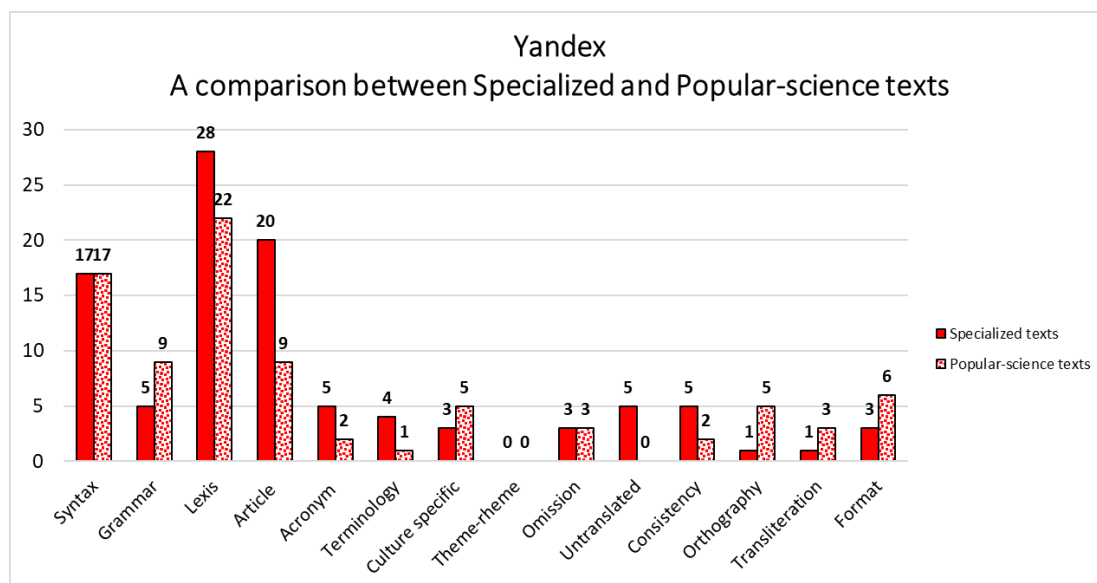
«Несмотря на доступность данных, пока не очень понятно, что стоит за ними и насколько они сопоставимы между странами».	"nonostante la disponibilità dei dati, non è ancora molto chiaro cosa c'è dietro di loro e quanto siano comparabili tra i paesi".	"Nonostante la disponibilità dei dati, non è ancora molto chiaro cosa ci sia dietro e quanto siano comparabili tra i vari paesi".															X
Взгляд демографа	Vista demografica	L'opinione di un demografo	X		X												
в зарубежных и русскоязычных СМИ	stranieri e russi dei MEDIA	sui media stranieri e in lingua russa	X		X												X
о «русском чуде»	«russo miracolo» e	sul "miracolo russo"	X	X													X
и «загадке поведения коронавируса» в России	e «l'enigma di comportamento coronavirus» in Russia	e sul "mistero del comportamento del coronavirus" in Russia	X		X	X											X
признания	il riconoscimento	le dichiarazioni			X												
министра здравоохранения	del ministro della sanità	del Ministro della Salute														X	
республики Дагестан	della repubblica del Daghestan	della Repubblica del Daghestan														X	
соотношении смертей от коронавируса	rapporto di morti da coronavirus	rapporto tra i decessi da coronavirus	X														
породили большую общественную дискуссию	generato un grande dibattito pubblico	hanno generato un importante dibattito pubblico		X													
о методах учёта	sui metodi di contabilità	sui metodi di contabilizzazione			X												
статистики	le statistiche	delle statistiche	X														
на эти и многие другие вопросы	queste e molte altre domande	a queste e a molte altre domande	X														
демографы Сергей Тимонин и Анатолий Вишневецкий.	demografi Sergey Tymonin e Anatoly Vishnevsky	i demografi Sergej Timonin e Anatolij Višnevskij				X											X
Согласно отчёту ВОЗ	Secondo il rapporto DELL'OMS	Secondo il rapporto dell'OMS															X
на 28 мая	il 28 maggio	al 28 maggio	X														
379 тысяч подтвержденных случаев COVID-19	379 migliaia di casi confermati COVID-19	379.000 casi confermati di COVID-19	X														
4,1 тысяча	4,1 mille	4,1 migliaia			X												

<p>обращает на себя внимание и вызывает неизбежные вопросы исключительно низкий уровень летальности от коронавируса в России</p>	<p>richiama l'attenzione e solleva questioni inevitabili eccezionalmente basso tasso di mortalità da coronavirus in Russia.</p>	<p>attira l'attenzione e solleva inevitabili domande il tasso di mortalità estremamente basso da coronavirus in Russia.</p>	<p>X</p>		<p>X</p>	<p>X</p>										
----------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------	----------	--	----------	----------	--	--	--	--	--	--	--	--	--	--



Graph 3 Yandex's translation performance

Yandex seems to distribute its total 184 errors quite uniformly, except for three error categories, namely Lexis, Syntax, and Article Usage, which display a significantly higher number of translation errors when compared to the other error categories. As above mentioned with regard to DeepL's translation performance, Syntax's and Articles' great amount of translation errors may be explained by the profound difference in syntactical structure and article usage between Russian and Italian. On the contrary, Lexis constitutes the error category displaying the greatest amount of translation errors and further analysis is needed, in this case, to determine whether this is due to a particular type of text or to the fact that the translation tool under examination encounters serious difficulties in properly rendering Italian lexis, independently of the text type. Apart from that, the other error categories do not seem to feature a great number of translation errors. Moving left to right across Graph 3 we can indeed easily notice that Grammar, Format, Culture-Specific References, Acronym, and Consistency, although collocated immediately after the Article Usage category, are divided from it by a huge gap, whereas they are clearly closer, by the number of translation errors, to the error categories occupying the right side of the graph, namely Omission, Orthography, Terminology, Untranslated Elements, and Transliteration, which display a small number of translation errors. Finally, no errors are marked as belonging to the Theme-Rheme Pattern category.

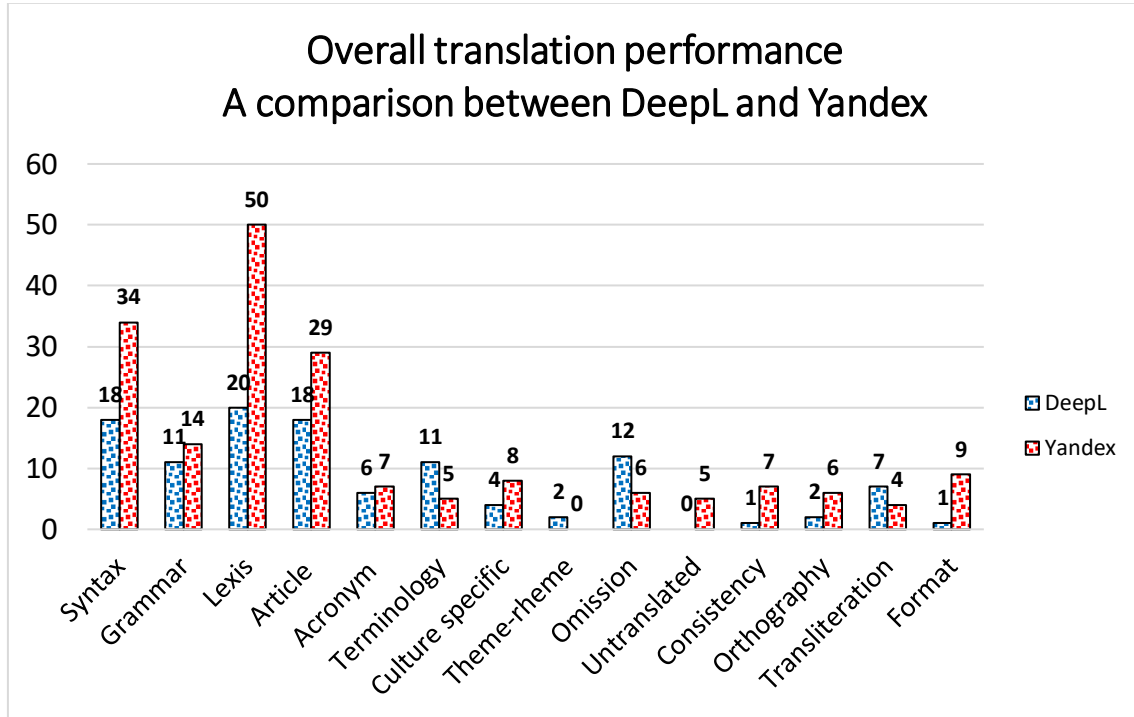


Graph 4 A comparison between Yandex's translations of specialized and popular-science texts

With a total of 100 errors in specialized texts' translations and 84 in popular science's ones, Yandex seems to provide a better translation performance with regard to popular-science texts. Nonetheless, it follows a significantly different pattern when compared to DeepL's translation performances. A small gap is indeed observed in each error category, except for Article Usage, which displays 20 errors in specialized texts and just 9 in popular-science ones. As for the other error categories, a better translation performance with regard to popular-science texts can be observed in Lexis, Acronym, Terminology, Untranslated Elements, and Consistency. On the other hand, Syntax, Omission, and Theme-Rheme display the same amount of translation errors, whereas Yandex seems to better render specialized texts' Grammar, Culture Specific References, Ortography, Transliteration, and Format. Starting from saying that the results obtained in Acronym, Terminology, Culture-Specific References, and Transliteration may be biased by the actual number of times that these elements occur throughout the texts under examination, the ones related to Grammar, Article Usage, Untranslated Elements, Consistency, Ortography, and Format undoubtedly constitute a good basis for a more detailed analysis of Yandex's behaviors and patterns in translating Russian-Italian specialized and popular-science medical texts.

4.3. A comparison between DeepL’s and Yandex’s translation performances

In the present paragraph, a comparison between DeepL’s and Yandex’s translation performances will be made.



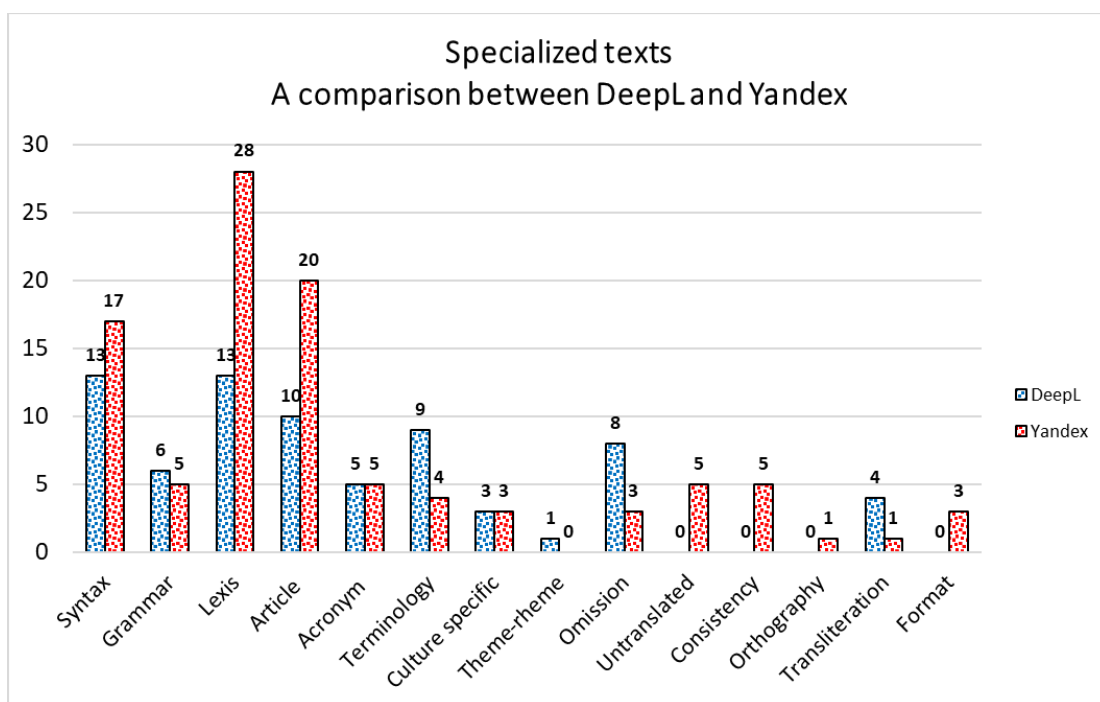
Graph 5 A comparison between DeepL and Yandex regarding the overall translation performance

When comparing the two translation tools under examination on the basis of their overall translation performances, it can be observed that they both collect a significant amount of translation errors in three error categories, namely Lexis, Syntax, and Article Usage, in order of the number of translation errors. However, these categories, first and foremost Lexis, display a wide gap between DeepL’s and Yandex’s translation performances. We can easily say that, although Syntax and Article Usage clearly constitute two of the major weakness of both translation tools, in all probabilities because of the profound difference between Russian and Italian syntactical structure and article usage, DeepL seems to be able to better render into Italian the Russian syntactical

structure and is proven to remain more compliant with Italian grammatical rules when it comes to properly insert the Italian articles, non-existent in Russian grammar. As for Lexis, DeepL, once again, provides a better translation performance, however, further analysis is needed to detect whether Yandex’s high amount of translation errors are related to a specific text type.

Moreover, Yandex seems to perform better with regard to Terminology, Omission, and Transliteration, whereas DeepL is observed to commit a smaller number of translation errors in Grammar, Acronym, Culture-Specific References, Untranslated Elements, Consistency, Ortography, and Format, even though the gap in these categories is not comparable to the one observed in Lexis, Syntax, and Article Usage.

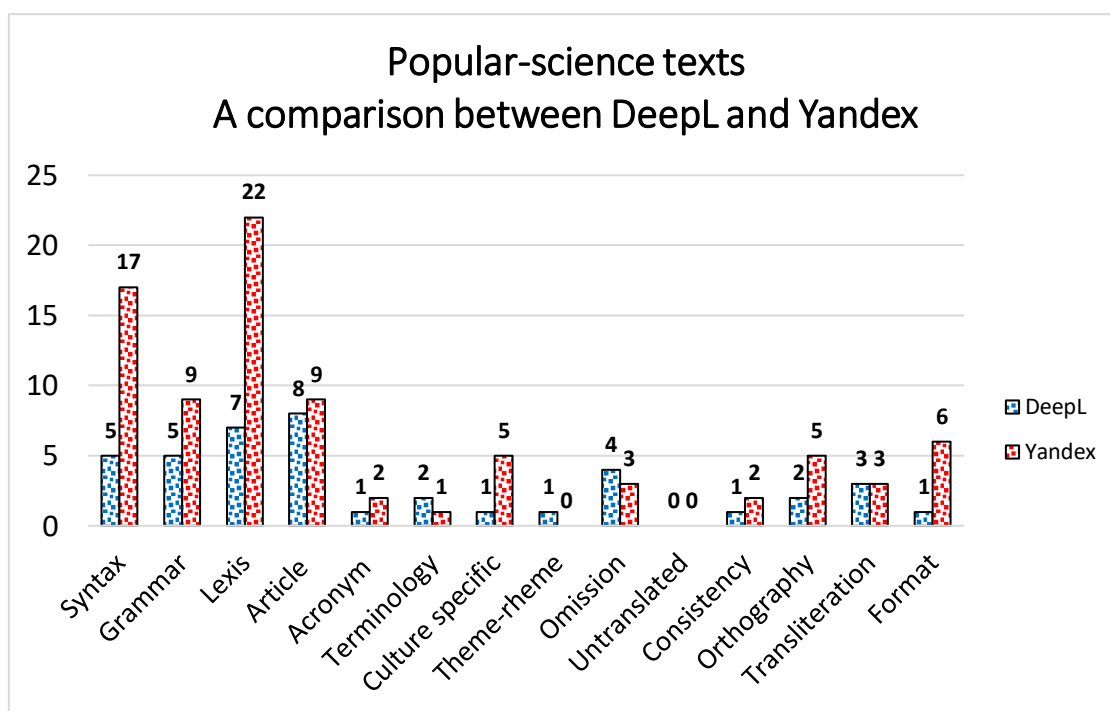
Generally speaking, a more comprehensive evaluation of DeepL’s and Yandex’s translation performances can be made by comparing the two translation tools’ behavior concerning specialized and popular-science texts separately.



Graph 6 A comparison between DeepL and Yandex regarding specialized texts' translation

When comparing DeepL’s and Yandex’s translation performances with regard to specialized texts, with a total of 72 and 100 translation errors, respectively, DeepL seems to perform better when compared to Yandex. Indeed, the latter collects a higher number

of translation errors in the majority of error categories, except for Grammar, Terminology, Theme-Rheme Pattern, Omission, and Transliteration. On the one hand, Lexis and Article Usage error categories undoubtedly stand out, as not only they do display a significantly higher amount of translation errors with regard to both translation tools, when compared to the other error categories but also a great gap can be observed between DeepL's and Yandex's translation performances. On the other hand, Yandex is proven to give a better translation performance when compared to DeepL in Terminology, Omission, and Transliteration, even though the gap is not comparable to the one observed in Lexis and Article Usage.



Graph 7 A comparison between DeepL and Yandex regarding popular-science texts' translation

With a total of 41 translation errors committed by DeepL and 84 by Yandex, both translation tools are proven to perform better in translating popular-science texts than specialized texts. Moreover, the gap between DeepL's and Yandex's translation performances is observed to increase from specialized texts to popular-science ones. On the one hand, DeepL seems to perform better in all error categories, except for Terminology, Theme-Rheme Pattern, and Omission, where Yandex collects a smaller

amount of translation errors, even though the gap between the translation performances under examination is quite narrow. On the other hand, Transliteration and Untranslated Elements are characterized by equal translation performances. Worth noting undoubtedly are Lexis and Syntax, which display a considerable gap between Yandex's and DeepL's translation performances. The two error categories, which constitute the major weaknesses of Yandex performance, do not represent an issue in DeepL's one. Moreover, we can easily notice that the two translation tools' performances do not follow the same pattern as with regard to specialized texts. In fact, while Lexis is characterized by a considerable gap for both text types, the same does not hold for Syntax, which displays similar translation performances in translating specialized texts and significantly different amounts of translation errors with regard to popular-science ones, and Article Usage, which presents a wide gap with regard to specialized texts and a narrow one when popular-science texts are considered.

4.4. BLEU metric's evaluation

Once displayed the comparative error analysis' results, in order to provide a comprehensive evaluation of DeepL's and Yandex's translation performances, we hereby conduct an automatic evaluation of the provided translations using the BLEU metric. As previously said, BLEU, which stands for Bilingual Evaluation Understudy, is a popular and widely used automatic evaluation metric, which assesses machine translation quality on the basis of its textual similarity with a number of human reference translations, also called golden translations. More specifically, it is an easily accessible platform where the users may upload the original document, the translations performed by the two machine translation systems under examination, and the corresponding reference translation, which serves as a benchmark to judge the quality of the machine translations. While the original document is optional, the other three texts are mandatory, and their absence may prevent the evaluation metric from properly completing the evaluation process. Once uploaded all the documents, a percentage, defining the quality of each translation texts is released by the BLEU metric.

Table 5 BLEU metric's evaluation

	<i>DeepL</i>	<i>Yandex</i>
Биологическая терапия в эру COVID-19	50.31	49.60
Ведение детей с заболеванием, вызванным новой	55.12	43.46
Коронавирус SARS-Cov 2: сложности патогенеза, поиски вакцин и будущие пандемии	45.14	51.18
Коронавирус завозили в Россию не менее 67 раз	36.66	52.88
Неизвестная летальность - Почему мы не знаем истинных масштабов COVID-19	71.58	45.34
Смертность от COVID 19 - Взгляд демографа на статистику причин смерти в России и мире	76.73	52.31

As shown in Table 5, the BLEU evaluation metric provides an automatic evaluation of DeepL's and Yandex's translation performances by means of a percentage, representing the textual similarity of each analyzed text with a specifically made human translation, used as reference translation. This undoubtedly constitutes an interesting basis for our discussion. Indeed, on the one hand, having provided, in the previous paragraphs, a manual error analysis, an automatic evaluation process is hereby displayed. On the other hand, assessing the BLEU metric the provided translations' quality at the article level, DeepL's and Yandex's translation performances with regard to each specific text can be observed. Generally speaking, the BLEU evaluation metric's results seem to reflect the ones shown by our comparative error analysis. In fact, not only popular-science articles display a higher percentage of textual similarity when compared to specialized ones but also DeepL's translations are observed to more accurately reflect human reference translations' features. Moreover, while a significantly wide gap can be noticed between the percentages of textual similarity assigned to DeepL's and Yandex's translations of popular-science texts, a minor difference is observed with regard to highly specialized texts. This perfectly applies to the first two highly specialized and the last two popular-science articles. An exception is undoubtedly constituted by the third specialized and the

first popular-science articles. Indeed, by showing a higher percentage of textual similarity with regard to Yandex's translation variants and a significantly wide gap when compared to the other articles, belonging to the same text types, they do not follow the pattern established in our previous analysis. In particular, an interesting case is represented by the first popular-science article, *Коронавирус завозили в Россию не менее 67 раз*, whose translation performed by DeepL is assigned a rather low percentage of textual similarity when compared not only to the one given to Yandex's translation performance concerning the same article but also the ones achieved by DeepL in the translation of the other highly specialized as well as popular-science texts.

Conclusion

The research conducted in the present thesis undoubtedly reflects the extraordinary development that neural machine translation systems have experienced over the last decades, reaching ever-increasingly quality standards with regard to Russian-Italian medical translation. However, the results of the comparative error analysis, as well as the BLEU metric evaluation, clearly shed light on several weaknesses, especially related to the fields of syntax, lexis, and article usage, that still prevent machine translation systems from providing natural sounding and accurate Russian-Italian medical translations. Moreover, by investigating and applying non-automatic and automatic machine translation evaluation methods, the role of the continuous evaluation of machine translation systems' performances has been remarked as crucial to assure their growth and enhancement. Further research in the field undoubtedly is highly needed. Starting from the results obtained in the present research, a possible direction to be taken in the future may regard an in-depth linguistic study of the error categories adopted in these pages as the main criteria to judge machine translation quality, aimed at determining the relevance of each one of them against the background of Russian-Italian medical translation. By doing so, a distinction between major and minor translation errors can be properly made and immediate and effective corrective measures can consequently be proposed in order to enhance machine translation systems performances. A crucial aspect of machine translation development is represented by the role of the highly diverse kinds of professionals that have taken part over the last years and still are at the very center of machine translation research, first and foremost computer scientists and linguists. In particular, in contrast to what is usually believed, linguists and translators will be asked, in the near future, to apply their extensive linguistic knowledge to accomplish a variety of tasks in the field of machine translation research and development. Among others, one field that doubtless requires human intellectual effort is represented by machine translation evaluation systems, be they non-automatic or automatic, with a particular focus on the establishment of the qualitative standards according to which machine translation systems' performances are assessed. In order to properly accomplish the required tasks, human translators will be asked to develop a whole set of new linguistic as well as technical skills. This may lead to a complete redefinition of the professional figures of translators and linguists, which is to be accompanied by continuous and

dedicated training. In fact, the extremely fast development of the research in the field of machine translation is experiencing an ever-increasingly close and successful collaboration between experts in scientific and linguistic disciplines, whose knowledge is equally required to accomplish the demanding tasks of incredibly fast-evolving fields like machine translation and machine learning.

Резюме

Темой этой диссертации является текущее развитие программ машинного перевода в области перевода с русского на итальянский. В частности, некоторые статьи как из специализированных, так и научно-популярных русскоязычных медицинских журналов, переведены с русского на итальянский с помощью двух крайне знаменитых нейронных систем машинного перевода, а именно DeepL и Яндекса. Ниже представляется быстрый обзор предмета и методов нашей исследования, а также основных причин, которые нас побудили исследовать эту тему.

Историческое развитие машинного перевода

В течение последних лет машинный перевод характеризуется постоянным и быстрым ростом, обусловленным обширными научными исследованиями и растущим интересом к области информатики. В его истории, начавшейся около девяноста лет назад, принимало участие множество экспертов из различных дисциплин, как научных, так и гуманитарных. В частности, четыре основных протагониста сыграли ключевую роль в зарождении и ранних стадиях развитии машинного перевода. В 1933 г. французский инженер Жорж Артсруони и русский ученый Петр Троянский подали патентные заявки на два механических устройства, называвшихся *Механический мозг* и *Машина для подбора и печатания слов при переводе с одного языка на другой или на несколько других одновременно*, признанных сегодня предшественниками систем машинного перевода. Из-за отсутствия заинтересования и доверия мировым научным сообществом, эти патенты остались нереализованными проектами, но они, безусловно, представили собой первую попытку приблизить научный мир к теме машинного перевода. Несколько лет спустя значительный вклад в разработку первых систем машинного перевода внес математик Уоррен Уивер, который первым выдвинул гипотезу о возможности использования компьютера для перевода с одного естественного языка на другой. Кроме того, благодаря своим исследованиям в области естественных языков, математик обратил внимание на то, что различные естественные языки имеют приблизительно схожую структуру, а также на важность контекста, который необходимо учитывать, чтобы правильно передать смысл исходного текста. Следующий шаг навстречу разработке первых программ

машинного перевода был сделан математиком Клодом Шенноном, который, в своей книге *Mathematical theory of communication* (Математическая теория коммуникации) иллюстрирует процесс коммуникации, впервые вводя концепцию декодирования и кодирования информации, которая лежит в основе компьютерной обработки информации. Кроме того, Шеннон исследует статистический характер языков, закладывая основы обработки естественного языка (Natural Language Processing, NLP). Обработка естественного языка является отраслью компьютерной науки, обучающей компьютеры понимать, обрабатывать и генерировать естественный язык так же, как это делает человек, для выполнения многочисленных человеческих задач. Поскольку естественный язык разрабатывается человеком и используется им в повседневной жизни, он характеризуется большим разнообразием и состоит из различных уровней, которые непрерывно взаимодействуют друг с другом. Следовательно, обработка естественного языка является очень сложной деятельностью, которая требует тщательного и постоянного изучения. Исследователи тренируют системы обработки естественного языка с помощью огромных объемов данных и разных методов, разработанных в течение долгого времени. В настоящее время, приложения для обработки естественного языка, в том числе для автоматического распознавания речи, виртуальные помощники, системы распознавания образов и генерации языков, очень распространены. Однако, первым применением обработки естественного языка безусловно является машинный перевод, родившийся в 1950-х годах благодаря первым исследованиям по обработке естественного языка. Положительный импульс развитию первых систем машинного перевода придал успешный Джорджтаунский эксперимент, проведенный Джорджтаунским университетом совместно с компанией IBM в 1954 г. В ходе этого эксперимента были полностью автоматически переведены шестидесяти предложений с русского на английский язык, с целью привлечения мирового внимания к успехам в области машинного перевода. Однако, публикация в 1966 г. доклада *Language and Machines*, также известного как доклад ALPAC по имени комиссии, которая его написала (Automatic Language Processing Advisory Committee, ALPAC), ослабила оптимизм, обусловленный Джорджтаунским экспериментом. Действительно, в докладе ставилась под сомнение не только

фактическая потребность в переводе в США, но и экономическая отдача от значительных инвестиций в исследования для разработки систем машинного перевода, которые в то время не могли обеспечивать выполнение целый процесс перевода. Поэтому рекомендовалось направить государственные средства на прогресс других дисциплин, в том числе компьютерной лингвистики, компьютерного перевода и ручного перевода. Несмотря на мировую актуальность доклада ALPAC, исследования в области машинного перевода продолжались, хотя более медленными темпами, и с течением времени и развитием компьютерных знаний достигли удивительных результатов.

На протяжении лет принимались разные подходы к разработке систем машинного перевода, чтобы достигнуть все более выгодного соотношения эффективности к производственными затратами. Машинный перевод на основе правил (Rule-Based Machine Translation) является методом, применявшимся к первым системам машинного перевода. В его основе лежит, главным образом, лингвистическая информация об исходном и целевом языках, содержащаяся в глоссариях, толковых и двуязычных словарях и базах данных. Машинный перевод на основе правил разделяется на три разных группы систем, а именно систем пословного перевода (Direct Machine Translation systems), трансферных систем (Transfer-based Machine Translation systems) и интерлингвистических систем (Interlingua-based Machine Translation systems). Системы пословного перевода выполняют дословный перевод исходного текста, который делится на несколько этапов. Во-первых, проводится морфологический анализ исходного текста. Во-вторых, дословный перевод исходного текста выполняется с помощью двуязычного автоматического словаря. В-третьих, полученный перевод адаптирован к грамматическим и синтаксическим правилам языка перевода. Системы дословного перевода не могут полностью передать контекстуальную информацию и поэтому только подходят таким документам, в которых представлено мало контекстуальной информации, как например технические пособия. Кроме того, системы дословного перевода запрограммированы на перевод только с одного языка на другой и не имеют возможности выполнять многоязычные переводы. С другой стороны, трансферные системы состоят из трех модулей, именно модуля анализа, модуля трансфера и модуля генерации целевого текста, каждый из которых выполняет определенную

задачу. Модуль анализа проводит морфологический анализ исходного текста, модуль трансфера содержит правила перевода одной языковой пары и одной направления перевода, а модуль генерации целевого текста генерирует целевой текст на основе грамматических и синтаксических правил языка перевода. По сравнению с системами дословного перевода, трансферные системы лучше передают контекстуальную информацию, однако они не запрограммированы на многоязычные переводы, которые потребовали бы использования слишком большого числа модулей перевода и, следовательно, чрезмерных интеллектуальных и экономических усилий. В заключение, интерлингвистические системы основаны на использовании нейтрального и универсального интерлингвы в качестве промежуточного языка между исходным и целевым языками. Процесс перевода происходит следующим образом: исходный текст переводится на интерлингву, придуманную для сохранения исходное значение и одновременного отказа от синтаксической структуры исходного текста. Впоследствии интерлингва, в свою очередь, переводится на целевой язык. Хотя интерлингвистические системы предположительно способны обеспечить многоязычные переводы, создание нейтрального и универсального интерлингвы, подобным образом подходящей всем языковым парам, несомненно является крайне сложной задачей. Несмотря на их относительно низкие производственные затраты, в системах машинного перевода на основе правил, наблюдаются некоторые критические недостатки, связанные главным образом с трудностями поиска и обновления большого объема лингвистической информации.

С целью преодоления проблем систем машинного перевода на основе правил, были разработаны системы машинного перевода на основе примеров. Этот новый подход изначально придумали как средство для поддержки систем машинного перевода на основе правил, а впоследствии стал самостоятельным. В системах машинного перевода на основе примеров используются базы данных, содержащие большое количество примеров перевода определенной языковой пары. Исходный текст разделяется на фрагменты сопоставимой длины, которые сравниваются с теми, находящимися в базе данных. Аналогичные фрагменты предлагаются пользователю вместе с них коэффициентом надежности. Фрагменты с наивысшим коэффициентом надежности выбираются и адаптируются адаптационным модулем

к структуре целевого текста. В целом, системы машинного перевода на основе примеров имеют более простую структуру и следовательно нуждаются в более дешёвой и простой разработке по сравнению с системами машинного перевода на основе правил. Причина этого в легкости, с которой можно найти большие количества новых примеров перевода и добавить их в базы данных систем машинного перевода. Тем не менее, качество переводов, выполненных системами машинного перевода на основе примеров, не оправдало ожидания. Поэтому системы машинного перевода на основе примеров были заменены системами статистического перевода, разработанными начиная с идеей Уоррена Уивера 1949 г.

Распределение вероятностей, сформулированный Клодом Э. Шенноном и Уорреном Уивером в книге *Mathematical theory of communication* (Математической теории коммуникации) является математическим законом для указания вероятности, с которой определенный выход будет получен из набора входов, на основе которого придуманы системы статистической машинной перевода. В частности, статистический машинный перевод основан на вероятности, с которой слово или группа слов в исходном тексте соответствуют слову или группе слов в целевом тексте. Системы статистического машинного перевода состоят из двух модулей, а именно модуля перевода и языкового модуля. Модуль перевода программируется на основе статистического анализа двуязычных баз данных, содержащих большое количество текстов и их соответствующие переводы. Статистический анализ проведен с помощью выравнивания исходных текстов с них переводами. Таким образом, статистические правила перевода определенной язычной парой получаются и используются при переводе. В ходе исторического развития статистического машинного перевода, выравнивание исходных текстов и переводов, содержащихся в базах данных систем статистического машинного перевода, выполнялось на текстовых строках разной длины. В частности, изначально оно сделано на уровне отдельных слов, а затем, с целью рассмотрения как можно большего количества лингвистических аспектов текстов, начали анализировать более длинные текстовые строки, как например синтагмы и синтаксические единицы. Это, безусловно, позволило не только лучше передать контекстуальную информацию, содержащуюся в исходных текстах, но и добиться

гораздо высших стандартов качества. Процесс перевода систем статистического перевода делается по следующим этапам: во-первых, исходный текст переводится модулем перевода; во-вторых, языковой модуль генерирует исходный текст, на основе грамматических и синтаксических правил языка перевода. На протяжении многих лет статистический машинный перевод являлся наиболее широко используемым подходом. Однако, недавно проявился новый подход, полностью революционизировавший стандарты качества машинного перевода с точки зрения эффективности и скорости обработки.

Отличные переводческие способности нейронного машинного перевода обусловлены, главным образом, применением концепции векторного представления слов при переводе с одного естественного языка на другой. Векторное представление слов является набором техник для представления слов с помощью направленных векторов. В частности, семантическое пространство задумано как геометрическое пространство, а евклидово расстояние между двумя векторами, т. е. между двумя словами, соответствует их семантическому расстоянию. Другими словами, чем ближе два слова находятся в семантическом пространстве, тем больше у них сходного значения. Существует множество разных типов систем нейронного машинного перевода. Один из наиболее популярных типов состоит из трех основных элементов, а именно декодера, кодера и модели внимания. Все трое являются искусственными нейронными сетями, имитирующими человеческие нейроны и запрограммированными на выполнение многочисленных человеческих задач, в том числе и перевод. Процесс перевода происходит следующим образом: во-первых, кодер анализирует фрагменты исходного текста и извлекает из них направленные векторы, представляющие их лингвистические особенности. Во-вторых, декодер использует векторные представления фрагментов исходного текста для создания целевого текста. В заключение модель внимания автоматически выполняет выравнивание слов исходного текста со словами целевого текста. В целом применение концепции векторного представления слов позволяет системам нейронного машинного перевода адекватно передать контекстуальную информацию и, таким образом, совершать более качественные переводы по сравнению с такими, выполненными системами машинного перевода на основе правил и системами статистического

перевода. Благодаря постоянному исследованию в области векторного представления слов и дистрибутивной семантики, нейронный машинный перевод развивается очень быстрыми темпами. Кроме того, по словам исследователей, на данный момент он представляется единственным подходом, на который следует инвестировать будущие интеллектуальные и экономические ресурсы.

Отличительные особенности медицинской терминологии

В течение последних лет лингвисты все больше интересовались определением основных особенностей языка науки. Это связано со значительным развитием научных исследований и ключевой ролью перевода в настоящей международной коммуникации. В целом язык науки является достаточно разнообразным. Действительно, для его лексики характерны три разных лексических группы, первая из которых состоит из терминов, полученных из повседневного языка, вторая, содержащая научные термины, которые широко распространены так в повседневном языке, как в языке науки, и третья, состоящая из научных терминов, использование которых ограничено профессиональной научной коммуникацией. Среди многочисленных отраслей языка науки, медицинский язык, безусловно, характеризуется особенно постоянной и быстрой эволюцией и с самого начала своего развития был связан с переводом. Медицинский язык можно определить как язык, используемый медицинскими работниками в профессиональном общении. В частности, в течение первых лет его развития, как и в настоящее время, перевод позволил разрушить языковые барьеры, которые препятствовали обмену медицинскими знаниями между специалистами и учеными из разных стран и культур. Поэтому при переносе медицинской терминологии из одной культуры в другую, из одного языка в другой, наблюдается начальный этап, на котором в медицинских книгах и трактатах сохраняются так термины на исходном языке, как и их эквиваленты на языке перевода. Медицинский язык обладает некоторыми характеристиками, способствующими его перевод с одного языка на другой. Во-первых, универсальность предмета медицинского языка, обсуждающего организм человека и его характеристики; во-вторых, наличие большого количества документов, которые могут быть использованы в качестве справочного материала для перевода медицинского языка; в-третьих, довольно распространена лексическая эквивалентность, обусловленная общим греческим и латинским

происхождением медицинского языка, используемого в большинстве стран мира. Тем не менее, другие особенности языка медицины могут вызывать проблемы при переводе с одного языка на другой. В первую очередь, язык медицины характеризуется, как указывалось выше, очень быстрым и постоянным развитием. Действительно, новые термины довольно часто придумываются для определения новых понятий, болезней, или симптомов, а устаревшие термины постепенно исчезают из медицинской лексики или заменяются новыми. Вторым важным аспектом, представляющим собой препятствие для выполнения качественных переводов, является синонимия. В частности, историческое развитие языка медицины обусловило наличие, в медицинском лексике, большого количества синонимичных пар. Эти синонимические пары могут состоять из разных элементов, таких как, в большинстве случаев, древних терминов и их современных эквивалентов или терминов греческого или латинского происхождения и их эквивалентов в национальных языках. Особенный случай представляет собой эпонимия, а именно сосуществование в медицинской терминологии, для определения патологий, имени ученого, который впервые открыл и изучил ее, и названия, описывающего патогенез такого заболевания. Это связано с тем, что при обнаружении патологии, ей дают название, прежде чем подробно изучить ее патогенез. Это название обычно является собственным именем ученого, открывшего патологию. В последствии, как только особенности патогенеза известны, выбирается другое, более значимое название, и в результате оба термина сохраняются в медицинской терминологии. Русский медицинский язык, который мы будем более подробно анализировать в ходе нашего исследования, безусловно имеет ряд типичных черт языка медицины, таких как высокий степень эпонимии и синонимии, обусловленного историческими истоками. Медицинские знания, совместно с медицинской терминологией, действительно ввозили в Россию из Греции в XVII веке. В результате этого, в настоящее время, медицинские термины греческого происхождения сосуществуют с русскими эквивалентами так в русских медицинских книгах и трактатах, как и в устной коммуникации. Имея более высокую степень понятности для людей, не владеющие русским языком, медицинские термины греческого происхождения чаще использованы в международной коммуникации по сравнению с их русскими эквивалентами.

Отличительными чертами русского медицинского языка являются также безличные предложения, склонность к употреблению пассивной формы и предпочтение существительных вместо глаголов. Кроме того, наблюдается довольно сложная синтаксическая структура, в которой выделяется широкое использование причастий и деепричастий.

Методы оценки произведений систем машинного перевода

С одной стороны, оценка эффективности систем машинного перевода крайне необходима экспертам по машинному переводу для детального определения сильных и слабых сторон систем машинного перевода и применения разумных мер с целью улучшения их общей производительности. С другой стороны, она безусловно является очень полезным способом для пользователей, чтобы выбрать систему машинного перевода, лучше всех удовлетворяющую их личные потребности. В прошлом, так же, как и в настоящее время, оценка машинного перевода была и остается открытым и спорным вопросом. Прежде чем перейти к описанию разных методов, используемых для оценки машинного перевода, необходимо перечислить основные критерии, которые учитываются при определении качества перевода, независимо от того, выполняется ли он человеком или автоматизированными системами. В частности, учитывается несколько универсальных характеристик перевода, в том числе беглость речи, адекватность, читабельность, понятность, и удобство использования. Дальнейшие параметры выбраны в зависимости специфических черт тестов и язычных пар проверяемого перевода. Кроме того, крайне важно подчеркнуть, что основополагающим понятием является то, что нет идеального перевода, а, по всей вероятности, существует несколько приемлемых переводов. Оценка машинного перевода выполняется автоматически, полуавтоматически и вручную. Однако, граница между разными методами еще не четко определена, и совместное использование этих методов встречается довольно часто. Кроме того, поскольку все методы имеют свои положительные, так и отрицательные стороны, вопрос о том, какой подход является наиболее эффективным, до сих пор остается предметом споров между учеными.

Автоматическая оценка машинного перевода включает все те методы, при которых интеллектуальные усилия человека минимальны или ограничены

второстепенными задачами, такими как сбор данных, подготовка справочных переводов и аннотирование ошибок. Основными представителями этой категории являются метрики, использующие справочные переводы. Они вычисляют расстояние редактирования, также называемое расстоянием Левенштейна, или текстовое сходство между переводом, выполняемым системой машинного перевода, и ручными справочными переводами. В первом случае, учитывается количество модификаций, таких как вставка, удаление, замещение, перемещение слов или групп слов, необходимые машинному переводу того, чтобы полностью совпасть со справочным переводом. Среди таких методов достаточно популярной и часто используемой стала метрика *WER*, которая расшифровывается как *Word Error Rate* (Ставка ошибок слов). Во втором случае вычисляются текстовые фрагменты машинного перевода, также появляющиеся в одном из справочных переводов. Сумма этих фрагментов впоследствии разделится на общее количество текстовых фрагментов, содержащихся в машинном переводе. Метрика *BLEU*, которая расшифровывается как *Bilingual Evaluation Understudy* (Двуязычная оценка), несомненно, является основным представителем этой категории. Напротив, человеческая оценка машинного перевода включает в себя все те методы, которые прямо или косвенно связаны с человеческим суждением. В частности, она разделяется на две основные группы, а именно на оценку, основанную на прямом человеческом суждении, и оценку, не основанную на прямом человеческом суждении. В первой группе принимают участие все те методы, при которых специалисты по оценке переводов, также известные как комментаторы, предложены дать свою оценку качества машинного перевода. В зависимости цели и процедуры оценки, комментатор может дать свою прямую оценку, выбрать более качественный перевод из группы машинных переводов, или составить рейтинговый список. Напротив, вторая группа состоит из таких методов, при которых оценка качества машинного перевода или группы машинных переводов дается косвенно. Другими словами, комментаторам предложено ответить на вопросы о понятности перевода или его полном соответствии с грамматическим и синтаксическим правилам языка перевода или выполнить анализ ошибок анализируемого машинного перевода. Анализ ошибок является очень интересным и популярным методом оценки переводов, выполненных

системами машинного перевода. Для того чтобы анализ ошибок был эффективным, он должен быть детально спланирован. Во-первых, исходный текст аккуратно проанализирован для полного понимания его наиболее релевантные лингвистические особенности. Во-вторых, на основе лингвистического анализа определяются категории ошибок, которые будут использованы в качестве критерий оценки перевода в ходе анализа ошибок. В-третьих, выполняется анализ ошибок машинного перевода, причём каждая ошибка помечается и маркируется как относящаяся к одной или больше из выбранных категорий ошибок. На сегодняшний день не существует универсально общепризнанного метода анализа ошибок, и каждый ученый предложил свой собственный вариант. Кроме того, каждый тип текста и язычной пары нуждается в индивидуально спланированном анализе. В целом выбор между автоматическими и человеческими методами оценки перевода не является простой задачей. Действительно, не только граница между двумя методами не четко определена, а также каждый из двух методов имеет свои плюсы и минусы. С одной стороны, автоматическая оценка, хотя гораздо быстрее и дешевле по сравнению с оценкой человеком, опирается на справочные переводы или аннотации, выполнены переводчиками и комментаторы, и, следовательно не является полностью автоматическим процессом. С другой стороны, ручная оценка постоянно критиковалась за то, что она является не только слишком субъективной, но и медленным и дорогим процессом. Напротив, так как перевод является человеческой деятельностью, используемой человечеством, люди лучше всего способны аккуратно оценить перевод. Как упоминалось выше, многочисленные методы оценки машинного перевода были придуманы в течение лет, но ни одна из них не была выбрана в качестве модели выполнения оценки переводов. Следовательно, оценка машинного перевода остается открытым вопросом, который, безусловно, требует дальнейшего изучения.

Подробнее о нашем исследовании

Для проведения нашего исследования выбираются три узкоспециализированные русскоязычные медицинские статьи и три научно-популярные русскоязычные медицинские статьи о коронавирусной пандемии. Выбор темы статей обусловлен желанием рассматривать недавно опубликованные статьи, как и анализировать влияние языкового феномена, вызванного эпидемией нового коронавируса, на

поведение двух крайне популярных систем нейронного машинного перевода, а именно DeepL и Яндекс. Действительно, помимо драматических санитарных и экономических последствий, которые наблюдались в течение последних месяцев, новый коронавирус ответственный за введение так в медицинскую терминологию, как и в общий язык, ряда ранее не существовавших или крайне редко употреблявшихся терминов. Наш анализ происходит следующим образом: во-первых, заголовки и рефераты узкоспециализированных медицинских статей, вместо с заголовками и первым абзацами научно-популярных медицинских статей разделяются на некоторые фрагменты одинаковой длины для обеспечения читателям полной удобочитаемости. Каждый фрагмент переводится с помощью DeepL и Яндекс. Фрагменты из исходного текста, совместно с их переводами, выполненными DeepL и Яндексом, показываются, впоследствии, в таблице, которая аккуратно представляет соответствия между оригинальными текстами и их переводами. Во-вторых, сравнительный анализ ошибок проводится так: на основе наиболее релевантных лингвистических особенностей исходных текстов выбираются и детально определяются следующие категории ошибок:

- Синтаксис
- Грамматика
- Лексика
- Использование артикля
- Аббревиатуры
- Терминология
- Культурные особенности
- Членение предложения
- Пропуск слова
- Непереведённые элементы
- Согласованность
- Орфография

- Транслитерация
- Формат

Ошибки, совершенные двумя системами машинного перевода, выявляются и кратко комментируются. Впоследствии каждая ошибка помечается как принадлежащая к одной или больше из выбранных категорий ошибок. Фрагменты переводов, содержащие ошибки, показываются в таблице, вместе с них эквиваленты на исходном языке, и фрагменты ручного перевода, специально выполненного для анализа ошибок и использованного в качестве справочного перевода. Для обеспечения читателям достаточной степени ясности и легкости чтения, создаются четыре таблицы, каждая из которых показывает детали переводов одного типа текста, выполненных одной системой машинного перевода. В заключение, на основе данных, содержащихся в таблицах, составляется ряд графиков, показывающих общее количество и характер ошибок, совершенных двумя системой машинного перевода. Цель использования графиков - привлечь внимание к некоторыми интересным переводческими тенденциям DeepL и Яндекса. В первой группе графиков дается полный обзор количества ошибок, совершенных каждой системой машинного перевода, и поделенных на категории ошибок. Кроме того, на графиках, категории ошибок отображаются в нисходящем порядке по количеству ошибок, и общее количество ошибок, относящихся к каждой категории, явно показывается. Во второй группе графиков переводы узкоспециализированных статей, выполненные каждой системы машинного перевода сравниваются с переводами научно-популярных статей. В третьей группе графиков сравнивается общая переводческая производительность DeepL и Яндекса. Таким образом, проводится прямое сравнение и оценка эффективности DeepL и Яндекса при переводе разных типов текста. Графики, очевидно, показывают, что DeepL и Яндекс, в целом, сталкиваются с наибольшими трудностями в тех же категориях ошибок, именно лексика, использование артикля и синтаксис. С одной стороны, большое количество ошибок в категориях синтаксиса и использования артикля обусловлено большей разницей в синтаксической структуре и использовании артикля между итальянским и русским языками. С другой стороны, для полного понимания лексических ошибок, необходимо учитывать их распределение между

двумя анализируемыми типами текстов. Сравнивая производительность каждой системы машинного перевода при переводе разных типов текстов, видно, что и DeepL, и Яндекс совершают меньше ошибок в ходе перевода научно-популярных статей. Кроме того, при переводе специализированных текстов, обе системы машинного перевода допускают многочисленные ошибки в категориях синтаксиса, лексики, использования артикля, терминологии, аббревиатур и транслитерации. В то время как в случае категорий терминологии, аббревиатур и транслитерации вероятно, что, на результаты влияет бóльшая частота этих элементов в специализированных статьях, так DeepL, как и Яндекс не способны аккуратно перевести синтаксическую структуру и специализированную лексику специализированных статей. В заключение сравнение переводческих способностей двух систем машинного перевода ясно показывает, что, в общем и целом, DeepL переводит как специализированные, так и научно-популярные тексты лучше, чем Яндекс. Действительно, хотя Яндекс добивается лучших результатов в категориях терминологии, пропущенных слов и транслитерации, наблюдается близкий разрыв с результатами DeepL. Напротив, DeepL допускает меньше ошибок, чем Яндекс во всех остальных категориях, а в некоторых из них, таких как использование артикля, синтаксис и лексика, видна очень заметная разница между производительностями двух систем машинного перевода. В заключение анализа, переводы, выполняемые двумя системами машинного перевода, оцениваются с использованием метрики BLEU, описанной выше, на основе текстового сходства каждого из них с переводами, выполненными человеком и используемыми в качестве справочных переводов. Результаты, представлены в таблице, не отличаются от такими, полученными от анализа ошибок. Тем не менее, использование метрики BLEU обеспечивает получение точного процента от текстового сходства между каждой статьей и её соответствующим справочным переводом. Таким образом, возможно анализировать поведение DeepL и Яндекса при переводе каждой статьи. Окажется, что особенно сложной для DeepL является первая научно-популярная статья, под названием *Коронавирус завозили в Россию не менее 67 раз*, перевод которой получает гораздо ниже процент от текстового сходства по сравнению с другими текстами. Этот анализ, безусловно, выявил огромные успехи, добившись в течение долгих лет научных исследований, нейронными системами машинного перевода, в

том числе DeepL и Яндекс. Тем не менее, очевидно, что, все еще наблюдаются, особенно в областях синтаксической структуры, лексики и использовании артикля, некоторые слабые стороны, препятствующие достижению желаемых стандартов качества при переводах с русского на итальянский язык. Дальнейшие исследования необходимо направить, начиная с результатов, полученных в данной работе, на лингвистический анализ разных категорий ошибок, рассматриваемых в данной диссертации, и попытаться определить значимость каждого из них в выполнении перевода с русского на итальянский. Таким образом, можно понять, какие ошибки систем машинного перевода необходимо срочно исправить, а какие не являются препятствием для полного понимания полученного перевода. Несомненно, в любой попытке увеличения эффективности систем машинного перевода активно принимают участие разные категории специалистов, в том числе лингвисты и программисты. Этот аспект приводит к постоянному переопределению важности и профессиональных черт работы переводчика, который будет во все большей мере способствовать своими лингвистическими знаниями улучшению качества систем машинного перевода.

REFERENCES

- Accademia della Crusca, <https://accademiadellacrusca.it/>, accessed on December 15th, 2020.
- Aleksandrovič, Ju.S., Bajbarina, E.N., Baranov, A.A., Višneva, E.A., Zvereva, N.N., Ivanov, D.O., Krjučko, D.S., Konovalov, I.V., Kuličenko, T.V., Lobsin, Ju.V., Mazankova, L.N., Namazova-Baranova, L.S., Petrenko, Ju.V., Prometnoj, D.V., Pšenisnov, K.V., Rtiščev, A.Ju., Sajfullin, M.A., Selimzjanova, L.R., Uskov, A.N., Fedoseenko, M.V., Harkin, A.V., Čumakova, O.V., Efendieva, K.E., Jakovlev, A.V. 2020. Vedenie detej s zabolevaniem, vyzvannym novoj koronavirusnoj infekciej (SARS-CoV-2). *Pediatričeskaja farmakologija*, 17/2.
- Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. 1966. *Language and machines, computers in translation and linguistics*. National Academy of Sciences. Washington, D. C.
- Bates, M. 1995. Models of natural language understanding, *Proc. Natl. Acad. Sci. USA*, 92 pp. 9977-9982, Colloquium Paper presented at a colloquium entitled "Human-Machine Communication by Voice," organized by Lawrence R. Rabiner, held by the National Academy of Sciences at The Arnold and Mabel Beckman Center in Irvine, CA, February 8-9, 1993.
- Callison-burch, C., Osborne, M., Koehn, P. 2006. Re-evaluating the role of BLEU in machine translation research. In *EACL*, volume 6, pp. 249–256.
- Castilho, S., Doherty S., Gaspari, F., Moorkens, J. 2018. *Approaches to Human and Machine Translation Quality Assessment. Translation Quality Assessment. From Principles to Practice*. Springer.
- Charoenpornawatt, P., Sornlertlamvanich, V., Charoenporn, T. 2002. *Improving Translation Quality of Rule-based Machine Translation*, Information Research and Development Division National Electronics and Computer Technology Center, Klong Luang.
- Chatzikoumi, E. 2020. How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering (2020)*, 26, pp. 137–161. Cambridge University Press.
- Collins English Dictionary, collins.co.uk, HarperCollins. Accessed on December 20th 2020.
- Fedina, E.A. 2011. Zaimstvivanija kak istočnik sinonimii v medicinskoj terminologii. *Vestnik IrGTU n. 4*, 2011. Irkutskij gosudarstvennij medicinskij universitet. Irkutsk.
- Fedina, E.A. 2017. K probleme sinonimii v medicinskoj terminologii. *Vestnik TGPU 2017*, 6. Irkutskij gosudarstvennij medicinskij universitet. Irkutsk.
- Fischbach, H. 1962. Problems of medical translation. *Bulletin of the Medical Library Association*, Vol. 50, N. 3, pp. 462-472.
- Fischbach, H. 1986. Some Anatomical and Physiological Aspects of Medical Translation. Lexical equivalence, ubiquitous references and universality of subject minimize misunderstanding and maximize transfer of meaning. *Journal des traducteurs*, Vol. 31, N. 1. Les Presses de l'Université de Montréal. Montréal.

- Gehring, J., Auli, M., Grangier, D., Dauphin, Y.N. 2017. A Convolutional Encoder Model for Neural Machine Translation. Facebook AI Research.
- Giovine, S. 2020. Coronavirus: un nome comune (di virus) per una malattia non comune. *Italiano digitale*, XII
- Glazunova, O.I. 2016. Grammatika russkogo jazyka v upražnenijach i kommentarijach. Zlatoust, Saint Petersburg.
- Guščina, L.N. 2005. Osobennosti jazika mediciny. *Žurnal GGLLU* 2005 n.1.
- Hadiwinoto, C. 2017. Syntax-Based Statistical Machine Translation, book review. *Computational Linguistics*, MIT Press.
- Harčenko, E.P. 2020. Koronavirus SARS-Cov-2: složnosti patogeneza, poiski vakcin i buduščie pandemii. *Epidemiologija I Vakcinoprofilaktika*, 19/3
- Hutchins, J. 1996. ALPAC: the (in)famous report. *MT News International*, 14, pp. 9-12.
- Hutchins, J. 2004. Two precursors of machine translation: Artsrouni and Trojanskij, *International Journal of Translation*.
- Hutchins, J. 2005. The Georgetown-IBM Demonstration, 7th January 1954. *MT News International* 8, pp. 15-18.
- Hutchins, J. 2005. Example based machine translation – a review commentary. *Machine Translation*, 19, pp 197-211.
- Hutchins, J. 2005. Towards a definition of example-based machine translation, The 10th Machine Translation Summit.
- Hutchins J. The history of machine translation in a nutshell. Website <http://ourworld.compuserve.com/homepages/WJHutchins/Nutshell.htm>, accessed on September 27th, 2020.
- HSE University, <https://www.hse.ru/en/>, accessed on January 3rd, 2020
- International Laboratory for Population and Health, <https://demogr.hse.ru/en/>, accessed on January 3rd, 2020
- Karwacka, W. 2014. Quality assurance in medical translation. *The Journal of Specialised Translation*, issue 21, pp. 19-34.
- Khurana, D., Koli, A., Khatter, K., Singh, S. 2017. Natural Language Processing: State of The Art, Current Trends and Challenges. Department of Computer Science and Engineering Manav Rachna International University.
- Korotaeva, T.V. 2015. Immunogennost, vyzvannaja genno-inženernymi biologičeskimi preparatami pri lečenii psoriaza i psoriatičeskovo artrita: vzljad na problemu. FGBNU Naučno-issledovatel'skij institut revmatologii im. V.A. Nasonovoj. Moscow, Russia.
- Kovalev, V. 2020. Il Kovalev. Dizionario russo-italiano, italiano-russo. Zanichelli, online resource, accessed on November 21st, 2020.
- Lagarda, A.L., Alabau, V., Casacuberta F., Silva, R., D'iaz-de-Liano, E. 2009. Statistical Post-Editing of a Rule-Based Machine Translation System. *Proceedings of NAACL HLT 2009. Short Papers*, Boulder, Association for Computational Linguistics, pp. 217-220. Colorado.
- Lenci, A. 2010. Modelli distribuzionali del lessico. Modelli computazionali per l'analisi semantica. *Informatica umanistica*. <https://www.ledonline.it/informatica-umanistica>, accessed on September 29th, 2020.

- Leusch, G., Ueffing, N., and Ney, H. 2006. Cder: Efficient mt evaluation using block movements. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, pp. 241–248. Trento, Italy.
- Lepschy, L., Lepschy, G. 1993. *La lingua italiana, Storia, varietà dell'uso, grammatica*. Bompiani, pp. 147, 148.
- Levenštejn, V. 1965. Dvoičnye kody s ispravleniem vypadenij. Vstavok I zameščenijsimbolov Doklady Akademij Nauk SSSR, 163, pp. 845-848. Moscow.
- Linguistic Data Consortium (LDC). 2005. Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations. Revision 1.5. Quoted by Monti, J., Montella, C. 2015. About adequacy, equivalence, and translatability in human and Machine Translation. Conference: New Horizons in Translation and Interpreting Studies. Malaga, Spain.
- Lommel, A. 2018. Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. Translation Quality Assessment. From Principles to Practice. Springer.
- Lvovič, V. 2020. Koronavirus Zavozili v Rossiju ne menee 67 raz. IQ HSE RU <https://iq.hse.ru/>. Accessed on October 15th, 2020
- Machine Translation.Yandex company.
<https://yandex.com/company/technologies/translation>. Accessed 14.11.2020.
Accessed on November 14th 2020
- Mandell, L.A., Wunderink, R.G., Anzueto, A., Bartlett, J.G., Campbell, D., Dean, N.C., Dowell, S.F., File, T.M., Musher, D.M., Niederman, M.S., Torres, A., Whitney, C.G. 2007. Infectious Diseases Society of America/American Thoracic Society Consensus Guidelines on the Management of Community-Acquired Pneumonia in Adults. *Clinical Infectious Diseases*, Volume 44, Issue Supplement_2.
- Mičić, S. 2013. Languages of medicine – present and future. *JAHHR*, Vol. 4, N. 7, pp. 217 –233.
- Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg, Dean, Jeffrey. 2013. Distributed representations of Words and Phrases and Their Compositionality.
- Munkova, D., Hajek, P., Munk, M., Skalka, J. 2020. Evaluation of Machine Translation Quality through the Metrics of Error Rate and Accuracy. Third International Conference on Computing and Network Communications (CoCoNet'19). Elsevier B.V.
- Nagao, M. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. In. *Artificial and Human Intelligence*. A. Elithorn and R. Banerji (ed.). Elsevier Science Publishers.
- Naldi, M. 2014. Traduzione automatica e traduzione assistita. pp 57-66. Esculapio.
- Namazova-Baranova, L.S., Myraškin, N.N., Ivanov, R.A. 2020. Biologičeskaja terapija v eru COVID-19. *Voprosy sovremennoj pediatrii*, 19/2
- Naumov, D.G. 2006. Filologeničeskij analiz semejstva belkov-gomologov. 2006. Gosudarstvennij naučno-issledovatel'skij institut genetiki i selekcii promyšlennyx mikroorganizmov. Moscow, Russia
- Och, F.J., Tillmann, C., Ney, H. 1999. Improved alignment models for statistical machine translation. In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, University of Maryland, College Park (1999), pp 20-28.

- Olekhovich O.G., Olshvang O.Y. 2017. The influence of European scientific thought on the development of Russian medical terminology. *Russian Linguistic Bulletin* 2(10).
- One model is better than two. Yandex.Translate launches a hybrid translation system. Yandex company. <https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system/>. Accessed on November 14th 2020
- Papineni, K., Roukos, S., Ward, T., Zhu W.J. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 311-318. Philadelphia.
- Papineni, K., Roukos, S., Ward, T., Henderson, J., Reeder, F. 2002. Corpus-based Comprehensive and Diagnostic MT Evaluation: Initial Arabic, Chinese, French, and Spanish Results” in *Proceedings of Human Language Technology 2002*, pp. 132–137. San Diego.
- Polackova, G. 2001. Synonymy of medical terminology from the point of view of comparative linguistics. *Bratisl Lek Listy*, 102 (3), pp. 174-177.
- Popovic', M. 2018. *Error Classification and Analysis for Machine Translation Quality Assessment. Translation Quality Assessment From Principles to Practice*. Springer.
- Reshamwala, A., Mishra, D., Pawar, P. 2013. Review of Natural Language Processing. *IRACST – Engineering Science and Technology: An International Journal (ESTIJ)*, 3(1).
- Salmanova, S. 2020. Neizvestnaja letalnost. Počemu my ne znaem istinnyh mashtabov COVID-19. IQ HSE RU <https://iq.hse.ru/>. Accessed on October 15th, 2020
- Shannon, C.E. 1948. *A Mathematical Theory of Communication*. Bell System Technical Journal.
- Skolkovo Institute of Science and Technology. <https://www.skoltech.ru/en/>, accessed on December 20th 2020.
- Smorodintsev Influenza Research Institute, <https://www.influenza.spb.ru/en/>, accessed on December 21st 2020.
- Sumita, E., Iida H. 1991. Experiments and Prospects of Example-Based Machine Translation. *29th Annual Meeting of the Association for Computational Linguistics*, pp. 185-192. Berkeley, California.
- Tiberii, P. 2018. *Dizionario delle collocazioni. Le combinazioni delle parole in italiano*. Zanichelli.
- Timonin, S. 2020. Smertnost ot COVID-19. Vzgljad demografa na statistiku pričn smerti v Rossii i mire. IQ HSE RU <https://iq.hse.ru/>. Accessed on October 15th, 2020
- Traduttore. DeepL Translate. <https://www.deepl.com/translator>. Accessed on November 14th 2020.
- Treccani. *Enciclopedia online*. <https://www.treccani.it/> Accessed on November, 21st, 2020.
- Trojanskij, P. 1935. *Opisanie mašiny dlja podbora i pečetanija slov pri perevode s odnovo jazika na neskolko drugih odnovenno*. Patent SSSR 40995.

- Van Hees, M., Kozłowska, P., Tian, N. 2015. Web-based automatic translation: the Yandex.Translate API. Leiden Institute of Advanced Computer Science (LIACS). Leiden.
- Van Merriënboer, K.C.B., Bahdanau, D., Bengio, Y. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. SSST@EMNLP.
- Wang, R., Finch, A., Utiyama, M., and Sumita E. 2017. Sentence Embedding for Neural Machine Translation Domain Adaptation. Conference paper, In Proceedings of the Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada.
- Weaver, W. 1947. Letter to Norbert Wiener. Quoted by Weaver W. 1949. Translation, Rockefeller Foundation.
- Weaver, W. 1949. Translation. Rockefeller Foundation.
- Wołk, K., Marasek, K. 2015. Neural-based machine translation for medical text domain. Based on European Medicines Agency leaflet texts. Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist.
- Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K. 2016. Google’s neural machine translation system: Bridging the gap between human and machine translation, arXiv preprint arXiv:1609.08144.
- Zens, R., Ney, H. 2004. Improvements in phrase-based statistical machine translation. Proceedings of HLT-NAACL 2004, pp. 257–264.
- Zingarelli N. 2020. Lo Zingarelli 2021. Vocabolario della lingua italiana. Zanichelli, online resource, accessed on November 21st, 2020.