**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 comparative study of machine translation engines: transliterating Russian anthroponyms and toponyms into Italian*

Relatrice
Prof.ssa Sabina Fata

Laureanda
Olga Gastaldello
n° matr. 2091337 / LMLCC

Anno Accademico 2023 / 2024

# TABLE OF CONTENTS

# INTRODUCTION

Throughout history, sharing information among people and cultures all over the world has become particularly important. To enhance cross-cultural communication translation strategies have always been used, so that people could read in their own native language what originally was written in a foreign language. Each language in the world has some words that are unique for the culture they belong to, or that have characteristics inherent to that language. Finding translation equivalents for these unique words, such as persons names and toponyms, may create some difficulties. Words falling under these categories can be considered "unique" in the sense that are deeply embedded in their original culture, making it challenging to render them accurately in the target language and culture. When these words need to undergo the process of translation, transliteration emerges as a useful translation strategy to be applied. Transliteration is a process which involves the "transformation of a word (normally a proper noun) into a language which has another phonology inventory and a different alphabet" (Benites, 2020). Transliteration is even more complex when languages taken into consideration have very different alphabets. An example can be provided by Russian and Italian languages, the focus languages of this thesis, which respectively use Cyrillic and Latin alphabet. The Latin alphabet, used in the Italian language, is made up of 21 letters. On the other hand, Cyrillic alphabet used in the Russian language is made up of 33 letters, including two signs that modify the pronunciation of the preceding letter.

Nowadays, the practice of translation performed by human translators is often aided by Machine Translation (from now on, the acronym MT will be used). This tool works through algorithms that have been trained to provide an automated translation result in the target language. The quality of MT and its outcomes, considering various language pairs, have been discussed by a number of researchers for quite a while now (Frederking and Nirenburg, 1994; Knight and Chander, 1994; Chatzikoumi 2019, Fitria 2021, Rivera-Trigueros 2021, Rossi and Carré, 2022). They adopted different views, also shifting the interest toward evaluating methods that can be later used to improve MT (Papineni, et al., 2002; Banerjee and Lavie, 2005; Snover et al., 2006). Despite some initially optimistic views, the general belief is that MT is useful nowadays to overcome the high demand for translation, but it is still considered just a helpful tool that cannot replace human translators. In spite of that, when it comes to considering a specific translation strategy, that is transliteration, with a further focus on transliteration performed by MT, research conducted in this field mainly focuses on creating transliteration resources. In fact, some researchers relied on the integration of knowledge between machine learning and computer science, and transliteration, suggesting how resources to enhance machine transliteration could be created (Azid et al., 2017; Benites, 2020). Therefore, the novelty of this thesis lies in adopting a different perspective when studying the practice of transliteration between Russian and Italian, focusing on transliteration results given by MT. Further investigation is necessary when considering the accuracy of transliteration performed by this tool, particularly when anthroponyms and toponyms are taken into consideration. Moreover, nowadays the use of Large Language Models (LLMs), such as ChatGPT, is increasingly common as they are able to perform a considerable number of Natural Language Processing (NLP) tasks (Naveed et al., 2024). Therefore, LLMs are also employed to generate translation outputs and to enhance NMT performance. More specifically, NMT systems can be guided through prompts that help shape the processing, so as to achieve the desired result (Li et al., 2022). Prompt engineering is an emerging field of study focused on investigating and analyzing prompt structures and outcomes to understand which prompt variations are necessary to obtain the best results (Amatrian, 2024; Naveed at al., 2024).

The aim of this study is to find the best MT tool that can be used when transliterating Russian anthroponyms and toponyms into Italian. In order to achieve this goal, these objectives are stated:

1. research of theoretical aspects related to transliteration strategy and transliteration of Russian into Italian, MT, transliteration in MT and LLMs;
2. definition of domain under investigation and generation of 22 passports;
3. selection of reference norms for each category;
4. insertion of generated passports in MT tools: Language Weaver, Modern MT, Intento, DeepL, Yandex;
5. selection of anthroponyms and toponyms;
6. insertion of generated passports in ChatGPT with instruction for translation;

7. identification and analysis of differences observed among the results obtained;
8. comparison on accuracy of these tools when transliterating Russian into Italian, with a focus on trained ChatGPT results.

The final results will be obtained following a qualitative method first, which will allow to identify and categorize the outcomes of the tools; secondly, a quantitative approach will be utilized, comparing the data gathered, so to provide a final consideration on which tool enables translators to obtain the best transliteration results, considering the norms taken as reference. Professional translators and translation students could benefit from this research. In fact, they would get to know which tool to rely on, helping them to speed up their work and to obtain better results for their jobs. In addition, this thesis leaves space for more research and investigation in various domains, with a focus on emerging technologies in the field of Artificial Intelligence (AI).

# 1. TRANSLITERATION

In the realm of language translation, the interaction between Russian and Italian languages requires an examination of translation processes and strategies used for texts in specific domains and having particular characteristics. This first part of literature review delves into translation, with a focus on transliteration, which is the main strategy employed in this research. Secondly, a brief overview on the historical evolution of transliteration is provided, so as to set the context for understanding the development and the standardization of one of the main strategies used to process anthroponyms and toponyms. Subsequently, the transliteration norms and the norms followed in this thesis are explained and briefly compared. Lastly, general knowledge about anthroponyms and toponyms is provided.

When dealing with a text that needs to be translated, translators need to take into consideration the text they are approaching, analyzing its characteristics and components. Higher attention has to be given to those lexical units that could create some translation problems[1]. By identifying recurrent problems and characteristics, translators are able to decide which is the best translation method to adopt. Translation methods are considered to be the actual realizations of translation approaches, which should be considered first. Subsequently, translation methods are transferred to translation techniques (Diadori, 2012). Posing an initial focus on translation methods, as stated in Diadori (2012) according to Vinay and Dalbernet (1958) and Malone (1988), it is possible to identify a double contraposition, between direct and oblique translation, and expansion and compression. In addition, some studies that consider translation techniques identify a categorization that allows to divide these techniques into lexical, grammatical, and syntactical translation techniques (Levickaja and Fiterman, 1976; Komissarov 1973). One of the direct, lexical translation techniques is transliteration: as stated by Bekbabayi and Amirzadeh (2019), scholars such as Catford, Newmark, and Harvey and Higgins generally consider it as the conversion of a word from one script into another. Newmark (1988) investigates the idea of translation procedure, opposed to the one of translation method, that needs to be introduced when dealing with smaller units of text, such as words

[1] This term was used by Peter Newmark in his book "A Textbook of Translation" published in 1988, when referring to lexical items that are highly intertwined with cultural features specifically belonging to the SL. This topic is further explained later in the paragraph.

and terms. In his categorization, he does not refer to transliteration as a translation procedure itself, but he claims that it falls under the category of transference. In addition, Newmark himself presented the concept of "translation problem" when considering words embedded in the cultural substrate of a language. When dealing with culture, particular attention has to be devoted to such cultural words that, once found in texts to be translated, often highlight the distance between source text (from now on, ST) and target text (from now on, TT). This phenomenon happens because of a gap between the two cultures, enhanced by the absence of a particular notion in the target culture (from now on, TC). Therefore, to deal with translation problems, Newmark suggests the adoption of transference, and consequently of transliteration. He is not the only one sustaining this idea, as also Vlahov and Florin, two years earlier than Newmark, in 1980 and the Estonian scholar Torop in 2000 (Staskevičiūtė and Baranauskienė, 2005) suggested transliteration, along with other techniques. Sometimes, being transliteration a mere transposition of characters, various techniques are suggested to be used at the same time, so to enhance comprehension and overcome the difficulties target readers can encounter in understanding the message. In this respect, Newmark (1988) develops the idea of couplet, pairing transliteration to techniques such as functional equivalence or notes. However, the strategy adopted has to take into account the readership and the text requirements, in addition to the cultural word itself. Thus, depending on the ST, adopting the couplet strategy may not always be possible; for instance, this might be the case for official documents, such as passports.

Transliteration is often presented alongside transcription, as another procedure to consider when dealing with different writing systems. Although these two methods are sometimes confused and used interchangeably, they differ significantly (Gerych, 1965; Mazzitelli, 2008). One of the first scholars to clearly differentiate between these two procedures was Alois Ševčik, whose considerations have been taken into account by the International Standard Organization. As stated in Gerych (1965), in the ISO R/9 some transliteration principles have been established, one of which makes a clear distinction between transliteration and transcription. Transliteration is employed to represent characters, focusing on how they are written, whereas transcription represents sounds, focusing on the pronunciation of words. Adhering to transcription norms often leads to inconsistent reproduction of SWs across different TLs. This topic, which is further explored in a later section where additional examples are provided, is supported by the idea that each language represents certain graphemes differently based on its phonetic rules. Consequently, adopting a strategy that relies on pronunciation and phonetic principles does not result in a one-to-one correspondence between a SW and a TW. Instead, as the following example illustrates, multiple TWs can correspond to the same original SW (Šimičević and Boljanović, 2009). The example, taken from the aforementioned authors' article, is the surname of the Russian author "Цветаева", which is rendered as "Tsvetaeva" in English, "Zwetajewa" in German, "Cvetaeva" in Italian, "Cvjetajeva" in Croatian and "Tswetaewa" in Polish.

## 1.1 HISTORICAL OVERVIEW

The practice of transliteration has long been adopted and was first introduced as a procedure to deal with different writing systems. According to Gerych (1965), one of the earliest works on this topic dates back to the Renaissance when, in 1548, Theodore Bibliander specifically discussed Cyrillic in one of his works on comparative philology and phonetics. However, major contributions occurred around the 18th and mid-19th centuries, when Lepsius, a German philologist, published an influential work which helped defining transliteration practices. By the late 19th-early 20th century, the question of transliteration practices became even more central due to cataloging and bibliographic needs. With the increase in bibliographical and documents exchange, the international community recognized the need for standardization and commonly agreed-upon norms. Until that time, many national systems for transliteration had been created, based on the characteristics of each language. Therefore, the International Phonetic Association in 1886 attempted to create a consistent phonetic alphabet that could be applied to all languages. Although this system initially failed because of the abundance of diacritics, this very solution was adopted by one of the first widely used transliteration systems, the Czech-style Roman alphabet, to avoid combinations of letters to reproduce Cyrillic characters. After the First World War the Roman alphabet became central and often adopted over the years also by countries that originally used other writing systems, so to further aim at standardization. Significant steps forward were made after the Second World War, which slightly delayed the discussion on transliteration systematization.

During a historical period that aimed at standardization in various fields, important decisions and actions were taken also in the area of transliteration. In 1954, the International Standard Organization continued the work of the previously founded International Standard Association (ISA), and approved the ISO/R9, which became the international scheme for transliteration of Cyrillic characters (Gerych, 1965). Although usually adopted by the member states of the organization, from the beginning ISO 9 was solely considered a recommendation to follow, which could eventually be replaced by national systems (Šimičević and Boljanović, 2009). The very first edition of the ISO 9 was issued by the Technical Committee ISO/TC 46 in 1986. However, as further explained in the following sections, at the present time the applicable standard is the ISO 9:1995 (ISO 9:1995). In addition to bibliographic and cataloging needs, the advent of the World Wide Web and the growing accessibility of information have expanded the application of transliteration also to the area of information retrieval (IR) and cross-language information retrieval (CLIR). This allows people to access information in languages that use different scripts from those they are familiar with (Chaudhary and Shekhar, 2023). As these authors affirm, Roman script is playing an increasingly important role in the field of transliteration and information accessibility. Other than this, it is an important aspect also in the field of MT, as will be further discussed in the following chapter.

## 1.2 TRANSLITERATION NORMS

Devoting further attention on the main translation technique that is investigated in this thesis, some norms providing guidelines that should be followed when transliterating from one language into another have been established. At academic level, the "scientific transliteration system" is usually taught (Cevese et al., 2018). Generally, it is used in this field to render names, surnames, toponyms, and realia. In Italy, guidelines that deal with transliteration of Cyrillic characters into Latin set by ISO are followed. As mentioned in the previous paragraph, the current standard is the ISO 9:1995 "Transliteration of Cyrillic characters into Latin characters. Slavic and non-Slavic Languages". Moreover, in March 2005 the standard has been translated into Italian and has been spread as UNI ISO 9 "Traslitterazione dei caratteri cirillici in caratteri latini: linguaggi slavi e non slavi" (UNI, 2005). One of the main advantages of the current standard, is the possibility to have a one-to-one correspondence between graphemes, thanks to the use of diacritics. Diacritical marks are introduced "when the number of characters used in the conversion system is smaller than the number of characters of the converted system" (ISO 9:1995)[2]. This idea was adopted by the Organization following the statements of Ševčik, who sustained the one-to-one correspondence and the use of diacritics (Gerych, 1965). However, the usage of diacritics when transliterating into Italian has been discussed for a while by scholars such as Šmurlo, Lo Gatto, Maver, Damiani, relying also on the knowledge provided by cataloging rules from 1921 until REICAT[3] (Mazzitelli, 2008). In fact, as mentioned in the historical section of this chapter, one of the first uses of transliteration was to address the problem of cataloging works by foreign authors in libraries. According to Damiani (1936), the main problems with the use of diacritics were linked to the difficulties for typographies in printing them and the habits to adapt Slavic names to phonetic rules of the target language. Therefore, to overcome the problem of various orthographic variants resulting from the adoption of different transliteration norms, cross-reference cards have been suggested, and even nowadays they are largely used (Mazzitelli, 2008). However, some discrepancies can be highlighted when considering ISO 9:1995 and widespread habits of transliteration in Italy that relies on Appendix 6 of RICA[4]. For example, the grapheme "щ" according to ISO 9:1995 is transliterated as "ŝ", whereas according to RICA is transliterated as "šč". In the second case, requirements established by the standard are not met because a digraph is being used. Nevertheless, as Mazzitelli and Garzaniti claim, it would be better to avoid marks as the circumflex accent ( ˆ ) that are unusual in Italian orthography[5]. Related to unusual orthography is also the problem of mispronunciation of transliterated words, which is likely to occur precisely due to unfamiliar spelling as a result of diacritical marks (Razran, 1959; Vlahov and Florin, 1980). Therefore, when transliterating, it is necessary to take into account that some adaptations may be required; TL rules tend to shape the rendering of the SW, which may

[2] The concepts of "converted system" and "conversion system" are introduced in the standard in the paragraph "General principles of conversion of writing systems" and they respectively refer to the given, original script and the different, target script.
[3] REICAT is the acronym for "Regole Italiane di Catalogazione".
[4] RICA is the former acronym for REICAT, that stood for "Regole Italiane di Catalogazione per Autori", replaced from 2009. (https://www.iccu.sbn.it/it/pagina/Commissione-RICA/)
[5] The debate did not find a final solution, even though as Mazzitelli claims in his work, it would be beneficial to move towards a common ground that takes into consideration both ISO and valuable academic traditions.

require phonological and graphical adaptations (Triberio, 2021), thus merging transliteration and transcription rules (Šimičević and Boljanović, 2009). The example suggested by Triberio (2021) is the one of the common Russian realia “матрёшка”. The author considers different monolingual Italian dictionaries, where the given results are the same (matrioska), except for Wikipedia that provides a double result (matrioska and matriosca), and Russian-Italian bilingual dictionaries, where the word appears as an adapted borrowing[6] “matrjoška” in the first, whereas in the second appears only in the RU-IT section and therefore is not encoded as a borrowing[7]. Phonological aspects need to be taken into account also when transliteration happens among different TL that have the same scripts. Due to how these different TL encode pronunciation, discrepancies amid languages and a resulting lack of homogeneity in transliteration norms can be noticed (Hsu et al., 2007; Šimičević and Boljanović, 2009). This is the reason why the surname “Ельцин” is transliterated as “Yeltsin” in English-speaking countries, but as “Eltsine” in French-speaking countries (Qizi et al., 2019), or “Горбачёв” becomes “Gorbachev”, “Gorbachov”, and “Gorbachyov” (Benites et al., 2020). Aspects as phonology and pronunciation are at the basis of national transliteration systems, which establish norms in specific countries; thus, they are considered unreliable in order to provide a homogeneous representation of SWs, being created to be suitable for national use and not international (Gerych, 1965). Therefore, finding an agreement and following norms that are given by institutions to provide a standardization may be helpful to avoid misunderstandings and an abundance of TWs, creating lack of homogeneity. Uniformity would be particularly appropriate mainly for those names, surnames, place names that are not certified, so as to prevent having different and ambiguous identifications for the same item, for example in different documents. Besides official documents, also bibliographic contexts would benefit from systematization, since usually titles and authors, for instance, need to be written rather than pronounced. Consequently, it would be preferable to prioritize character precision, rather than sound precision. Furthermore, for sake of accuracy, an internationally agreed-upon system would justify any phonetic compromises that might occur to achieve exact spelling reproduction (Gerych, 1965).

[6] As the author states in her work, by “borrowing” she means the results of the transliteration from SL to TL.

[7] Monolingual dictionaries: Treccani, Dizionario Italiano (Olivetti), Wikipedia; bilingual dictionaries: Dobrovol’skaja, Kovalev.

In this thesis two sets of norms are suggested, one provided by the Russian Ministry of Foreign Affairs and the second one by the UNGEGN Working Group on Romanization Systems. The former published in 2020 a decree regulating the issue of Russian passports, informing on the correspondence between Latin and Cyrillic characters. The table that can be found in the decree is considered when transliterating anthroponyms. This suggestion does not meet the principles behind the creation of ISO 9:1995, because the grapheme "щ", for instance, is rendered as "shch", thereby involving a quadrigraph. The latter is the division of United Nations Group of Experts on Geographical Names, whose objective is to suggest and agree on a romanization system for each non-Roman writing system[8]. In 2007, the "Technical reference manual for the standardization of geographical names" was published, containing various information on toponyms and their transfer and identifications, including tables for many languages with scripts different from the Latin script. For the Russian language, in 2016 a report on the current status of the system was released, which is based on the GOST 1983 system. This system involves letters with diacritics, rendering "щ" as "šč", for example. However, being a conventional system, it does not provide phonetic conformity of the names. In the following tables, the differences among graphemes are presented. Only the characters that vary across the four systems mentioned are shown, in order to provide an overview of the differences in rendering the same graphemes between the norms considered in this thesis and two of the most commonly used systems.

| CYRILLIC GRAPHEMES | SCIENTIFIC TRANSLITERATION | ISO 9:1995 | MINISTRY'S DECREE | UNGEGN REPORT |
|---|---|---|---|---|
| ё | ё | ё | e | ё |
| ж | ž | ž | zh | ž |
| й | j | j | i | j |
| х | ch | h | kh | h |
| ц | c/ts | c | ts | c |
| ч | č | č | ch | č |
| ш | š | š | sh | š |

[8] Romanization is used as a synonym for transliteration of Cyrillic script into Latin script; it is defined as "conversion from non-Roman into Roman script" (Glossary of Terms for the Standardization of Geographical Names, 2002).

| | | | | |
|---|---|---|---|---|
| щ | šč | ŝ | shch | šč |
| ъ | ” | ” | ie | ” |
| э | e | è | e | è |
| ю | ju | û | iu | ju |
| я | ja | â | ia | ja |

*Norms' comparison.*

## 1.3 TRANSLATION PROBLEMS: ANTHROPONYMS AND TOPONYMS

According to the definition of cultural realia by Vlahov and Florin (1980) provided in the introduction, it is possible to consider the terms “cultural word” and “cultural realia” as synonyms. Generally speaking, in both cases the reference is directed towards cultural elements shaping a certain culture. Therefore, it could be possible to conclude that these terms are used as “umbrella terms” under which some categorizations can be identified. For example, Vlahov and Florin themselves talk about geographic realia, ethnographic realia, and political and social realia. A similar classification was adopted by Newmark (1988), who identifies the following categories: proper names; geographical and topographical names that do not already have a recognized translation; names of periodicals and newspapers; titles of untranslated literary works, plays, films; names of private companies and institutions; names of public or nationalized institutions, unless they have recognized translations; street names, addresses. Another classification was suggested by Tomakhin in 1988 (Rasulova, 2020) who talked about geography, ethnography, folklore, mythology, everyday life, politics and society, and history. As it can be seen from these three categorizations mentioned, classifications overlap among different authors. Following these classifications, in this thesis focus is particularly devoted to anthroponyms and toponyms of Russian, as stated in the introduction. Whereas persons’ names do not require a particular classification and explanation, when it comes to toponyms, it is necessary to identify the categories included in these concepts. According to Gornostay and Skadiņa (2009), four categories make up the concept of toponyms: hydronyms, oronyms, geonyms and oeconyms. According to Adrian Room (Zgusta, 1998), in the field of names studies, Greek and Latin languages have had a great impact, as there are many signs of their presence in denominations and categories of names. In particular, even the denomination "study of names" comes from the Greek "onomastics", where "onoma" is the Greek word for "name". Therefore, the productivity and the inflection of this morpheme led to the creation of many derivations, such as anthroponym (name of a person) and toponym (name of a place), which are the individual names, and anthroponymy and toponymy, which refer to the range and study of such names. The present study narrows down the focus only to the last category of toponyms mentioned earlier, oeconyms[9], which involves among the others, city and regions name; in this category also denominations of inhabited locations are found (Gornostay and Skadiņa, 2009) and few of them were found in the passports considered.

[9] This term can be usually found in the variant of "econym".

# 2. MACHINE TRANSLATION

Moving towards the second part of this literature review, attention needs to be paid to machine translation (MT), with a further focus on machine transliteration, since it is a practice that can be used to improve MT performance (Oh et al., 2006). Firstly, a brief overview of the main points of MT history is provided; secondly, a short comparison of different types of MT is proposed to illustrate their evolution, leading to the following paragraph devoted to a more focused discussion on NMT and its functioning. Next, a paragraph on evaluation of MT has been included, to provide a comprehensive view of the various aspects that need to be considered when studying this tool. Moreover, special attention is given to machine transliteration, being it the main focus point of this thesis; besides explaining its functioning and development, some tools used to perform automated transliteration are explained. In addition, a section devoted to Large Language Models (LLMs) and their functionality is added towards the end of the chapter, with a focus on the engineering behind prompting. Given that the final part of this project work is conducted using ChatGPT, which is based on LLMs, it was considered to be valuable to include an overview on how these models operate. Finally, although not utilized in this thesis, a brief paragraph on CAT tools is included, so as to provide an initial overview of other technologies that support translators' work.

## 2.1 HISTORY

The history of machine translation has experienced various fluctuations in its research and study, shaped by the historical context in which it initially developed. The concept of machine translation was first born in 1947 with Warren Weaver and from that time it has been long considered and studied in the domain of natural language processing (NLP). Ever since MT was born, certain principles have been established and remain valid today, shaping the development of MT and its functionalities. An example of these principles is decoding foreign languages, which can be regarded as a direct development from code breaking practices; in particular, the crack of the German Enigma code in World War Two can be considered the very first example of decoding language codes. Currently, governments like that of the USA provide substantial funding to support code-breaking efforts for languages of nations deemed threats to national security (Koehn, 2020). Initially, there was great optimism in the research field of MT, also brought by the Georgetown experiment carried out in 1954 that suggested a well-functioning Russian-English MT system. However, in 1966, MT funding was abruptly stopped due to the ALPAC report, which claimed that MT systems were inadequate and did not fulfill the promises of being cost-effective or time-efficient. At the beginning, rule-based methods that relied on bilingual dictionaries and manually written rules to translate ST into TT were used to train machines. On this basis, around the late 1960s to 1970s, the first commercial systems that relied on rule-based machine translation (RBMT) started to be released. Examples include SYSTRAN in 1968, which began with the Russian-English language pair and later expanded to other European languages, as well as Logos and Metal (Koehn, 2020; Wang et al., 2021). However, the growing availability of bilingual and parallel corpora, along with the idea that it is better to rely on previous translations[10], resulted in a shift towards a new methodology: corpus-based methods, which became predominant after the 2000s. In particular, through the years three corpus-based methods have been developed: example-based machine translation (EBMT), statistical machine translation (SMT) and neural machine translation (NMT). The first one functioned well only when similar sentence pairs were present in the corpora. Despite that, the corpora's inability to cover all linguistic phenomena belonging to a language posed a great limitation to this approach. Thus, in 1990 SMT was introduced. According to Brown et al. (1990), SMT involves machines automatically acquiring translation knowledge from a vast amount of data, and not relying on humans to write rules. However, SMT started to be adopted more frequently only after some implementations, leading to the establishment of new companies, for example Language Weaver in 2002, and major software companies, such as Microsoft and Google, that developed MT systems based on implemented SMT. In these years, SMT became widely adopted and popular also for increasing translators' productivity. Even though with this method the quality of the outcomes started to improve, further studies were needed. In fact, around the mid-2010s, also thanks to fundings by EU, USA and China researchers deepened their studies in deep learning and applied deep learning technologies to MT, until the third above-mentioned method was created. At first, this method was developed as the integration of neural

[10] This idea gave rise to the concept of TM, which is a fundamental feature of CAT tools.

language models in SMT, but further studies were aiming at pure NMT. Accordingly, Sutskever et al. in 2014 and Bahdanau et al. in 2014 suggested neural network translation models that operate end-to-end and officially introduced the acronym NMT. The general idea relies on the possibility "to map the source language into a dense semantic representation, and then generate the translation by using an attention mechanism" (Wang et al., 2021). In particular, the attention mechanism developed by Bahdanau et al. in 2014 allows an increasingly better performance that can provide appropriate outcomes also for longer sentences. In fact, one of the main criticisms of convolutional or sequence-to-sequence models, that were the first developments for pure NMT, was related to their effectiveness limited to shorter sentences.

While SMT was advancing in the 1990s, the development of CAT tools also progressed, stimulated by the spreading of desktop computer systems. For the past decades, MT has been integrated into CAT tools to enhance translators' productivity. A notable example of this phenomenon is the acquisition of Language Weaver by SDL, the developer of Trados, one of the main translation tools on the market (Koehn, 2020). In this thesis, MT will not be integrated in CAT tools, as the focus is on analyzing stand-alone MT engines. However, a brief, dedicated section will be included, to provide an overview on how MT and CAT tools together can further support translators' work.

## 2.2 RBMT, EBMT, STM: A COMPARISON

In order to show the rationale behind the development of various MT approaches over history, as mentioned in the previous paragraph, a general overview on the architectures developed prior to the advent of NMT is provided.

As stated above, until the late 1980s, RBMT dominated the field of machine translation. However, since that time, corpus-based approaches have become more prevalent, with a primary distinction existing between EBMT and SMT. RBMT operates on predefined linguistic rules and bilingual dictionaries to transform SL structures into TL equivalents. According to Wang and Sawyer (2023), the process can be divided into three steps. It begins with an analysis phase where the SL input is parsed and understood according to grammatical rules. The analysis can include various actions, such as POS tagging, morphological, semantic and syntactic analysis. Secondly, the transfer step through algorithms creates representations of rules and lexical items gained from the analysis, which become useful in the synthesis phase, generating the TL output based on these representations. While high accuracy for language pairs with structured grammar can be achieved, the main downside of this approach is the dependance on extensive manual rules to develop and implement the architecture, which consequently restricts its ability to adapt to new languages and new domains (Khenglawt, 2018; Ganesh et al., 2023). Moreover, encoding all linguistic rules into computer programs is very time-consuming and makes the system hard to maintain. The appearance of data-driven approaches, rather than rule-driven, shifted the focus towards a paradigm relying on bilingual corpora. EBMT is said to be translating through analogy (Hutchins, 2005), revolving around the extraction and combination of phrases or short segments of texts, and not producing a translation from scratch. Broadly speaking, EBMT decomposes input sentences into fragments and then searches a database of previously translated sentences for segments that can match the ones considered. Once the lookup is complete, the identified segments are recombined by the EBMT, so to create the output translation that conveys the same meaning of the input sentence. Even though this model enhances the accuracy, especially for less formal and domain-specific language constructs, a drawback relates to the insufficient coverage of examples including all phenomena and phrases that may be considered for translation. To address the demands for flexibility in handling diverse language pairs, domains, and varying sentence structures, SMT approaches started to be adopted. This architecture relies predominantly on word frequency and combinations (Hutchins, 2005). The approach involves aligning sentences from bilingual corpora and subsequently aligning individual words between ST and TT to establish correspondences. Based on these alignments, SMT constructs a translation model that includes SL-TL frequencies and employs a language model that arranges TL word sequences. More precisely, according to Somers (Banitz, 2020) the translation model calculates the probability of each word in the SL being translated into possible words in the TL and selects the most likely translation from the bilingual corpus. Consequently, the language model ensures that the chosen words in the TL form a correct and natural TS. The reliance on statistical data may pose some challenges when it comes to facing translations of less common words and phrases, or specific language structures. The core distinction between these two primary corpus-based approaches lies in their basic units and methodologies: SMT, which operates at word level, primarily employs statistical methods, while EBMT, which works with phrases as its basic units, relies more on linguistic fragments and text examples. In order to transform the approach to translation, NMT models do not rely on predefined rules or statistical alignments of individual words. Rather, TTs are produced by learning and predicting translations at the level of entire sentences. Being NMT the new state-of-art approach (Koehn, 2020), the following paragraph will be fully devoted to a further and broader explanation of the architecture at its basis.

## 2.3 NEURAL MACHINE TRANSLATION

Neural Machine Translation represents a significant advancement in the field of MT, using neural networks to improve translation quality and efficiency. This section explores the working of NMT, focusing on its architecture and key components.

To understand the architecture and the functioning of NMT, an important concept needs to be highlighted: NMT is based on mathematical systems, known as artificial neural networks (ANNs), that transform sentences into numerical representations and perform mathematical operations to generate translations (Pérez-Ortiz et al., 2022). They are usually compared to the functioning of biological neurons. Biological neurons receive signals, process them, and transmit the output. Similarly, ANNs combine inputs, activate, and generate outputs (Koehn, 2020). These networks are organized in layers, where each layer processes data sequentially to produce the final output. The early models of neural networks are known as perceptrons and were created by Rosenblatt in 1958 (Forcada, 2017). They consisted of a single processing layer with neurons having binary inputs and outputs. However, perceptrons were limited in their ability to perform complex operations, leading to a pause in research. The resurgence in neural network research, particularly with the advent of multilayer networks, paved the way for more sophisticated models like Recurrent Neural Networks (RNNs) and the eventual development of deep learning techniques (Koehn, 2020). Deep learning techniques are named for the depth of layers in NNs, specifically the number of layers between the input and output layers. According to the IBM website, if there is more than one hidden layer (the layers between the input and output layers), the network is considered a deep neural network. This means that the system has more than the usual three layers, forming a more complex architecture. Each layer is made of hundreds of neural units, also called "nodes", or simply "neurons" that are interconnected through weights in one layer, to all neural units in the next layer. Consequently, when nodes are activated, they are not activated by themselves, as it would not be useful; the activation of individual neurons is grouped with the activations of the other neurons, so to form distributed representations of words and their context, that is what allows to have actual translation outputs. Both the context of the source sentence processed and of the target sentence produced is taken into consideration (Forcada, 2017).

A typical NMT model comprises two main components: the encoder network and the decoder network (Cho et al., 2014). However, prior to these two steps, the system needs to be trained. NMT falls under the category of corpus-based MT (also known as data-driven) and therefore is trained on large parallel corpora. The corpora provide reference translations, which are considered when creating the output at the end of the decoding step. To ensure that the output closely resembles the reference translations, a training algorithm is used to modify and determine the strength of the connections between neurons. The aim is to reduce the error function to the lowest possible levels; therefore, the process is repeated until this goal is achieved (Pérez-Ortiz et al., 2022). When it comes to the actual encoding-decoding process, adopting the model built on RNN, the encoder network maps the source sentence into a real-valued vector. This vector representation captures the semantic information of the source sentence, compressing it into a format that the decoder can use to generate the translation. This process is done recursively, with part of the output calculated in one step influencing the next step, as the model's name suggests. This continues until the last unit is encoded, meaning that the last hidden state contains the representation of the ST and is ready for processing by the decoder. Improved RNNs work with gating structures that allow NNs to remember or forget past inputs. This means that a neuron can either stimulate or inhibit the next neuron, determining whether information is transmitted or retained. This process involves the use of a threshold: if a neuron's activation exceeds this threshold, it stimulates the next neuron; if not, the neuron is inhibited, and information is not transmitted. Two of the most common gating structures are: long-short term memory (LSTM) and gated recurrent units (GRU) (Forcada, 2017, Wang et al., 2021). In order to clarify how this process works, an explanatory example taken from Forcada (2017) is provided below. The language pair considered is English-Spanish and the source sentence is "My flight is delayed.". In the encoding process, firstly all the units making up the sentence are identified and represented as embeddings: e('my'), e('flight'), e('is'), e('delayed'), e('.'). Secondly, the encoder processes the embeddings sequentially, updating the above-mentioned hidden states with each word, until the ST is fully encoded:

1. encoder combines preexisting encoding for the empty sentence E('  ') + embedding of first word e('my')⇛ production of the encoding E('My')
2. encoder combines representation of E('My') + embedding of e('flight')⇛ production of the encoding E('My flight')
3. E('My flight') + e('is')⇛ E('My flight is')
4. E('My flight is') + e('delayed')⇛ E('My flight is delayed')
5. E('My flight is delayed') + e('.')⇛ E('My flight is delayed.')

The concept of embeddings, first introduced by Mikolov et al. in 2013 (Pérez-Ortiz et al., 2022) is useful and brings many advantages when NMT encounters new sentences. Words with analogous meaning or that usually co-occur within the same context are assigned similar embeddings, meaning that their representations have similar dimensions[11]. Therefore, representing semantically related words through similar numerical vectors, facilitates the network in translating novel texts.

The decoder network now generates the translation from the real-valued vector produced by the encoder, following a probability algorithm.

1. initially, the decoder takes into consideration the encoding of the ST and an empty sequence of target words (TW) represented by ' ', and creates two vectors
   1.1. initial decoder state: D('My flight is delayed', ' ')
   1.2. vector of probabilities for all possible words *x* in the first position of the TT: p(xl'My flight is delayed', ' ')

   ⇛ decoder assigns the maximum likelihood to the Spanish word 'Mi' for the source word (SW) 'My', which becomes the output⇛ x='Mi'

[11] Pérez-Ortiz et al. (2022) in their work present a little exercise that allows the reader, who is supposedly less familiar with engineering architecture, to better understand the rationale behind assigning coordinates to words. However, they also point out that the coordinates representing the words are not just two numbers as it happened in the exercise, having considered a bidimensional space. In NMT, embeddings have hundreds of dimensions.

2. decoder reads D(‘My flight is delayed’, ‘ ’) + word ‘Mi’ and creates two vectors
    2.1. next decoder state D(‘My flight is delayed’, ‘Mi’)
    2.2. vector of probabilities for all possible target words *x* in the second position of the sentence: p(xl’My flight is delayed’, ‘Mi’)

    ⇛ decoder assigns the maximum likelihood to the Spanish word ‘vuelo’ for the SW ‘flight’, which becomes the output⇛ x=‘vuelo’
3. D(‘My flight is delayed’, ‘Mi’) + word ‘vuelo’ and creates two vectors
    3.1. D(‘My flight is delayed’, ‘Mi vuelo’)
    3.2. p(xl’My flight is delayed’, ‘Mi vuelo’)

    ⇛ x=‘lleva’, to replace ‘is’
4. D(‘My flight is delayed’, ‘Mi vuelo’) + ‘lleva’
    4.1. D(‘My flight is delayed’, ‘Mi vuelo lleva’)
    4.2. p(xl’My flight is delayed’, ‘Mi vuelo lleva’)

    ⇛ x=retraso, to replace ‘delayed’
5. D(‘My flight is delayed’, ‘Mi vuelo lleva’) + ‘retraso’
    5.1. D(‘My flight is delayed’, ‘Mi vuelo lleva retraso’)
    5.2. p(xl’My flight is delayed’, ‘Mi vuelo lleva retraso’)

    ⇛ x=., to replace ‘.’
6. ‘Mi vuelo lleva retraso.’

NMT models, particularly those based on RNNs, face challenges with long sentences as information tends to be lost during transmission. Besides gating structures mentioned above, other enhancements have been proposed to address the issue of long sentences and the possible loss of information during transmission. An improvement suggested by Bahdanau et al. in 2014 is the attention mechanism that works through an additional set of neural connections and layers. The attention mechanism improves translation quality relying not only on the last hidden state produced by the encoder, i.e. E(‘My flight is delayed.’), but also on the whole sequence of representations built during the encoding E(‘My), E(‘My flight’), E(‘My flight is’), etc. Based on the attention mechanism, in 2017 Vaswani and other colleagues published an article on a model they implemented, which nowadays became fundamental in the domain of NLP (Amatriain et al., 2024). These scholars suggested the Transformer, as a model relying entirely on the attention mechanism to further enhance and speed up machine translation performance (Vaswani, 2017), but which could also be implemented in various machine learning applications (Peng and Sawyer, 2023). More specifically, a self-attention model is utilized, which allows to look and process words within sentences simultaneously, and not processing words one-by-one. Prior to being utilized in MT, self-attention was also employed in other fields, such as reading comprehension and summarization (Vaswani, 2017). Another improvement suggested is the bidirectional encoding (Wang et al. 2021). Traditional encoders usually process ST in a unidirectional manner from left to right, following the typical reading direction of European languages. On the other hand, to provide more context that helps improving probability calculations and therefore translation quality, bidirectional encoding has been suggested, so to read the source sentence both from left to right and right to left[12]. Despite being an absolute breakthrough in the field of machine translation, NMT requires high computational resources and extensive knowledge that could pose some challenges in terms of training, and improvement of its architecture, especially for users such as translators, who presumably do not have those necessary skills (Forcada, 2017). Comparing the functioning of NMT to human translators, it is analogous in that information and translation knowledge are first read and understood and then processed to generate a TT. Human translators follow the principle of semantic compositionality when approaching a new sentence to be translated. Starting from the comprehension of single words in the SL, they build the meaning of the whole sentence, and later form the target sentence with the corresponding meaning. Similarly, during the encoding phase representations of words are created and are combined into a sentence. The decoder then predicts each word from the previous representations, so to form the TT (Pérez-Ortiz et al., 2022). An interesting aspect of this modern architecture is its relation to the older concept of interlingua, introduced by Vauquois in 1968. Based on the theorization of Jurafsky and Martin, as discussed in Wang and Sawyer's book (2023), the encoder-decoder network can be viewed as an interlingua approach. Vauquois suggested the concept of universal interlingua representation of a SL, embodying an ideal level of understanding that is independent of any specific linguistic

[12] For the purpose of this thesis, improvements in NMT systems are briefly mentioned and explained, as an extensive explanation would shift the perspective to engineering, which is beyond the main scope of this thesis.

forms. In the field of MT, the systems are formed by an analyzer parsing the SL and a generator for each TL that transfer the interlingua representations between languages. Thus, the interlingua serves at an intermediate level. The comparison with NMT is clear as the encoder processes input sentences and creates their representations, and the decoder generates specific output sentences. Vauquois illustrated the concept of interlingua within the Vauquois triangle, so to show the increasing depth of analysis required when approaching a text to be translated, and at the same time a decreasing dependance on the language form. Starting from a direct approach, where almost all knowledge is found at word level, it decreases to the transfer approach and to the subsequent interlingua approach, where there is no specific transfer of knowledge.

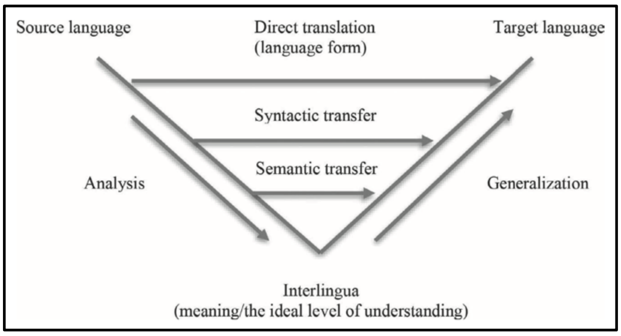



*Figure 1. Vauquois triangle. Taken from Wang and Sawyer (2023), adapted from Vauquois (1968).*

## 2.4 QUALITY EVALUATION

As stated in the introduction, one of the practical parts of this thesis is comparing and evaluating the accuracy of the different MT taken into consideration. The main metric used for evaluating the performance of MT tools considered in this thesis involves a comparison between the output given by the machine and a reference translation performed using established norms. This comparison will be done manually but resembles one of the methodologies adopted by an AEM, which is explained in this chapter. Therefore, a general overview with a brief explanation of various MT quality evaluation metrics and an example that provides some data to justify the shift towards new MT systems are considered to be relevant for this paper, so as to give a broader and more complete view on the topic.

The growing demand for translations and the necessity to overcome language barriers across different cultures have led human translators to increasingly rely on MT. Machine translation evaluation is a topic that has been discussed ever since the 1990s (Frederking and Nirenburg, 1994; Knight and Chander, 1994; King, 1996), proposing metrics and approaches to assess the performance and the outputs of this tool, with improvements over the years to match the development of different MT systems (Rei et al., 2020). Their goal is to use evaluation metrics to determine which systems translators should use to increase productivity and enhance their work (Rivera-Trigueros, 2021). When evaluating the output of a machine translation it is important to consider the tailoring of the engine used, as different domains and text types have different approaches and requirements when translating (Rossi and Carré, 2022). Evaluation metrics are divided in two major categories: automated evaluation metrics (AEMs) and human evaluations. In literature, scholars have been trying to understand which metrics perform better when assessing quality thus coming to the general idea that combining both methods would provide more accurate results (Chatzikoumi, 2019; Rivera-Trigueros, 2021; Rossi and Carré, 2022). In fact, to compensate for the potential high-level of subjectivity that might arise from human evaluation alone, which would not allow for a generalized conclusion on which systems perform better, automatic algorithm-based evaluations are also employed. They are less time-consuming and faster. However, an advantage of human evaluations is that they can be objective when using productivity measures (Rossi and Carré, 2022). The principal parameters considered in human evaluation metrics are adequacy and fluency: the former considering how much of the meaning of the ST is expressed in the TT[13], the latter considering the adherence of the output to grammar rules and norms of TL. Evaluations for both parameters are assessed on a four-point scale, where 1 stands for no meaning transferred to the output and/or lack of fluency, and 4 stands for complete transfer of meaning and/or complete adherence to the rules. Being a quite time-consuming method, a faster comparison can be carried out by simply ranking outputs, without specifying the reasons why such a decision is taken. On the contrary, a much greater effort is required to identify and classify errors found in

[13] Even though in this explanation the word "text" is used (both alone and in the acronym for "source text" (ST) and "target text" (TT)), in this case it does not refer to a whole, single text considered; rather, most of the times these evaluations are carried out sentence-by-sentence (Rossi and Carré, 2022).

TTs, which prevents them from simply assigning a score to the machine-translated output. On the other hand, as previously mentioned, AEMs enable a faster assessment of MT quality. By using algorithms and engine training, AEMs allow for frequent testing to evaluate MT quality as often as needed. They work following the principle of similarity; thus, they consider reference translations performed by human translators and candidate translations created by MT, comparing and computing similarities and discrepancies, following a series of parameters. Some of these metrics, such as BLEU (Papineni, et al., 2002), METEOR (Banerjee and Lavie, 2005) and TER (Snover et al., 2006), have been known and used for a long time. Others, like COMET (Rei et al., 2020), have been newly created and adopted to address the evolving field of NMT. According to Papineni et al. (2002), BLEU (Bilingual Evaluation Understudy) score evaluates the quality of candidate machine translation outputs by comparing them to human reference translations, in terms of n-word overlap. It employs modified n-gram precision to consider multiple reference translations and legitimate variations in word choice and order. The scores range from 0 to 1, with higher scores indicating a closer match to the reference translation. The chances of getting higher scores increase with a higher number of reference translations. On the other hand, as claimed by Banerjee and Lavie (2005), METEOR metric was created to better match humans' judgment on translation quality, compared to older metrics such as BLEU. It focuses on the two main parameters presented above as the parameters of human evaluations: accuracy and frequency. To calculate the scores, this approach is based on recall (i.e., how many correct words are found in candidate translation compared to the number of words in the reference translation) and precision (i.e., how many correct words are found in the candidate translation compared to the total number of words in candidate translation), penalizing the translations that do not keep the same word order as the original text. Finally, the last of the older metrics mentioned is the TER approach, investigated by Snover et al. in 2006. This approach takes into consideration how many changes are necessary to transform the MT output into a reference translation. These changes vary from insertions and omissions to word order rearrangements, leading to a quite straightforward method. In fact, changes are calculated by dynamic programming that finds the minimum number of edits required between the candidate and the reference translation. However, there are some limitations mainly related to the semantics of candidate translation, because it does not take into account whether the meaning has been correctly reproduced. Therefore, HTER is introduced, which considers human annotators who have to create reference translations matching closely the meaning and fluency of the MT output. In comparison with the previous two evaluation metrics, in this case measuring the editing needed, a lower score indicates that the tool performed better, while a higher score means that much more edits are necessary. A downside of HTER approach lies in its high resource costs, thus not completely aligning with the initial idea of AEMs being less time-consuming than human evaluations.

A concrete example that used these methods to perform quality evaluation of three different MT systems was presented by Stasimioti et al. (2020) and is here mentioned to provide some concrete knowledge that can be helpful in better understanding how these tests work. The authors take into consideration four semi-specialized articles taken from the newspaper The Guardian about 2019 EU elections, that are processed by three MT tools: SMT, NMT and a tailored-NMT that is specifically focusing on the language pair of this research, English-Greek. In general, the conclusion drawn from their study is that NMT systems perform better than SMT, according to both human and automatic evaluations. In particular, tailored-NMT is even better than generic NMT, as it obtained higher scores in any of the parameters considered[14]. For example, looking at adequacy and fluency, SMT scores are 3.49 and 2.96, respectively; in NMT, 4.01 and 3.73; in tailored-NMT, 4.26 and 3.96. Looking at automatic means, BLEU results in 0.34, 0.39 and 0.46 for SMT, NMT, and tailored-NMT respectively; METEOR results in 0.48, 0.52 and 0.56; TER results in 0.52, 0.51 and 0.39. Their findings are in line with results presented by other scholars who researched this field. They mentioned authors like Toral and Sánchez-Cartagena (2017); Klubička et al. (2017, 2018); Jia et al. (2019); Koponen et al. (2019) who investigated various language pairs, concluding that NMT systems outperform previous MT engines.

## 2.5 MACHINE TRANSLITERATION

In the context of MT, the performance of this tool can be enhanced by machine transliteration, mainly when it is necessary to deal with proper nouns and terms, so as to

[14] For a more detailed explanation of the parameters and the study conducted, refer to the study cited.

prevent possible errors. Some research has been carried out in this field both to investigate how machine transliteration works, and which are some of the tools used to carry out this task. In this section an initial overview of machine transliteration functionality will be provided, followed by a paragraph devoted to some examples of tools created.

Researchers such as Knight and Graehl (1997) and Al-Onaizan and Knight (2002), Oh et al. (2006) have conducted studies that put in relation MT and machine transliteration, which can be defined as "a method for automatically converting words in one language into phonetically equivalent ones in another language" (Oh et al., 2006). Machine transliteration is usually investigated following two distinct paths: generative transliteration, that works with algorithms to transliterate new terms, and transliteration extraction, that extracts transliterated words from existing corpora (Karimi et al., 2011). As Karimi and his colleagues state, under the category of generative transliteration, some methods can be identified according to attributes influencing transliteration tasks. Direction of transliteration and scripts of different languages are important, but higher focus is posed on a categorization that relies on units taken into consideration for the transliteration task. These units build four different models according to which machine transliteration work: phoneme-based model, grapheme-based model, hybrid model and correspondence-based model (Oh et al., 2006; Karimi et al., 2011; Kaur and Singh, 2014; Prabhakar and Pal, 2018). The phoneme-based model was initially adopted in early studies on transliteration, where speech recognition was used as a phonetic representation. This approach was based on the insight that phonetic representation, being a common feature across all languages, could serve as intermediate between SL and TL (Karimi et al., 2011). Exactly for this reason, this method was also called "pivot method", because the transformation started from SL graphemes, converted in source language phonemes, converted in target language graphemes (Oh et al., 2006). However, because of these many steps, a great disadvantage identified in phoneme-based methods is the facility in making errors and their consequent propagation (Lee, 1999; Oh et al., 2006; Karimi et al., 2011). As showed in Oh and his colleagues' research, Lee (1999) conducted a study between English and Korean, converting English graphemes to English phonemes first, step in which many mistakes were made, and later English phonemes were converted in Korean graphemes according to EKSCR[15]. However, this study did not result in a good

[15] EKSCR: acronym that stands for English-to-Korean standard conversion rules.

transliteration performance, due to propagated errors and limitations in the conversion rules. Other works carried out also combined different methods in generative transliteration, in the framework of phoneme-based model, which was combined with back transliteration (Knight and Graehl, 1997, 1998 for Japanese and English; Stalls and Knight, 1998 for Arabic and English). Further studies shifted their attention to the grapheme-based model, which started to become more popular (also known as "spelling-based method" or direct method according to Oh et al., 2006; Karimi et al., 2011; Kaur and Singh, 2014). It is considered to be more reliable than the phoneme-based method by the researchers above-mentioned, because it involves fewer steps when converting SL grapheme to TL grapheme. Nevertheless, in some models based on this approach some mistakes can still be made and propagated, such as in the source-channel model. This method involves segmenting SL words into graphemes, producing all possible TL graphemes corresponding to the SL ones and finally identifying relevant sequences of TL graphemes with the model. Other transliteration methods that use this grapheme-based model and have been investigated are decision tree and joint source-channel model[16]. Being based on graphemes, the model does not pay particular attention to pronunciation and phonetic aspects of words (Karimi et al., 2011; Kaur and Singh, 2014). In fact, a disadvantage may arise for languages in which the pronunciation of words differs significantly from the written form. For example, in English, the place name "Edinburgh" is pronounced /ˈedənbərə/[17] and in this case the "gh" sound is highly different from its normal pronunciation (Karimi et al., 2011). To overcome the disadvantages identified in both previous models, researchers thought about merging them to gather their advantages in the hybrid and correspondence-base models. While hybrid models work with probability of grapheme-based and phoneme-based models, correspondence-based one can be divided in glass-box combination and black-box combination, the difference of which relies on whether the combination happens in the internal functions or on the final output (Karimi et al., 2011). The basic idea, however, is that it "makes use of the correspondence between a source grapheme and a source phoneme when it produces

[16] Lee & Choi (1998); Lee (1999); Jeong et al. (1999); Kim et al. (1999) conducted more exhaustive studies with source-channel model; Kang & Choi, 2000; Kang, 2001 conducted more exhaustive studies with decision tree; Li et al., 2004 2001 conducted more exhaustive studies with joint source-channel model, which are not explained here to respect the purpose of this thesis.
[17] Longman online dictionary (https://www.ldoceonline.com/dictionary/edinburgh).

target language graphemes" (Oh et al., 2006). The correspondence method is still developing but one of the studies conducted is the one by Oh and Isahara (2007a), who considered two language pairs, English-Japanese and English-Korean, and based their research on support vector machine (SVM) and maximum entropy models (MEM). To evaluate the final outcome, they used two corpora by means of which they found an accuracy of 87.8% for English-Japanese and of 88.2% for English-Korean with SVM tool[18]. Even though much research focused on single models, as demonstrated by Oh and his colleagues in their work published in 2006 combining the four models together would provide better results for correct transliteration. As shown from the examples mentioned above, much work in this field has been conducted taking into consideration English and Asian languages (Oh et al., 2006, Karimi et al., 2011). A general overview of major studies is presented in Karimi et al. (2011), where the abundance of Asiatic languages paired with English is clear.

### 2.5.1 MACHINE TRANSLITERATION TOOLS

Some research about tools that can perform automated transliteration has been carried out in recent years. Being machine transliteration a topic that merges language and engineering, when it comes to investigating and building machine transliteration tools, research is clearly filled with mathematics, engineering, computational writings that make articles difficult to understand if a background knowledge of these topics is not known. Some research about these developed tools will be presented, so as to introduce the practical use of generative transliteration. To present knowledge, little research has been carried out between Russian and Italian; here just one example will be later presented. The above-mentioned study with SVM is relevant to this thesis because this method was also used by Azid et al. (2017) in transliteration from Russian Cyrillic alphabet to Latin alphabet in Indonesia. Researchers created the application "Capture to Translate" that works through SVM algorithms, using image processing and AI. Firstly, an image with Russian text is processed, secondly the systems that had been previously trained with SVM algorithm runs the automatic transliteration. Researchers found an accuracy of 93.8% (Azid et al., 2017). Another tool built is TRANSLIT, which is a dataset

[18] Further and more detailed explanations are not explored here for the sake of this thesis but can be found in their article.

for name transliteration (Benites et al., 2020). In this new dataset, already existing corpora and a new one were merged together to have a large-scale corpus that would help training automated transliteration. Training is necessary because ISO rules may not be followed in automatic transliteration and therefore huge variation may happen (Benites et al., 2020). The accuracy achieved was 92%. One last example may be relevant for this study because it involves Russian and Italian language. Anselma et al. (2009) thought about a method accessible also to non-expert in IT that deals with spreadsheets. They based their work on the ISO 9:1995, following the one-to-one correspondence between characters. This method involves using the "find" and "replace" commands and can be easy to adapt to operative systems. In the article they also provide further explanation on how to deal with capital letters, considering whether only the first one needs to be a capital letter, or all the letters are capitalized. However, this research would need further investigations so to pair this tool with word processing programs and an implementation in JavaScript language so to use it online.

Using these tools involves copying and pasting transliterated words and terms, rather than finding them as an output of a whole translation process. On the contrary, using MT to carry out translation, and transliteration of those terms that require it, would allow translators to use only one tool to perform both. Moreover, MT, differently from machine transliteration tools presented here, can be integrated in CAT tools, so to avoid the process of copying and pasting, and translators can also take advantage of the opportunities offered by such tools. In a study conducted by Grundkiewicz and Heafield and presented to the Named Entities Workshop in 2018 a transliteration system based on deep attentional RNN encoder-decoder model was proposed. The scholars applied NMT techniques, that are regularization, model ensembling, re-scoring with right-left models and back-translation[19] to investigate transliteration of personal and place names in language pairs that involved English and nine other Asiatic languages. The primary evaluation metric they used was word accuracy, calculated using a mathematical formula. The results they obtained allowed them to conclude that their approach can be used to develop effective machine transliteration systems for various language pairs, reaching cutting-edge results.

[19] In this thesis NMTs are only mentioned, but further explanation can be found in the article cited.

In the growing field of literature investigating machine transliteration, different approaches and models have been developed to deal with this technique. One of the most recent comprehensive reviews of these systems was conducted by Chaudhary and Shekhar in 2023. In addition to discussing the four approaches outlined in the previous paragraphs and referencing several scholars who investigated this field, they also highlighted some limitations that machine transliteration tools may encounter. Specifically, their work presents a detailed ten-point classification of challenges observed across different models and approaches analyzed. This classification is briefly presented here to provide a general overview of some aspects that might be considered and implemented in future tools. Most of the time transliteration models are tailored for specific languages, limiting their adaptability to other sets of languages. Furthermore, since training requires a high amount of data, systems are less properly trained for low-resource languages, or they even tend to be excluded, resulting in more mistransliterations and inaccuracies. The models are usually trained on specific vocabulary, which can lead to difficulties in processing "out-of-vocabulary" (OOV) words, those that are new or uncommon in STs. Additionally, texts that are highly specialized and contain jargon and technical terminology may result in inaccurate TTs, as the systems may struggle to process and apply the appropriate rules required for that specific context. Moreover, machine transliteration tools are designed to map sounds, which might result in a lack of contextual understanding and a subsequent misinterpretation of word meanings. This challenge is evident also when dealing with homographs, that are words written with identical spelling but having different meanings. If the system is not able to distinguish and recognize the correct meaning, it might again result in inaccurate transliteration. Despite being designed to deal also with phonetics, the authors note that these systems may also encounter some difficulties in mapping and processing sounds of languages having different phonetic structures, particularly when a TL lacks a direct equivalent for a sound in the SL. The tools also need to be adequately trained to identify and manage possible different scripts within a single text, especially when the text is characterized by code-mixing or a mixture of languages. Moreover, the systems also need to be able to account for national transliteration systems or users' preferences. Lastly, while there is extensive literature about MT quality evaluation, when it comes to machine transliteration a widely accepted evaluation metrics is missing. Developing and training adequate tools for machine transliteration is particularly challenging and demanding, as substantial data, time, and resources are needed.

## 2.6 LARGE LANGUAGE MODELS

Despite being central to the topic of MT, the previously discussed Artificial Neural Networks are also key in the broader field of Artificial Intelligence (AI). More specifically, a variant of the traditional AI, known as the Generative Artificial Intelligence (GAI), is capable of generating human-like content. ANNs are at the basis of Large Language Models (LLMs), which can be considered a special type of GAI, able to generate human-like natural language texts in response to natural language prompts (Bridgelall, 2024). Thus, they are designed to carry out a wide range of NLP tasks, such as machine translation, text summarization, or conversational interactions (Naveed et al., 2024). According to Naveed and his colleagues, to achieve present day status of NLP, over history a series of evolutions and improvements occurred. Initially, the architectures shifted from statistical to neural language modeling, which later evolved in pre-trained language models (PLMs) and the current LLMs. The transition from PLMs to LLMs is explained by the broader amount of training data and parameters required to obtain better results. The main difference between traditional language modeling and advanced models, lies in the need to have systems that are trained to be used in a variety of NLP tasks. Therefore, the models also undergo fine-tuning, to enhance their performance. This is particularly relevant for first models released in recent years, for which the initial transfer learning was insufficient. On the other hand, more recent systems like GPT-3 were supposed to function properly and align with users' intent simply by providing instructions and examples, a technique known as "zero-shot learning". However, despite this initial design, it was still considered necessary to include additional steps that could better understand users' needs and improve the generalization to new tasks (Naveed et al., 2024).

At the core of Generative AI, and consequently of LLMs, is the Transformer model, developed by Vaswani and his colleagues (Amatrian et al., 2024). In their work, Vaswani and his colleagues accurately describe the Transformer model, including mathematical expressions and functions, schemes and explanations, to provide an accurate representation of the model they implemented. However, in this thesis, only the aspects considered relevant to this study will be briefly presented. Specifically, attention is particularly devoted to the Transformer, as it represents the main novelty from the previously presented architecture of NMT. While the encoder-decoder architecture is also used in LLMs, it is implemented using the Transformer (Vaswani et al. 2017; Naveed et al, 2024; Amatrian et al., 2014). The Transformer has been a revolutionary development, as it can highly enhance the outputs of the machines. It operates using self-attention, relying solely on the attention mechanism (Vaswani et al., 2017), which was originally proposed by Bahdanau and his colleagues in 2014 and is previously introduced in the section about NMT. The attention function considers a query and some related information, maps and converts them into vectors, so that they can be processed. The final aim is to identify the most relevant information that will be found in the output. To achieve this, the function has to check the compatibility between the question and the piece of information identified and uses this match to determine how much each piece of information should influence the final answer. However, the Transformer model relies on the advanced Multi-Head Attention, which involves several attention layers working independently but simultaneously to compute the representations for each token, and then gathering the results for the final representation (Vaswani et al. 2017; Amatrian et al., 2014). There are three main ways in which the Transformer uses the Multi-Head Attention: besides the "encoder-decoder attention" setup that follows the mechanism previously explained for NMT systems, in this model the encoder and the decoder have also their own self-attention layers, which help understanding the input and the output respectively processed and generated in the previous layer, so to handle various tasks more effectively (Vaswani et al. 2017). The feature of self-attention is what distinguishes the Transformer from RNNs. One of the main consequences of the self-attention mechanism is related to the Transformer not understanding the order of words in a sentence. Therefore, to address this issue, positional encoding is added. It involves pairing the token embeddings with positional embeddings to provide information on sentence structure, helping the model identify where each word should go in the final output (Naveed et al., 2024). Although the Transformer architecture was mainly designed for language tasks, it has been applied also to various domains, such as image or audio generation. However, it became particularly well-known with the growing popularity of

ChatGPT (Amatrian et al., 2024). ChatGPT, an AI chatbot developed by OpenAI, allows users to interact by sending messages, acting as prompts, enabling it to perform a wide range of tasks. Thus, nowadays people use natural language to easily interact with machines, much as they would communicate with other humans (Naveed et al., 2024). The success of ChatGPT has also driven great interest in the Transformer architecture and in LLMs (Amatrian et al., 2024). As an LLM, the Transformer is one of the fundamental components of ChatGPT. This aspect is represented also in the acronym GPT, which stands for "Generative Pre-trained Transformer". The "Generative" component refers to the chatbot's ability to generate coherent sentences in response to humans' inputs, while "Pre-trained" indicates the approach used to develop the model, which undergoes extensive training before being fine-tuned for specific tasks (Bridgelall, 2024). To be more precise, ChatGPT-4Omni, which is the model used in this thesis, can be classified under the category of Multimodal LLMs (OpenAI, 2023). In comparison to traditional LLMs, they can handle more complex tasks and process various types of inputs, such as images, audio, or video. Dealing with a broader type of information, they have a better understanding of the contexts and can provide more accurate results (Naveed et al., 2024).

While LLMs represent a significant breakthrough in NLP tasks, new and advanced LLMs also pose great challenges and limitations that should not be unnoticed. In their overview of LLMs, Naveed and his colleagues (2024) identify and discuss various challenges that cover a range of aspects, which could be categorized into ethical concerns and technological issues. Specifically, new LLMs require advanced computational resources, which can also increase energy consumption, thereby contributing to environmental harm. In addition, developing and training these models requires considerable financial investments, limiting participation only to well-funded companies. Focusing on the technical side, many challenges have been identified. Two of the most meaningful are hallucinations and overfitting. The former involves the model providing responses that seem likely and plausible but are actually incorrect. The latter is related to learning capabilities of models; in fact, they may have some difficulties in finding the right balance between memorization and generalization. This means that models might learn too many specific, irrelevant details, becoming less effective when dealing with new inputs and real-world data (Naveed et al., 2024). Moreover, there is a growing field of study, prompt engineering, which focuses on analyzing and implementing the prompts given to these models so that they can provide accurate results. In fact, their syntax and semantics highly impact the models' outputs (Amatriain, 2024; Bridgelall, 2024; Naveed et al., 2024). Ultimately, prompting is likely one of the main methods through which users interact with LLMs. Therefore, limitations and challenges previously discussed, some of which are also highlighted in Amatriain's article (2024), further emphasize the need for sophisticated prompt engineering to improve the effectiveness of LLMs.

### 2.6.1 PROMPT ENGINEERING

As mentioned at the end of the previous section, prompt engineering is a discipline rapidly evolving in the field of Generative AI and LLMs. Despite focusing on prompts and their structure, to generate the best possible model's output there has to be also a deep understanding of the context the model is operating in, knowledge of the AI model's limitations and capabilities and, lastly, a systematic approach to adapt prompts for various contexts (Amatriain, 2024). Models can be fed with different types of natural language prompts, varying from basic ones, to more advanced. According to Bridgelall (2024), however, there are four key components necessary to create what is defined as a "good prompt". Firstly, as mentioned earlier, it is essential to set the context of the task or topic, enabling the model to provide more accurate responses by understanding relevant domains. Secondly, to avoid irrelevant responses, ambiguity should be avoided, and the requirements should be clear. In addition, the model can be instructed with desired format, so that the responses are structured according to the user's needs. Lastly, examples should be included, so as to guide the system in understanding how the final output should be (Bridgelall, 2024). Amatriain (2024) in his work presents a variety of different prompt structures, providing examples and explanations on their structure and functioning. Other scholars, who are occasionally mentioned, also contributed to define this developing field of study. Considering basic prompts, their structure should contain at least instructions or questions that could eventually be paired with input data and examples. An instance of prompting techniques that contains examples is the "Few-shot" learning technique, which involves providing demonstrations and examples in the prompt, to guide the output generation. This method is opposed to the "Zero-shot" learning technique, which does not involve any example in the prompt (Naveed et al., 2024). On the other hand, advanced prompts include more complex and elaborated structures. For example, the model could be instructed to provide the reasoning process behind the output produced, so to demonstrate the various steps it had to perform to generate that answer. This specific prompt is known as "Chain-of-Thought" prompting (CoT) and allows the system to show its generative, predictive skills, but also forces it to deal with one of its limitations, that is related to lack of reasoning skills (Naveed et al., 2024), explicitly stating the steps followed, which would usually be just implicit. Another prompting method that aligns with humans' cognitive abilities is the "Tree-of-Thought" (ToT) technique. This method considers a series of possible solutions to a task or question, before generating the most plausible one. At the core of the approach there are the "thought trees", which consist of a series of branches representing various hypotheses. These hypotheses are analyzed before providing the most relevant output (Amatriain, 2024).

One of the NLP tasks LLMs can perform is machine translation. Despite NMT systems significantly improving this process, there are still some issues and inconsistencies that need to be addressed. Instead of manually training the engines, over past years the practice of prompting NMT for enhancing its outcomes has become more common. The prompt-driven neural machine translation proposed by Li and his colleagues in 2022 provides an example of a model improving its outputs by being instructed on how to translate certain words or how the TS should be structured. The model works through translation constraints, which were inspired by instance-level constraints used in other studies cited in their article. However, despite improving the final output, instance-level constraints also had some limitations. The model implemented resulted in effective translations and also improved the efficiency of post-editing, lowering the time professional translators spent in editing the texts (Li et al., 2022). In addition to TS structure and words' translation, multilingual models might be prompted to translate into a specific TL, including dialects or variants, by explicitly stating the language name. Garcia and Firat (2022) conducted a preliminary study that explored formal and informal second person pronoun alternance in translations. They demonstrated that by adding a sentence that presupposes a formal expression to a neutral sentence, the output results in a TS with formal features. Consequently, they focused on using a template in the prompt, formulated as "Translate to {language_name}: {input_slot}", which would replace language tags introduced in earlier studies. This approach addressed the limitations of previous methods, which did not include specific dialects or languages missing from the fine-tuning stage. Their study found that adapting already-trained translation models was more successful than training models from scratch (Garcia and Firat, 2022).

## 2.7 CAT TOOLS

Although not directly relevant to the realization of this research, MT can be integrated in CAT tools to further support the work of translators. Thus, a brief overview of these tools is provided below, solely to offer general information on the potential integration of MT within CAT tools.

CAT tools are helpful software programs designed to improve the translation process, fastening the workflow and aiding translators in carrying out more accurate translations (Fernandes Silva et al., 2017). The core components that make CAT tools stand out compared to traditional human translation methods are their functionalities, such as Translation Memory (TM), Term Base (TB) and quality assurance. CAT tools segment texts into manageable units, allowing translators to reuse previous translations. This not only speeds up the process, but also helps maintain the original document's format. Previous translations are stored in the TM, which acts as a database that stores the translations in pairs, known as Translation Units (TU). These units can be used again later in the text or in other tasks using the same TM and are particularly beneficial for standardized or repetitive content. Similarly, the TB acts as a specialized dictionary, storing specific terms or expressions to ensure a consistent use of terminology. Moreover, TM can identify various types of matches between the ST and previously translated segments. This helps translators assess the similarity and applicability of previously translated content to the current segment. The matches vary in degree, including context matches, 100% matches, fuzzy matches and fragment matches, which can be adapted by translators, if necessary, to fit the current context and ensure consistency (Nwanjoku & Zayyanu, 2021). When MT is integrated into CAT tools, the process can be even more enhanced and made efficient. MT provides an initial translation that human translators can later refine, if necessary, adopting a strategy called "post-editing". This approach combines the efficiency of MT and the accuracy of human translators, speeding up the whole translation process and resulting particularly useful for large scale projects (Weng and Sawyer, 2023).

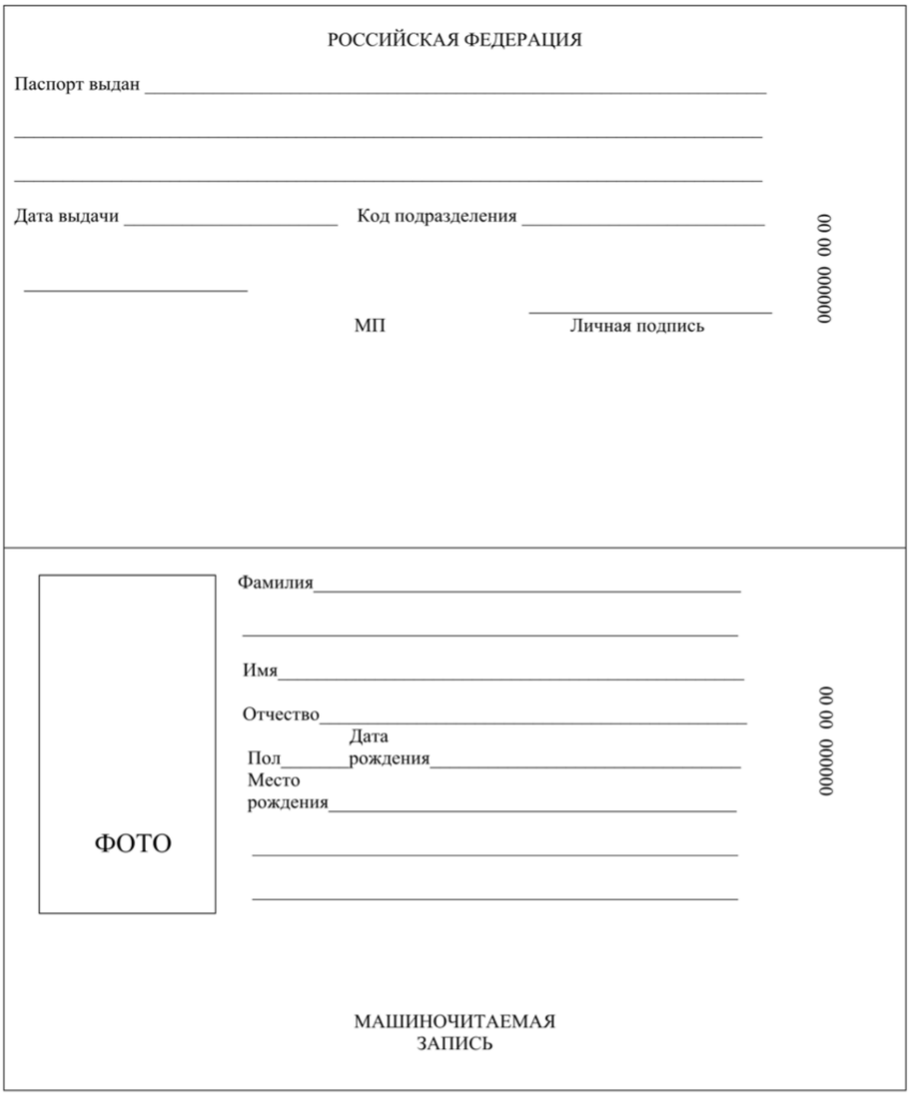
# 3. PRACTICAL FRAMEWORK

Translators working with CAT tools and MT are aided in performing their work, because these tools can help enhance the whole translation process. However, these tools sometimes do not provide the outputs translators wished for, leading to slowing down their job, rather than improving the process. This issue could arise particularly when systems are not adequately trained to handle words which do not have a straightforward equivalent in the TL, or for which translation is not an appropriate strategy. This is the case of anthroponyms and toponyms that do not have a recognized translation, for which transliteration strategies tend to be adopted, both by MT engines and human translators. However, the main problem arises when there is a lack of consistency in the processing of identical graphemes and SWs, which generates discrepancies and decreases MT's accuracy. As mentioned in the paragraph of the first chapter about transliteration norms, transliteration is not standardized and globally agreed, but there are different systems that have been proposed and used. In regards of this project work, the outputs seem to not align with the norms taken as reference for transliteration of anthroponyms and toponyms, respectively, that are the Decree published in 2020 by Russian Ministry of Foreign Affairs on issuing passports, and the manual on standardization of geographical names published in 2007 by the UNGEN Working Group on Romanization Systems. These norms have been considered, hoping to propose a possible transliteration of anthroponyms and toponyms found in Russian passports, so as to produce uniform results that could avoid creating differences among various documents belonging to a single person. Although the TL is always selected when translating, adopting a national system for transliteration would not be an appropriate choice. In fact, in official documents rendering the name "Александр" as Aleksandr, according to the decree, or Alexander (or other variants), according to a different transliteration system, would create confusion and lack of clarity. Even though it could still be possible to understand that reference is made to the same SW, the inconsistency could lead to misunderstandings and errors. Between anthroponyms and toponyms there are some differences in transliteration, due to the decision of following the UNGEGN directories for standardization of geographical names, that clearly could not be suitable for proper names of people. Therefore, in the results gained from this study it can be seen that the Russian “x”, for instance, is rendered as “kh” for persons names and as a simple “h” for place names.

## 3.1 METHODOLOGY AND DATA

Cultural words are found in a variety of fields and domains and according to some authors they belong to many different categories and subcategories. At the beginning, the broader category of realia was taken into consideration, but not narrowing it down to specific categories would have made it particularly difficult to carry out a coherent and comprehensive study. To ensure accuracy and consistency, the domain of interest was limited to passports, as they provide both anthroponyms and toponyms. Given that the aim of this thesis is to show and assess MT outputs, it was not necessary to analyze existing passports that contained real personal information. Instead, the STs used as material for the project were generated by ChatGPT-4 Omni, based on an initial model included in the prompt given to the system. The template, which is provided here below, is an example of the form a Russian Federation citizen’s passport should have according to the Decree of the Russian Federation Government published in 2023 titled "Об утверждении Положения о паспорте гражданина Российской Федерации, образца и описания бланка паспорта гражданина Российской Федерации"[20].

[20] “On the approval of the Regulations on the Passport of a citizen of the Russian Federation, model and description of the form of the passport of a citizen of the Russian Federation”, personal translation.


РОССИЙСКАЯ ФЕДЕРАЦИЯ

Паспорт выдан ______

Дата выдачи ______ Код подразделения ______

МП ______ Личная подпись

00 00 000000

ФОТО

Фамилия ______

Имя ______

Отчество ______

Пол ______ Дата рождения ______

Место рождения ______

00 00 000000

МАШИНОЧИТАЕМАЯ ЗАПИСЬ

*Figure 2. Sample form of a passport of a citizen of the Russian Federation.*

The prompt given to the system was "Generate 22 Russian passports by filling out this sample passport, providing 22 different anthroponyms and toponyms. These proper names must include graphemes that are difficult to transliterate into the Italian language." This request specifically included choosing words with unusual graphemes for the Italian language so as to further test MT abilities or difficulties in rendering items that do not have direct equivalents in the Italian alphabet. In fact, lots of proper nouns contained consonants such as "х", "ш", "ч", the hard vowel "ы", and the soft vowels "ю" and "я". Nonetheless, words containing the soft sign "ь" were considered. In addition, the request to avoid repetitions across passports allowed for the broadest and most varied sample possible. According to the model, each passport provides five proper nouns to be considered for transliteration. Besides the person's name, which consists of a first name, a patronymic and a last name, there are two place names, one found in the place of birth field and one in the field of the authority that issued the passport. However, two aspects need to be considered: firstly, these fields sometimes included not only the larger city but also regions and/or specific, smaller towns or villages; secondly, most of the times, within passports the toponyms in these fields corresponded. Therefore, the resulting material to be transliterated could vary for each text. In addition, almost all the passports considered (17) had specific denominations for Russian administrative units and a rural locality, lacking direct translations that could convey the same meaning without the need for further explanation. Therefore, also these few items, which can effectively be considered realia, underwent transliteration[21]. To be more precise, the term "область" figured at least once in all these 17 documents, but for seven times it figured twice; the term "станица" occurred twice in two documents; the term "хутор" occurred once in one document. The first two terms were usually written as abbreviations (i.e., "обл." and "ст." or "ст-ца"). Moreover, while anthroponyms were always written in the nominative case, toponyms sometimes appeared also in the genitive and prepositional cases. Since the Italian language uses prepositions instead of cases to indicate grammatical complements, prepositions within the strings of toponyms are also included in the analysis to provide the correct representation of the ST.

After generating the STs, each sample passport was pasted into a text file and subsequently processed by the five different generic MT systems considered. Finally, each output was saved to create five TL corpora. These Italian translations primarily serve to compare automated transliterations with the norms taken as reference; however, firstly the outputs of each category, for each MT, were commented. Subsequently, besides the five MT corpora, a corpus of texts manually translated, and transliterated following reference norms, was created, so as to have the reference material for comparison. The six corpora can be found in the appendix of the thesis. Furthermore, the outputs were compared with each other to provide an even more accurate assessment of the MT systems performance. When personally working with the texts, choices were made to translate rather than transliterate some place names, so as to act in accordance with what

[21] They are included in the category of oeconyms, as stated by Gornostay and Skadiņa (2009), which is the actual category of toponyms considered in the present study. Again, to ensure a smoother reading, the personal transliteration of the terms is provided here: oblast', stanitsa, hutor.

stated Newmark in 1988. In fact, for toponyms that already had a recognized translation, transliteration would have been an inappropriate strategy to follow. There are several evaluation metrics, such as BLEU, TER, COMET, METEOR, which are commonly used to assess and rate machine translation engines. However, for the purpose of this thesis, a parameter similar to those suggested by these metrics is applied. Specifically, the results of the engines are compared to transliterations performed according to the norms identified, adopting an approach that reflects these metrics but focuses only on specific terms rather than syntax or semantics. Moreover, no scores or diagram are used to evaluate the outputs. Instead, the aspects considered are transliterations that align with the identified norms, transliterations that does not align with the norm, and translations. Engines producing a higher number of translations are considered the worst performing, as in this case translation is acceptable only if a documented and established term exists in the TL, as would happen for cities such as Moscow or Saint Petersburg. Consequently, translation also requires greater effort from the human translator to refine the MT output. Subsequently, transliterations that deviate from the norms show inaccuracies and inconsistencies which diminish the performance of the MT engine. Conversely, outputs which do not show inaccuracies indicate that the engine is adequate and well performing, leading to the conclusion that the engine producing the highest number of adequate transliterations is ranked as the best performer.

According to recent studies presented in the second chapter of this thesis, instead of training each MT engine with manually translated texts to finetune the outputs, AI and LLMs have been employed to improve the transliterations of anthroponyms and toponyms considered. Specifically, in this thesis ChatGPT-4 Omni was used to generate transliterated proper nouns that adhered to standardized norms. As previously discussed, ChatGPT works throughout prompts filled with instructions and requests. Thus, the following, second prompt was given to ChatGPT: "Translate the {text} to Italian using {table1} for transliteration of anthroponyms and {table2} for transliteration of toponyms". The tag {text} was replaced with passports' texts, the tags {table1} and {table2} were respectively replaced with the Decree's table and the UNGEGN's table[22]. Initially, the idea was to provide two separate prompts to prevent any potential confusion in identifying which terms should be transliterated according to the relevant norms.

[22] This prompt, in its extended version, can be found in the appendix.

However, it was assumed that with a precise prompt, the system should have been able to accurately distinguish between toponyms and anthroponyms. The results obtained from the given prompt were commented and analyzed, so as to determine whether ChatGPT could provide accurate transliterations when given guidelines to follow and whether it could be a suitable alternative to the time-consuming practice of training MT engines.

## 3.2 MT TOOLS AND CHAT GPT-4 OMNI

In order to choose the machine translations to be investigated, two criteria were taken into account: firstly, the possibility to fully access some of these MTs, namely Language Weaver, ModernMT and Intento, thanks to an agreement between the University of Padua and the companies that developed them, which led to the decision to explore their functionality in this field. Secondly, reference was made to the report published by Intento, "The State of Machine Translation 2023", that evaluated 37 MT engines considering 22 language pairs and nine content domains. Taking into account the assessments of this report, it was decided to investigate DeepL and Yandex. They both resulted in being language leaders, meaning that they had the highest ratings in a language pair across numerous domains. Additionally, DeepL gained the denomination of domain leader, meaning that in various language pairs was ranked best for a specific domain.

- Language Weaver (LW). It was first launched in 2002 as an independent company but joined SDL in 2010. In 2020 both became part of RWS, which is the company that developed the CAT tool Trados. In fact, it is the reference MT used in this CAT tool. It offers neural machine translation systems that can be installed or used in the cloud, combining them with AI, to provide high-quality outputs. An interesting aspect of this tool is its name that directly reminds of Warren Weaver, a pioneer in the field of MT (https://www.rws.com/language-weaver/our-history/).
- ModernMT (mMT). This engine was also investigated in Intento's report "The State of Machine Translation 2023", standing out as one of the MTs obtaining the "Language Coverage" distinction. Along with five other engines, it is one of the providers that offers the broadest range of language pairs (Intento, 2023). According

to the company, ModernMT distinguishes itself from other providers thanks to the "document-level translation" feature, as they named it. This approach allows the machine to translate by considering the entire context of the document. Additionally, the outputs can be improved by learning from human corrections in real time, meaning that the engine does not need to be trained again. The company distinguishes the subscriptions available for enterprises and the ones available for translators and language service providers (LSPs); the latter subscriptions allow for the integration of the ModernMT plug-in into CAT tools (https://www.modernmt.com/translators).

- Intento (IN). A platform that offers MT via an API[23], enabling users to choose the engine they prefer to carry out the translation task, among 77 engines available on the platform (https://inten.to/api-platform/ai/text/translate#mt-reports). Intento is a language hub that aims to enhance global companies' job and customers' experience through MT and AI. In fact, they believe that combining these two tools can help in achieving better results in translation and optimizing businesses' multilingual workflows (Intento, 2023). To carry out this project, the "Smart Routing" mode has been used. When using this mode, translations are processed according to the top-performing MT engines which stood out in Intento's report (https://help.inten.to/hc/en-us/articles/360016598320-Intento-Smart-Routing).
  Intento also provides "Domain-Specific Routing" that chooses among the best MT engines for specific domains. At the present time, they offer nine domains. As stated above, at the time of MT testing, the most recent report was the 2023 one. However, as of now the 2024 report has been published.
- DeepL (DL). This MT system was launched by DeepL GmbH in 2017 and uses NNs and deep learning algorithms; this engine is recognized as one of the main MT systems in the whole MT industry (Intento, 2023). The company offers different subscription possibilities so that customers can find the option that most suits their needs. The free version, which is the one used for this thesis, is also available; clearly, it has some limitations in comparison to paid services, but it remains useful for academic-level projects. Through the years the company developed a broader offer,

[23] According to Treccani, API (application programming interface) provides rules so that different software can communicate with each other, facilitating their interaction (https://www.treccani.it/enciclopedia/api_(Lessico-del-XXI-Secolo)/).

introducing DeepL Write that can be used to enhance communication. It works through AI that revises and improves written texts, so that they can be appropriate for the context (https://www.deepl.com/it/pro#team).

- Yandex (YA). At first, when it was launched in 2011 Yandex Translate adopted a SMT approach. However, from 2017 the approach shifted to a hybrid method combining SMT and NMT. According to the company, adopting a hybrid approach would improve the quality of MT outputs. (https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system/). Being founded by a Russian company, the initial idea on adopting this system was also driven by the hypothesis that it could provide better outputs in comparison to the rest of MT considered.
- ChatGPT-4 Omni. This system was first launched by the company OpenAI in 2022. It was a chatbot application based on GPT-3.5 model which quickly reached a large audience thanks to its accessibility (Amatrian et al., 2024). The acronym GPT, as mentioned in the previous chapter, stands for "generative pre-trained transformer", which is a type of LLM trained on a vast set of data. As of today, the company has released the GPT-4 version of the system and, in 2024, launched the updated GPT-4 Omni version, which is used in the current study. According to the OpenAI blog, this version significantly improved response times, making it even more similar to human capabilities. Although major improvements occurred in the fields of vision and audio understanding, this system is generally considered a further step forward in the interaction between humans and computers (https://openai.com/index/hello-gpt-4o/).

## 3.3 RESULTS ANALYSIS OF STAND-ALONE MTs

In this section, the outputs of the five MT engines are analyzed. To provide a better organized explanation and representation of results, two main sections have been created. The first section, which focuses on the analysis of anthroponyms, is further divided into three subsections: first names, patronymics and last names. The second section is devoted to toponyms. In addition, within these sections, each set of items is analyzed considering MTs separately.

### 3.3.1 ANTHROPONYMS: FIRST NAMES

From Table 1 found below some common patterns followed by Language Weaver can be noticed. Firstly, out of 22 first names, half of them, 11, have been translated, rather than transliterated. Additionally, in many cases (i.e., Alexander, Andrew, Eugene, Cathrine, Christine, Jacob) an English name was preferred over the Italian name. "Michael" and "Nicholas" may be added too to this list, even though they are usually used also in Italian and therefore for the purpose of this thesis do not fall under the categorization of "English names". Also, the names Veronica and Victoria can be considered a translation of the original source names, adopting the common Italian grapheme "c" when representing the corresponding Russian grapheme "k". As will be shown in the following chapter, proper transliteration of this letter would have kept the grapheme "k", which is also known in the Italian language. Therefore, proper transliteration would have resulted in "Veronika" and "Viktoria". Another aspect worth to be highlighted is the digraph "кс" found in the names Александр and Алексей, which resemble the sound corresponding to the grapheme "x", which is in fact chosen in the transliteration[24]. When adopting this strategy, a digraph in the ST becomes a single grapheme in the TT, thus all the original letters are not represented. Going back to names that have been translated, there are two cases in which the MT produced two interesting outputs: firstly, the name Ираида was incorrectly translated as "Israele", the Italian name for the country of Israel. Upon consideration, it could be concluded that the MT considered it appropriate to translate rather than transliterate the name, possibly due to inappropriate training of the system in recognizing Ираида as a Russian name requiring transliteration, or because of the phonetic similarity of the initial sounds. Secondly, the name Оксана was not transliterated at all, retaining the same letters in the TT. A hypothesis is that, since the name consists of graphemes which are all present in the TL, Language Weaver did not recognize the ST as requiring transliteration, and thus reproduced the exact same graphemes. However, some of these graphemes represent different sounds in the TL compared to the SL, again resulting in incorrect processing of the SW. One last point to be considered is that the grapheme "ю", representing a soft vowel, is rendered in two different ways, probably depending on its position in the word.

[24] In the case of Александр > Alexander the choice is most probably related to the translation of the name, rather than transliteration, which follows the spelling of the English name.

When it is found after a consonant, as in the name Людмила, it is transliterated with the grapheme “yu”; whereas, when found at the beginning of the word, as in the name Юлия, it is transliterated as “ju”.

| SOURCE FIRST NAMES | TARGET FIRST NAMES – LANGUAGE WEAVER |
|---|---|
| АЛЕКСАНДР | ALEXANDER |
| АЛЕКСЕЙ | ALEXEY |
| АНДРЕЙ | ANDREW |
| БОРИС | BORIS |
| ВЕРОНИКА | VERONICA |
| ВИКТОРИЯ | VICTORIA |
| ВЛАДИМИР | VLADIMIR |
| ДАРЬЯ | DARIA |
| ЕВГЕНИЙ | EUGENE |
| ЕКАТЕРИНА | CATHERINE |
| ИГОРЬ | IGOR |
| ИРАИДА | ISRAELE |
| КРИСТИНА | CHRISTINE |
| ЛЮДМИЛА | LYUDMILA |
| МИХАИЛ | MICHAEL |
| НИКОЛАЙ | NICHOLAS |
| ОКСАНА | OKCAHA |
| СВЕТЛАНА | SVETLANA |
| СЕРГЕЙ | SERGEY |
| ТАТЬЯНА | TATIANA |
| ЮЛИЯ | JULIA |
| ЯКОВ | JACOB |

*Table 1. Language Weaver transliteration of first names.*

The following table shows the output of first names produced by ModernMT. There are some similarities with the previous table, as also in this case some names have been translated with English names rather than transliterated. In particular, Александр and Евгений are rendered as Alexander and Eugene. Another similarity is that the name Оксана again is not correctly processed, as none of the graphemes was changed. The hypothesis made is consistent with the one made for Language Weaver output. Lastly, the digraph "кс" again is transliterated as the single grapheme "x", following the English representation of the sound "ks". When it comes to the name Ираида, MT fails in representing all the letters, as the final "a" is missing. One final aspect to be considered is the difference in rendering the same grapheme in different words, probably because of the position the grapheme has in the word, whether it is found at the beginning, in the middle or at the end of the word. Considering the grapheme "k" found in the Russian names Вероника, Виктория, Кристина, Екатерина, Николай and Яков while it is transliterated as "k" when it appears at the beginning of the word, as in the last name cited (Кристина > Kristina), the grapheme is rendered as "c" in the first two cases, that is when it is found inside the name (Вероника > Veronica, Виктория > Victoria). However, this pattern is not consistent in the last three name cited; although the grapheme is found inside the word, it is still rendered as "k", and not as "c". This inconsistency may be related to the fact that in the first two cases the transliteration could also resemble the Italian-sounding names Veronica and Victoria and the system may be trained to follow that rule, whereas in the Italian-sounding name corresponding to Екатерина, the first vowel "e" is dropped, becoming Caterina, which involves an additional step that would not be appropriate to consider in this case. For what it concerns the last two male names, they do not resemble Italian-sounding names, therefore the system has probably been trained to simply follow transliteration rules that include rendering the Cyrillic grapheme "k" as the Latin grapheme "k", considering the sound that lies behind it. Another example of the abovementioned case is the grapheme "я", which appears mainly in female names and just once in a male name. In the four female names, except for Татьяна, "я" is always found at the end of the name (Виктория, Дарья, Юлия) and is always transliterated as "a". On the contrary, in the male name Яков, the grapheme is found at the beginning of the name and is transliterated with the digraph "ya". However, the explanation for this difference may be related also to the presence of the vowel "и" or the soft sign "ь" preceding "я", which influence its representation. Therefore, the system may be trained to preserve the phonetic representation of Italian language, which matches "ia" to the combination "ья" and "ия" rather than considering the grapheme "я" on its own. Also in the earlier outputs of Language Weaver there is this inconsistency between the four female names and the male name, but in that case, it is motivated by the fact that the male name Яков is translated into the English name Jacob, thus it follows English spelling.

| SOURCE FIRST NAMES | TARGET FIRST NAMES – MODERN MT |
|---|---|
| АЛЕКСАНДР | ALEXANDER |
| АЛЕКСЕЙ | ALEXEY |
| АНДРЕЙ | ANDREY |
| БОРИС | BORIS |
| ВЕРОНИКА | VERONICA |
| ВИКТОРИЯ | VICTORIA |
| ВЛАДИМИР | VLADIMIR |
| ДАРЬЯ | DARIA |
| ЕВГЕНИЙ | EUGENE |
| ЕКАТЕРИНА | EKATERINA |
| ИГОРЬ | IGOR |
| ИРАИДА | IRAID |
| КРИСТИНА | KRISTINA |
| ЛЮДМИЛА | LYUDMILA |
| МИХАИЛ | MIKHAIL |
| НИКОЛАЙ | NIKOLAY |
| ОКСАНА | OKCAHA |
| СВЕТЛАНА | SVETLANA |
| СЕРГЕЙ | SERGEY |
| ТАТЬЯНА | TATIANA |
| ЮЛИЯ | YULIA |
| ЯКОВ | YAKOV |

*Table 2. ModernMT transliteration of first names.*

The third table presented displays the transliteration of Russian first names performed by Intento, using the Smart Routing functionality and DeepL. Considering that the results obtained are identical, they are analyzed together. Although not related to transliteration, what can be immediately noticed is that two proper names were formatted differently in the TL and not rendered with capital letters but in small letters (Ekaterina, Yulia). In certain situations,  and DeepL act similarly as the other two MTs already analyzed: firstly, it fails in transliterating the name Оксана, which is simply reproduced with the same letters; secondly, it adopts the English representation of “ks” sound in the name Александр, which is again rendered as the single grapheme “x”; lastly, probably depending on the position the letters have in the names, they are sometimes transliterated differently. In particular, the grapheme “к” is again transliterated both as “c” in the names Veronica and Victoria, which probably resemble Italian-sounding names, and as “k” in five other names. In contrast to the previous case in Modern MT, also the name “Алексей” is included in this categorization, as the grapheme “к” is not rendered with the English grapheme “x”. Among these five other names, only Кристина resembles an Italian-sounding name, Cristina; nevertheless, the grapheme, which is found at the beginning of the name, unlike the previous two, is still transliterated as “k”. Considering another grapheme previously discussed, “я”, in this case transliteration also depends on the grapheme that precedes it. When the grapheme is found at the beginning of the name as in Яков or is preceded by the soft sign “ь” as in Дарья and Татьяна in Latin alphabet is rendered as “ya”. On the other hand, it is transliterated as “ia” when preceded by the vowel “и”, such as in Виктория and Юлия. In this last name mentioned also the semivowel “ю” needs to be commented; whereas in this name it was transliterated with the digraph “yu”, in the name Людмила it was rendered with the simple vowel “u”. Again, this difference may be motivated by the position the grapheme has in the SW. One last difference spotted, and worth highlighting is the different transliteration of the grapheme “e” at the beginning of a name. Even though the sounds do not differ, in the two names that are here considered once the grapheme is transliterated with the digraph “ye” (Евгений > Yevgeniy), once with the grapheme “e” (Екатерина > Ekaterina).

| SOURCE FIRST NAMES | TARGET FIRST NAMES – INTENTO and DEEPL |
|---|---|
| АЛЕКСАНДР | ALEXANDER |
| АЛЕКСЕЙ | ALEKSEY |
| АНДРЕЙ | ANDREY |
| БОРИС | BORIS |

| | |
|---|---|
| ВЕРОНИКА | VERONICA |
| ВИКТОРИЯ | VICTORIA |
| ВЛАДИМИР | VLADIMIR |
| ДАРЬЯ | DARYA |
| ЕВГЕНИЙ | YEVGENIY |
| ЕКАТЕРИНА | Ekaterina |
| ИГОРЬ | IGOR |
| ИРАИДА | IRAIDA |
| КРИСТИНА | KRISTINA |
| ЛЮДМИЛА | LUDMILA |
| МИХАИЛ | MIKHAIL |
| НИКОЛАЙ | NIKOLAY |
| ОКСАНА | OKCAHA |
| СВЕТЛАНА | SVETLANA |
| СЕРГЕЙ | SERGEY |
| ТАТЬЯНА | TATYANA |
| ЮЛИЯ | Yulia |
| ЯКОВ | YAKOV |

*Table 3. Intento and DeepL transliteration of first names.*

In table 4 are shown the results gathered from transliteration performed in Yandex. As happened in Intento and DeepL there are names with different formatting; in this case, however, the majority of first names were written in small letters and only two of them kept the original formatting. Moreover, one of these two names that were still written in capital letters seemed to not be processed at all, as all the graphemes of the SW are also found in the TW. Therefore, Yandex is no exception compared to the other four MT systems when it comes to processing the Russian name Оксана. As happened in Language Weaver, some names have been translated rather than transliterated, mainly adopting an English-sounding name (Alexander, Andrew, Eugene, Catherine, Michael, Nicholas). There are three more names, which are Veronica, Victoria and Cristina, that can be considered translations; in this case, an Italian-sounding name was preferred. One aspect that these tree names have in common is that they all contain the grapheme “к” in the SW, which was transliterated as “c” in the TW. As already stated in the analysis performed for Language Weaver, this grapheme corresponds to the sound represented by “k” in Russian, thus this explains why the names sound Italian. On the other hand, the grapheme “к” was transliterated as “k” in other two names (Aleksey and Yakov), probably because they do not sound Italian and thus the MT was trained to opt for transliteration rather than translation. In any case, this creates inconsistency in the processing of this specific grapheme. Some more similarities with the outputs given by the previous MTs can be highlighted. In particular, Yandex processed the graphemes “ю” and “я” as Language Weaver did. More specifically, the transliteration of the semivowel “ю” again differs between the names Людмила and Юлия: when the grapheme is found within the name is transliterated with the digraph “yu”, whereas when it is found at the beginning of the name is rendered as “ju”. Secondly, the grapheme “я” in the female names is transliterated as “a” both when it is preceded by the vowel “и” and the soft sign “ь”; whereas in the male name Яков is found at the beginning of the word and is transliterated with the digraph “ya”.

| SOURCE FIRST NAMES | TARGET FIRST NAMES – YANDEX |
|---|---|
| АЛЕКСАНДР | Alexander |
| АЛЕКСЕЙ | ALEKSEY |
| АНДРЕЙ | Andrew |
| БОРИС | Boris |
| ВЕРОНИКА | Veronica |
| ВИКТОРИЯ | Victoria |
| ВЛАДИМИР | Vladimir |
| ДАРЬЯ | Daria |
| ЕВГЕНИЙ | Eugene |
| ЕКАТЕРИНА | Catherine |
| ИГОРЬ | Igor |
| ИРАИДА | Iraida |
| КРИСТИНА | Cristina |
| ЛЮДМИЛА | Lyudmila |
| МИХАИЛ | Michael |

| | |
|---|---|
| НИКОЛАЙ | Nicholas |
| ОКСАНА | ОКСАНА |
| СВЕТЛАНА | Svetlana |
| СЕРГЕЙ | Sergei |
| ТАТЬЯНА | Tatiana |
| ЮЛИЯ | Julia |
| ЯКОВ | Yakov |

*Table 1. Yandex transliteration of first names.*

### 3.3.2 ANTHROPONYMS: PATRONYMICS

The following tables show the transliteration of patronymics performed by each MT. In comparison to transliteration of first names, there are less differences among the outputs of single graphemes. Table 5 represents the transliteration of Language Weaver. The question regarding the digraph "кс" can be noted again in the patronymics Александрович and Максимович, where it is transliterated with the single grapheme "x" (Alexandrovich and Maximovich), again probably because of the resemblance with the English sound. In fact, the machine probably understood the connection between the patronymics and the male names they came from (Александр and Максим), one of which was consistently processed by Language Weaver and rendered as Alexander, an English name. Another comparison with first names can be done for the grapheme "я", which is here rendered as "ya" (Вячеславович > Vyacheslavovich), whereas in the table of first names it mainly figures as "ia" (Татьяна > Tatiana). One last aspect to be highlighted but not compared to first names' processing is the different transliteration of the grapheme "е" when found after the soft sign "ь". These two graphemes figure together in five patronymics, which are Васильевич, Геннадьевна, Григорьевич, Евгеньевич and Леонтьевич. What can be noticed from the Italian transliteration is that in the first two patronymics the digraph "ье" is rendered as "ye" (Vasilyevich, Gennadyevna), whereas in the subsequent three it is rendered with the digraph "ie" (Grigorievich, Evgenievich, Leontievich). This inconsistency creates ambiguity when analyzing the outputs, because the role of the soft sign is identical in each patronymic and the pronunciation of the words does not vary between patronymics that have been transliterated with "ye" or "ie".

| SOURCE PATRONYMICS | TARGET PATRONYMICS – LANGUAGE WEAVER |
|---|---|
| АЛЕКСАНДРОВИЧ | ALEXANDROVICH |
| АНДРЕЕВИЧ | ANDREEVICH |
| ВАЛЕНТИНОВНА | VALENTINOVNA |
| ВАСИЛЬЕВИЧ | VASILYEVICH |
| ВИКТОРОВНА | VIKTOROVNA |
| ВЛАДИМИРОВНА | VLADIMIROVNA |
| ВЯЧЕСЛАВОВИЧ | VYACHESLAVOVICH |
| ГЕННАДЬЕВНА | GENNADYEVNA |
| ГЕОРГИЕВИЧ | GEORGIEVICH |
| ГРИГОРЬЕВИЧ | GRIGORIEVICH |
| ДМИТРИЕВИЧ | DMITRIEVICH |
| ЕВГЕНЬЕВИЧ | EVGENIEVICH |
| КОНСТАНТИНОВИЧ | KONSTANTINOVICH |
| ЛЕОНИДОВНА | LEONIDOVNA |
| ЛЕОНТЬЕВИЧ | LEONTIEVICH |
| МАКСИМОВНА | MAXIMOVNA |
| МИХАЙЛОВНА | MIKHAILOVNA |
| НИКОЛАЕВНА | NIKOLAEVNA |
| ОЛЕГОВИЧ | OLEGOVICH |
| ПЕТРОВНА | PETROVNA |
| СЕРГЕЕВНА | SERGEEVNA |
| СТЕПАНОВНА | STEPANOVNA |

*Table 2. Language Weaver transliteration of patronymics.*

The table presented below displays Modern MT transliteration of patronymics. The first thing that can be immediately seen is that for the first time, considering the anthroponyms analyzed up until now, the digraph "кс" has not been rendered with the grapheme "x". Modern MT kept the two separate graphemes in both patronymics that presented this digraph, i.e. Александрович and Максимович, which have been transliterated with the Latin digraph "ks" (Aleksandrovich and Maksimovich).

Considering the grapheme “e” found after the soft sign “ь”, in this case there is a difference between the patronymic Васильевич and the remaining four. While the four patronymics are transliterated as in Language Weaver, that is with the digraph “ie”, the patronymic Васильевич adopts the apostrophe to represent the soft sign and the “e” remains the same grapheme. Therefore, in this case the soft sign is not represented by a letter, but by a punctuation mark, as it is usually done according to scientific transliteration and ISO 9. Additionally, this patronymic presents another difference from the other ten male patronymics considered; the last grapheme “ч”, which is the typical ending of male patronymics, is transliterated with a single grapheme representing the diacritical sign “č”, and not with the usual digraph “ch” (Vasil’evič). Making again a comparison with transliteration of first names for the grapheme “я” found in the patronymic Вячеславович, the considerations are identical to the ones made for Language Weaver. In fact, both MTs in this case processed the grapheme found in first names and in the patronymic in the same way.

| SOURCE PATRONYMICS | TARGET PATRONYMICS – MODERN MT |
|---|---|
| АЛЕКСАНДРОВИЧ | ALEKSANDROVICH |
| АНДРЕЕВИЧ | ANDREEVICH |
| ВАЛЕНТИНОВНА | VALENTINOVNA |
| ВАСИЛЬЕВИЧ | VASIL' EVIČ |
| ВИКТОРОВНА | VIKTOROVNA |
| ВЛАДИМИРОВНА | VLADIMIROVNA |
| ВЯЧЕСЛАВОВИЧ | VYACHESLAVOVICH |
| ГЕННАДЬЕВНА | GENNADIEVNA |
| ГЕОРГИЕВИЧ | GEORGIEVICH |
| ГРИГОРЬЕВИЧ | GRIGORIEVICH |
| ДМИТРИЕВИЧ | DMITRIEVICH |
| ЕВГЕНЬЕВИЧ | EVGENIEVICH |
| КОНСТАНТИНОВИЧ | KONSTANTINOVICH |
| ЛЕОНИДОВНА | LEONIDOVNA |
| ЛЕОНТЬЕВИЧ | LEONTIEVICH |
| МАКСИМОВНА | MAKSIMOVNA |

| | |
|---|---|
| МИХАЙЛОВНА | MIKHAILOVNA |
| НИКОЛАЕВНА | NIKOLAEVNA |
| ОЛЕГОВИЧ | OLEGOVICH |
| ПЕТРОВНА | PETROVNA |
| СЕРГЕЕВНА | SERGEEVNA |
| СТЕПАНОВНА | STEPANOVNA |

*Table 3. Modern MT transliteration of patronymics.*

Intento and DeepL again processed the anthroponyms in the same manner, thus the following table represents the outputs given by both of them. Although it is not important for transliteration purposes, as happened for some first names, a single patronymic has also been rendered in lowercase letters rather than in capital letters. Besides this inconsistency of formatting, attention again needs to be devoted to the grapheme "e" when it is preceded by the soft sign. In five occurrences, the graphemes have been processed in three different ways: once it was simply rendered as "e" (Васильевич > Vasilevich) and once with the digraph "ie" (Григорьевич > Grigorievich), whereas for three times it was transliterated as "ye" (Геннадьевна > Gennadyevna, Евгеньевич > Evgenyevich, Леонтьевич > Leontyevich). Additionally, there are two more inconsistencies to be noticed. Firstly, the digraph "кс", which was represented by the corresponding Latin digraph "ks" for the first time only in the previous table, here is rendered differently in the patronymics Александрович and Максимовна. In the former case, the two graphemes are unified under the Latin grapheme "x" (Alexandrovich), likely for the same reasons given for the previous similar processing. On the other hand, in the second patronymic, two separate graphemes have been retained, rendering the digraph with its corresponding Latin graphemes "ks" (Maksimovna). The second inconsistency found is related to the repetition of "e" in the patronymics Андреевич and Сергеевна. Whilst in the second patronymic cited the two letters are simply transliterated with the same, corresponding Latin graphemes "ee" (Sergeevna), in the first case, the second "e" is rendered with the digraph "ye", thus resulting in the initial digraph "ee" being rendered as the trigraph "eye" (Andreyevich). This happened probably because of the influence the sound may have had when processing these words. However, all these inconsistencies create some difficulties when processing the same graphemes in different words, because it may seem that they are transliterations of different original graphemes, which in reality do not vary.

| SOURCE PATRONYMICS | TARGET PATRONYMICS – INTENTO and DEEPL |
|---|---|
| АЛЕКСАНДРОВИЧ | ALEXANDROVICH |
| АНДРЕЕВИЧ | ANDREYEVICH |
| ВАЛЕНТИНОВНА | VALENTINOVNA |
| ВАСИЛЬЕВИЧ | VASILEVICH |
| ВИКТОРОВНА | VIKTOROVNA |
| ВЛАДИМИРОВНА | VLADIMIROVNA |
| ВЯЧЕСЛАВОВИЧ | VYACHESLAVOVICH |
| ГЕННАДЬЕВНА | GENNADYEVNA |
| ГЕОРГИЕВИЧ | GEORGIEVICH |
| ГРИГОРЬЕВИЧ | GRIGORIEVICH |
| ДМИТРИЕВИЧ | DMITRIEVICH |
| ЕВГЕНЬЕВИЧ | Evgenyevich |
| КОНСТАНТИНОВИЧ | KONSTANTINOVICH |
| ЛЕОНИДОВНА | LEONIDOVNA |
| ЛЕОНТЬЕВИЧ | LEONTYEVICH |
| МАКСИМОВНА | MAKSIMOVNA |
| МИХАЙЛОВНА | MIKHAILOVNA |
| НИКОЛАЕВНА | NIKOLAEVNA |
| ОЛЕГОВИЧ | OLEGOVICH |
| ПЕТРОВНА | PETROVNA |
| СЕРГЕЕВНА | SERGEEVNA |
| СТЕПАНОВНА | STEPANOVNA |

*Table 4. Intento and DeepL transliteration of patronymics.*

Table 8 is the last table dedicated to patronymics and it shows the outputs of Yandex. Again, immediately it can be seen that most patronymics have been written in lowercase letters and only two kept the original formatting. However, from the transliteration perspective, Yandex’s handling of patronymics could be considered quite consistent across all the anthroponyms considered, as none of the aforementioned inconsistencies are present. The only aspect to be commented on is related to the rendering of the digraph “кс”, which is again transliterated as the single grapheme “x” in both patronymics that contain it, consistent with most cases in this study. Regarding the digraph “ье”, all five occurrences have been transliterated as “ie” in the Italian outputs. In addition, the processing of the digraph “ee” and of the ending “ч” of male patronymics is consistent in each TW.

| SOURCE PATRONYMICS | TARGET PATRONYMICS – YANDEX |
|---|---|
| АЛЕКСАНДРОВИЧ | Alexandrovich |
| АНДРЕЕВИЧ | Andreevich |
| ВАЛЕНТИНОВНА | Valentinovna |
| ВАСИЛЬЕВИЧ | Vasilievich |
| ВИКТОРОВНА | Viktorovna |
| ВЛАДИМИРОВНА | Vladimirovna |
| ВЯЧЕСЛАВОВИЧ | Vyacheslavovich |
| ГЕННАДЬЕВНА | Gennadievna |
| ГЕОРГИЕВИЧ | Georgievich |
| ГРИГОРЬЕВИЧ | Grigorievich |
| ДМИТРИЕВИЧ | Dmitrievich |
| ЕВГЕНЬЕВИЧ | Evgenievich |
| КОНСТАНТИНОВИЧ | Konstantinovich |
| ЛЕОНИДОВНА | Leonidovna |
| ЛЕОНТЬЕВИЧ | LEONTIEVICH |
| МАКСИМОВНА | MAXIMOVNA |
| МИХАЙЛОВНА | Mikhailovna |
| НИКОЛАЕВНА | Nikolaevna |
| ОЛЕГОВИЧ | Olegovich |
| ПЕТРОВНА | Petrovna |
| СЕРГЕЕВНА | Sergeevna |
| СТЕПАНОВНА | Stepanovna |

*Table 5. Yandex transliteration of patronymics.*

### 3.3.3 ANTHROPONYMS: LAST NAMES

This last section dedicated to the analysis of anthroponyms examines last names transliteration performed by each MTs investigated. The first table presents the outputs of Language Weaver, which sometimes did not performed transliteration but rather translated four last names. Specifically, the machine may not have been well-trained to recognize that in this case Баранов, Воронов and Козлов are not the genitive plural forms of the corresponding nominative common nouns of animals: баран (IT: pecora, EN: sheep), ворон (IT: corvo, EN: crow) and козлы (IT: capra, EN: goat). Instead, MT translated the surnames as "della pecora", "dei corvi" and "delle capre", which literally mean "of a sheep", of crows" and "of goats", rather than transliterating them. Additionally, Language Weaver translated the last name "Имярек" as "nome", which is Italian for "name". A consideration on the reasons why the machine may have given this output is related to the idea that the system may have processed the surname based on the similarity in sound of the initial graphemes, confusing "Имярек" with the Russian word for "name", "имя". Another aspect to highlight is the omission of the "-a" suffix at the end of each female surname, which is the common ending of female anthroponyms. The table below shows that none of the nine target last names ends with the grapheme "a", which the machine may not have considered as part of the base noun. Instead, a consistent pattern across all surnames is that the system seems to have interpreted the suffix as expressing the genitive singular form, rendering it in Italian as "di Agapov" (Агапова), "di Andreev" (Андреева), and "di Zverintsev" (Зверинцева), etc., which translate to English as "of Agapov", "of Andreev", and "of Zverintsev", etc. As it can be seen from the collection of text in the appendix, the prepositions can also be found before some male surnames and before the two surnames ending with "-o" belonging to a female person. However, the reasons for this cannot be related to the genitive form of words, as male surnames do not end with "-a" and "-o" is not a suffix for the genitive case. Therefore, since the preposition "di" (EN: of) is found in almost all the documents, regardless of whether the surnames belong to a male or female person, another conclusion that could be drawn is that the omission of "-a" at the end of female surnames is solely related to how the MT has been trained. To conclude the analysis of Language Weaver outputs, it is important to highlight that once again, a SW has not been transliterated at all: this is the case of the surname Сорока which retained all its graphemes in the Italian output.

The observations can be consistent with those made for the first name Оксана, which was not transliterated by any of the MT systems considered. Again, the system likely recognized that all the graphemes also existed in the Italian language and did not opt for transliterating them. However, the Italian graphemes do not correspond in sound to the Russian ones, indicating a failure in the MT's performance. Lastly, even though not related to transliteration, two surnames out of the 22 analyzed have been written in lowercase letters, rather than in capital letters.

| SOURCE LAST NAMES | TARGET LAST NAMES – LANGUAGE WEAVER |
|---|---|
| АГАПОВА | di AGAPOV |
| АНДРЕЕВА | DI ANDREEV |
| БАРАНОВ | DELLA PECORA |
| БРАГИН | BRAGIN |
| ВОРОНОВ | dei CORVI |
| ЗВЕРИНЦЕВА | di ZVERINTSEV |
| ИМЯРЕК | NOME |
| КОЗЕРОДОВ | KOZERODOV |
| КОЗЛОВ | DELLE CAPRE |
| КРЫЛОВ | KRYLOV |
| КУЗЬМЕНКО | KUZMENKO |
| МАРИНИНА | DI MARININ |
| САЗОНОВА | di SAZONOV |
| СИДОРОВ | SIDOROV |
| СОРОКА | COPOKA |
| СТОЛЬНИКОВА | di Stolnikov |
| ТАШКИНА | di Tashkin |
| ТРОФИМЕНКО | TROFIMENKO |
| ХЕРЛО | HERLO |
| ЧУМАКОВ | CHUMAKOV |
| ШАПОШНИКОВА | DI SHAPOSHNIKOV |
| ЩЕРБАКОВА | DI SHCHERBAKOV |

*Table 6. Language Weaver transliteration of last names.*

The next table displays Modern MT transliteration of surnames. Firstly, as with Language Weaver, Modern MT also fails to transliterate the surname Сорока, which retains the same graphemes of the SL in the TL. This system works similarly to the previous one also when it comes to processing the surnames Баранов and Козлов; they have been literally translated rather than transliterated, adopting the genitive form "della pecora" and "delle capre". It is interesting to note that even though the "-ов" suffix represents the plural form, Баранов is translated as "of a sheep" rather than "of sheep". In this specific case, the systems were probably not properly trained at all. However, Modern MT performed better for the surname Воронов by avoiding rendering the "-ов" suffix as possessive form, as happened with three other last names presenting this ending digraph (Крылов > Krylov, Сидоров > Sidorov, Чумаков > Chumakov). On the other hand, in the last name Козеродов, the suffix is once again not considered part of the base noun but instead is interpreted as representing the possessive form, thus becoming "di Kozerod" (EN: of Kozerod). Considering graphemes that might be interpreted as not being part of the base noun, Modern MT omitted the final "a" in two female surnames, Агапова and Зверинцева, transliterating them as Agapov and Zverintsev. However, there are seven other female surnames ending with -a for which this omission did not occur. These different transliteration results show inconsistency in processing words with similar characteristics. One final aspect to notice, even though it is not that important in relation to transliteration, is that in this case the surname Маринина has been rendered in lowercase letters, rather than capital letters.

| SOURCE LAST NAMES | TARGET LAST NAMES – MODERN MT |
|---|---|
| АГАПОВА | AGAPOV |
| АНДРЕЕВА | ANDREEVA |
| БАРАНОВ | DELLA PECORA |
| БРАГИН | BRAGIN |
| ВОРОНОВ | VORONOV |
| ЗВЕРИНЦЕВА | ZVERINTSEV |
| ИМЯРЕК | IMYAREK |
| КОЗЕРОДОВ | DI KOZEROD |
| КОЗЛОВ | DELLE CAPRE |

| | |
|---|---|
| КРЫЛОВ | KRYLOV |
| КУЗЬМЕНКО | KUZMENKO |
| МАРИНИНА | Marinina |
| САЗОНОВА | SAZONOVA |
| СИДОРОВ | SIDOROV |
| СОРОКА | COPOKA |
| СТОЛЬНИКОВА | STOLNIKOVA |
| ТАШКИНА | TASHKINA |
| ТРОФИМЕНКО | TROFIMENKO |
| ХЕРЛО | HERLO |
| ЧУМАКОВ | CHUMAKOV |
| ШАПОШНИКОВА | SHAPOSHNIKOVA |
| ЩЕРБАКОВА | SHCHERBAKOVA |

*Table 7. Modern MT transliteration of last names.*

As occurred for first names and patronymics, Intento and DeepL gave identical outputs also for last names. In contrast to Language Weaver and Modern MT, among these outputs there are less inconsistencies to be spotted. In fact, aside from the formatting of the last surname, which again is not a fundamental aspect, there is just one other discrepancy to highlight. The grapheme "з" is found in five surnames: Зверинцева, Козеродов, Козлов, Кузьменько, and Сазонова. In four of these surnames, the grapheme was rendered as "z", but once the system transliterated it as "s" (Козеродов > Koserodov). Moreover, the surname Сорока retained the same graphemes, probably because the system did not recognize them as Cyrillic graphemes, but as Latin ones. For sake of completeness, it is worth making a consideration on surnames that resemble common animal nouns; these two systems chose to transliterate rather than translate the last names.

| SOURCE LAST NAMES | TARGET LAST NAMES – INTENTO and DEEPL |
|---|---|
| АГАПОВА | AGAPOVA |
| АНДРЕЕВА | ANDREEVA |
| БАРАНОВ | BARANOV |

| | |
|---|---|
| БРАГИН | BRAGIN |
| ВОРОНОВ | VORONOV |
| ЗВЕРИНЦЕВА | ZVERINTSEVA |
| ИМЯРЕК | IMYAREK |
| КОЗЕРОДОВ | KOSERODOV |
| КОЗЛОВ | KOZLOV |
| КРЫЛОВ | KRYLOV |
| КУЗЬМЕНКО | KUZMENKO |
| МАРИНИНА | MARININA |
| САЗОНОВА | SAZONOVA |
| СИДОРОВ | SIDOROV |
| СОРОКА | COPOKA |
| СТОЛЬНИКОВА | STOLNIKOVA |
| ТАШКИНА | TASHKINA |
| ТРОФИМЕНКО | TROFIMENKO |
| ХЕРЛО | HERLO |
| ЧУМАКОВ | CHUMAKOV |
| ШАПОШНИКОВА | SHAPOSHNIKOVA |
| ЩЕРБАКОВА | Shcherbakova |

*Table 8. Intento and DeepL transliteration of last names.*

The last table of anthroponyms displays the outputs of Yandex in transliterating surnames. As it can be seen from the right column below, the machine failed in transliterating the surnames Воронов, Имярек and Козеродов. The "-ов" suffix in Воронов has not been considered as part of the base nouns but as indicating the genitive plural form of ворон (crow), thus resulting in the translation "dei corvi" in the TL (EN: of crows). Regarding the other surname with the "-ов" suffix, the machine did not interpret the suffix as representing the genitive case. Probably, it likely noticed a resemblance to the Russian word for the astrological sign "козерог", thus translating the surname into Italian as "capricorno". A possible reason for this could be that the machine was more familiar with the zodiac sign than with the surname, thus opting for what it considered the more suitable translation. Finally, to conclude on the last names that have not been transliterated, the machine processed the surname “Имярек” as “nome”, which is the Italian translation of the Russian word for “name”, “имя”. As happened for the previous surnames, the two words share some graphemes which probably led Yandex to consider “nome” the appropriate rendering for this surname. Besides the abovementioned surnames, the last name “Херло” was also translated. However, in this case, the TL adopted a completely different TW from the SW; with the final combination of graphemes “-ough” the last name appears to resemble an English-sounding word. This is particularly interesting because the other anthroponyms of this document, that are “Дарья” and “Геннадьевна”, were not altered by this MT in the TL. Had they been, it could be concluded that the system aimed for a consistent approach to processing the entire passport. In addition, among the nine surnames with a feminine ending that refer to female individuals, the MT transliterated all graphemes in only three cases (Агапова > Agapova, Зверинцева > Zverintseva, Маринина > Marinina). Conversely, the surnames Андреева, Сазонова, Стольникова, Ташкина, Шапошникова and Щербакова were rendered in the TL in their masculine form, omitting the “-a” suffix. The surname Сорока has been excluded from this analysis because, despite having the “-a” suffix, it does not refer to a female individual. Additionally, the system failed to transliterate the last name “Сорока”, retaining all the original graphemes in the TL. Lastly, only 6 last names out of 22 kept the original formatting.

| SOURCE LAST NAMES | TARGET LAST NAMES – YANDEX |
|---|---|
| АГАПОВА | Agapova |
| АНДРЕЕВА | Andreev |
| БАРАНОВ | Baranov |
| БРАГИН | Bragin |
| ВОРОНОВ | dei corvi |
| ЗВЕРИНЦЕВА | ZVERINTSEVA |
| ИМЯРЕК | nome |
| КОЗЕРОДОВ | Capricorno |
| КОЗЛОВ | Kozlov |
| КРЫЛОВ | Krylov |
| КУЗЬМЕНКО | Kuzmenko |

| | |
|---|---|
| МАРИНИНА | Marinina |
| САЗОНОВА | Sazonov |
| СИДОРОВ | Sidorov |
| СОРОКА | СОРОКА |
| СТОЛЬНИКОВА | STOLNIKOV |
| ТАШКИНА | TASHKIN |
| ТРОФИМЕНКО | TROFIMENKO |
| ХЕРЛО | HURLOUGH |
| ЧУМАКОВ | CHumakov |
| ШАПОШНИКОВА | Shaposhnikov |
| ЩЕРБАКОВА | Shcherbakov |

*Table 9. Yandex transliteration of last names.*

## 3.3.4 TOPONYMS

As explained above, the template used to create the 22 passports analyzed in this thesis included two sections requiring toponyms. Usually, each section contains at least two place names, and occasionally up to three. Consequently, not considering repetitions, this collection of texts comprises approximately 60 terms that could potentially be transliterated and analyzed. Moreover, unlike anthroponyms which were consistently written in the nominative case, toponyms appeared in the nominative, genitive and prepositional cases, sometimes even with prepositions. This variability necessitates a broader and more detailed explanation and analysis, leading to the decision not to address such a large number of toponyms. Instead, 25 proper nouns were selected to allow for a more focused analysis. In addition to these 25 toponyms considered, four common nouns, also categorized as toponyms even though not as proper nouns of cities, regions or villages, were included. These terms refer to different types of administrative units unique to Russian culture and history and lack direct translations into Italian. Consistently with the purpose of this thesis and the methodology adopted, the selection aims at highlighting the challenges that the systems encountered in processing Russian-specific words and graphemes. To ensure a comprehensive representation of all passports, at least one proper noun from each passport was included. During the selection process, numerous linguistic challenges were identified; however, only the most recurrent issues were ultimately considered. This approach was taken to demonstrate the potential variations in the processing of the same graphemes and words within a specific system. When analyzing the toponyms, it must be considered that in the TL, they are typically rendered in their base form. This means that the declinations are not included into the word itself; instead, cases are indicated by prepositions, as Italian does not employ a case system. Therefore, significant variation can be observed between SL and TL, involving an intermediate, implicit step in transitioning from the SW to the final TW. The toponyms in the following tables are presented individually, rather than within complete sentences, to save space and ensure clarity of the analysis. However, the complete source and target documents can be found in the appendix.

The first table of toponyms presented below shows Language Weaver's outputs for selected terms. Unlike previous analyses, there are few discrepancies in the processing of graphemes. In fact, only two main differences have been identified. The first is a minor variation in the rendering of the grapheme "я", which, except when following the vowel "и" (as in the toponym "Казачия"), has always been transliterated as "ya". In this particular toponym, however, it was rendered as "a", likely because the presence of the preceding vowel led the machine to process the graphemes together, resulting in "ia" (Kazachia). The second difference, potentially more significant due to the lack of influence from surrounding letters, involves the rendering of the grapheme "к". While this consonant has typically been rendered as "k" in the TL, a variation can be observed within the toponym "в городе Семикаракорске". Specifically, Language Weaver rendered this term as "nella città di Semicarakorsk", resulting in the first "к" being transliterated as "c", whereas the other two are transliterated as "k". Although this inconsistency appeared in only one of the 25 toponyms analyzed, it introduces discrepancies in the processing of identical graphemes. In addition, there are instances of terms that have been literally translated rather than transliterated. For example, two toponyms have been partially translated into English, suggesting a possible intermediate step in Language Weaver's processing, where texts are first translated into English and then from English into Italian. Specifically, "Тепло-Огаревского р-на" was rendered as "del Heat-Ogarevskogo r-on", with "тепло" literally translated as "heat", and the genitive case indicated by the Russian suffix "-ого" translated as the Italian preposition "del" (instead of the English preposition “of”). Another example is “п. Серебряный бор”, which resulted in “p. Silver bor”, with “silver” being the literal translation of “серебряный”. Another (mis)translation occurred for “гор. Грязи”, which resulted in “delle montagne. Terra”. Focusing on the proper noun, which is the one posing problems of transliteration, the machine interpreted it as “terra” probably in the sense of “fango” (EN: mud), demonstrating the machine’s literal translation approach. Besides transliteration inconsistencies, some observations could be made regarding the processing of the abbreviations “гор.” and its variants with the preposition “в г.” and “в гор.”, which stand for “город”, that is “city”. However, while this is not the primary focus of this thesis, it is worth noting them briefly for sake of completeness. For instance, in “гор. Грязи”, the abbreviation was literally translated as “delle montagne” (EN: of the mountains), thus rendering the genitive plural form of“гора”, the nominative form for “mountain”. This mistranslation occurred again with “гор. Пенза”, processed as “delle montagne. Penza”. In another case, the common noun was transliterated with an additional final letter, possibly to reflect the genitive form of the SL, as seen in Гор. Нижнего Тагила > gora. Nizhny Tagil. In the toponym “в гор. Устюжна”, the abbreviation was translated in the singular form, thus resulting is “in montagna. Ustyuzhna”, as the form indicated the prepositional case. Finally, to conclude on the processing of these abbreviations, of the eight instances where “в г.” appeared, the system processed them as “a” half of the time and as “in” the other half. Another consideration on abbreviations can be made on the abbreviation of the word “район”, here translated as “distretto” (EN: district), particularly for the one found in the toponym “Тепло-Огаревского р-на”. The system failed to recognize the abbreviation “р-на” in the toponym, which consequently was neither transliterated nor translated, in comparison to the other three cases where the noun was written in the full form, in both the prepositional and genitive case (в Куйбышевском районе, в Кирово-Чепецком районе, Любытинского района). Lastly, attention should be given to the toponyms with the suffix “-ский/цкий”, indicating the adjectival form. In the SL, nominative forms never occur, thus the toponyms use the singular, masculine adjectival form in the genitive (-ого) or prepositional case (-ом) because they refer to the masculine nouns район and округ, both rendered as “district”. When processing these terms, the system twice transliterated and transformed them into the nominative form (в Кирово-Чепецком районе > nel distretto Kirov-Chepetsky, Любытинского района > nel distretto Lyubytinsky), once retained the genitive form (Тепло-Огаревского р-на > del heat-ogarevskogo r-on), and twice rendered the noun derived from the adjective (в Куйбышевском районе > nel distretto di Kuibyshev, в Прикубанском округе > nel distretto Prikuban).

| SOURCE TOPONYMS | TARGET TOPONYMS - LANGUAGE WEAVER |
|---|---|
| В Г. ЧЕРНЯХОВСК | A CHERNYAKHOVSK |
| В Г. КАЗАНЬ | IN KAZAN |
| ГОР. НИЖНЕГО ТАГИЛА | GORA. NIZHNY TAGIL |
| В Г. СЫКТЫВКАР | IN SYKTYVKAR |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | NEL DISTRETTO DI KUIBYSHEV |
| В Г. ВЯЗЬМА | IN VYAZMA |
| ЧЕЛЯБИНСКОЙ ОБЛ. | REGIONE DI CHELYABINSK |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | NEL DISTRETTO DI KIROV-CHEPETSKY |
| ГОР. ПЕНЗА | delle MONTAGNE. PENZA |
| ХУТОР КАЗАЧИЯ | IL VILLAGGIO DI KAZACHIA |
| В Г. КИРИШИ | A KIRISHI |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | NELLA REGIONE DI TYUMEN |
| ЛЮБЫТИНСКОГО РАЙОНА | NEL DISTRETTO DI LYUBYTINSKY |
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | DEL HEAT-OGAREVSKOGO R-ON |
| В Г. ШАДРИНСК | A SHADRINSK |
| П. СЕРЕБРЯНЫЙ БОР | P. SILVER BOR |
| В ПРИКУБАНСКОМ ОКРУГЕ | NEL DISTRETTO PRIKUBAN |
| В Г. КЕМЬ | IN KEM |
| ВО ЧУВАШСКОЙ АССР | NELL'ASSR DI CHUVASH |
| ГОР. ГРЯЗИ | delle MONTAGNE. TERRA |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | NELLA CITTÀ DI SEMICARAKORSK |
| В Г. ЖЕЛЕЗНОГОРСК | A ZHELEZNOGORSK |

| | |
|---|---|
| СТ. БАГАЕВСКАЯ | S... BAGAEVSKAYA |
| ЛИПЕЦКОЙ ОБЛ. | DELLA REGIONE DI LIPETSK |
| В ГОР. УСТЮЖНА | IN MONTAGNA. USTYUZHNA |

*Table 10. Language Weaver transliteration of toponyms.*

The following table illustrates Modern MT results for the 25 selected toponyms. When considering inconsistencies among graphemes, two major discrepancies can be identified. Firstly, the consonant "ц", was transliterated differently in the two toponyms where it appears: "в Кирово-Чепецком районе" and "Липецкой обл.". In the first instance, the grapheme was rendered with the digraph "ts", resulting in "nel distretto di Kirovo-Chepetsky". However, in the second term, it was transliterated with the single grapheme "c", leading to the toponym "della regione di Lipeck". The second major inconsistency relates to the soft vowel "ю". This grapheme appeared three times in different toponyms, and Modern MT processed it in two different ways: once, it was transliterated with the digraph "ju" (по Тюменской области > per la regione di Tjumen) and twice with the digraph "yu" (Любытинского района > del distretto di Lyubytinsky, в гор. Устюжна > nella città di Ustyuzhna). In addition, there are three toponyms that were translated into English, even though the TL was set to Italian. Specifically, the toponym "п. Серебряный бор" was literally translated into English as "p. Silver forest", with only the abbreviation for the common noun "посёлок" being transliterated. Similarly, in the toponym "гор. Грязи" the proper name of the city was literally translated into English, even adopting the genitive form. The system likely identified a resemblance to the genitive form of the common noun "грязь", which, as mentioned in the previous paragraph, means "mud". Moreover, it interpreted the abbreviation "гор." as if it were the short form for the genitive "города", resulting in "della città di Mud". Finally, the system's performance with toponyms translated into English was inconsistent when compared to the previous two instances. While it transliterated the proper noun, it rendered the preposition for the prepositional case in English, resulting in the toponym being rendered as "in the Chuvash ASSR". Focusing briefly on abbreviations, of the eight instances "в г." appeared, it was extended in the TL as "nella città" twice, and both times it even was formatted differently from the original. In the other instances, the output was consistent. Additionally, there is a discrepancy in the processing of the abbreviation “гор.”. When it appeared in the toponym “гор. Пенза”, it was simply omitted in the TL, whereas in the aforementioned case of “гор. Грязи”, it was mistranslated.

| SOURCE TOPONYMS | TARGET TOPONYMS – MODERN MT |
|---|---|
| В Г. ЧЕРНЯХОВСК | A CHERNYAKHOVSK |
| В Г. КАЗАНЬ | A KAZAN |
| ГОР. НИЖНЕГО ТАГИЛА | DELLA CITTÀ DI NIZHNY TAGIL |
| В Г. СЫКТЫВКАР | A SYKTYVKAR |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | NEL DISTRETTO DI KUIBYSHEVSKY |
| В Г. ВЯЗЬМА | NELLA città DI VYAZMA |
| ЧЕЛЯБИНСКОЙ ОБЛ. | REGIONE DI CHELYABINSK |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | NEL DISTRETTO DI KIROVO-CHEPETSKY |
| ГОР. ПЕНЗА | PENZA |
| ХУТОР КАЗАЧИЯ | KHUTOR KAZACHIYA |
| В Г. КИРИШИ | A KIRISHI |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | per LA REGIONE DI TJUMEN |
| ЛЮБЫТИНСКОГО РАЙОНА | DEL DISTRETTO DI LYUBYTINSKY |
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | DISTRETTO DI TEPLO-OGAREVSKY |
| В Г. ШАДРИНСК | NELLA città DI SHADRINSK |
| П. СЕРЕБРЯНЫЙ БОР | P. SILVER FOREST |
| В ПРИКУБАНСКОМ ОКРУГЕ | NEL DISTRETTO Prikubansky |
| В Г. КЕМЬ | A KEM |
| ВО ЧУВАШСКОЙ АССР | IN THE CHUVASH ASSR |
| ГОР. ГРЯЗИ | DELLA CITTÀ DI MUD |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | NELLA CITTÀ DI SEMIKARAKORSK |
| В Г. ЖЕЛЕЗНОГОРСК | A ZHELEZNOGORSK |
| СТ. БАГАЕВСКАЯ | Bagaevskaya St |

| | |
|---|---|
| ЛИПЕЦКОЙ ОБЛ. | della regione DI LIPECK |
| В ГОР. УСТЮЖНА | NELLA CITTÀ DI USTYUZHNA |

*Table 11. Modern MT transliteration of toponyms.*

As happened for anthroponyms, Intento and DeepL gave the same outputs also for the 25 toponyms selected. Therefore, the results have been unified in a single table. There are two transliteration mistakes that can easily be identified in the table: the first one occurred in the first toponym, where the vowel "е" was transliterated as "o" in the TL; the second one can be seen in the toponym "п. Серебряный бор", which in the TL was rendered as "P. Serebryaniyaniy bor", thus involving a duplicate of the final part of the word (-aniy). Furthermore, in the third toponym "гор. Нижнего Тагила", the proper noun was transliterated retaining the genitive form, rather than providing the nominative, base form. As a result, the output in the TL included the genitive case the suffixes for the masculine adjective (-его) and for the masculine noun (-а). Some inconsistencies can be identified also in the processing of the soft sign "ь". While it was omitted when found inside the toponym "в г. Вязьма" (a Vyazma), in the common noun "область" found in the toponym "по Тюменской области" it was rendered with the apostrophe (per l'Oblast' di Tumen). On the other hand, in the third toponym it is rendered with the consonant "y", resulting in в г. Кемь > a Kemy. The abovementioned toponym "по Тюменской области" presents another discrepancy when considering the soft vowel "ю". This grapheme is found also in the toponyms "Любытинского района" and "в гор. Устюжна". While the processing of this grapheme is consistent across the first two toponyms, where it is rendered as "u" (per l'Oblast' di Tumen, distretto di Lubytinsk) in the third toponym the systems transliterated it with the digraph "yu" (nella città di Ustyuzhna). Another discrepancy can be seen between the toponym "ст. Багаевская" where the vowel "е" has been transliterated with the digraph "ye", resulting in "st. Bagayevskaya", and all the other terms where this grapheme is found, which the system simply transliterated as "e". As happened with the other systems, the toponym "гор. Грязи" was translated rather than transliterated. The machine interpreted the proper noun as the plural form of the common noun "грязь", which was literally translated into Italian as "fanghi" (EN: mud). Moreover, in this toponym it is worth noticing also the abbreviation, which contrarily to the other two instances where "гор." appeared, it was transliterated rather than translated. Consequently, the whole toponym resulted in being "gor. fanghi". One final inconsistency in transliteration involves toponyms ending with the adjectival suffix in the prepositional or genitive case (-ском/-ской/-ского). Specifically, while in the cases of "Тепло-Огаревского р-на" and "во Чувашской АССР" the systems rendered the nouns in the nominative case of adjectival form, in the other four instances the suffix was simply omitted (в Куйбышевском районе > nel distretto di Kuibyshev, в Кирово-Чепецком районе > nel distretto di Kirovo-Chepetsk, Любытинского района > distretto di Lubytinsk, в Прикубанском округе > nel distretto di Prikuban). In addition, there is a discrepancy in the transliterations of the two toponyms that retained the adjectival form: in the former, the suffix was transliterated with "iy", while in the latter, it was rendered with "y". To conclude the comments on this table, it is worth noting that the abbreviations "в г." and "в гор." were mostly rendered in full form in the TL, explicitly translating the term "city". However, in three instances, the output only included the preposition "a". Finally, as anticipated earlier, the abbreviation "гор." once has been transliterated, once omitted and once simply translated as "città".

| SOURCE TOPONYMS | TARGET TOPONYMS – INTENTO and DEEPL |
|---|---|
| В Г. ЧЕРНЯХОВСК | nella città di Chornyakhovsk |
| В Г. КАЗАНЬ | nella città di Kazan |
| ГОР. НИЖНЕГО ТАГИЛА | NIZHNEGO TAGILA |
| В Г. СЫКТЫВКАР | nella città di Syktyvkar |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | nel distretto di Kuibyshev |
| В Г. ВЯЗЬМА | a Vyazma |
| ЧЕЛЯБИНСКОЙ ОБЛ. | REGIONE DI CHELYABINSK. |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | nel Distretto di Kirovo-Chepetsk |
| ГОР. ПЕНЗА | CITTÀ. PENZA |
| ХУТОР КАЗАЧИЯ | KHUTOR KAZACHIYA |
| В Г. КИРИШИ | nella città di Kirishi |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | per l'Oblast' di Tumen |
| ЛЮБЫТИНСКОГО РАЙОНА | Distretto di Lubytinsk |

| | |
|---|---|
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | DISTRETTO DI TEPLO-OGAREVSKIY |
| В Г. ШАДРИНСК | nella città di Shadrinsk |
| П. СЕРЕБРЯНЫЙ БОР | P. Serebryaniyaniy BOR |
| В ПРИКУБАНСКОМ ОКРУГЕ | nel DISTRETTO DI PRIKUBAN |
| В Г. КЕМЬ | a KEMY |
| ВО ЧУВАШСКОЙ АССР | VO CHUVASHSKY ASSR |
| ГОР. ГРЯЗИ | GOR. FANGHI |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | NELLA CITTÀ DI SEMIKARAKORSK |
| В Г. ЖЕЛЕЗНОГОРСК | a Zheleznogorsk |
| СТ. БАГАЕВСКАЯ | ST. BAGAYEVSKAYA |
| ЛИПЕЦКОЙ ОБЛ. | DELLA REGIONE DI LIPETSK. |
| В ГОР. УСТЮЖНА | nella città di Ustyuzhna |

*Table 12. Intento and DeepL transliteration of toponyms.*

This last table shows Yandex transliteration of toponyms. The outputs given by the system do not present many discrepancies in terms of graphemes inconsistencies. The main difference that can be noticed is related to those nouns having adjectival suffixes in prepositional or genitive cases (-ском/-цком/-ской/-ского). In four of the instances the adjectival form was not retained in the TL, rendering the toponyms as simple nouns (в Куйбышевском районе > nel distretto di Kuibyshev, по Тюменской области > nella regione di Tyumen, в Прикубанском округе > nel distretto di Prikuban, во Чувашской АССР > in Chuvash ASSR); on the other hand, the other three place names were rendered in the adjectival nominative form with the suffix "y" (в Кирово-Чепецком районе, Любытинского района, Тело-Огаревского р-на). In this last toponym mentioned there are two more aspects to be taken into consideration. Firstly, Yandex failed to transliterate the first part of the name, rendering the final "o" of "Тепло" as "aya". Secondly, the system misinterpreted the abbreviation "р-на", treating the final part as if it were a preposition equivalent to the Italian "sulla" (EN: on). Consequently, the final output in the TL was rendered as "Teplaya-Ogarevsky r-sulla". There are three additional instances where the system opted for translation rather than transliteration of place names. The toponym “хутор Казачия” was literally translated into “della fattoria cosacca”, which is the Italian for “of the Cossack farm”; the system might have mistaken it for the adjective “казацкий”, meaning “Cossack”. The second translated toponym is “п. Серебряный бор”, which was rendered as “p. Silver bor”, thus literally translating only the first part of the proper noun, and incorrectly into English instead of Italian. The third term translated is “гор. Грязи”; in this case the system translated both the common noun and the proper noun of the city. For the proper noun, the machine likely interpreted the name as the common noun “грязь”. Additionally, the abbreviation for “город” was interpreted as the genitive plural form of the word “гора” (EN: mountain). This same misinterpretation occurred also for another toponym, “гор. Пенза”, which was rendered as “delle montagne. Penza”. There is a third instance of this abbreviation among the 25 toponyms, гор. Нижнего Тагила, which was totally omitted in the TL. Lastly, the abbreviation “в гор.” in the last toponym was transliterated, adding a final “e”, likely due to the influence of the preposition in the SL, indicating the prepositional case.

| SOURCE TOPONYMS | TARGET TOPONYMS – YANDEX |
|---|---|
| В Г. ЧЕРНЯХОВСК | a Chernyakhovsk |
| В Г. КАЗАНЬ | a Kazan |
| ГОР. НИЖНЕГО ТАГИЛА | NIZHNY TAGIL |
| В Г. СЫКТЫВКАР | a Syktyvkar |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | nel distretto di Kuibyshev |
| В Г. ВЯЗЬМА | a Vyazma |
| ЧЕЛЯБИНСКОЙ ОБЛ. | REGIONE DI CHELYABINSK. |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | nel distretto di Kirovo-CHEPETSKY |
| ГОР. ПЕНЗА | delle montagne. PENZA |
| ХУТОР КАЗАЧИЯ | della fattoria cosacca |
| В Г. КИРИШИ | a Kirishi |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | nella regione di Tyumen |
| ЛЮБЫТИНСКОГО РАЙОНА | Distretto di Lyubytinsky |
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | TEPLAYA-OGAREVSKY R-SULLA |
| В Г. ШАДРИНСК | a Shadrinsk |
| П. СЕРЕБРЯНЫЙ БОР | P. Silver Bor |

| | |
|---|---|
| В ПРИКУБАНСКОМ ОКРУГЕ | nel distretto di PRIKUBAN |
| В Г. КЕМЬ | a Kem |
| ВО ЧУВАШСКОЙ АССР | in Chuvash ASSR |
| ГОР. ГРЯЗИ | delle montagne. FANGO |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | nella città di Semikarakorsk |
| В Г. ЖЕЛЕЗНОГОРСК | a Zheleznogorsk |
| СТ. БАГАЕВСКАЯ | St. BAGAEVSKAYA |
| ЛИПЕЦКОЙ ОБЛ. | REGIONE DI LIPETSK. |
| В ГОР. УСТЮЖНА | a Gore. Ustyuzhna |

*Table 13. Yandex transliteration of toponyms.*

## 3.4 COMPARISON ON ACCURACY

The machine translation systems analyzed show numerous inconsistencies, both within each system when transliterating identical graphemes, as discussed in the previous paragraphs, and across different systems. Therefore, it is crucial to compare them against the selected norms to determine which one, despite its discrepancies, has performed most effectively. The following sections are dedicated to a quantitative and qualitative comparison between the MT outputs and the norms considered, highlighting discrepancies and differences among graphemes.

Prior to the qualitative analysis, a quantitative analysis was also conducted to provide a more comprehensive investigation of these tools. Regarding the performance of Language Weaver eleven first names were translated, eight of which were incorrectly translated into English rather than Italian. Four names present one discrepancy in comparison with the norm, whereas only six names fully adhere to the norm. Considering patronymics, only seven of them do not align with the norm. Across surnames, there is more variation, as all female surnames were transliterated as genitive forms, four were literally translated into Italian and only one did not align with the norm, whereas seven of them fully adhere with the Decree's table. In toponyms, paying attention mainly to proper nouns, only three out of twenty-five align with the norm. However, only three more had part of them which was translated instead of transliterated, and the remaining seventeen were incorrectly transliterated. Besides literal translation and correct or incorrect transliteration, another difference that was highlighted in the category of toponyms is the rendering of adjectival endings present in seven proper nouns; three remained adjectives and four were converted into nouns. The processing of common nouns in toponyms is also another topic that presented some inconsistencies. Language Weaver translated into Italian the abbreviation "гор." three times, as if it was the genitive, plural form of the Russian term for "mountains". Once, this abbreviation was simply transliterated, as happened also for the other abbreviations "ст." and "р-на".

Modern MT translated only five first names, four of which match Language Weaver's translations; it is the engine that performed better when transliterating first names aligned with the norm, but there are still eight names incorrectly transliterated. When it comes to surnames, it is the second-best engine, since it only translated two surnames and transliterated fourteen of them according to the norm. However, among the six inconsistently transliterated surnames, one was rendered with the simple preposition "di" to indicate a genitive construction, while in two cases the final "-a" of the female surname is missing, leading to the perception that the last names belong to males. In addition, similarly to the remaining three engines, it inconsistently transliterated only seven patronymics out of twenty-two. Considering toponyms, Modern MT performed few translations and instead it is the engine that performed the majority of aligned transliteration across all the engines, although they are just four transliterations. Among the nineteen transliterations performed, also the seven terms with adjectival suffixes are included; five of them retained the original endings and two were converted into nouns. Except for the abbreviation "в г." which was again always rendered with the Italian preposition to indicate locations, the abbreviations were not literally translated but rather either simply transliterated, as happened for "п." and "ст." or written in the full form.

Intento and DeepL performed forty-one aligned transliterations of anthroponyms. There are only three instances of literal translation, which can all be found in first names. In terms of incorrect transliterations, these engines failed to properly perform transliteration in eleven first names and only four surnames. The remaining surnames were properly transliterated according to the norm; thus, none were literally translated. As happened for Modern MT, fifteen patronymics were transliterated aligning with the norm and seven were not. In toponyms, only one place name was literally translated. The number of correctly transliterated toponyms aligns with Language Weaver: three toponyms adhered to the norm, while the remaining twenty-one showed incorrect transliterations. Focusing on adjectival endings, they processed them contrarily to Modern MT, as only two adjectival endings were retained, and the other five toponyms were converted into nouns. Considering the abbreviations, only once the abbreviation "гор." was transliterated and not provided in the full form or omitted.

Lastly, Yandex performed more literal translations, nine, than proper transliterations, eight, for first names. In contrast, for last names it properly transliterated eleven surnames and translated only four. Consequently, seven were not aligned with the norm, as happened for patronymics; in fact, in this case fifteen were consistent with the Decree's norm, and none were translated, similar to the previous three systems mentioned. Focusing on toponyms, in Yandex only three toponyms were consistent with the UNGEGN report, while three more were literally translated. Thus, nineteen toponyms did not align with the report. Among the ones which did not align there were also the toponyms with adjectival suffix: three of them retained the adjectival form, and the remaining four were rendered as nouns. In addition, Yandex also failed in recognizing and rendering the abbreviations. For instance, "гор." was literally translated twice, omitted once, and simply transliterated once, retaining the inflected suffix. The tables below summarize the data presented, providing a concise overview of the qualitative analysis conducted. In the table of anthroponyms, each cell contains the numbers of target terms that either followed or deviated from the transliteration norms, as well as the number of terms that were translated. They are divided according to the category they belong to. On the other hand, the table of toponyms shows a simpler representation since there are no different classifications. However, to avoid providing an excessive amount of data, this table only contains the rendering of proper place names, leaving out the common nouns.

| | | LW | mMT | IN | DL | YA |
|---|---|---|---|---|---|---|
| ALIGNED TRANSLITERATIONS | FIRST NAMES | 6 | 9 | 8 | 8 | 8 |
| | PATRONYMICS | 14 | 15 | 15 | 15 | 15 |
| | LAST NAMES | 7 | 14 | 18 | 18 | 11 |

| | | | | | | |
|---|---|---|---|---|---|---|
| NOT ALIGNED TRANSLITERTIONS | FIRST NAMES | 5 | 8 | 11 | 11 | 5 |
| | PATRONYMICS | 8 | 7 | 7 | 7 | 7 |
| | LAST NAMES | 11 | 6 | 4 | 4 | 7 |
| TRANSLATIONS | FIRST NAMES | 11 | 5 | 3 | 3 | 9 |
| | PATRONYMICS | 0 | 0 | 0 | 0 | 0 |
| | LAST NAMES | 4 | 2 | 0 | 0 | 4 |

*Table 14. Anthroponyms qualitative analysis results.*

| | LW | mMT | IN | DL | YA |
|---|---|---|---|---|---|
| ALIGNED TRANSLITERATIONS | 3 | 4 | 3 | 3 | 3 |
| NOT ALIGNED TRANSLITERTIONS | 19 | 19 | 21 | 21 | 19 |
| TRANSLATIONS | 3 | 2 | 1 | 1 | 3 |

*Table 15. Toponyms qualitative analysis results.*

### 3.4.1 LANGUAGE WEAVER

The system acquired by RWS presented notable inaccuracies when processing anthroponyms. One of the most significant inconsistencies involved the translation of some first and last names, which occurred mainly in English, despite the TL being Italian. Additionally, it failed to transliterate the first name "Оксана" and the last name "Сорока", retaining all original graphemes. Specifically, the graphemes "к", "с" and "н", as well as "с", "р" and "к" were not transliterated as "k", "s", and "n", and "s", "r" and "k", respectively, as they were in all other anthroponyms containing these letters. Furthermore, the vowels were also not transliterated, despite the Cyrillic and Latin letters having identical graphic forms. Focusing on graphemes, the digraph "кс" introduced some inconsistencies both in first names and patronymics, as it was transliterated as "x", merging the sounds of individual letters. However, according to the Decree's table, the source digraph should have been rendered with the target digraph "ks". Moreover, also the single grapheme "к" was inconsistently transliterated in first names, sometimes being rendered as “c” instead of “k”. Other inconsistencies related to vowels are also identified. In particular, the soft vowel “ю” in first names was never transliterated as “iu”, as suggested by the table, but was instead consistently rendered as either “yu” or “ju”. Similarly, the soft vowel “я” in patronymics was transliterated as “ya”, rather than as “ia”, contrary to its consistent transliteration according to the Decree’s table in first names. The corresponding hard vowel “a”, when functioning as a suffix in female surnames, was never reflected in the TL. Instead, the system appeared to interpret it as a genitive form of masculine nouns, introducing the preposition “di” and omitting the suffix. However, as previously discussed, the preposition also appears before some masculine surnames. Consequently, there are numerous discrepancies, many of which lack a clear rationale. Lastly, an additional inconsistency identified in Language Weaver is the processing of “ье”. When the vowel appeared after the soft sign, it was always rendered either with “ie” or “ye” and never simply as “e”.

Focusing on toponyms, the primary transliteration inconsistencies involve two graphemes previously discussed in relation to anthroponyms. The soft vowel “я” should have been transliterated as “ja”, but it was instead inconsistently rendered as either “ya” or “a”. Moreover, the consonant “к” was variably transliterated as either “k” or “c”, whereas the UNGEGN Report’s table only permits the grapheme “k”. Further considerations apply to “ь” and “ю”, which were not consistently transliterated according to the table. In this system, the soft sign was always omitted, and the soft vowel was always rendered as “yu”. There were also discrepancies concerning toponyms with adjectival endings, which should have been rendered in the nominative form or as nouns. However, one retained the adjectival inflected form, while the others were either converted into nouns or transformed in the nominative case. Lastly, considering abbreviations of designations, the machine failed in rendering them in the full form and in translating or transliterating them when necessary. The abbreviations standing for “город” when paired with the preposition “в” that indicates the prepositional case, were rendered either with “a” or “in”.

### 3.4.2 MODERN MT

Analyzing the outputs provided from Modern reveals several inconsistencies similar to those found in Language Weaver. First, Modern MT failed in transliterating the first name “Оксана” and the last name “Сорока”. Additionally, the system inconsistently translated some first names into English and literally translated some surnames into Italian. However, Modern MT performed better than Language Weaver, as it prioritized transliteration over translation for most names. One notable issue is related to the digraph “кс”, which was rendered as “x” in first names but as “ks” in patronymics, consistently with the Decree’s recommendations. In contrast, the system transliterated the grapheme “я” as “ya” across all categories of anthroponyms considered, except when it was preceded by the soft sign “ь” or the vowel “и”. In these cases, which appeared only in first names, the digraph was rendered as “ia”. The table suggests that this digraph corresponds solely to “я”, because the vowel “и” also has its own corresponding letter, “i”. Therefore, the combinations “ья” and “ия” should not be transliterated simply as “ia”. This issue can also be partially identified in Language Weaver’s processing of patronymics. Another similarity can be identified in the processing of “ье”, which was transliterated as either “ie” or “’e”, thus representing the soft sign with an apostrophe. According to the table, the soft sign should be omitted, and the vowel should be represented by “e”. Furthermore, the letter “ч” at the end of male patronymics was transliterated inconsistently as either “ch” or “č”, despite the table’s recommendation for “ch”. Lastly, the consonant “к” was transliterated inconsistently in first names: it was correctly rendered as “k” at the beginning of names, as suggested by the table, but as “c” in other positions, as also occurred in Language Weaver.

In toponyms, Modern MT consistently omitted the soft sign “ь” in the TL, contrarily to anthroponyms, as occurred in Language Weaver. Another similar approach involved the transliteration of the soft vowels “я” and “ю”. In Language Weaver, “я” was transliterated as either “ya” or “a”, while in Modern MT it was rendered only as “ya”, still not aligning with the table’s requirements. Similarly, “ю” was transliterated with the variants “ju” and “yu”, whereas the table only allows “ju”. An additional inconsistency in Modern MT concerns the grapheme “ц”, which was rendered as either “ts” and “c”, even though the table only allows “c”. Furthermore, Modern MT performed similarly to Yandex and Language Weaver, as they literally translated certain terms that should have been transliterated. Another aspect worth highlighting is the discrepancy in processing nouns having adjectival suffixes, which were mostly rendered as adjectives in the nominative form, except for two that were transformed into nouns. Interestingly, as noted in the next section, the following systems processed these terms in the opposite manner. Lastly, regarding abbreviations, none were literally translated; the main difference lies in the rendering of the preposition, which, however, does not create major inaccuracies.

### 3.4.3 INTENTO AND DEEPL

Since Intento and DeepL produced identical results across all categories of proper nouns, also the performance on their accuracy is discussed in one paragraph. Like the previous systems, neither Intento nor DeepL correctly transliterated the first name "Оксана" and the last name "Сорока". Moreover, these systems translated some first names in a manner consistent with Language Weaver and Modern MT. However, Intento and DeepL performed fewer literal translations in comparison to the other three systems. For instance, none of the surnames was translated. On the other hand, there are more inconsistencies related to the digraph "ье". Although it was sometimes transliterated as "e", as specified by the table, there were other instances where it was processed similarly to Language Weaver, thus rendered as "ye" or "ie". Additionally, a discrepancy in contrast with the other three systems is the vowel "е" at the beginning of first names, which once was transliterated as "ye" and once as "e". A further issue involved the letter "е" when paired with another "е" in two patronymics. They were not consistently translated, since once were rendered as the table would suggest, that is as "ee", while in the other instance they appeared as the trigraph "eye". Discrepancies were also present in the handling of the digraph "кс" and of the grapheme "к". The digraph was rendered either as "x" or as "ks" both in first names and patronymics, whereas the grapheme was transliterated as "c" or "k", the latter being consistent with the Decree's table. Other considerations on graphemes that have already been discussed in previous paragraphs include, for example, the semi vowel "ю", which was rendered as either "iu" or "u" in first names. However, the table suggests using the digraph "iu". Likely, the soft vowel "я" was rendered as "ya" when found alone or after the soft sign, and "ia" when combined with the vowel "и". These considerations are similar to those made for Modern MT. Finally, in last names, these systems once transliterated the consonant "з" as "s" rather than "z", which was used consistently in other contexts and systems differently from the other categories and from the other systems and is aligned with what the table suggests.

As anticipated in the previous paragraph, Intento and DeepL processed adjectival suffixes in the opposite manner compared with Modern MT, retaining only two adjectival forms and converting the others into nouns. However, even though only two adjectival forms were kept, they still displayed inconsistent transliteration, as one ended only with "y" and the other with "iy". Moreover, as in Language Weaver, an inflected toponym was not rendered in the nominative case, thus retaining the genitive suffixes "-ego" and "-a" for the adjective and the noun, respectively. Another similarity with Modern MT can be identified in the processing of abbreviations, which were not literally translated. In addition, these two systems performed better when dealing with proper nouns that resemble common nouns, with only one term being literally translated. In terms of graphemes, as with previous systems, the soft vowel "я" was transliterated as "ya" instead of "ja". There was also variability in the rendering of "ю", which in this case was either transliterated as "u" or "yu", thus never aligning with the UNGEGN table's suggestions. Moreover, while the soft sign is usually omitted or rarely rendered with an apostrophe, in these systems its rendering varied significantly: it was represented by an apostrophe, omitted, or once even rendered as the consonant "y". One last unique to Intento and DeepL, involves the vowel "e", which is supposed to be transliterated as "e", but was also rendered as "ye" and once even as "o". This latter case likely resulted from the system misinterpreting the beginning of the SW, which might have resembled the Russian word for the color "red".

### 3.4.4 YANDEX

Yandex system does not introduce new inconsistencies beyond those present in previous systems. The most significant difference can be identified when considering last names. Yandex literally translated three surnames into Italian, one of which was never transliterated by the other systems. In addition, it transformed the surname "Херло" to resemble an English surname. Besides these inconsistencies, all other discrepancies have already been discussed in previous systems. More specifically, nine first names were translated into English or Italian. The digraph "кс" was always rendered as "x" in patronymics, and the grapheme "к" was transliterated either as "k" or "c", as occurred in all other systems. Also, the processing of "ье" and "ю" can be considered consistent with other systems, which do not follow the table's suggestions: Yandex respectively rendered them as “ie”, as all other systems, and “yu” or “ju”, as in Language Weaver, Intento and DeepL. One last consideration on discrepancies in anthroponyms can be done for the soft vowel “я”, and the hard vowel “a” as a suffix in female surnames. The former was transliterated as “ya”, instead of “ia” when found alone in first names, and simply as “ia” when found after “ь” or “и”, thus merging the two sounds of the digraph “ия”. The latter was omitted in some female surnames, as happened in Language Weaver, but not rendered as genitive case. Unlike the previous systems, Yandex consistently transliterated “ee” as “ee” and “ч” as “ch”, as the table would suggest.

Devoting the attention to toponyms, Yandex also fails to consistently process terms with adjectival suffix, rendering them similarly to Language Weaver: nearly half were transformed into nouns, while the others retained the adjectival form. However, in this case, none kept their inflected form, and all were rendered in the nominative case. Nonetheless, the system might have merged the final sounds of the nominative case of adjectives, rendering it only as “y” and not as “ij”, as the UNGEGN would suggest. Yandex shares a considerable number of similarities with Language Weaver. For instance, “ю” was transliterated as “yu” instead of “ju”, and the soft sign was always omitted. In addition, “я” was rendered as “ya”. In contrast, there are no discrepancies for the consonant “к”, which was always transliterated as “k”. As happened in Language Weaver, three proper nouns were literally translated rather than transliterated, although only two of these correspond. Lastly, also the processing of abbreviations in Yandex resembles that of Language Weaver for certain designations. Specifically, “гор.” was literally translated, and the abbreviation “р-на”, standing for the genitive form of “район”, was literally rendered as “r-sulla”, thus misinterpreting the last part of the abbreviation as a preposition.

## 3.5 CHAT GPT-4 OMNI PROMPT

After analyzing the outputs from different MTs and identifying numerous discrepancies in the consistency of transliteration within the same systems and across the various systems, a prompt was given to ChatGPT-4 Omni. As anticipated earlier, the system generated 22 TTs that present anthroponyms and toponyms transliterated according to the selected norms.

The following tables illustrate ChatGPT anthroponyms transliterated according to the norm suggested by the Decree of the Russian Ministry of Foreign Affairs. Few discrepancies can be identified among the outputs generated for first names. Considering the grapheme "й", it was transliterated with two different graphemes, depending on the preceding vowel. When found after the vowel "e", as in "Алексей", "Андрей" and "Сергей", it was rendered as "j" (Aleksej, Andrej, Sergej). On the other hand, when the grapheme followed the vowels "и" or "а", as in "Евгений" and "Николай", it was transliterated as "i" (Evgenii, Nikolai). The second inconsistency observed is related to the grapheme "я". Three distinctions can be made based on its position in a name: when it appeared at the beginning of a first name, it was transliterated as "ia" (Яков > Iakov); when it appeared at the end of a first name, preceded by the vowel "и", it was rendered either with "ia" or "iia", with the latter rendering preserving the distinction between the two original letters (Виктория > Viktoria, Юлия > Iuliia). Lastly, when it appeared after the soft sign "ь", it was rendered either with the digraph "ia" or "ja" (Дарья > Daria, Татьяна > Tatjana).

| SOURCE FIRST NAMES | TARGET FIRST NAMES |
|---|---|
| АЛЕКСАНДР | ALEKSANDR |
| АЛЕКСЕЙ | ALEKSEJ |
| АНДРЕЙ | ANDREJ |
| БОРИС | BORIS |
| ВЕРОНИКА | VERONIKA |
| ВИКТОРИЯ | VIKTORIA |
| ВЛАДИМИР | VLADIMIR |
| ДАРЬЯ | DARIA |
| ЕВГЕНИЙ | EVGENII |
| ЕКАТЕРИНА | EKATERINA |
| ИГОРЬ | IGOR |
| ИРАИДА | IRAIDA |
| КРИСТИНА | KRISTINA |
| ЛЮДМИЛА | LIUDMILA |
| МИХАИЛ | MIKHAIL |
| НИКОЛАЙ | NIKOLAI |

| | |
|---|---|
| ОКСАНА | OKSANA |
| СВЕТЛАНА | SVETLANA |
| СЕРГЕЙ | SERGEJ |
| ТАТЬЯНА | TATJANA |
| ЮЛИЯ | IULIIA |
| ЯКОВ | IAKOV |

*Table 16. ChatGPT-4 Omni transliteration of first names.*

As happened for first names, also in the table of patronymics few inconsistencies can be observed. Specifically, the first discrepancy involves the last letter of the suffix in male patronymics. Of the eleven instances of male patronymics, the grapheme "ч" was transliterated with the digraph "ch" eighth times, while in the remaining instances, it was transliterated with the diacritic "č". The second discrepancy can be identified in four of these eleven male patronymics, as well as in the female patronymic "Геннадьевна". In these instances, the soft sign "ь" appears before the vowel "е". However, the rendering of this sign varies: in two cases, it was rendered with an apostrophe and the vowel "e" was simply rendered as "e" (Григорьевич > Grigor'evič, Леонтьевич > Leont'evich); in the other three cases, the soft sign was omitted, and the vowel was transliterated with the digraph "ie" (Васильевич > Vasilievich, Геннадьевна > Gennadievna, Евгеньевич > Evgenievich).

| SOURCE PATRONYMICS | TARGET PATRONYMICS |
|---|---|
| АЛЕКСАНДРОВИЧ | ALEKSANDROVICH |
| АНДРЕЕВИЧ | ANDREEVIČ |
| ВАЛЕНТИНОВНА | VALENTINOVNA |
| ВАСИЛЬЕВИЧ | VASILIEVICH |
| ВИКТОРОВНА | VIKTOROVNA |
| ВЛАДИМИРОВНА | VLADIMIROVNA |
| ВЯЧЕСЛАВОВИЧ | VIACHESLAVOVICH |
| ГЕННАДЬЕВНА | GENNADIEVNA |
| ГЕОРГИЕВИЧ | GEORGIEVICH |
| ГРИГОРЬЕВИЧ | GRIGOR'EVIČ |
| ДМИТРИЕВИЧ | DMITRIEVIČ |

| | |
|---|---|
| ЕВГЕНЬЕВИЧ | EVGENIEVICH |
| КОНСТАНТИНОВИЧ | KONSTANTINOVICH |
| ЛЕОНИДОВНА | LEONIDOVNA |
| ЛЕОНТЬЕВИЧ | LEONT'EVICH |
| МАКСИМОВНА | MAKSIMOVNA |
| МИХАЙЛОВНА | MIKHAILOVNA |
| НИКОЛАЕВНА | NIKOLAEVNA |
| ОЛЕГОВИЧ | OLEGOVICH |
| ПЕТРОВНА | PETROVNA |
| СЕРГЕЕВНА | SERGEEVNA |
| СТЕПАНОВНА | STEPANOVNA |

*Table 17. ChatGPT-4 Omni transliteration of patronymics.*

Finally, to conclude the discussion on ChatGPT-4 Omni outputs, the following table displays the transliteration of last names. There are some inconsistencies in the processing of the soft sign "ь" among last names as well. In two instances, the soft sign was treated differently: once it was omitted (Кузьменко > Kuzmenko) and once it was represented by the apostrophe (Стольникова > Stol'nikova). Another discrepancy is observed with the surname "Сорока", which was left untransliterated. In this case, all source graphemes were preserved in the TL, likely because the system might have interpreted them as Latin characters and thus did not require transliteration.

| SOURCE LAST NAMES | TARGET LAST NAMES |
|---|---|
| АГАПОВА | AGAPOVA |
| АНДРЕЕВА | ANDREEVA |
| БАРАНОВ | BARANOV |
| БРАГИН | BRAGIN |
| ВОРОНОВ | VORONOV |
| ЗВЕРИНЦЕВА | ZVERINTSEVA |
| ИМЯРЕК | IMIAREK |
| КОЗЕРОДОВ | KOZERODOV |
| КОЗЛОВ | KOZLOV |
| КРЫЛОВ | KRYLOV |

| | |
|---|---|
| КУЗЬМЕНКО | KUZMENKO |
| МАРИНИНА | MARININA |
| САЗОНОВА | SAZONOVA |
| СИДОРОВ | SIDOROV |
| СОРОКА | COPOKA |
| СТОЛЬНИКОВА | STOL'NIKOVA |
| ТАШКИНА | TASHKINA |
| ТРОФИМЕНКО | TROFIMENKO |
| ХЕРЛО | KHERLO |
| ЧУМАКОВ | CHUMAKOV |
| ШАПОШНИКОВА | SHAPOSHNIKOVA |
| ЩЕРБАКОВА | SHCHERBAKOVA |

*Table 18. ChatGPT-4 Omni transliteration of last names.*

The following table presents the toponyms transliterated by GPT-4 Omni according to the table created by UNGEGN Working Group on Romanization Systems. The discrepancies observed are related to the graphemes "ч" and "я", as well as to the rendering of the abbreviations "в г." and "в гор.", although the latter mainly involves stylistic differences. Regarding the first grapheme, it appears in four different instances: in three cases, it is found at the beginning of the toponym and is transliterated with the diacritic "č", whereas in one case, within the toponym "Казачия", it is rendered with the digraph "ch" (Kazachia). The second grapheme appears six times. Even in this case the position of the grapheme likely influences its transliteration. When found within the term, it was transliterated as "ja" (e.g., Вязьма > Vjaz'ma, Серебряный бор > Serebrjanyj bor), whereas at the end of a term, it was rendered "ia" (Багаевская > Bagaevskaia). An interesting case to highlight is the rendering of "во Чувашской АССР", where the abbreviation was expanded in its full form and the whole toponym was translated. This is the only instance of translation among those seen until now which cannot be considered a mistake. The reason this toponym should be translated and not transliterated is the existence of documented forms of the proper noun in online maps and resources concerning Autonomous Soviet Socialist Republics, even though some variations of the name are present. The variations in the name are likely related to the terms used to refer to its inhabitants. For instance, in Sergio Salvi's work "La disunione sovietica" (1990), this proper noun is recorded as “Ciuvassia”, with its inhabitants called “ciuvassi”. In contrast, other sources, such as the online encyclopedia Treccani, use the term “ciuvasci”, from which the toponym “Ciuvascia” is likely derived. Lastly, two final brief considerations can be made to mention the aspects analyzed in the previous sections, so as to provide a more comprehensive investigation and comparison of the outputs. Firstly, starting with toponyms having adjectival suffix, none was retained, and each term was converted into the noun form. Secondly, briefly focusing on abbreviations, the system recognized all abbreviations, translated them and, most importantly, rendered them in the full form. Moreover, the system opted for translating also the terms which might have required transliteration, such as “хутор”, which was rendered as “villaggio”. Paying attention at consistency, the abbreviations for “city” (“г.” and “гор.”) paired with the preposition “в” to indicate prepositional case were all rendered as “nella città”, except for three instances where the Italian preposition is missing. However, this would be a minor inconsistency as it is only related to the understanding and the processing of cases and prepositions, and not of actual toponyms.

| SOURCE TOPONYMS | TARGET TOPONYMS |
|---|---|
| В Г. ЧЕРНЯХОВСК | nella città di Černjachovsk |
| В Г. КАЗАНЬ | nella città di Kazan’ |
| ГОР. НИЖНЕГО ТАГИЛА | della città di Nižnij Tagil |
| В Г. СЫКТЫВКАР | nella città di Syktyvkar |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | nel distretto Kujbyšev |
| В Г. ВЯЗЬМА | nella città di Vjaz'ma |
| ЧЕЛЯБИНСКОЙ ОБЛ. | Regione di Čeljabinsk. |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | nel distretto Kirovo-Čepeck. |
| ГОР. ПЕНЗА | Città di Penza |
| ХУТОР КАЗАЧИЯ | Villaggio di Kazachija |
| В Г. КИРИШИ | nella città di Kiriši. |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | nella regione di Tjumen |
| ЛЮБЫТИНСКОГО РАЙОНА | nel distretto di Ljubytino |
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | Distretto di Teplo-Ogarev |
| В Г. ШАДРИНСК | nella città di Šadrinsk |

| | |
|---|---|
| П. СЕРЕБРЯНЫЙ БОР | Insediamento di Serebrjanyj Bor |
| В ПРИКУБАНСКОМ ОКРУГЕ | nel distretto di Prikubansk |
| В Г. КЕМЬ | città di Kem' |
| ВО ЧУВАШСКОЙ АССР | nella Repubblica Socialista Sovietica Autonoma Ciuvascia. |
| ГОР. ГРЯЗИ | Città di Grjazi |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | nella città di Semikarakorsk |
| В Г. ЖЕЛЕЗНОГОРСК | città di Železnogorsk |
| СТ. БАГАЕВСКАЯ | Villaggio di Bagaevskaia |
| ЛИПЕЦКОЙ ОБЛ. | regione di Lipeck |
| В ГОР. УСТЮЖНА | città di Ustjužna |

*Table 19. ChatGPT-4 Omni transliteration of toponyms.*

## 3.5.1 CHAT GPT-4 OMNI ACCURACY

Despite being instructed to transliterate anthroponyms and toponyms according to specified norms, ChatGPT-4 Omni did not produce outputs fully consistent with the established requirements. This section shows an analysis of the few inconsistencies generated, providing some reasoning to try finding an explanation for the system's outputs. Firstly, focusing on the quantitative aspect, the following tables sum up the performance of ChatGPT, providing the data for each category. Considering anthroponyms, none were literally translated and only eleven did not align with the instructions provided in the prompt. Among the three categories, last names provided the best outputs, as twenty surnames out of twenty-two adhered to the norm. However, in the category of first names, which is the one where there are less correct transliterations, only three less names did not adhere to the norm. Looking at toponyms, there is one instance of translation, which is not considered a mistake but rather the correct rendering of the toponym, as previously mentioned. In addition, only four toponyms showed some differences in comparison to the standard considered; thus ChatGPT correctly transliterated the remaining twenty toponyms.

| CHAT GPT-4 OMNI | | | |
|---|---|---|---|
| | FIRST NAMES | PATRONYMICS | LAST NAMES |
| ALIGNED TRANSLITERATIONS | 17 | 18 | 20 |
| NOT ALIGNED TRANSLITERATIONS | 5 | 4 | 2 |
| TRANSLATIONS | 0 | 0 | 0 |

*Table 20. ChatGPT's anthroponyms qualitative analysis results.*

| CHAT GPT-4 OMNI | |
|---|---|
| | TOPONYMS |
| ALIGNED TRANSLITERATIONS | 20 |
| NOT ALIGNED TRANSLITERATIONS | 4 |
| TRANSLATIONS | 1 |

*Table 21. ChatGPT 's toponyms qualitative analysis results.*

Subsequently, paying attention to the qualitative aspect, one of the most evident inconsistencies in anthroponyms concerns the processing of the soft sign “ь”. The Decree’s table does not include the soft sign, which might suggest that it can be omitted when encountered in personal names. However, while it does not appear in any of the TL first names containing it in the SL, the system treated it differently in patronymics and last names. In these cases, the soft sign was either omitted or replaced with an apostrophe. Another discrepancy involves the semivowel “й”, which, according to the table, should have been transliterated as “i”. In practice, however, first names containing this character were transliterated inconsistently, using both “j” and “i”, while in patronymics it was consistently transliterated as “i”. Additionally, the soft vowel “я”, which the table indicates should be rendered as “ia”, was correctly transliterated in patronymics but appeared as both “ia” and “ja” in first names. Furthermore, the consonant “ч” displayed notable inconsistencies in patronymics. Although it should have consistently been rendered as “ch”, according to the table, as happened in some instances, the system also transliterated it as “č”. In contrast, the consonant was correctly transliterated as “ch” when it appeared as the initial letter in a surname. Lastly, a significant inaccuracy in GPT-4

Omni's performance pertains to the surname "Сорока". The graphemes "с", "р" and "к" were not transliterated as "s", "r" and "k", respectively, as they were in all other anthroponyms containing these letters. Moreover, the vowels were also not transliterated, despite the Cyrillic and Latin letters having identical graphic forms.

Considering the toponyms, which should have been consistently transliterated according to the table provided in the UNGEGN Report, GPT-4 Omni shows two major transliteration discrepancies. These discrepancies are similar to those found in the anthroponyms, involving the letters "ч" and "я". However, for toponyms, the transliterations suggested by the Report are opposite to those recommended by the Decree for anthroponyms. Specifically, the grapheme "ч" was again either transliterated as "č" or "ch", whereas the table indicates that it should have been consistently rendered as "č". Similarly, the soft vowel "я" was transliterated both as "ia" and "ja", although only the latter is allowed according to the table. In comparison with MT engines, none of the abbreviations were literally translated or simply transliterated. In this case, the system was able to recognize and expand the abbreviations, providing the correct translation of words.

# CONCLUSION

The primary aim of this thesis was to provide concrete insights into the domain of machine translation, analyzing and assessing the outputs of five MT engines that translated twenty-two Russian passports into Italian. More specifically, the focus was narrowed to transliteration of anthroponyms and toponyms, so as to offer a different perspective from the more common studies on MT quality. Moreover, to include a further step in the evaluation and future developments of MT engines, an example of possible improvements following recent studies is suggested, basing the initial investigation on ChatGPT-4Omni abilities to comprehend requests and generate desired outputs. Additionally, as the adoption of CAT tools alongside MT applications has revolutionized the translation industry, improving not only efficiency but also accuracy and consistency of translations, one of the initial motivations to carry out this study was the possibility to further facilitate human translators' job in performing translations, also with the perspective to further improving MT engines within CAT tools, so to avoid having to copy and paste automated transliterations from external resources. Consequently, tools that can automatically suggest appropriate transliterations would significantly streamline the translation process.

Personal and place proper nouns are considered a subcategory of the broader classification of realia. These are terms for which there is not a direct equivalent in the TL or for which the direct translation is not appropriate considering the context. Thus, the most suitable strategy in such cases would be transliteration. The main issues identified, which motivated this project, were the discrepancies and inconsistencies observed among words with similar features when processed by the same engine. One of the possible reasons explaining inconsistent outputs might be related to lack of appropriate training of the engines for Russian and Italian languages. MT engines are often trained with limited corpora for specific language pairs, such as Russian and Italian, which might not include sufficient examples of proper nouns. Thus, the systems' ability to generalize and apply previously learned knowledge might be affected when it encounters less common names. In this specific instance, the five different MTs, more or less frequently used English phonetics in the target terms, likely because these engines might be primarily trained with English, which thus serves as an intermediary between Russian and Italian. As a result, the STs underwent an additional step rather than being directly processed into the selected TL. Possible evidence of this statement could be seen in the overall rendering of the source texts; while the sections that required translation were rendered into Italian, anthroponyms and toponyms occasionally adopted an anglicized form. Moreover, transliteration is not a standardized strategy, despite systems and organizations that try to suggest globally agreed norms to avoid misinterpretations and issues given by different rendering of the same SW. One of the most common transliteration norms is the Scientific Transliteration system which is usually taught at academic level. However, this norm does not align with ISO 9:1995, which should instead provide a globally agreed norm considering its international scope. The main differences concern whether digraphs are used to represent single graphemes, and the subsequent use of diacritics to compensate for digraphs' absence. More importantly, standardization would avoid different representations of the same grapheme across different languages. In fact, languages adapt foreign sounds to their own phonetic systems, which consequently result in one or more graphical signs for just one source grapheme. Therefore, it is not surprising that one grapheme may be transliterated with a digraph or a quadrigraph in two different languages. One aspect which could explain the existence and adoption of various transliteration norms is that different word categories might effectively require different transliteration norms. The key aspect though, is that there must not be inconsistencies within the same category. Considering that the focus terms of this thesis are anthroponyms and toponyms, two different sets of norms have been considered. On the one hand, the table in the Ministry's Decree used to transliterate anthroponyms could not be suitable for toponyms; on the other hand, the norm suggested by the United Nations Group of Experts on Geographical, as evident from the same name, could be suitable only for toponyms. The main difference between the table in the Decree and the report by the UNGEGN is the use of one or more graphemes in the former, while in the same instances the latter opted for diacritics. The following list displays the graphemes that are rendered differently: ж > zh/ž, х > kh/h, ц > ts/c, ч > ch/č, ш > sh/š, щ > shch/šč. Thus, in line with the aim of this thesis, the entire project involved translating the passports into Italian, while adhering to standardized norms for processing anthroponyms and toponyms, providing a scientific foundation to rely on. By following this methodology, the issue of having different transliterations based on national phonetic systems for the same term would be overcome, leading to fewer difficulties in rendering the original STs into various TLs.

Therefore, also when machine transliterating terms the results may be inconsistent because of various transliteration systems with which the engines have been trained. In addition, these inconsistencies slow down the translator's work, who is consequently forced to resort to post-editing. Post-editing, through which translators revise MT engines output, is useful to overcome the increasingly high demand for translations. However, the need to post-edit inconsistent transliterations significantly affects the efficiency of human translators, decreasing the time-saving benefits of MT engines, as translators are required to manually correct numerous small mistakes. Nowadays, NMT is the most widespread and accurate system in performing translation. It is based on deep neural networks, and despite the challenges they may bring, advancements and ongoing research are fundamental to improve the functioning of these architectures in this field. Enhancements in the performance and accuracy of NMT systems brought them to be highly required and effectively useful to help human translators. One of the downsides of NMT is the high-cost demands when training the engines. Over the past few years instead of continuously training the engines, new strategies have been adopted. The increasing use of AI systems and LLMs to carry out NLP tasks started to spread also in the domain of MT and there are some studies that proved the efficiency in prompting the engines, providing instructions to guide the process, so as to obtain the desired result. The attention mechanism, included also in NMT engines, and found at the basis of LLMs allows systems to focus on specific graphemes, thus improving the likelihood of accurate transliteration. This method was used also in this thesis, after observing many discrepancies in the MT outputs. However, as will be subsequently explained, despite being given the norms to follow, the system still generated some mistakes in processing the focus items. Thus, while LLMs might excel in comprehension and generation of grammatically correct answers, they may lack specialized training to handle specific tasks, such as transliteration across languages with significant differences.

When translating the passports with the five MT engines, the systems did not adopt a consistent strategy to process the words. In fact, while the majority of words were transliterated, some were literally translated, others were translated into a different TL, and some had their graphemes inconsistently transliterated. Although ChatGPT-4Omni was asked to generate twenty-two passports, following the provided template five sections required the insertion of anthroponyms or toponyms. As a result, the material to be analyzed was extended beyond just twenty-two simple personal and place names. In Russia, each person has three anthroponyms, which consequently led to the analysis of sixty-six anthroponyms. On the other hand, a different choice was made for toponyms. Since most toponyms were repeated across the two sections where they appeared, they were not analyzed twice. In addition, both sections included at least two toponyms each, thus there would have been a large number of toponyms for analysis. However, only a selection of twenty-five toponyms was analyzed, since they did not belong to different categories like anthroponyms did. The only classification that could be made was identifying cities, regions or villages names, where the primary difference only lies in the administrative units they represent. Additionally, three realia represented by common nouns (хутор, область, станица) were included. Considering their nature as realia, they should have been transliterated.

The norms considered are included in the appendix at the end of the thesis work. In addition, there are eight tables displaying the SWs transliterated, including the column with transliteration according to the relevant norms, in order to provide a direct comparison. Making some general considerations, among the two categories, in toponyms there are fewer literal translations, but their transliterations align less with the norms, thus requiring more post-editing. Patronymics in all engines are the category of anthroponyms which deviate less from the standard and present less problems when being processed. Out of twenty-two, more than half of them, thirteen, were properly transliterated by all engines. In addition, there are two terms, the first name "Оксана" and the last name "Сорока" which were never transliterated by the engines, but always rendered as the original SW, probably because the systems interpreted the graphemes as belonging to Latin alphabet, thus not requiring transliteration. Moreover, there are some consistent patterns across engines. For example, the first names "Вероника" and "Виктория" were rendered with the Italian spelling "Veronica" and "Victoria" in all five engines, thus preferring the translation over the transliteration. This might suggest that the engines gave more importance to frequency or familiarity of the terms, as they appeared in training data, rather than to actual transliteration. Considering the TL, this adoption might not be considered completely wrong, but it must be generally stated and agreed that in such cases phonetic adaptation is preferred over transliteration according to standardized norms. Language Weaver and Yandex frequently align in inconsistent processing of terms, usually providing translations rather than transliterations. Similarly, ModernMT, Intento and DeepL provide similar transliterations which do not align with the chosen norm, or on the other hand, provide proper transliteration of terms. In general, however, only four first names and six last names have been properly transliterated by all engines. On the other hand, only three toponyms were correctly transliterated by all engines. The following tables sum up the data presented up until now, so as to provide a comprehensive overview of the discussion.

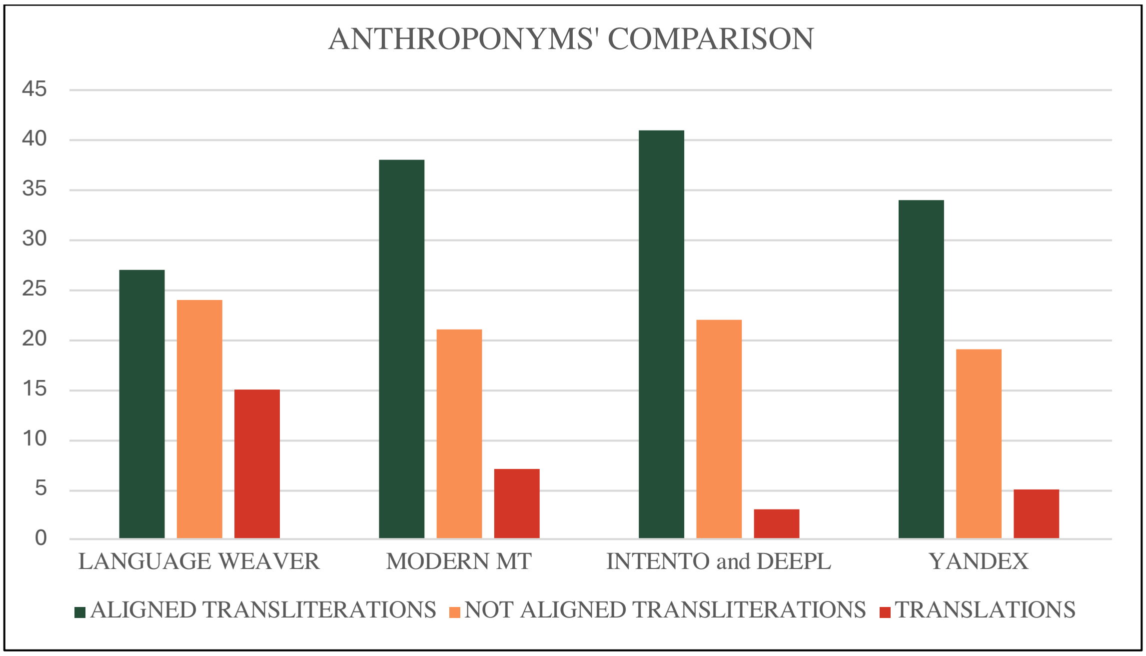



*Figure 3. Anthroponyms' comparison diagram.*

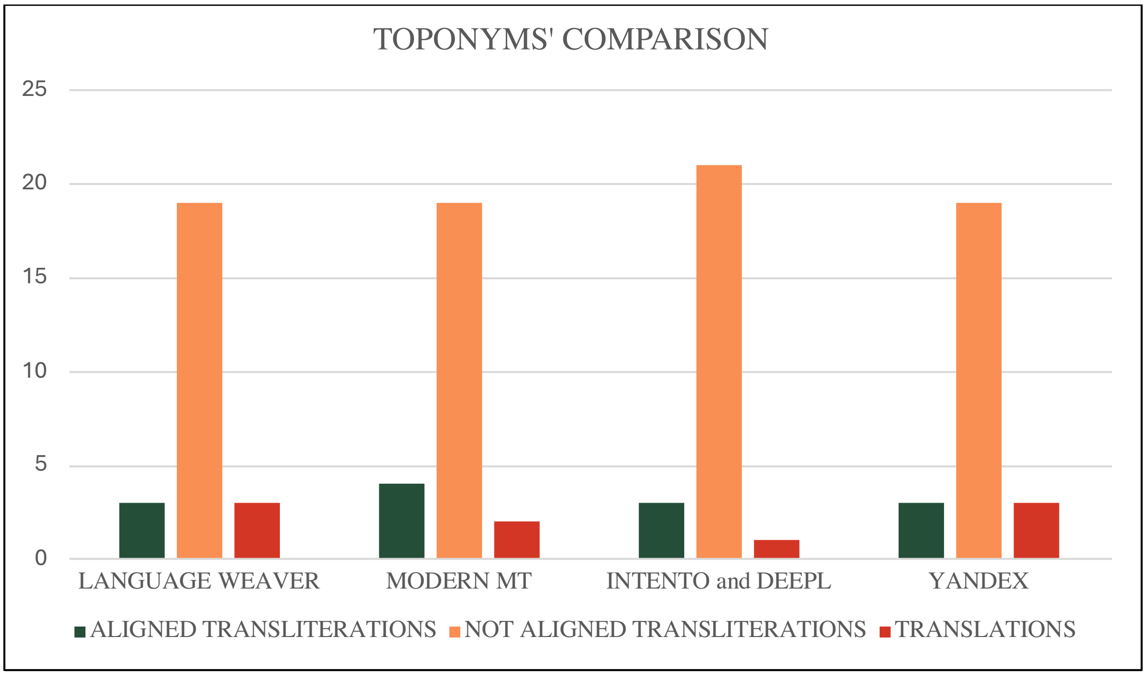



*Figure 4. Toponyms' comparison diagram.*

The parameters used to rank the five engines when dealing with the outputs for anthroponyms and toponyms in Russian passports do not totally follow traditional evaluation metrics such as BLEU, TER, COMET or METEOR. They are not considered fully appropriate for the intended analyses of this thesis, due to the fact that they mainly focus on syntax and semantics (Intento, 2023). However, there are some similarities between their approach and the approach adopted in this thesis, as both methods involve comparing candidate translations with reference translations. The key parameter for ranking the engines was the number of transliterated terms that align with the norms. Subsequently, between not aligned transliterations and translations, more importance was given to the former, since providing inaccurate transliterations would require less post-editing than translations. In addition, given that these terms are supposed to be transliterated, performing translations is a completely inappropriate approach that diminishes the performance of the engines.

Considering anthroponyms and toponyms together, Intento and DeepL were the best-performing engines, as out of the sixty-six anthroponyms, forty-one were properly transliterated and only three were translated, while the remaining twenty-two showed inconsistencies among graphemes. Considering toponyms, although they did not perform the most accurate transliteration according to the norm, the difference from ModernMT, which performed better, is small; only one fewer toponym was inaccurately transliterated. In addition, only one toponym was literally translated. However, what must be noted of Intento and DeepL is that they sometimes repeated the last toponym in the full passport formatting. For instance, although "Сыктывкар" was correctly transliterated, the city's name was repeated at the end of the string, resulting "per la Repubblica di Komi nella città di Syktyvkar. SYKTYVKAR", as can be seen from the texts in the appendix. Therefore, despite the correct transliteration, post-editing is still required to refine the final output. The outputs of these two engines have always been analyzed and commented on together, as they produced identical outputs across all the investigated categories. This is likely due to the Smart Routing mechanism adopted by Intento, which probably relied on DeepL's automated translation. ModernMT ranked second overall, not only for surnames, but also when considering both anthroponyms and toponyms. In fact, it accurately transliterated thirty-eight anthroponyms and four toponyms, while only seven anthroponyms and two toponyms were translated. Therefore, twenty-one anthroponyms did not adhere to the norm, one fewer than Intento and DeepL. Across toponyms, nineteen deviated from the norm, two fewer than the best-performing engines identified. Yandex, which could potentially be the best-performing engine, was the second-worse engine when considering both anthroponyms and toponyms. It literally translated thirteen anthroponyms, only two fewer than Language Weaver, but it properly transliterated thirty-four anthroponyms, just four fewer than ModernMT. Consequently, Yandex produced the fewest transliterations that were inconsistent with the norm. For toponyms, its performance is aligned with Language Weaver. Language Weaver performed worse than the other engines across all categories, correctly transliterating only twenty-seven anthroponyms and three toponyms. It also performed the highest number of literal translations in both major categories, fifteen for anthroponyms and three for toponyms. Although three literal translations might seem small, it was still the highest number among the engines. If the category of toponyms is separated from anthroponyms, it can be noticed also from the diagram above that ModernMT outperformed the other engines, since it has the highest number of correct transliterations, four. Intento and DeepL rank second, as they produced the fewest translations among all engines, and Language Weaver and Yandex rank third.

As previously mentioned, recent studies suggest using prompts with LLMs to refine outputs and achieve desired results, without the need for retraining the engines. In the present study, ChatGPT-4Omni was instructed both with the Ministry's norm and UNGEGN's report, so as to generate accurate transliterations. Despite being given rules to follow, some inconsistencies remained, though they were clearly fewer than those found in the MT engines. The diagram below shows that the number of terms adopting correct transliteration is much higher than terms with not aligned transliteration. Among sixty-six anthroponyms, eleven did not adhere to the norm, showing inconsistencies mainly for the soft sign, the semivowel "й", the soft vowel "я", and the consonant "ч". However, generally speaking, the three categories of anthroponyms are quite well aligned, since the number of correct transliterations range only from seventeen to twenty. Another aspect worth highlighting is the rendering of the surname "Сорока", which posed some difficulties for the MT engines and was not correctly transliterated by ChatGPT either, despite being instructed with specific transliteration rules. It is likely that the system again considered the letters as Latin characters, which did not need to be transliterated. In contrast, the first name "Оксана", which encountered similar issues in MT engines processing, did not show any inconsistency in this processing. Regarding toponyms, four of them showed some differences in comparison to the standard considered. An interesting instance in this category is the translation of the toponym "во Чувашской АССР", which cannot be considered a mistake, since the Italian rendering generated by ChatGPT is documented and established. In fact, transliteration should be adopted only for terms which do not already have a documented translation in the target language. Thus, this is the only instance when translation is not considered a mistake. A comparison could be made with well-known Russian cities, such as "Москва" and "Санкт Петербург", rendered in English as "Moscow" and "Saint Petersburg", and in Italian as "Mosca" and "San Pietroburgo". Since these terms are widely recognized, transliterating them would be a mistake. Lastly, consistency in GPT's outputs is also evident in its handling of abbreviations, all of which were expanded to their full form. Additionally, none of the toponyms with adjectival suffixes retained these endings in the outputs, which instead presented only the noun forms.

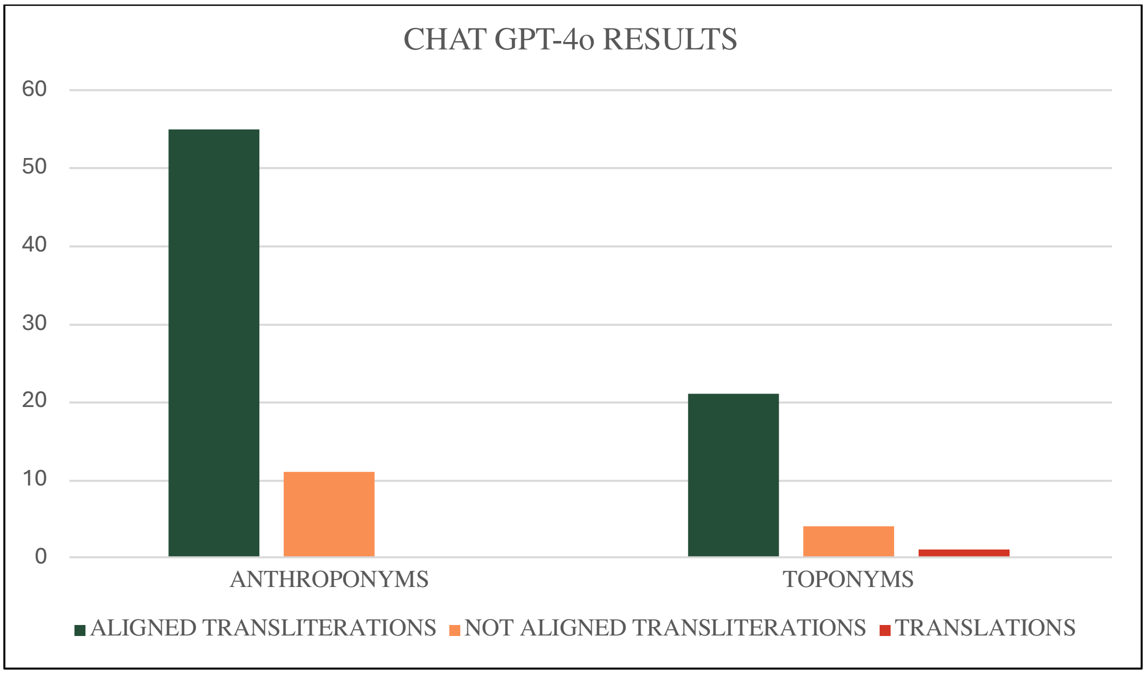



*Figure 5. Chat GPT-4o results diagram.*

This study explored the challenges some MT engines might face when handling "translation problems" commonly encountered in everyday documents, specifically Russian passports, where such problems were addressed through romanization, using modern, widely spread tools that could assist translators' work. However, further research is needed, as the rapidly evolving field of AI continues to offer new possibilities when dealing with NLP tasks, making this area of study subject to ongoing developments and improvements. Future studies could explore a wider range of realia, beyond just personal and place names. For instance, further focus could be devoted to common nouns falling under the category of realia. Words such as "область" or "округ" (these are Russian terms indicating territorial units usually recalling concepts like a region or province, and a district, county, or region, respectively) could either be translated or transliterated, taking into consideration that the translation could provide a simplification or adaptation to TC, which might not necessarily be the right strategy for this domain. In the outputs of ChatGPT, some common nouns which could be considered realia, were translated. In this context adopting transliteration rather than translation might not be the correct strategy, since in the domain of passports it might be more important to provide an administrative unit, rather than retaining the original term. However, the topic of common noun realia is not sufficiently investigated in the present study to draw accurate conclusions; therefore, it is only mentioned for sake of completeness, so as not to deviate from the original aim. The task of deciding whether to translate or transliterate becomes even more complex when there are cultural and legal considerations that play a role together. Also retaining or converting adjectival suffixes may be another aspect to be investigated. In Russian, many place names present adjectival endings such as "-ский", which pose challenges for MT engines. If not handled properly, these suffixes can lead to incomplete or inaccurate transliterations, influencing the accuracy of the final output. Therefore, further studies could focus on the strategies MT engines use to process these items and suggest some solutions on how to overcome these discrepancies.

Considering the topic of prompt engineering, finetuning the prompt given to ChatGPT specifying that all anthroponyms and toponyms are written in Cyrillic characters, would probably avoid terms not transliterated, such as the surname "Сорока". For instance, after adding to the original prompt the sentence "All anthroponyms and toponyms are written in Cyrillic letters.", the system effectively rendered the surname as "Soroka". However, this serves only as a potential starting point for developing the topic, because it is also necessary to understand why the first name "Оксана", which contains all letters belonging also to Latin alphabet, was correctly interpreted and processed by the system, whereas the surname was not.

The issue of transliteration standardization remains a central challenge, as there is not a commonly agreed norm for training the system. Thus, this inconsistent application of norms results in multiple possible outputs, which can lead to confusion. Future efforts to enhance the accuracy of the systems could involve developing transliterations models which can distinguish between anthroponyms, toponyms, and other cultural words, so that appropriate transliteration norms are applied contextually.

Furthermore, to enhance the efficiency of translators' work, the integration of automated transliteration resources within CAT tools would be beneficial to overcome the difficulties and the time-consuming practice of manual transliteration of terms. In addition to training machine translation engines integrated into CAT tools to generate accurate translations and transliterations, which ideally would require least possible post-editing, the development and integration of transliteration resources based on standardized transliteration norms would provide tools that are suitable for a broader, global audience.

Clearly, these studies and enhancements would require massive work and updating of applications and maps that use transliteration and translations strategy to make them accessible to a wide audience. In any case, developing new studies and strategies to address these aspects would be helpful to understand which is the path that should be followed to aim for a universal standardization, in such instances where it is required.

# APPENDIX

1. Norms' tables.
   1.1. Table from the 2020 Decree of Russian Ministry of Foreign Affairs.

| Транслитерация кириллических знаков (извлечение) | | |
|---|---|---|
| N п/п | Национальный знак | Рекомендуемая транслитерация |
| 1. | А | A |
| 2. | Б | B |
| 3. | В | V |
| 4. | Г | G |
| 5. | Д | D |
| 6. | Е | E |
| 7. | Ё | E |
| 8. | Ж | ZH |
| 9. | З | Z |
| 10. | И | I |
| 11. | Й | I |
| 12. | К | K |
| 13. | Л | L |
| 14. | М | M |
| 15. | Н | N |
| 16. | О | O |
| 17. | П | P |
| 18. | Р | R |
| 19. | С | S |
| 20. | Т | T |

| | | |
|---|---|---|
| 21. | У | U |
| 22. | Ф | F |
| 23. | Х | KH |
| 24. | Ц | TS |
| 25. | Ч | CH |
| 26. | Ш | SH |
| 27. | Щ | SHCH |
| 28. | Ы | Y |
| 29. | Ъ | IE |
| 30. | Э | E |
| 31. | Ю | IU |
| 32. | Я | IA |

1.2 Report on the current status of United Nations Romanization systems for Geographical Names.

| | | |
|---|---|---|
| 1. | А а | a |
| 2. | Б б | b |
| 3. | В в | v |
| 4. | Г г | g |
| 5. | Д д | d |
| 6. | Е е | e |
| 7. | Ё ё | ë |
| 8. | Ж ж | ž |
| 9. | З з | z |
| 10. | И и | i |
| 11. | Й й | j |
| 12. | К к | k |
| 13. | Л л | l |
| 14. | М м | m |
| 15. | Н н | n |
| 16. | О о | o |
| 17. | П п | p |
| 18. | Р р | r |
| 19. | С с | s |
| 20. | Т т | t |
| 21. | У у | u |
| 22. | Ф ф | f |
| 23. | Х х | h |
| 24. | Ц ц | c |

| | | |
|---|---|---|
| 25. | Ч ч | č |
| 26. | Ш ш | š |
| 27. | Щ щ | šč |
| 28. | Ъ ъ | " |
| 29. | Ы ы | y |
| 30. | Ь ь | ’ |
| 31. | Э э | è |
| 32. | Ю ю | ju |
| 33. | Я я | ja |

2. Tables with source and target anthroponyms.
   2.1. Table of first names: Language Weaver, Modern MT, Intento and DeepL.

| FIRST NAMES | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | LANGUAGE WEAVER | MODERN MT | INTENTO and DEEPL |
| АЛЕКСАНДР | ALEXANDER | ALEXANDER | ALEXANDER |
| АЛЕКСЕЙ | ALEXEY | ALEXEY | ALEKSEY |
| АНДРЕЙ | ANDREW | ANDREY | ANDREY |
| БОРИС | BORIS | BORIS | BORIS |
| ВЕРОНИКА | VERONICA | VERONICA | VERONICA |
| ВИКТОРИЯ | VICTORIA | VICTORIA | VICTORIA |
| ВЛАДИМИР | VLADIMIR | VLADIMIR | VLADIMIR |
| ДАРЬЯ | DARIA | DARIA | DARYA |
| ЕВГЕНИЙ | EUGENE | EUGENE | YEVGENIY |
| ЕКАТЕРИНА | CATHERINE | EKATERINA | Ekaterina |
| ИГОРЬ | IGOR | IGOR | IGOR |
| ИРАИДА | ISRAELE | IRAID | IRAIDA |
| КРИСТИНА | CHRISTINE | KRISTINA | KRISTINA |
| ЛЮДМИЛА | LYUDMILA | LYUDMILA | LUDMILA |
| МИХАИЛ | MICHAEL | MIKHAIL | MIKHAIL |
| НИКОЛАЙ | NICHOLAS | NIKOLAY | NIKOLAY |
| ОКСАНА | OKCAHA | OKCAHA | OKCAHA |

| | | | |
|---|---|---|---|
| СВЕТЛАНА | SVETLANA | SVETLANA | SVETLANA |
| СЕРГЕЙ | SERGEY | SERGEY | SERGEY |
| ТАТЬЯНА | TATIANA | TATIANA | TATYANA |
| ЮЛИЯ | JULIA | YULIA | Yulia |
| ЯКОВ | JACOB | YAKOV | YAKOV |

2.2 Table of first names: Yandex, ChatGPT-4Omni, transliteration according to the decree.

| FIRST NAMES | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | YANDEX | CHAT GPT-4Omni | TRANSLITERATION ACCORDING TO THE DECREE |
| АЛЕКСАНДР | Alexander | ALEKSANDR | ALEKSANDR |
| АЛЕКСЕЙ | ALEKSEY | ALEKSEJ | ALEKSEI |
| АНДРЕЙ | Andrew | ANDREJ | ANDREI |
| БОРИС | Boris | BORIS | BORIS |
| ВЕРОНИКА | Veronica | VERONIKA | VERONIKA |
| ВИКТОРИЯ | Victoria | VIKTORIA | VIKTORIA |
| ВЛАДИМИР | Vladimir | VLADIMIR | VLADIMIR |
| ДАРЬЯ | Daria | DARIA | DARIA |
| ЕВГЕНИЙ | Eugene | EVGENII | EVGENIJ |
| ЕКАТЕРИНА | Catherine | EKATERINA | EKATERINA |
| ИГОРЬ | Igor | IGOR | IGOR |
| ИРАИДА | Iraida | IRAIDA | IRAIDA |
| КРИСТИНА | Cristina | KRISTINA | KRISTINA |
| ЛЮДМИЛА | Lyudmila | LIUDMILA | LIUDMILA |
| МИХАИЛ | Michael | MIKHAIL | MIKHAIL |
| НИКОЛАЙ | Nicholas | NIKOLAI | NIKOLAI |
| ОКСАНА | ОКСАНА | OKSANA | OKSANA |

| | | | |
|---|---|---|---|
| СВЕТЛАНА | Svetlana | SVETLANA | SVETLANA |
| СЕРГЕЙ | Sergei | SERGEJ | SERGEI |
| ТАТЬЯНА | Tatiana | TATJANA | TATIANA |
| ЮЛИЯ | Julia | IULIIA | IULIIA |
| ЯКОВ | Yakov | IAKOV | IAKOV |

2.3 Table of patronymics: Language Weaver, Modern MT, Intento and DeepL.

| PATRONYMICS | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | LANGUAGE WEAVER | MODERN MT | INTENTO and DEEPL |
| АЛЕКСАНДРОВИЧ | ALEXANDROVICH | ALEKSANDROVICH | ALEXANDROVICH |
| АНДРЕЕВИЧ | ANDREEVICH | ANDREEVICH | ANDREYEVICH |
| ВАЛЕНТИНОВНА | VALENTINOVNA | VALENTINOVNA | VALENTINOVNA |
| ВАСИЛЬЕВИЧ | VASILYEVICH | VASIL' EVIČ | VASILEVICH |
| ВИКТОРОВНА | VIKTOROVNA | VIKTOROVNA | VIKTOROVNA |
| ВЛАДИМИРОВНА | VLADIMIROVNA | VLADIMIROVNA | VLADIMIROVNA |
| ВЯЧЕСЛАВОВИЧ | VYACHESLAVOVICH | VYACHESLAVOVICH | VYACHESLAVOVICH |
| ГЕННАДЬЕВНА | GENNADYEVNA | GENNADIEVNA | GENNADYEVNA |
| ГЕОРГИЕВИЧ | GEORGIEVICH | GEORGIEVICH | GEORGIEVICH |
| ГРИГОРЬЕВИЧ | GRIGORIEVICH | GRIGORIEVICH | GRIGORIEVICH |
| ДМИТРИЕВИЧ | DMITRIEVICH | DMITRIEVICH | DMITRIEVICH |
| ЕВГЕНЬЕВИЧ | EVGENIEVICH | EVGENIEVICH | Evgenyevich |
| КОНСТАНТИНОВИЧ | KONSTANTINOVICH | KONSTANTINOVICH | KONSTANTINOVICH |
| ЛЕОНИДОВНА | LEONIDOVNA | LEONIDOVNA | LEONIDOVNA |
| ЛЕОНТЬЕВИЧ | LEONTIEVICH | LEONTIEVICH | LEONTYEVICH |
| МАКСИМОВНА | MAXIMOVNA | MAKSIMOVNA | MAKSIMOVNA |
| МИХАЙЛОВНА | MIKHAILOVNA | MIKHAILOVNA | MIKHAILOVNA |
| НИКОЛАЕВНА | NIKOLAEVNA | NIKOLAEVNA | NIKOLAEVNA |

| | | | |
|---|---|---|---|
| ОЛЕГОВИЧ | OLEGOVICH | OLEGOVICH | OLEGOVICH |
| ПЕТРОВНА | PETROVNA | PETROVNA | PETROVNA |
| СЕРГЕЕВНА | SERGEEVNA | SERGEEVNA | SERGEEVNA |
| СТЕПАНОВНА | STEPANOVNA | STEPANOVNA | STEPANOVNA |

2.4 Table of patronymics: Yandex, ChatGPT-4Omni, transliteration according to the decree.

| PATRONYMICS | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | YANDEX | CHAT GPT-4Omni | TRANSLITERATION ACCORDING TO THE DECREE |
| АЛЕКСАНДРОВИЧ | Alexandrovich | ALEKSANDROVICH | ALEKSANDROVICH |
| АНДРЕЕВИЧ | Andreevich | ANDREEVIČ | ANDREEVICH |
| ВАЛЕНТИНОВНА | Valentinovna | VALENTINOVNA | VALENTINOVNA |
| ВАСИЛЬЕВИЧ | Vasilievich | VASILIEVICH | VASILEVICH |
| ВИКТОРОВНА | Viktorovna | VIKTOROVNA | VIKTOROVNA |
| ВЛАДИМИРОВНА | Vladimirovna | VLADIMIROVNA | VLADIMIROVNA |
| ВЯЧЕСЛАВОВИЧ | Vyacheslavovich | VIACHESLAVOVICH | VIACHESLAVOVICH |
| ГЕННАДЬЕВНА | Gennadievna | GENNADIEVNA | GENNADEVNA |
| ГЕОРГИЕВИЧ | Georgievich | GEORGIEVICH | GEORGIEVICH |
| ГРИГОРЬЕВИЧ | Grigorievich | GRIGOR'EVIČ | GRIGOREVICH |
| ДМИТРИЕВИЧ | Dmitrievich | DMITRIEVIČ | DMITRIEVICH |
| ЕВГЕНЬЕВИЧ | Evgenievich | EVGENIEVICH | EVGENEVICH |
| КОНСТАНТИНОВИЧ | Konstantinovich | KONSTANTINOVICH | KONSTANTINOVICH |
| ЛЕОНИДОВНА | Leonidovna | LEONIDOVNA | LEONIDOVNA |
| ЛЕОНТЬЕВИЧ | LEONTIEVICH | LEONT'EVICH | LEONTEVICH |
| МАКСИМОВНА | MAXIMOVNA | MAKSIMOVNA | MAKSIMOVNA |
| МИХАЙЛОВНА | Mikhailovna | MIKHAILOVNA | MIKHAILOVNA |

| | | | |
|---|---|---|---|
| НИКОЛАЕВНА | Nikolaevna | NIKOLAEVNA | NIKOLAEVNA |
| ОЛЕГОВИЧ | Olegovich | OLEGOVICH | OLEGOVICH |
| ПЕТРОВНА | Petrovna | PETROVNA | PETROVNA |
| СЕРГЕЕВНА | Sergeevna | SERGEEVNA | SERGEEVNA |
| СТЕПАНОВНА | Stepanovna | STEPANOVNA | STEPANOVNA |

2.5 Table of last names: Language Weaver, Modern MT, Intento and DeepL.

| LAST NAMES | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | LANGUAGE WEAVER | MODERN MT | INTENTO and DEEPL |
| АГАПОВА | di AGAPOV | AGAPOV | AGAPOVA |
| АНДРЕЕВА | DI ANDREEV | ANDREEVA | ANDREEVA |
| БАРАНОВ | DELLA PECORA | DELLA PECORA | BARANOV |
| БРАГИН | BRAGIN | BRAGIN | BRAGIN |
| ВОРОНОВ | dei CORVI | VORONOV | VORONOV |
| ЗВЕРИНЦЕВА | ZVERINTSEV | ZVERINTSEV | ZVERINTSEVA |
| ИМЯРЕК | NOME | IMYAREK | IMYAREK |
| КОЗЕРОДОВ | KOZERODOV | DI KOZEROD | KOSERODOV |
| КОЗЛОВ | DELLE CAPRE | DELLE CAPRE | KOZLOV |
| КРЫЛОВ | KRYLOV | KRYLOV | KRYLOV |
| КУЗЬМЕНКО | KUZMENKO | KUZMENKO | KUZMENKO |
| МАРИНИНА | DI MARININ | Marinina | MARININA |
| САЗОНОВА | di SAZONOV | SAZONOVA | SAZONOVA |
| СИДОРОВ | SIDOROV | SIDOROV | SIDOROV |
| СОРОКА | COPOKA | COPOKA | COPOKA |
| СТОЛЬНИКОВА | di Stolnikov | STOLNIKOVA | STOLNIKOVA |
| ТАШКИНА | di Tashkin | TASHKINA | TASHKINA |
| ТРОФИМЕНКО | TROFIMENKO | TROFIMENKO | TROFIMENKO |

| | | | |
|---|---|---|---|
| ХЕРЛО | HERLO | HERLO | HERLO |
| ЧУМАКОВ | CHUMAKOV | CHUMAKOV | CHUMAKOV |
| ШАПОШНИКОВА | DI SHAPOSHNIKOV | SHAPOSHNIKOVA | SHAPOSHNIKOVA |
| ЩЕРБАКОВА | DI SHCHERBAKOV | SHCHERBAKOVA | Shcherbakova |

2.6 Table of last names: Yandex, ChatGPT-4Omni, transliteration according to the decree.

| LAST NAMES | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | YANDEX | CHAT GPT-4Omni | TRANSLITERATION ACCORDING TO THE DECREE |
| АГАПОВА | Agapova | AGAPOVA | AGAPOVA |
| АНДРЕЕВА | Andreev | ANDREEVA | ANDREEVA |
| БАРАНОВ | Baranov | BARANOV | BARANOV |
| БРАГИН | Bragin | BRAGIN | BRAGIN |
| ВОРОНОВ | dei corvi | VORONOV | VORONOV |
| ЗВЕРИНЦЕВА | ZVERINTSEVA | ZVERINTSEVA | ZVERINTSEVA |
| ИМЯРЕК | nome | IMIAREK | IMIAREK |
| КОЗЕРОДОВ | Capricorno | KOZERODOV | KOZERODOV |
| КОЗЛОВ | Kozlov | KOZLOV | KOZLOV |
| КРЫЛОВ | Krylov | KRYLOV | KRYLOV |
| КУЗЬМЕНКО | Kuzmenko | KUZMENKO | KUZMENKO |
| МАРИНИНА | Marinina | MARININA | MARININA |
| САЗОНОВА | Sazonov | SAZONOVA | SAZONOVA |
| СИДОРОВ | Sidorov | SIDOROV | SIDOROV |
| СОРОКА | СОРОКА | СОРОКА | SOROKA |
| СТОЛЬНИКОВА | STOLNIKOV | STOL'NIKOVA | STOLNIKOVA |
| ТАШКИНА | TASHKIN | TASHKINA | TASHKINA |

| | | | |
|---|---|---|---|
| ТРОФИМЕНКО | TROFIMENKO | TROFIMENKO | TROFIMENKO |
| ХЕРЛО | HURLOUGH | KHERLO | KHERLO |
| ЧУМАКОВ | CHumakov | CHUMAKOV | CHUMAKOV |
| ШАПОШНИКОВА | Shaposhnikov | SHAPOSHNIKOVA | SHAPOSHNIKOVA |
| ЩЕРБАКОВА | Shcherbakova | SHCHERBAKOVA | SHCHERBAKOVA |

3. Tables with source and target toponyms.
3.1. Table of toponyms: Language Weaver, Modern MT, Intento and DeepL.

| TOPONYMS | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | LANGUAGE WEAVER | MODERN MT | INTENTO AND DEEPL |
| В Г. ЧЕРНЯХОВСК | A CHERNYAKHOVSK | A CHERNYAKHOVSK | nella città di Chornyakhovsk |
| В Г. КАЗАНЬ | IN KAZAN | A KAZAN | nella città di Kazan |
| ГОР. НИЖНЕГО ТАГИЛА | GORA. NIZHNY TAGIL | DELLA CITTÀ DI NIZHNY TAGIL | NIZHNEGO TAGILA |
| В Г. СЫКТЫВКАР | IN SYKTYVKAR | A SYKTYVKAR | nella città di Syktyvkar |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | NEL DISTRETTO DI KUIBYSHEV | NEL DISTRETTO DI KUIBYSHEVSKY | nel distretto di Kuibyshev |
| В Г. ВЯЗЬМА | IN VYAZMA | NELLA città DI VYAZMA | a Vyazma |
| ЧЕЛЯБИНСКОЙ ОБЛ. | REGIONE DI CHELYABINSK | REGIONE DI CHELYABINSK | REGIONE DI CHELYABINSK. |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | NEL DISTRETTO DI KIROV-CHEPETSKY | NEL DISTRETTO DI KIROVO-CHEPETSKY | nel Distretto di Kirovo-Chepetsk |
| ГОР. ПЕНЗА | delle MONTAGNE. PENZA | PENZA | CITTÀ. PENZA |
| ХУТОР КАЗАЧИЯ | IL VILLAGGIO DI KAZACHIA | KHUTOR KAZACHIYA | KHUTOR KAZACHIYA |
| В Г. КИРИШИ | A KIRISHI | A KIRISHI | nella città di Kirishi |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | NELLA REGIONE DI TYUMEN | per LA REGIONE DI TJUMEN | per l'Oblast' di Tumen |
| ЛЮБЫТИНСКОГО РАЙОНА | NEL DISTRETTO DI LYUBYTINSKY | DEL DISTRETTO DI LYUBYTINSKY | Distretto di Lubytinsk |
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | DEL HEAT-OGAREVSKOGO R-ON | DISTRETTO DI TEPLO-OGAREVSKY | DISTRETTO DI TEPLO-OGAREVSKIY |

| | | | |
|---|---|---|---|
| В Г. ШАДРИНСК | A SHADRINSK | NELLA città DI SHADRINSK | nella città di Shadrinsk |
| П. СЕРЕБРЯНЫЙ БОР | P. SILVER BOR | P. SILVER FOREST | P. Serebryaniyaniy BOR |
| В ПРИКУБАНСКОМ ОКРУГЕ | NEL DISTRETTO PRIKUBAN | NEL DISTRETTO Prikubansky | nel DISTRETTO DI PRIKUBAN |
| В Г. КЕМЬ | IN KEM | A KEM | a KEMY |
| В ГОР. УСТЮЖНА | IN MONTAGNA. USTYUZHNA | NELLA CITTÀ DI USTYUZHNA | nella città di Ustyuzhna |
| ВО ЧУВАШСКОЙ АССР | NELL'ASSR DI CHUVASH | IN THE CHUVASH ASSR | VO CHUVASHSKY ASSR |
| ГОР. ГРЯЗИ | delle MONTAGNE. TERRA | DELLA CITTÀ DI MUD | GOR. FANGHI |
| ЛИПЕЦКОЙ ОБЛ. | DELLA REGIONE DI LIPETSK | della regione DI LIPECK | DELLA REGIONE DI LIPETSK. |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | NELLA CITTÀ DI SEMICARAKORSK | NELLA CITTÀ DI SEMIKARAKORSK | NELLA CITTÀ DI SEMIKARAKORSK |
| СТ. БАГАЕВСКАЯ | S... BAGAEVSKAYA | Bagaevskaya St. | ST. BAGAYEVSKAYA |
| В Г. ЖЕЛЕЗНОГОРСК | A ZHELEZNOGORSK | A ZHELEZNOGORSK | a Zheleznogorsk |

3.2 Table of toponyms: Yandex, ChatGPT-4Omni, transliteration according to the report.

| TOPONYMS | | | |
|---|---|---|---|
| RUSSIAN SOURCE TEXT | ITALIAN TARGET TEXT | | |
| | YANDEX | CHAT GPT-4Omni | TRANSLITERATION ACCORDING TO THE REPORT |
| В Г. ЧЕРНЯХОВСК | a Chernyakhovsk | nella città di Černjachovsk | CITTÀ DI ČERNJAHOVSK |
| В Г. КАЗАНЬ | a Kazan | nella città di Kazan’ | CITTÀ DI KAZAN' |
| ГОР. НИЖНЕГО ТАГИЛА | NIZHNY TAGIL | della città di Nižnij Tagil | CITTÀ DI NIŽNIJ TAGIL |
| В Г. СЫКТЫВКАР | a Syktyvkar | nella città di Syktyvkar | CITTÀ DI SYKTYVKAR |
| В КУЙБЫШЕВСКОМ РАЙОНЕ | nel distretto di Kuibyshev | nel distretto Kujbyšev | TERRITORIO KUJBYŠEVSKIJ |
| В Г. ВЯЗЬМА | a Vyazma | nella città di Vjaz'ma. | CITTÀ DI VJAZ'MA |
| ЧЕЛЯБИНСКОЙ ОБЛ. | REGIONE DI CHELYABINSK. | Regione di Čeljabinsk. | OBLAST' DI ČELJABINSK |
| В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ | nel distretto di Kirovo-CHEPETSKY | nel distretto Kirovo-Čepeck. | TERRITORIO KIROVO-ČEPECKIJ |
| ГОР. ПЕНЗА | delle montagne. PENZA | Città di Penza | CITTÀ PENZA |
| ХУТОР КАЗАЧИЯ | della fattoria cosacca | Villaggio di Kazachija | HUTOR KAZAČIJA |
| В Г. КИРИШИ | a Kirishi | nella città di Kiriši. | CITTÀ DI KIRIŠI |
| ПО ТЮМЕНСКОЙ ОБЛАСТИ | nella regione di Tyumen | nella regione di Tjumen | PER L'OBLAST' DI TJUMEN' |
| ЛЮБЫТИНСКОГО РАЙОНА | Distretto di Lyubytinsky | nel distretto di Ljubytino. | TERRITORIO LJUBYTINSKIJ |
| ТЕПЛО-ОГАРЕВСКОГО Р-НА | TEPLAYA - OGAREVSKY R-SULLA | Distretto di Teplo-Ogarev | TERRITORIO TEPLO-OGAREVSKIJ |

| | | | |
|---|---|---|---|
| В Г. ШАДРИНСК | a Shadrinsk | nella città di Šadrinsk. | CITTÀ DI ŠADRINSK |
| П. СЕРЕБРЯНЫЙ БОР | P. Silver Bor | Insediamento di Serebrjanyj Bor | BORGO SEREBRJANYJ BOR |
| В ПРИКУБАНСКОМ ОКРУГЕ | nel distretto di PRIKUBAN | nel distretto di Prikubansk | CIRCONDARIO PRIKUBANSKIJ |
| В Г. КЕМЬ | a Kem | città di Kem'. | CITTÀ DI KEM' |
| В ГОР. УСТЮЖНА | a Gore. Ustyuzhna | città di Ustjužna. | CITTÀ DI USTJUŽNA |
| ВО ЧУВАШСКОЙ АССР | in Chuvash ASSR | nella Repubblica Socialista Sovietica Autonoma Ciuvascia. | REPUBBLICA SOCIALISTA SOVIETICA AUTONOMA CIUVASCIA |
| ГОР. ГРЯЗИ | delle montagne. FANGO | Città di Grjazi | CITTÀ DI GRJAZI |
| ЛИПЕЦКОЙ ОБЛ. | REGIONE DI LIPETSK. | regione di Lipeck. | OBLAST' DI LIPECK |
| В ГОРОДЕ СЕМИКАРАКОРСКЕ | nella città di Semikarakorsk | nella città di Semikarakorsk. | CITTÀ DI SEMIKARAKORSK |
| СТ. БАГАЕВСКАЯ | St. BAGAEVSKAYA | Villaggio di Bagaevskaia | STANITSA BAGAEVSKAJA |
| В Г. ЖЕЛЕЗНОГОРСК | a Zheleznogorsk | città di Železnogorsk. | CITTÀ DI ŽELEZNOGORSK |

4. Corpora.
    4.1. Russian passports.

**Passport 1.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО РЕСПУБЛИКЕ ТАТАРСТАН В Г. КАЗАНЬ
Дата выдачи 10.10.2013
Код подразделения 710-016
Личная подпись
Фамилия ИМЯРЕК
Имя АЛЕКСАНДР
Отчество КОНСТАНТИНОВИЧ
Пол МУЖ.
Дата рождения 18.01.1995
Место рождения ГОР. КАЗАНЬ РЕСПУБЛИКИ ТАТАРСТАН

**Passport 2.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО КАЛИНИНГРАДСКОЙ ОБЛАСТИ В Г. ЧЕРНЯХОВСК
Дата выдачи 15.02.2016
Код подразделения 390-007
Личная подпись
Фамилия СОРОКА
Имя АЛЕКСЕЙ
Отчество ГРИГОРЬЕВИЧ
Пол МУЖ.
Дата рождения 14.01.1987
Место рождения ГОР. ЧЕРНЯХОВСК КАЛИНИНГРАДСКОЙ ОБЛ.

**Passport 3.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО СВЕРДЛОВСКОЙ ОБЛАСТИ В ДЗЕРЖИНСКОМ Р-НЕ ГОР. НИЖНЕГО ТАГИЛА
Дата выдачи 12.05.2017
Код подразделения 160-020
Личная подпись
Фамилия БАРАНОВ
Имя АНДРЕЙ
Отчество АНДРЕЕВИЧ
Пол МУЖ.
Дата рождения 02.12.1980
Место рождения ГОР. НИЖНИЙ ТАГИЛ СВЕРДЛОВСКОЙ ОБЛ.

**Passport 4.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ

Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО СМОЛЕНСКОЙ ОБЛАСТИ В Г. ВЯЗЬМА
Дата выдачи 03.06.2014
Код подразделения 670-009
Личная подпись
Фамилия СИДОРОВ
Имя БОРИС
Отчество ВЯЧЕСЛАВОВИЧ
Пол Муж
Дата рождения 13.01.1999
Место рождения ГОР. ВЯЗЬМА СМОЛЕНСКОЙ ОБЛ.

**Passport 5.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт видан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО ЛЕНИНГРАДСКОЙ ОБЛАСТИ В Г. КИРИШИ
Дата выдачи 13.03.2014
Код подразделения 450-016
Личная подпись
Фамилия ЗВЕРИНЦЕВА
Имя ВЕРОНИКА
Отчество ВИКТОРОВНА
Пол ЖЕН
Дата рождения 12.05.1996
Место рождения ГОР. КИРИШИ ЛЕНИНГРАДСКОЙ ОБЛ.

**Passport 6.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО ТЮМЕНСКОЙ ОБЛАСТИ В Г. ИШИМ
Дата выдачи 09.11.2012
Код подразделения 570-008
Личная подпись
Фамилия АГАПОВА
Имя ВИКТОРИЯ
Отчество ЛЕОНИДОВНА
Пол ЖЕН.
Дата рождения 20.06.2000
Место рождения ГОР. ИШИМ ТЮМЕНСКОЙ ОБЛ.

**Passport 7.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО РЕСПУБЛИКЕ КОМИ В Г. СЫКТЫВКАР
Дата выдачи 08.04.2012
Код подразделения 540-012
Личная подпись
Фамилия БРАГИН

Имя ВЛАДИМИР
Отчество ДМИТРИЕВИЧ
Пол МУЖ.
Дата рождения 11.08.2007
Место рождения ГОР. СЫКТЫВКАР РЕСПУБЛИКИ КОМИ

**Passport 8.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО САМАРСКОЙ ОБЛАСТИ В КУЙБЫШЕВСКОМ РАЙОНЕ ГОР. САМАРЫ
Дата выдачи 18.01.2011
Код подразделения 500-005
Личная подпись
Фамилия ХЕРЛО
Имя ДАРЬЯ
Отчество ГЕННАДЬЕВНА
Пол ЖЕН.
Дата рождения 17.01.1996
Место рождения ГОР. САМАРА САМАРСКОЙ ОБЛ.

**Passport 9.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО ЧЕЛЯБИНСКОЙ ОБЛАСТИ В Г. МИАСС
Дата выдачи 17.12.2004
Код подразделения 740-018
Личная подпись
Фамилия КУЗЬМЕНКО
Имя ЕВГЕНИЙ
Отчество АЛЕКСАНДРОВИЧ
Пол МУЖ.
Дата рождения 06.03.1994
Место рождения ГОР. МИАСС ЧЕЛЯБИНСКОЙ ОБЛ.

**Passport 10.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ТЕРРИТОРИАЛЬНЫМ ПУНКТОМ УФМС РОССИИ ПО ПЕНЗЕНСКОЙ ОБЛ. В ПЕНЗЕНСКОМ РАЙОНЕ
Дата выдачи 14.05.2019
Код подразделения 330-010
Личная подпись
Фамилия ЩЕРБАКОВА
Имя ЕКАТЕРИНА
Отчество МИХАЙЛОВНА
Пол ЖЕН.
Дата рождения 12.09.2003
Место рождения ГОР. ПЕНЗА ПЕНЗЕНСКОГО Р-НА ПЕНЗЕНСКОЙ ОБЛ.

**Passport 11.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛОМ УФМС РОССИИ ПО КИРОВСКОЙ ОБЛАСТИ В КИРОВО-ЧЕПЕЦКОМ РАЙОНЕ
Дата выдачи 22.07.2015
Код подразделения 470-011
Личная подпись
Фамилия ВОРОНОВ
Имя ИГОРЬ
Отчество ЛЕОНТЬЕВИЧ
Пол МУЖ.
Дата рождения 16.10.1966
Место рождения ГОР. СЛОБОДСКОЙ КИРОВСКОЙ ОБЛ.

**Passport 12.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ТЕРРИТОРИАЛЬНЫМ ПУНКТОМ УФМС РОССИИ ПО ВОЛГОГРАДСКОЙ ОБЛАСТИ В СТ-ЦЕ КЛЕТСКАЯ
Дата выдачи 05.08.2016
Код подразделения 720-014
Личная подпись
Фамилия ТАШКИНА
Имя ИРАИДА
Отчество СТЕПАНОВНА
Пол ЖЕН
Дата рождения 22.05.1998
Место рождения ХУТОР КАЗАЧИЯ КЛЕТСКОГО Р -НА ВОЛГОГРАДСКОЙ ОБЛ.

**Passport 13.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО ТУЛЬСКОЙ ОБЛАСТИ В КИРЕЕВСКОМ РАЙОНЕ
Дата выдачи 16.06.2011
Код подразделения 480-007
Личная подпись
Фамилия АНДРЕЕВА
Имя КРИСТИНА
Отчество ВЛАДИМИРОВНА
Пол ЖЕН.
Дата рождения 14.09.1995
Место рождения ДЕР. КРЮКОВКА ТЕПЛО -ОГАРЕВСКОГО Р-НА ТУЛЬСКОЙ ОБЛ.

**Passport 14.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО НОВГОРОДСКОЙ ОБЛАСТИ ЛЮБЫТИНСКОГО РАЙОНА

Дата выдачи 20.01.2013
Код подразделения 630-019
Личная подпись
Фамилия МАРИНИНА
Имя ЛЮДМИЛА
Отчество ПЕТРОВНА
Пол ЖЕН.
Дата рождения 24.08.2000
Место рождения С. ПОДВИНОГРАДОВ ВИНОГРАДОВСКОГО Р -НА ЗАКАРПАТСКОЙ ОБЛ.

**Passport 15.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО КУРГАНСКОЙ ОБЛАСТИ В Г. ШАДРИНСК
Дата выдачи 07.12.2014
Код подразделения 700-015
Личная подпись
Фамилия ЧУМАКОВ
Имя МИХАИЛ
Отчество ВАСИЛЬЕВИЧ
Пол МУЖ.
Дата рождения 13.09.1986
Место рождения ГОР. ШАДРИНСК КУРГАНСКОЙ ОБЛ.

**Passport 16.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛ УФМС РОССИИ ПО КРАСНОДАРСКОМУ КРАЮ В ПРИКУБАНСКОМ ОКРУГЕ Г. КРАСНОДАРА
Дата выдачи 30.07.2013
Код подразделения 580-013
Личная подпись
Фамилия КОЗЛОВ
Имя НИКОЛАЙ
Отчество ОЛЕГОВИЧ
Пол МУЖ
Дата рождения 13.02.1959
Место рождения ГОР. КРАСНОДАР, КРАСНОДАРСКОГО КРАЯ

**Passport 17.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО САРАТОВСКОЙ ОБЛАСТИ ГОР. БАЛАКОВО И БАЛАКОВСКОГО РАЙОНА
Дата выдачи 29.09.2018
Код подразделения 760-021
Личная подпись
Фамилия ШАПОШНИКОВА
Имя ОКСАНА

Отчество МАКСИМОВНА
Пол ЖЕН
Дата рождения 02.04.1975
Место рождения П. СЕРЕБРЯНЫЙ БОР Г. НЕРЮНГРИ РЕСПУБЛИКА САХА/ЯКУТИЯ/

**Passport 18.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО КАРЕЛИИ В Г. КЕМЬ
Дата выдачи 03.10.2015
Код подразделения 320-011
Личная подпись
Фамилия ТРОФИМЕНКО
Имя СВЕТЛАНА
Отчество НИКОЛАЕВНА
Пол ЖЕН.
Дата рождения 26.09.1992
Место рождения ГОР. КЕМЬ КАРЕЛИИ

**Passport 19.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО ЛИПЕЦКОЙ ОБЛАСТИ В Г. ГРЯЗИ
Дата выдачи 24.01.2012
Код подразделения 460-009
Личная подпись
Фамилия КОЗЕРОДОВ
Имя СЕРГЕЙ
Отчество ЕВГЕНЬЕВИЧ
Пол МУЖ.
Дата рождения 02.03.1996
Место рождения ГОР. ГРЯЗИ ЛИПЕЦКОЙ ОБЛ.

**Passport 20.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛОМ УФМС РОССИИ ПО ГОР. ЙАРЦЕВО ПО РАЙОНУ КАЗЫГ
Дата выдачи 05.12.2017
Код подразделения 340-018
Личная подпись
Фамилия САЗОНОВА
Имя ТАТЬЯНА
Отчество ВАЛЕНТИНОВНА
Пол ЖЕН
Дата рождения 19.12.1968
Место рождения С. БИЧУРЧА БАНТЕВО ШЕМУРШИНСКОГО Р -НА ВО ЧУВАШСКОЙ АССР

**Passport 21.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО РОСТОВСКОЙ ОБЛ. В ГОРОДЕ СЕМИКАРАКОРСКЕ
Дата выдачи 18.05.2014
Код подразделения 100-007
Личная подпись
Фамилия СТОЛЬНИКОВА
Имя ЮЛИЯ
Отчество СЕРГЕЕВНА
Пол ЖЕН.
Дата рождения 25.11.2007
Место рождения СТ. БАГАЕВСКАЯ РОСТОВСКОЙ ОБЛ. РОССИЯ

**Passport 22.**
РОССИЙСКАЯ ФЕДЕРАЦИЯ
Паспорт выдан ОТДЕЛЕНИЕМ УФМС РОССИИ ПО КУРСКОЙ ОБЛАСТИ В Г. ЖЕЛЕЗНОГОРСК
Дата выдачи 09.09.2016
Код подразделения 680-012
Личная подпись
Фамилия КРЫЛОВ
Имя ЯКОВ
Отчество ГЕОРГИЕВИЧ
Пол МУЖ.
Дата рождения 24.02.2004
Место рождения ГОР. ЖЕЛЕЗНОГОРСК КУРСКОЙ ОБЛ.

4.2. Language Weaver translated passports.

**Passport 1.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA PER LA REPUBBLICA DEL TATARSTAN IN KAZAN
Data di rilascio 10.10.2013
Codice divisione 710-016
Firma personale
Cognome NOME
Mi chiamo ALEXANDER
Patronimico KONSTANTINOVICH
MARITO sessuale.
Data di nascita 18.01.1995
La nascita delle MONTAGNE. KAZAN DELLA REPUBBLICA DEL TATARSTAN

**Passport 2.**
FEDERAZIONE RUSSA

IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA NELLA REGIONE DI KALININGRAD A CHERNYAKHOVSK
Data di rilascio 15.02.2016
Codice divisione 390-007
Firma personale
Nome della COPOKA
Mi chiamo ALEXEY
Patronimico GRIGORIEVICH
MARITO sessuale.
Data di nascita 14.01.1987
Il luogo di nascita delle MONTAGNE. CHERNYAKHOVSK, REGIONE DI KALININGRAD

**Passport 3.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA PER LA REGIONE DI SVERDLOVSK NEL DZERZHINSKY R-NE GORA. NIZHNY TAGIL
Data di rilascio 12.05.2017
Codice divisione 160-020
Firma personale
Il nome DELLA PECORA
Nome ANDREW
Patronimico ANDREEVICH
MARITO sessuale.
Data di nascita 02.12.1980
Il luogo di nascita delle MONTAGNE. NIZHNY TAGIL, REGIONE DI SVERDLOVSK

**Passport 4.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DELL'FMS DELLA RUSSIA NELLA REGIONE DI SMOLENSK IN VYAZMA
Data di rilascio 03.06.2014
Codice divisione 670-009
Firma personale
Nome di SIDOROV
Mi chiamo BORIS
VYACHESLAVOVICH patronimico
Marito sessuale
Data di nascita 13.01.1999
Il luogo di nascita delle MONTAGNE. REGIONE DI VYAZMA SMOLENSK

**Passport 5.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO VISTO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA NELLA REGIONE DI LENINGRADO A KIRISHI

Data di rilascio 13.03.2014
Codice divisione 450-016
Firma personale
Nome di ZVERINTSEV
Nome VERONICA
Patronimico VIKTOROVNA
Paul MOGLI
Data di nascita 12.05.1996
Il luogo di nascita delle MONTAGNE. KIRISHI, REGIONE DI LENINGRADO

**Passport 6.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA NELLA REGIONE DI TYUMEN A ISHIM
Data di rilascio 09.11.2012
Codice divisione 570-008
Firma personale
Nome di AGAPOV
Mi chiamo VICTORIA
Il patronimico di LEONIDOVNA
Paul MOGLI.
Data di nascita 20.06.2000
La nascita delle MONTAGNE. REGIONE ISHIM TYUMEN

**Passport 7.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER L'IMMIGRAZIONE DELLA RUSSIA PER LA REPUBBLICA DI KOMI IN SYKTYVKAR
Data di rilascio 08.04.2012
Codice divisione 540-012
Firma personale
Nome BRAGIN
Nome VLADIMIR
Patronimico DMITRIEVICH
MARITO sessuale.
Data di nascita 11.08.2007
Il luogo di nascita delle MONTAGNE. SYKTYVKAR DELLA REPUBBLICA DI KOMI

**Passport 8.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DELL'FMS DELLA RUSSIA NELLA REGIONE DI SAMARA NEL DISTRETTO DI KUIBYSHEV DELLE MONTAGNE. SAMARA
Data di rilascio 18.01.2011
Codice divisione 500-005

Firma personale
Nome di HERLO
Il nome di DARIA
Il patronimico di GENNADYEVNA
Paul MOGLI.
Data di nascita 17.01.1996
Il luogo di nascita delle MONTAGNE. REGIONE DI SAMARA

**Passport 9.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA NELLA REGIONE DI CHELYABINSK NEL MIASS
Data di rilascio 17.12.2004
Codice divisione 740-018
Firma personale
Nome KUZMENKO
Nome EUGENE
Patronimico ALEXANDROVICH
MARITO sessuale.
Data di nascita 06.03.1994
Il luogo di nascita delle MONTAGNE. MIASS REGIONE DI CHELYABINSK

**Passport 10.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL PUNTO TERRITORIALE DELL'UFMS DELLA RUSSIA NELLA REGIONE DI PENZA NEL QUARTIERE DI PENZA
Data di rilascio 14.05.2019
Codice divisione 330-010
Firma personale
Nome DI SHCHERBAKOV
Nome di CATHERINE
Il patronimico di MIKHAILOVNA
Paul MOGLI.
Data di nascita 12.09.2003
Il luogo di nascita delle MONTAGNE. PENZA PENZA R-ON REGIONE PENZA

**Passport 11.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA PER LA REGIONE DI KIROV NEL DISTRETTO DI KIROV-CHEPETSKY
Data di rilascio 22.07.2015
Codice divisione 470-011
Firma personale
Il nome dei CORVI
Nome IGOR

LEONTIEVICH patronimico
MARITO sessuale.
Data di nascita 16.10.1966
Il luogo di nascita delle MONTAGNE. SLOBODSKAYA REGIONE DI KIROV

**Passport 12.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL PUNTO TERRITORIALE DEL SERVIZIO FEDERALE PER L'IMMIGRAZIONE DELLA RUSSIA NELLA REGIONE DI VOLGOGRAD A ST. KLETSKAYA
Data di rilascio 05.08.2016
Codice divisione 720-014
Firma personale
Nome di Tashkin
Il nome di ISRAELE
STEPANOVNA e' patronimico
Paul MOGLI
Data di nascita 22.05.1998
IL VILLAGGIO DI KAZACHIA KLETSKY R-ON REGIONE DI VOLGOGRAD È NATO

**Passport 13.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER L'IMMIGRAZIONE DELLA RUSSIA NELLA REGIONE DI TULA NEL DISTRETTO DI KIREEVSKOM
Data di rilascio 16.06.2011
Codice divisione 480-007
Firma personale
Nome DI ANDREEV
Mi chiamo CHRISTINE
Il patronimico di VLADIMIROVNA
Paul MOGLI.
Data di nascita 14.09.1995
Luogo di nascita DEL DER. KRYUKOVKA HEAT-OGAREVSKOGO R-ON REGIONE DI TULA

**Passport 14.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DELL'FMS DELLA RUSSIA NELLA REGIONE DI NOVGOROD NEL DISTRETTO DI LYUBYTINSKY
Data di rilascio 20.01.2013
Codice divisione 630-019
Firma personale
Nome DI MARININ
Il nome di LYUDMILA
Il patronimico di PETROVNA

Paul MOGLI.
Data di nascita 24.08.2000
LUOGO DI NASCITA DELLA REGIONE TRANSCARPATICA S. PODVINOGRADOV VINOGRADOVSKY R-ON

**Passport 15.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER L'IMMIGRAZIONE DELLA RUSSIA NELLA REGIONE DI KURGAN A SHADRINSK
Data di rilascio 07.12.2014
Codice divisione 700-015
Firma personale
Nome di CHUMAKOV
Mi chiamo MICHAEL
Patronimico VASILYEVICH
MARITO sessuale.
Data di nascita 13.09.1986
Il luogo di nascita delle MONTAGNE. SHADRINSK, REGIONE DI KURGAN

**Passport 16.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE DI MIGRAZIONE DELLA RUSSIA NEL TERRITORIO DI KRASNODAR NEL DISTRETTO PRIKUBAN DI KRASNODAR
Data di rilascio 30.07.2013
Codice divisione 580-013
Firma personale
NOME DELLE CAPRE
Nome NICHOLAS
Patronimico OLEGOVICH
MARITO sessuale
Data di nascita 13.02.1959
Il luogo di nascita delle MONTAGNE. KRASNODAR, TERRITORIO DI KRASNODAR

**Passport 17.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA PER LA REGIONE DI SARATOV DELLE MONTAGNE. BALAKOVO E BALAKOVSKY
Data di rilascio 29.09.2018
Codice divisione 760-021
Firma personale
Nome DI SHAPOSHNIKOV
Nome di OKCAHA
Il secondo nome È MAXIMOVNA
Paul MOGLI

Data di nascita 02.04.1975
LUOGO DI NASCITA P. SILVER BOR G. NERUNGRI REPUBBLICA DI SAKHA/YAKUTIA/

**Passport 18.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DELL'FMS DELLA RUSSIA IN CARELIA IN KEM
Data di rilascio 03.10.2015
Codice divisione 320-011
Firma personale
Nome DI TROFIMENKO
Il nome di SVETLANA
Il patronimico di NIKOLAEVNA
Paul MOGLI.
Data di nascita 26.09.1992
Il luogo di nascita delle MONTAGNE. KEM KARELIA

**Passport 19.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA NELLA REGIONE DI LIPETSK NELLA CITTÀ DI FANGO
Data di rilascio 24.01.2012
Codice divisione 460-009
Firma personale
Nome di KOZERODOV
Nome SERGEY
EVGENIEVICH patronimico
MARITO sessuale.
Data di nascita 02.03.1996
Il luogo di nascita delle MONTAGNE. TERRA DELLA REGIONE DI LIPETSK

**Passport 20.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DELL'FMS DELLA RUSSIA NELLA REGIONE DI VOLOGDA, IN MONTAGNA. USTYUZHNA
Data di rilascio 05.12.2017
Codice divisione 340-018
Firma personale
Nome di SAZONOV
Nome di TATIANA
Il patronimico di VALENTINOVNA
Paul MOGLI
Data di nascita 19.12.1968
LA NASCITA DI S. BICHURCH BANTEVO SHEMURSHINSKY R-N NELL'ASSR DI CHUVASH

**Passport 21.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER L'IMMIGRAZIONE DELLA RUSSIA NELLA REGIONE DI ROSTOV NELLA CITTÀ DI SEMICARAKORSK
Data di rilascio 18.05.2014
Codice divisione 100-007
Firma personale
Nome di Stolnikov
Il nome di JULIA
Il patronimico di SERGEEVNA
Paul MOGLI.
Data di nascita 25.11.2007
Luogo di nascita di S... BAGAEVSKAYA REGIONE DI ROSTOV RUSSIA

**Passport 22.**
FEDERAZIONE RUSSA
IL PASSAPORTO È STATO RILASCIATO DAL DIPARTIMENTO DEL SERVIZIO FEDERALE PER LE MIGRAZIONI DELLA RUSSIA NELLA REGIONE DI KURSK A ZHELEZNOGORSK
Data di rilascio 09.09.2016
Codice divisione 680-012
Firma personale
Nome KRYLOV
Nome di JACOB
Patronimico GEORGIEVICH
MARITO sessuale.
Data di nascita 24.02.2004
Il luogo di nascita delle MONTAGNE. ZHELEZNOGORSK REGIONE DI KURSK

4.3. Modern MT translated passports.

**Passport 1.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA per LA REPUBBLICA DEL TATARSTAN A KAZAN
Data di emissione 10.10.2013
Codice suddivisione 710-016
Firma personale
Cognome IMYAREK
Nome ALEXANDER
Patronimico KONSTANTINOVICH
Sesso MASCHILE
Data di nascita 18.01.1995
Il luogo di NASCITA DELLA CITTÀ di KAZAN DELLA REPUBBLICA DEL TATARSTAN

**Passport 2.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI KALININGRAD A CHERNYAKHOVSK
Data di emissione 15.02.2016
Codice suddivisione 390-007
Firma personale
Cognome COPOKA
Nome ALEXEY
Patronimico GRIGORIEVICH
Sesso MASCHILE
Data di nascita 14.01.1987
Luogo di NASCITA CHERNYAKHOVSK, REGIONE DI KALININGRAD

**Passport 3.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI SVERDLOVSK NEL DISTRETTO DI DZERZHINSKY DELLA CITTÀ DI NIZHNY TAGIL
Data di emissione 12.05.2017
Codice suddivisione 160-020
Firma personale
Cognome DELLA PECORA
Nome ANDREY
Patronimico ANDREEVICH
Sesso MASCHILE
Data di nascita 02.12.1980
Luogo di NASCITA: NIŽNIJ TAGIL, OBLAST 'DI SVERDLOVSK

**Passport 4.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del servizio federale DI MIGRAZIONE della RUSSIA per LA REGIONE DI SMOLENSK NELLA città DI VYAZMA
Data di emissione 03.06.2014
Codice suddivisione 670-009
Firma personale
Cognome SIDOROV
Nome BORIS
Patronimico VYACHESLAVOVICH
Sesso Maschile
Data di nascita 13.01.1999
Luogo di NASCITA VYAZMA, REGIONE DI SMOLENSK

**Passport 5.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI LENINGRADO A KIRISHI
Data di emissione 13.03.2014

Codice suddivisione 450-016
Firma personale
Cognome ZVERINTSEV
Nome VERONICA
Patronimico VIKTOROVNA
MOGLI di genere
Data di nascita 12.05.1996
Luogo di NASCITA KIRISHI, REGIONE DI LENINGRADO

**Passport 6.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del servizio federale DI MIGRAZIONE della RUSSIA per LA REGIONE DI TJUMEN NELLA CITTÀ DI ISHIM
Data di emissione 09.11.2012
Codice suddivisione 570-008
Firma personale
Cognome AGAPOV
Nome VICTORIA
Patronimico LEONIDOVNA
MOGLI di genere
Data di nascita 20.06.2000
Luogo di NASCITA DELLA CITTÀ DI ISHIM, REGIONE DI TYUMEN

**Passport 7.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione DELLA RUSSIA per LA REPUBBLICA DI KOMI A SYKTYVKAR
Data di emissione 08.04.2012
Codice suddivisione 540-012
Firma personale
Cognome BRAGIN
Nome VLADIMIR
Patronimico DMITRIEVICH
Sesso MASCHILE
Data di nascita 11.08.2007
Luogo di NASCITA SYKTYVKAR KOMI REPUBLIC

**Passport 8.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI SAMARA NEL DISTRETTO DI KUIBYSHEVSKY DELLA CITTÀ DI SAMARA
Data di emissione 18.01.2011
Codice suddivisione 500-005
Firma personale
Cognome HERLO
Nome DARIA
Patronimico GENNADIEVNA

MOGLI di genere
Data di nascita 17.01.1996
Luogo di NASCITA SAMARA, REGIONE DI SAMARA

**Passport 9.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI CHELYABINSK IN MIASS
Data di emissione 17.12.2004
Codice suddivisione 740-018
Firma personale
Cognome KUZMENKO
Nome EUGENE
Patronimico ALEKSANDROVICH
Sesso MASCHILE
Data di nascita 06.03.1994
Luogo di NASCITA MIASS, REGIONE DI CHELYABINSK

**Passport 10.**
FEDERAZIONE RUSSA
Passaporto rilasciato dall'UFFICIO TERRITORIALE della direzione DEL servizio federale DI migrazione della RUSSIA per LA regione DI PENZA NELLA REGIONE DI Penza
Data di emissione 14.05.2019
Codice suddivisione 330-010
Firma personale
Cognome SHCHERBAKOVA
Nome EKATERINA
Patronimico MIKHAILOVNA
MOGLI di genere
Data di nascita 12.09.2003
Luogo di NASCITA: PENZA, DISTRETTO DI PENZA, REGIONE DI PENZA.

**Passport 11.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI KIROV NEL DISTRETTO DI KIROVO-CHEPETSKY
Data di emissione 22.07.2015
Codice suddivisione 470-011
Firma personale
Cognome VORONOV
Nome IGOR
Patronimico LEONTIEVICH
Sesso MASCHILE
Data di nascita 16.10.1966
Luogo di NASCITA SLOBODSKOY, REGIONE DI KIROV

**Passport 12.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal PUNTO TERRITORIALE del servizio federale DI migrazione della RUSSIA PER LA REGIONE DI VOLGOGRAD A ST-CE KLETSKAYA
Data di emissione 05.08.2016
Codice suddivisione 720-014
Firma personale
Cognome TASHKINA
Nome dell'IRAID
Patronimico STEPANOVNA
MOGLI di genere
Data di nascita 22.05.1998
Luogo DI nascita KHUTOR KAZACHIYA KLETSKY DISTRETTO REGIONE DI VOLGOGRAD

**Passport 13.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI TULA NEL DISTRETTO DI KIREEV
Data di emissione 16.06.2011
Codice suddivisione 480-007
Firma personale
Cognome ANDREEVA
Nome KRISTINA
Patronimico VLADIMIROVNA
MOGLI di genere
Data di nascita 14.09.1995
Luogo di nascita: VILLAGGIO DI KRYUKOVKA, DISTRETTO DI TEPLO-OGAREVSKY, REGIONE DI TULA.

**Passport 14.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del servizio federale DI MIGRAZIONE della RUSSIA per LA REGIONE DI NOVGOROD DEL DISTRETTO DI LYUBYTINSKY
Data di emissione 20.01.2013
Codice suddivisione 630-019
Firma personale
Cognome Marinina
Nome LYUDMILA
Patronimico PETROVNA
MOGLI di genere
Data di nascita 24.08.2000
Luogo di nascita: S. PODVINOGRADOV, DISTRETTO DI VINOGRADOVSKYI, REGIONE DI ZAKARPATTIA

**Passport 15.**
FEDERAZIONE RUSSA

Passaporto rilasciato dal Dipartimento del servizio federale DI MIGRAZIONE della RUSSIA per LA REGIONE DI KURGAN NELLA città DI SHADRINSK
Data di emissione 07.12.2014
Codice di suddivisione 700-015
Firma personale
Cognome CHUMAKOV
Nome MIKHAIL
Patronimico VASIL' EVIČ
Sesso MASCHILE
Data di nascita 13.09.1986
Luogo di NASCITA Shadrinsk, REGIONE DI KURGAN

**Passport 16.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione DELLA RUSSIA PER LA REGIONE DI KRASNODAR NEL DISTRETTO Prikubansky DELLA CITTÀ DI KRASNODAR
Data di emissione 30.07.2013
Codice suddivisione 580-013
Firma personale
Cognome DELLE CAPRE
Nome NIKOLAY
Patronimico OLEGOVICH
Sesso MARITO
Data di nascita 13.02.1959
Luogo di NASCITA KRASNODAR, TERRITORIO DI KRASNODAR

**Passport 17.**
FEDERAZIONE RUSSA
IL passaporto È stato rilasciato DAL DIPARTIMENTO del servizio federale DI migrazione DELLA RUSSIA per LA REGIONE DI SARATOV DELLA CITTÀ DI DISTRETTO DI BALAKOVO E BALAKOVSKY
Data di emissione 29.09.2018
Codice suddivisione 760-021
Firma personale
Cognome SHAPOSHNIKOVA
Nome OKCAHA
MAKSIMOVNA patronimico
MOGLI di genere
Data di nascita 02.04.1975
Luogo di nascita P. SILVER FOREST G. NERYUNGRI REPUBBLICA DI SAKHA/YAKUTIA/

**Passport 18.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA per LA CARELIA A KEM
Data di emissione 03.10.2015

Codice suddivisione 320-011
Firma personale
Cognome TROFIMENKO
Nome SVETLANA
Patronimico NIKOLAEVNA
MOGLI di genere
Data di nascita 26.09.1992
Luogo di NASCITA KEM, CARELIA

**Passport 19.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del servizio federale DI MIGRAZIONE della RUSSIA per LA REGIONE DI LIPETSK NELLA città DI MUD
Data di emissione 24.01.2012
Codice suddivisione 460-009
Firma personale
Cognome DI KOZEROD
Nome SERGEY
Patronimico EVGENIEVICH
Sesso MASCHILE
Data di nascita 02.03.1996
Luogo DI NASCITA DELLA CITTÀ DI MUD della regione DI LIPECK

**Passport 20.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI VOLOGDA NELLA CITTÀ DI USTYUZHNA
Data di emissione 05.12.2017
Codice suddivisione 340-018
Firma personale
Cognome SAZONOVA
Nome TATIANA
Patronimico VALENTINOVNA
MOGLI di genere
Data di nascita 19.12.1968
Luogo DI nascita S. BICHURCHA BANTEVO SHEMURSHINSKY DISTRICT IN THE CHUVASH ASSR

**Passport 21.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA per LA REGIONE DI ROSTOV NELLA CITTÀ DI SEMIKARAKORSK
Data di emissione 18.05.2014
Codice suddivisione 100-007
Firma personale
Cognome STOLNIKOVA
Nome YULIA
Patronimico SERGEEVNA

MOGLI di genere
Data di nascita 25.11.2007
Luogo di nascita Bagaevskaya St., REGIONE DI ROSTOV, RUSSIA

**Passport 22.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal DIPARTIMENTO DEL servizio federale DI migrazione della RUSSIA PER LA REGIONE DI KURSK A ZHELEZNOGORSK
Data di emissione 09.09.2016
Codice suddivisione 680-012
Firma personale
Cognome KRYLOV
Nome YAKOV
Patronimico GEORGIEVICH
Sesso MASCHILE
Data di nascita 24.02.2004
Luogo di NASCITA ZHELEZNOGORSK, REGIONE DI KURSK

4.4. Intento translated passports.

**Passport 1.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per la Repubblica del Tatarstan nella città di Kazan. KAZAN
Data di rilascio 10.10.2013
Codice unità 710-016
Firma personale
Cognome IMYAREK
Nome ALEXANDER
Patronimico KONSTANTINOVICH
Sesso Maschio
Data di nascita 18.01.1995
Luogo di nascita CITTÀ. KAZAN DELLA REPUBBLICA DEL TATARSTAN

**Passport 2.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Kaliningrad nella città di Chornyakhovsk. CHERNYAKHOVSK
Data di rilascio 15.02.2016
Codice unità 390-007
Firma personale
Cognome COPOKA
Nome ALEKSEY
Patronimico GRIGORIEVICH
Sesso Maschio
Data di nascita 14.01.1987

Luogo di nascita CITTÀ. CHERNYAKHOVSK KALININGRAD OBLAST.

**Passport 3.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per SVERDLOVSK OBLAST a DZERZHINSK RANE, NIZHNEGO TAGILA. NIZHNEGO TAGIL
Data di emissione 12.05.2017
Codice unità 160-020
Firma personale
Cognome BARANOV
Nome ANDREY
Patronimico ANDREYEVICH
Sesso Maschio
Data di nascita 02.12.1980
Luogo di nascita GOR. REGIONE DI TAGIL INFERIORE SVERDLOVSK.

**Passport 4.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Smolensk a Vyazma. VYAZMA
Data di rilascio 03.06.2014
Codice unità 670-009
Firma personale
Cognome SIDOROV
Nome BORIS
Patronimico VYACHESLAVOVICH
Sesso Marito
Data di nascita 13.01.1999
Luogo di nascita CITTÀ. REGIONE DI SMOLENSK VYAZMA.

**Passport 5.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Leningrado nella città di Kirishi. KIRISHI
Data di rilascio 13.03.2014
Codice unità 450-016
Firma personale
Cognome ZVERINTSEVA
Nome VERONIKA
Patronimico VIKTOROVNA
Sesso Donna
Data di nascita 12.05.1996
Luogo di nascita CITTÀ. KIRISHI LENINGRAD OBLAST.

**Passport 6.**
FEDERAZIONE RUSSA

Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Tumen nella città di Ishim. ISHIM
Data di rilascio 09.11.2012
Codice unità 570-008
Firma personale
Cognome AGAPOVA
Nome VICTORIA
Patronimico LEONIDOVNA
Sesso Donna
Data di nascita 20.06.2000
Luogo di nascita GOR. ISHIM, REGIONE DI TYUMEN.

**Passport 7.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per la Repubblica di Komi nella città di Syktyvkar. SYKTYVKAR
Data di rilascio 08.04.2012
Codice unità 540-012
Firma personale
Cognome BRAGIN
Nome VLADIMIR
Patronimico DMITRIEVICH
Sesso MASCHIO
Data di nascita 11.08.2007
Luogo di nascita GOR. SYKTYVKAR REPUBBLICA DI KOMI

**Passport 8.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Samara nel distretto di Kuibyshev della città di Samara. SAMARA
Data di rilascio 18.01.2011
Codice unità 500-005
Firma personale
Cognome HERLO
Nome DARYA
Patronimico GENNADYEVNA
Genere DONNA
Data di nascita 17.01.1996
Luogo di nascita GOR. REGIONE DI SAMARA SAMARA.

**Passport 9.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Chelyabinsk nella città di MIASS. MIASS
Data di rilascio 17.12.2004
Codice unità 740-018
Firma personale
Cognome KUZMENKO

Nome YEVGENIY
Patronimico ALEXANDROVICH
Sesso Maschio
Data di nascita 06.03.1994
Luogo di nascita GOR. MIASS REGIONE DI CHELYABINSK.

**Passport 10.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal PUNTO TERRITORIALE dell'FMS della Russia per l'Oblast' di Penza. NEL DISTRETTO DI PENZA
Data di rilascio 14.05.2019
Codice unità 330-010
Firma personale
Cognome Shcherbakova
Nome Ekaterina
Patronimico MIKHAILOVNA
Sesso Donna
Data di nascita 12.09.2003
Luogo di nascita CITTÀ. PENZA DEL DISTRETTO DI PENZA, REGIONE DI PENZA, OBLAST' DI PENZA.

**Passport 11.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Kirov nel Distretto di Kirovo-Chepetsk
Data di rilascio 22.07.2015
Codice unità 470-011
Firma personale
Cognome VORONOV
Nome IGOR
Patronimico LEONTYEVICH
Sesso MASCHIO
Data di nascita 16.10.1966
Luogo di nascita GOR. SLOBODSKOY KIROV OBLAST.

**Passport 12.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal PUNTO TERRITORIALE DELLA FMS DELLA RUSSIA PER L'OBLASTRO DI VOLGOGRADO IN ST-TS KLETSKAYA
Data di rilascio 05.08.2016
Codice unità 720-014
Firma personale
Cognome TASHKINA
Nome IRAIDA
Patronimico STEPANOVNA
Sesso Donna
Data di nascita 22.05.1998

Luogo di nascita KAZACHIYA KHUTOR KAZACHIYA KLETSKY RAN di VOLGOGRAD OBL.

**Passport 13.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Tula nel distretto di Kireevsky
Data di rilascio 16.06.2011
Codice unità 480-007
Firma personale
Cognome ANDREEVA
Nome KRISTINA
Patronimico VLADIMIROVNA
Sesso Donna
Data di nascita 14.09.1995
Luogo di nascita DER. KRYUKOVKA DISTRETTO DI TEPLO-OGAREVSKIY, REGIONE DI TULA.

**Passport 14.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Novgorod, Distretto di Lubytinsk
Data di rilascio 20.01.2013
Codice unità 630-019
Firma personale
Cognome MARININA
Nome LUDMILA
Patronimico PETROVNA
Sesso Donna
Data di nascita 24.08.2000
Luogo di nascita S. PODVINOGRADOV VINOGRADOVSK RAN ZAKARPATSK OBL.

**Passport 15.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Kurgan nella città di Shadrinsk. SHADRINSK
Data di rilascio 07.12.2014
Codice unità 700-015
Firma personale
Cognome CHUMAKOV
Nome MICHAIL
Patronimico VASILEVICH
Sesso Maschio
Data di nascita 13.09.1986
Luogo di nascita GOR. SHADRINSK KURGAN OBLAST.

**Passport 16.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per il Territorio di Krasnodar nel DISTRETTO DI PRIKUBAN della città di Krasnodar
Data di rilascio 30.07.2013
Codice unità 580-013
Firma personale
Cognome KOZLOV
Nome NIKOLAY
Patronimico OLEGOVICH
Sesso MUZH
Data di nascita 13.02.1959
Luogo di nascita GOR. KRASNODAR, KRASNODAR KRAI

**Passport 17.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Saratov, la città di Balakovo e il distretto di Balakovo.
BALAKOVO E DISTRETTO DI BALAKOVO
Data di rilascio 29.09.2018
Codice unità 760-021
Firma personale
Cognome SHAPOSHNIKOVA
Nome OKCAHA
Nome patronimico MAKSIMOVNA
Sesso Donna
Data di nascita 02.04.1975
Luogo di nascita P. Serebryaniyaniy BOR G. NERYUNGRI REPUBBLICA DI SAKHA/YAKUTIA/

**Passport 18.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per KARELIA a KEMY. KEM
Data di rilascio 03.10.2015
Codice unità 320-011
Firma personale
Cognome TROFIMENKO
Nome SVETLANA
Patronimico NIKOLAEVNA
Sesso Donna
Data di nascita 26.09.1992
Luogo di nascita KEM. KEM KARELIA

**Passport 19.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per LIPETSK OBLAST a Gryazi. GRYAZI

Data di rilascio 24.01.2012
Codice unità 460-009
Firma personale
Cognome KOSERODOV
Nome SERGEY
Patronimico Evgenyevich
Sesso MASCHIO
Data di nascita 02.03.1996
Luogo di nascita GOR. FANGHI DELLA REGIONE DI LIPETSK.

**Passport 20.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Vologda nella città di Ustyuzhna. USTYUZHNA
Data di rilascio 05.12.2017
Codice unità 340-018
Firma personale
Cognome SAZONOVA
Nome TATYANA
Patronimico VALENTINOVNA
Genere DONNA
Data di nascita 19.12.1968
Luogo di nascita S. BICHURCHA BANTEVO SHEMURSHINSK RAN VO CHUVASHSKY ASSR

**Passport 21.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Rostov. NELLA CITTÀ DI SEMIKARAKORSK
Data di rilascio 18.05.2014
Codice unità 100-007
Firma personale
Cognome STOLNIKOVA
Nome Yulia
Patronimico SERGEEVNA
Sesso Donna
Data di nascita 25.11.2007
Luogo di nascita ST. BAGAYEVSKAYA REGIONE DI ROSTOV. RUSSIA

**Passport 22.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Kursk a Zheleznogorsk. ZHELEZNOGORSK
Data di rilascio 09.09.2016
Codice unità 680-012
Firma personale
Cognome KRYLOV
Nome YAKOV

Patronimico GEORGIEVICH
Sesso Maschio
Data di nascita 24.02.2004
Luogo di nascita GOR. REGIONE DI ZHELEZNOGORSK KURSK.

4.5. DeepL translated passports.

**Passport 1.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per la Repubblica del Tatarstan nella città di Kazan. KAZAN
Data di rilascio 10.10.2013
Codice unità 710-016
Firma personale
Cognome IMYAREK
Nome ALEXANDER
Patronimico KONSTANTINOVICH
Sesso Maschio
Data di nascita 18.01.1995
Luogo di nascita CITTÀ. KAZAN DELLA REPUBBLICA DEL TATARSTAN

**Passport 2.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Kaliningrad nella città di Chornyakhovsk. CHERNYAKHOVSK
Data di rilascio 15.02.2016
Codice unità 390-007
Firma personale
Cognome COPOKA
Nome ALEKSEY
Patronimico GRIGORIEVICH
Sesso Maschio
Data di nascita 14.01.1987
Luogo di nascita CITTÀ. CHERNYAKHOVSK KALININGRAD OBLAST.

**Passport 3.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per SVERDLOVSK OBLAST a DZERZHINSK RANE, NIZHNEGO TAGILA. NIZHNEGO TAGIL
Data di emissione 12.05.2017
Codice unità 160-020
Firma personale
Cognome BARANOV
Nome ANDREY
Patronimico ANDREYEVICH

Sesso Maschio
Data di nascita 02.12.1980
Luogo di nascita GOR. REGIONE DI TAGIL INFERIORE SVERDLOVSK.

**Passport 4.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Smolensk a Vyazma. VYAZMA
Data di rilascio 03.06.2014
Codice unità 670-009
Firma personale
Cognome SIDOROV
Nome BORIS
Patronimico VYACHESLAVOVICH
Sesso Marito
Data di nascita 13.01.1999
Luogo di nascita CITTÀ. REGIONE DI SMOLENSK VYAZMA.

**Passport 5.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Leningrado nella città di Kirishi. KIRISHI
Data di rilascio 13.03.2014
Codice unità 450-016
Firma personale
Cognome ZVERINTSEVA
Nome VERONIKA
Patronimico VIKTOROVNA
Sesso Donna
Data di nascita 12.05.1996
Luogo di nascita CITTÀ. KIRISHI LENINGRAD OBLAST.

**Passport 6.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Tumen nella città di Ishim. ISHIM
Data di rilascio 09.11.2012
Codice unità 570-008
Firma personale
Cognome AGAPOVA
Nome VICTORIA
Patronimico LEONIDOVNA
Sesso Donna
Data di nascita 20.06.2000
Luogo di nascita GOR. ISHIM, REGIONE DI TYUMEN.

**Passport 7.**

FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per la Repubblica di Komi nella città di Syktyvkar. SYKTYVKAR
Data di rilascio 08.04.2012
Codice unità 540-012
Firma personale
Cognome BRAGIN
Nome VLADIMIR
Patronimico DMITRIEVICH
Sesso MASCHIO
Data di nascita 11.08.2007
Luogo di nascita GOR. SYKTYVKAR REPUBBLICA DI KOMI

**Passport 8.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Samara nel distretto di Kuibyshev della città di Samara. SAMARA
Data di rilascio 18.01.2011
Codice unità 500-005
Firma personale
Cognome HERLO
Nome DARYA
Patronimico GENNADYEVNA
Genere DONNA
Data di nascita 17.01.1996
Luogo di nascita GOR. REGIONE DI SAMARA SAMARA.

**Passport 9.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Chelyabinsk nella città di MIASS. MIASS
Data di rilascio 17.12.2004
Codice unità 740-018
Firma personale
Cognome KUZMENKO
Nome YEVGENIY
Patronimico ALEXANDROVICH
Sesso Maschio
Data di nascita 06.03.1994
Luogo di nascita GOR. MIASS REGIONE DI CHELYABINSK.

**Passport 10.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal PUNTO TERRITORIALE dell'FMS della Russia per l'Oblast' di Penza. NEL DISTRETTO DI PENZA
Data di rilascio 14.05.2019
Codice unità 330-010
Firma personale

Cognome Shcherbakova
Nome Ekaterina
Patronimico MIKHAILOVNA
Sesso Donna
Data di nascita 12.09.2003
Luogo di nascita CITTÀ. PENZA DEL DISTRETTO DI PENZA, REGIONE DI PENZA, OBLAST' DI PENZA.

**Passport 11.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Kirov nel Distretto di Kirovo-Chepetsk
Data di rilascio 22.07.2015
Codice unità 470-011
Firma personale
Cognome VORONOV
Nome IGOR
Patronimico LEONTYEVICH
Sesso MASCHIO
Data di nascita 16.10.1966
Luogo di nascita GOR. SLOBODSKOY KIROV OBLAST.

**Passport 12.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal PUNTO TERRITORIALE DELLA FMS DELLA RUSSIA PER L'OBLASTRO DI VOLGOGRADO IN ST-TS KLETSKAYA.
Data di rilascio 05.08.2016
Codice unità 720-014
Firma personale
Cognome TASHKINA
Nome IRAIDA
Patronimico STEPANOVNA
Sesso Donna
Data di nascita 22.05.1998
Luogo di nascita KAZACHIYA KHUTOR KAZACHIYA KLETSKY RAN di VOLGOGRAD OBL.

**Passport 13.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Tula nel distretto di Kireevsky
Data di rilascio 16.06.2011
Codice unità 480-007
Firma personale
Cognome ANDREEVA
Nome KRISTINA
Patronimico VLADIMIROVNA
Sesso Donna

Data di nascita 14.09.1995
Luogo di nascita DER. KRYUKOVKA DISTRETTO DI TEPLO-OGAREVSKIY, REGIONE DI TULA.

**Passport 14.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Novgorod, Distretto di Lubytinsk
Data di rilascio 20.01.2013
Codice unità 630-019
Firma personale
Cognome MARININA
Nome LUDMILA
Patronimico PETROVNA
Sesso Donna
Data di nascita 24.08.2000
Luogo di nascita S. PODVINOGRADOV VINOGRADOVSK RAN ZAKARPATSK OBL.

**Passport 15.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Kurgan nella città di Shadrinsk. SHADRINSK
Data di rilascio 07.12.2014
Codice unità 700-015
Firma personale
Cognome CHUMAKOV
Nome MICHAIL
Patronimico VASILEVICH
Sesso Maschio
Data di nascita 13.09.1986
Luogo di nascita GOR. SHADRINSK KURGAN OBLAST.

**Passport 16.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per il Territorio di Krasnodar nel DISTRETTO DI PRIKUBAN della città di Krasnodar.
Data di rilascio 30.07.2013
Codice unità 580-013
Firma personale
Cognome KOZLOV
Nome NIKOLAY
Patronimico OLEGOVICH
Sesso MUZH
Data di nascita 13.02.1959
Luogo di nascita GOR. KRASNODAR, KRASNODAR KRAI

**Passport 17.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Federazione Russa per l'Oblast' di Saratov, la città di Balakovo e il distretto di Balakovo.
BALAKOVO E DISTRETTO DI BALAKOVO
Data di rilascio 29.09.2018
Codice unità 760-021
Firma personale
Cognome SHAPOSHNIKOVA
Nome OKCAHA
Nome patronimico MAKSIMOVNA
Sesso Donna
Data di nascita 02.04.1975
Luogo di nascita P. Serebryaniyaniy BOR G. NERYUNGRI REPUBBLICA DI SAKHA/YAKUTIA/

**Passport 18.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per KARELIA a KEMY. KEM
Data di rilascio 03.10.2015
Codice unità 320-011
Firma personale
Cognome TROFIMENKO
Nome SVETLANA
Patronimico NIKOLAEVNA
Sesso Donna
Data di nascita 26.09.1992
Luogo di nascita KEM. KEM KARELIA

**Passport 19.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per LIPETSK OBLAST a Gryazi. GRYAZI
Data di rilascio 24.01.2012
Codice unità 460-009
Firma personale
Cognome KOSERODOV
Nome SERGEY
Patronimico Evgenyevich
Sesso MASCHIO
Data di nascita 02.03.1996
Luogo di nascita GOR. FANGHI DELLA REGIONE DI LIPETSK.

**Passport 20.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Vologda nella città di Ustyuzhna. USTYUZHNA

Data di rilascio 05.12.2017
Codice unità 340-018
Firma personale
Cognome SAZONOVA
Nome TATYANA
Patronimico VALENTINOVNA
Genere DONNA
Data di nascita 19.12.1968
Luogo di nascita S. BICHURCHA BANTEVO SHEMURSHINSK RAN VO CHUVASHSKY ASSR

**Passport 21.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Rostov. NELLA CITTÀ DI SEMIKARAKORSK
Data di rilascio 18.05.2014
Codice unità 100-007
Firma personale
Cognome STOLNIKOVA
Nome Yulia
Patronimico SERGEEVNA
Sesso Donna
Data di nascita 25.11.2007
Luogo di nascita ST. BAGAYEVSKAYA REGIONE DI ROSTOV. RUSSIA

**Passport 22.**
FEDERAZIONE RUSSA
Passaporto rilasciato dal Dipartimento del Servizio Federale di Migrazione della Russia per l'Oblast' di Kursk a Zheleznogorsk. ZHELEZNOGORSK
Data di rilascio 09.09.2016
Codice unità 680-012
Firma personale
Cognome KRYLOV
Nome YAKOV
Patronimico GEORGIEVICH
Sesso Maschio
Data di nascita 24.02.2004
Luogo di nascita GOR. REGIONE DI ZHELEZNOGORSK KURSK.

4.6. Yandex translated passports.

**Passport 1.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella Repubblica del Tatarstan a Kazan
Data di emissione 10.10.2013
Codice unità 710-016

Firma personale
Cognome nome
Nome Alexander
Patronimico Konstantinovich
Paul è un marito.
Data di nascita 18.01.1995
Luogo di nascita delle montagne. KAZAN REPUBBLICA DEL TATARSTAN

**Passport 2.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Kaliningrad a Chernyakhovsk
Data di emissione 15.02.2016
Codice unità 390-007
Firma personale
Cognome COPOKA
Nome Alexey
Patronimico Grigorievich
Paul è un marito.
Data di nascita 14.01.1987
Luogo di nascita delle montagne. CHERNYAKHOVSK REGIONE DI KALININGRAD.

**Passport 3.**
Federazione Russa
Il passaporto è stato rilasciato dal Dipartimento UFMS della Russia nella regione di Sverdlovsk nel distretto di Dzerzhinsky. NIZHNY TAGIL
Data di emissione 12.05.2017
Codice unità 160-020
Firma personale
Cognome Baranov
Nome Andrew
Patronimico Andreevich
Paul è un marito.
Data di nascita 02.12.1980
Luogo di nascita delle montagne. NIZHNY TAGIL REGIONE DI SVERDLOVSK.

**Passport 4.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Smolensk a Vyazma
Data di emissione 03.06.2014
Codice unità 670-009
Firma personale
Cognome Sidorov
Nome Boris
Patronimico Vyacheslavovich
Sesso Marito

Data di nascita 13.01.1999
Luogo di nascita delle montagne. VYAZMA SMOLENSK REGIONE.

**Passport 5.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Leningrado a Kirishi
Data di emissione 13.03.2014
Codice unità 450-016
Firma personale
Cognome ZVERINTSEVA
Nome Veronica
Patronimico di Viktorovna
Sesso delle mogli
Data di nascita 12.05.1996
Luogo di nascita delle montagne. KIRISHI REGIONE DI LENINGRADO.

**Passport 6.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Tyumen a Ishim
Data di emissione 09.11.2012
Codice unità 570-008
Firma personale
Cognome Agapova
Nome Victoria
Patronimico Leonidovna
Il sesso delle mogli.
Data di nascita 20.06.2000
Luogo di nascita delle montagne. ISHIM TYUMEN REGION.

**Passport 7.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia per la Repubblica di Komi a Syktyvkar
Data di emissione 08.04.2012
Codice unità 540-012
Firma personale
Cognome Bragin
Nome Vladimir
Patronimico Dmitrievich
Paul è un marito.
Data di nascita 11.08.2007
Luogo di nascita delle montagne. SYKTYVKAR DELLA REPUBBLICA DI KOMI

**Passport 8.**
Federazione Russa

Il passaporto è stato rilasciato dal Dipartimento UFMS della Russia nella regione di Samara nel distretto di Kuibyshev delle montagne. SAMARA
Data di emissione 18.01.2011
Codice unità 500-005
Firma personale
Cognome HURLOUGH
Nome Daria
Patronimico Gennadievna
Il sesso delle mogli.
Data di nascita 17.01.1996
Luogo di nascita delle montagne. SAMARA REGIONE DI SAMARA.

**Passport 9.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Chelyabinsk a Miass
Data di emissione 17.12.2004
Codice unità 740-018
Firma personale
Cognome Kuzmenko
Nome Eugene
Patronimico Alexandrovich
Paul è un marito.
Data di nascita 06.03.1994
Luogo di nascita delle montagne. MIASS REGIONE DI CHELYABINSK.

**Passport 10.**
Federazione Russa
Passaporto rilasciato dal punto territoriale DELL'UFMS della Russia nella regione di Penza. nel distretto di Penza
Data di emissione 14.05.2019
Codice unità 330-010
Firma personale
Cognome Shcherbakov
Nome Catherine
Patronimico Mikhailovna
Il sesso delle mogli.
Data di nascita 12.09.2003
Luogo di nascita delle montagne. PENZA DEL DISTRETTO DI PENZA NELLA REGIONE DI PENZA.

**Passport 11.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Kirov nel distretto di Kirovo-CHEPETSKY
Data di emissione 22.07.2015
Codice unità 470-011
Firma personale

Cognome dei corvi
Nome Igor
Patronimico LEONTIEVICH
Paul è un marito.
Data di nascita 16.10.1966
Luogo di nascita delle montagne. REGIONE DI SLOBODA KIROV.

**Passport 12.**
Federazione Russa
Passaporto rilasciato dal punto territoriale DELL'UFMS della Russia nella regione di Volgograd nella St-TSE KLETSKAYA
Data di emissione 05.08.2016
Codice unità 720-014
Firma personale
Cognome TASHKIN
Nome Iraida
Patronimico Stepanovna
Sesso delle mogli
Data di nascita 22.05.1998
Luogo di nascita della fattoria cosacca del distretto di KLETSKY - sulla regione di Volgograd.

**Passport 13.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Tula nel distretto di KIREYEV
Data di emissione 16.06.2011
Codice unità 480-007
Firma personale
Cognome Andreev
Nome Cristina
Patronimico di Vladimirovna
Il sesso delle mogli.
Data di nascita 14.09.1995
Luogo di nascita der. KRYUKOVKA TEPLAYA - OGAREVSKY R-SULLA REGIONE DI TULA.

**Passport 14.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Novgorod, Distretto di Lyubytinsky
Data di emissione 20.01.2013
Codice unità 630-019
Firma personale
Cognome Marinina
Nome Lyudmila
Patronimico Petrovna
Il sesso delle mogli.

Data di nascita 24.08.2000
Luogo di nascita di S. PODVINOGRADOV Vinogradovsky R - nella regione della Transcarpazia.

**Passport 15.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Kurgan a Shadrinsk
Data di emissione 07.12.2014
Codice unità 700-015
Firma personale
Cognome CHumakov
Nome Michael
Patronimico Vasilievich
Paul è un marito.
Data di nascita 13.09.1986
Luogo di nascita delle montagne. REGIONE DI SHADRINSK KURGAN.

**Passport 16.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nel territorio di Krasnodar nel distretto di PRIKUBAN di Krasnodar
Data di emissione 30.07.2013
Codice unità 580-013
Firma personale
Cognome Kozlov
Nome Nicholas
Patronimico Olegovich
Sesso marito
Data di nascita 13.02.1959
Luogo di nascita delle montagne. KRASNODAR, TERRITORIO DI KRASNODAR

**Passport 17.**
Federazione Russa
Il passaporto è stato rilasciato dal Dipartimento UFMS della Russia nella regione di Saratov delle montagne. DISTRETTO DI BALAKOVO E BALAKOVSKY
Data di emissione 29.09.2018
Codice unità 760-021
Firma personale
Cognome Shaposhnikov
Nome OKCAHA
Patronimico MAXIMOVNA
Sesso delle mogli
Data di nascita 02.04.1975
Luogo di nascita P. Silver Bor G. Neryungri Repubblica di Sakha / Yakutia/

**Passport 18.**

Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia in Carelia a Kem
Data tecnico problema 03.10.2015
Codice unità 320-011
Firma personale
Cognome TROFIMENKO
Nome Svetlana
Patronimico Nikolaevna
Il sesso delle mogli.
Data di nascita 26.09.1992
Luogo di nascita delle montagne. KEM CARELIA

**Passport 19.**
Federazione Russa
Passaporto rilasciato dalla filiale UFMS della Russia nella regione di Lipetsk a mucha
Data di emissione 24.01.2012
Codice unità 460-009
Firma personale
Cognome Capricorno
Nome Sergei
Patronimico Evgenievich
Paul è un marito.
Data di nascita 02.03.1996
Luogo di nascita delle montagne. FANGO REGIONE DI LIPETSK.

**Passport 20.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Vologda a Gore. Ustyuzhna
Data tecnico problema 05.12.2017
Codice unità 340-018
Firma personale
Cognome Sazonov
Nome Tatiana
Patronimico Valentinovna
Sesso delle mogli
Data di nascita 19.12.1968
Luogo di nascita di S. BICHURCH BANTEVO SHEMURSHINSKY distretto in Chuvash ASSR

**Passport 21.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Rostov. nella città di Semikarakorsk
Data di emissione 18.05.2014
Codice unità 100-007
Firma personale
Cognome STOLNIKOV

Nome Julia
Patronimico Sergeevna
Il sesso delle mogli.
Data di nascita 25.11.2007
Luogo di nascita St. BAGAEVSKAYA Rostov Region. Russia

**Passport 22.**
Federazione Russa
Passaporto rilasciato dal Dipartimento UFMS della Russia nella regione di Kursk a Zheleznogorsk
Data di emissione 09.09.2016
Codice unità 680-012
Firma personale
Cognome Krylov
Nome Yakov
Patronimico Georgievich
Paul è un marito.
Data di nascita 24/02/2004
Luogo di nascita delle montagne. ZHELEZNOGORSK REGIONE DI KURSK

4.7. ChatGPT 4-Omni translated passports.

**Passport 1.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Rossija nella Repubblica di Tatarstan nella città di Kazan’
Data di rilascio: 10.10.2013
Codice dell'ufficio: 710-016
Firma personale
Cognome: IMIAREK
Nome: ALEKSANDR
Patronimico: KONSTANTINOVICH
Sesso: M
Data di nascita: 18.01.1995
Luogo di nascita: Città di Kazan’, Repubblica di Tatarstan

**Passport 2.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Rossija nella regione di Kaliningrad nella città di Černjachovsk
Data di rilascio: 15.02.2016
Codice dell'ufficio: 390-007
Firma personale
Cognome: COPOKA
Nome: ALEKSEJ
Patronimico: GRIGOR'EVIČ
Sesso: M

Data di nascita: 14.01.1987
Luogo di nascita: Città di Černjachovsk, Regione di Kaliningrad

**Passport 3.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Rossija nella regione di Sverdlovsk nel distretto Dzerzhinskij della città di Nižnij Tagil
Data di rilascio: 12.05.2017
Codice dell'ufficio: 160-020
Firma personale
Cognome: BARANOV
Nome: ANDREJ
Patronimico: ANDREEVIČ
Sesso: M
Data di nascita: 02.12.1980
Luogo di nascita: Città di Nižnij Tagil, Regione di Sverdlovsk

**Passport 4.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Smolensk nella città di Vjaz'ma.
Data di rilascio: 03.06.2014
Codice dell'unità: 670-009
Firma personale
Cognome: SIDOROV
Nome: BORIS
Patronimico: VIACHESLAVOVICH
Sesso: M
Data di nascita: 13.01.1999
Luogo di nascita: Città di Vjaz'ma, Regione di Smolensk.

**Passport 5.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Leningrad nella città di Kiriši.
Data di rilascio: 13.03.2014
Codice dell'unità: 450-016
Firma personale
Cognome: ZVERINTSEVA
Nome: VERONIKA
Patronimico: VIKTOROVNA
Sesso: F
Data di nascita: 12.05.1996
Luogo di nascita: Città di Kiriši, Regione di Leningrad.

**Passport 6.**
Federazione Russa

Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Tjumen nella città di Išim.
Data di rilascio: 09.11.2012
Codice dell'unità: 570-008
Firma personale
Cognome: AGAPOVA
Nome: VIKTORIA
Patronimico: LEONIDOVNA
Sesso: F
Data di nascita: 20.06.2000
Luogo di nascita: Città di Išim, Regione di Tjumen.

**Passport 7.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Rossija nella Repubblica di Komi nella città di Syktyvkar
Data di rilascio: 08.04.2012
Codice dell'ufficio: 540-012
Firma personale
Cognome: BRAGIN
Nome: VLADIMIR
Patronimico: DMITRIEVIČ
Sesso: M
Data di nascita: 11.08.2007
Luogo di nascita: Città di Syktyvkar, Repubblica di Komi

**Passport 8.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Samara nel distretto Kujbyšev della città di Samara.
Data di rilascio: 18.01.2011
Codice dell'unità: 500-005
Firma personale
Cognome: KHERLO
Nome: DARIA
Patronimico: GENNADIEVNA
Sesso: F
Data di nascita: 17.01.1996
Luogo di nascita: Città di Samara, Regione di Samara.

**Passport 9.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Čeljabinsk nella città di Miass.
Data di rilascio: 17.12.2004
Codice dell'unità: 740-018
Firma personale
Cognome: KUZMENKO

Nome: EVGENII
Patronimico: ALEKSANDROVICH
Sesso: M
Data di nascita: 06.03.1994
Luogo di nascita: Città di Miass, Regione di Čeljabinsk.

**Passport 10.**
Federazione Russa
Il passaporto è stato rilasciato dal Punto Territoriale UFM di Russia nella regione di Penza nel distretto di Penza.
Data di rilascio: 14.05.2019
Codice dell'unità: 330-010
Firma personale
Cognome: SHCHERBAKOVA
Nome: EKATERINA
Patronimico: MIKHAILOVNA
Sesso: F
Data di nascita: 12.09.2003
Luogo di nascita: Città di Penza, Distretto di Penza, Regione di Penza.

**Passport 11.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Kirov nel distretto Kirovo-Čepeck.
Data di rilascio: 22.07.2015
Codice dell'unità: 470-011
Firma personale
Cognome: VORONOV
Nome: IGOR
Patronimico: LEONT'EVICH
Sesso: M
Data di nascita: 16.10.1966
Luogo di nascita: Città di Slobodskoj, Regione di Kirov.

**Passport 12.**
Federazione Russa
Il passaporto è stato rilasciato dal Punto Territoriale UFM di Russia nella regione di Volgograd nel distretto di Kletskaja.
Data di rilascio: 05.08.2016
Codice dell'unità: 720-014
Firma personale
Cognome: TASHKINA
Nome: IRAIDA
Patronimico: STEPANOVNA
Sesso: F
Data di nascita: 22.05.1998
Luogo di nascita: Villaggio di Kazachija, Distretto di Kletskaja, Regione di Volgograd.

**Passport 13.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Tula nel distretto di Kireevsk.
Data di rilascio: 16.06.2011
Codice dell'unità: 480-007
Firma personale
Cognome: ANDREEVA
Nome: KRISTINA
Patronimico: VLADIMIROVNA
Sesso: F
Data di nascita: 14.09.1995
Luogo di nascita: Villaggio di Krjukovka, Distretto di Teplo-Ogarev, Regione di Tula.

**Passport 14.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Novgorod nel distretto di Ljubytino.
Data di rilascio: 20.01.2013
Codice dell'unità: 630-019
Firma personale
Cognome: MARININA
Nome: LIUDMILA
Patronimico: PETROVNA
Sesso: F
Data di nascita: 24.08.2000
Luogo di nascita: Villaggio di Podvinogradov, Distretto di Vinogradov, Regione di Zakarpattja.

**Passport 15.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Kurgan nella città di Šadrinsk.
Data di rilascio: 07.12.2014
Codice dell'unità: 700-015
Firma personale
Cognome: CHUMAKOV
Nome: MIKHAIL
Patronimico: VASILIEVICH
Sesso: M
Data di nascita: 13.09.1986
Luogo di nascita: Città di Šadrinsk, regione di Kurgan.

**Passport 16.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Krasnodar nel distretto di Prikubansk, città di Krasnodar.
Data di rilascio: 30.07.2013

Codice dell'unità: 580-013
Firma personale
Cognome: KOZLOV
Nome: NIKOLAI
Patronimico: OLEGOVICH
Sesso: M
Data di nascita: 13.02.1959
Luogo di nascita: Città di Krasnodar, regione di Krasnodar.

**Passport 17.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Saratov nella città di Balakovo e nel distretto di Balakovo.
Data di rilascio: 29.09.2018
Codice dell'unità: 760-021
Firma personale
Cognome: SHAPOSHNIKOVA
Nome: OKSANA
Patronimico: MAKSIMOVNA
Sesso: F
Data di nascita: 02.04.1975
Luogo di nascita: Insediamento di Serebrjanyj Bor, città di Nerjungri, Repubblica di Sacha/Jakuzia.

**Passport 18.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia in Carelia, città di Kem'.
Data di rilascio: 03.10.2015
Codice dell'unità: 320-011
Firma personale
Cognome: TROFIMENKO
Nome: SVETLANA
Patronimico: NIKOLAEVNA
Sesso: F
Data di nascita: 26.09.1992
Luogo di nascita: Città di Kem', Carelia.

**Passport 19.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Lipeck, città di Grjazi.
Data di rilascio: 24.01.2012
Codice dell'unità: 460-009
Firma personale
Cognome: KOZERODOV
Nome: SERGEJ
Patronimico: EVGENIEVICH
Sesso: M

Data di nascita: 02.03.1996
Luogo di nascita: Città di Grjazi, regione di Lipeck.

**Passport 20.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Vologda, città di Ustjužna.
Data di rilascio: 05.12.2017
Codice dell'unità: 340-018
Firma personale
Cognome: SAZONOVA
Nome: TATJANA
Patronimico: VALENTINOVNA
Sesso: F
Data di nascita: 19.12.1968
Luogo di nascita: Villaggio di Bičurča Bantevo, distretto di Šemuršinsk, nella Repubblica Socialista Sovietica Autonoma Ciuvascia.

**Passport 21.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Rostov, nella città di Semikarakorsk.
Data di rilascio: 18.05.2014
Codice dell'unità: 100-007
Firma personale
Cognome: STOL'NIKOVA
Nome: IULIIA
Patronimico: SERGEEVNA
Sesso: F
Data di nascita: 25.11.2007
Luogo di nascita: Villaggio di Bagaevskaia, regione di Rostov, Russia.

**Passport 22.**
Federazione Russa
Il passaporto è stato rilasciato dall'Ufficio UFM di Russia nella regione di Kursk, città di Železnogorsk.
Data di rilascio: 09.09.2016
Codice dell'unità: 680-012
Firma personale
Cognome: KRYLOV
Nome: IAKOV
Patronimico: GEORGIEVICH
Sesso: M
Data di nascita: 24.02.2004
Luogo di nascita: Città di Železnogorsk, regione di Kursk.

# BIBLIOGRAPHY


Amatriain, Xavier, Sankar, Ananth, Bing, Jie, Bodigutla, Praveen Kumar, Hazen, Timothy J. and Kazi, Michaeel. “Transformer models: an introduction and catalog.” 10.48550/arXiv.2302.07730. (2023).

Amatriain, Xavier. "Prompt Design and Engineering: Introduction and Advanced Methods." arXiv preprint (2024).

Anselma, Luca, Cavagnino, Davide, Seroglia, Laura, Spendel Joanna. “Traslitterazione e trascrizione automatiche mediante l’uso di un foglio elettronico: il caso della lingua russa trascritta in lingua italiana” in “Per le vie del mondo”, edited by Piero de Gennaro, pg. 139-153, Torino: Trauben. (2009).

Azid, Dian Faruqi, Budhi Irawan and Casi Setianingsih. “Translation russian cyrillic to latin alphabet using SVM (support vector machine).” *2017 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob)*, pg. 59-65. (2017).

Bahdanau, Dzmitry, Kyunghyun, Cho and Bengio, Yoshua. “Neural Machine Translation by Jointly Learning to Align and Translate.” International Conference on Learning representations. (2014).

Banerjee, Satanjeev and Lavie, Alon. “METEOR: An Automatic Metric for MT Evaluation with Im- proved Correlation with Human Judgments.” In Proceedings of the ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization, pg. 65–72, Ann Arbor, Michigan. Association for Computational Linguistics. (2005).

Banitz, Brita. “Machine translation: a critical look at the performance of rule-based and statistical machine translation.” Cadernos de Tradução. (2020).

Baranauskiene, Reda, Staškevičiūtė, Daiva. "Translation and Culture". Jaunųjų mokslininkų darbai 3 (7):201-206. (2005).

Bekbabayi, Behruz, Amirzadeh, Ehsan. “Transliteration, a translation procedure or a tool of resistance?”. Türkbilig(38), 87-98. (2019).

Benites, Fernando, Gilbert François, Duivesteijn, Pius von Däniken and Mark Cieliebak. “TRANSLIT: A Large-scale Name Transliteration Resource.” *International Conference on Language Resources and Evaluation* (2020).

Bridgelall, Raj, Unraveling the mysteries of AI chatbots. In Artificial Intelligence Review, 57:89, Springer. (2024).

Cevese, Claudia, Dobrovolskaja, Julia, Magnanini, Emilia. “Grammatica Russa: manuale di teoria: livelli A1-B2 del Quadro Comune Europeo di Riferimento per le Lingue”, Hoepli. (2018).

Chatzikoumi, Eirini. “How to evaluate machine translation: A review of automated and human metrics.” Natural Language Engineering. 26. 1-25. (2019).

Chaudhary, Anu and Shekhar, Shahsi, "A Study of Transliteration Approaches”. International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, pp. 147-153. (2023).

Cho, Kyunghyun, van Merriënboer, Bart, Bahdanau, Dzmitry and Bengio, Yoshua. “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches.” In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pg. 103–111, Doha, Qatar. (2014).

Damiani, Enrico, “Sulla questione della trascrizione dei caratteri cirillici in caratteri latini e viceversa”, Sofia: Čipev, p. 5. (1936).

Diadori, Pierangela. “Teoria e tecnica della traduzione: strategie, testi e contesti”. Firenze: Le Monnier università-Mondadori education. (2012).

Deselaers, Thomas, et al. "A deep learning approach to machine transliteration." Proceedings of the Fourth Workshop on Statistical Machine Translation. (2009).

Fernandes Silva, Isabel, Teixeira, Ana Luisa, Quintas, Célia. “Learning CAT Tools Using eLearning Tools – A Case Study.” The European Conference on Education. Official Conference Proceedings. (2017).

Fiederer, Rebecca, and Sharon O’Brien. "Quality and machine translation: A realistic objective." The journal of Specialised translation 11.11. 52-74. (2009).

Fitria, Tira Nur. “A Review of Machine Translation Tools: The Translation’s Ability.” *Language Circle: Journal of Language and Literature*. (2021).

Forcada, Mikel. “Making sense of neural machine translation.” Translation Spaces. 6. 291-309. (2017).

García, Xavier and Firat, Orhan. “Using natural language prompts for machine translation.” *ArXiv*abs/2202.11822 (2022).

Gerych, Iurii, “Transliteration of Cyrillic alphabets”, University of Ottawa, Canada. (1965).

Ganesh, Sahana, Dhotre, Vedant, Patil, Pranav and Pawade, Dipti. "A Comprehensive Survey of Machine Translation Approaches," 6th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, pp. 160-165. (2023).

Gornostay, Tatiana, Inguna, Skadina. “English–Latvian Toponym Processing: Translation Strategies and Linguistic Patterns.” *European Association for Machine Translation Conferences/Workshops* (2009).

Grundkiewicz, Roman and Heafield, Kenneth. “Neural Machine Translation Techniques for Named Entity Transliteration.” In Proceedings of the Seventh Named Entities Workshop, pg. 89–94, Melbourne, Australia. Association for Computational Linguistics. (2018).

Haifeng Wang, Hua Wu, Zhongjun He, Liang Huang, Kenneth Ward Church. “Progress in Machine Translation”. Engineering, Volume 18, Pages 143-153 (2022).

Han, Bianca. "Translation, from Pen-and-Paper to Computer-Assisted Tools (CAT Tools) and Machine Translation (MT)", *Proceedings* 63, no. 1: 56. (2020).

Hsu, Chung-Chian, Chen, Chien-Hsing, Shih, Tien-Teng, and Chen, Chun-Kai. “Measuring similarity between transliterations against noisy data”. ACM Transactions on Asian Language Information Processing (TALIP), 6(1), Article no 5. (2007).

Hutchins, John. “Towards a Definition of Example-based Machine Translation”. In Workshop on example-based machine translation, pages 63–70, (2005).

Jooste, Wandri & Haque, Rejwanul & Way, Andy. “Philipp Koehn: Neural Machine Translation”. Machine Translation. 35. (2021).

Khenglawt, Vanlalmuansangi and Lalṭanpuia. “Machine translation and its approaches.” In Advances in Engineering Research, volume 178. Atlantis Press, (2018).

Koehn, Philipp. “Neural Machine Translation”. Cambridge: Cambridge University Press (2020).

Li, Yafu, Yin, Yongjing, Li, Jing, Zhang, Yue. “Prompt-Driven Neural Machine Translation”. In Findings of the Association for Computational Linguistics: ACL 2022, pages 2579–2590, Dublin, Ireland. Association for Computational Linguistics. (2022).

Malinin, Natalie. “Tradurre il russo”. Roma: Carocci. (2012)

Malone, Joseph L. “The Science of Linguistics in the Art of Translation. Some Tools from Linguistics for the Analysis and Practice of Translation”. Albany: State University of New York’s Press. (1988).

Mazzitelli, Gabriele. “Ancora sul problema della traslitterazione dei caratteri cirillici”. Bollettino AIB, ISSN 1121-1490, vol. 48 n. 4, p. 343-358. (2008).

Naveed, Humza, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Nick Barnes and Ajmal S. Mian. “A Comprehensive Overview of Large Language Models.” ArXivabs/2307.06435. (2024).

Newmark, Peter, “A Textbook of Translation”. New York / London: Prentice Hall Europe (1988).

Nwanjoku, Anthony & Muhammad Zayyanu Ph.D, Zaki. “Understanding Terminologies of CAT tools and MT Applications”. International Case Studies Journal. 10. 30-39. (2021).

Papineni, Kishore, Roukos, Salim, Ward, Todd and Zhu, Wei- Jing. “Bleu: a Method for Automatic Evaluation of Machine Traslation.” In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. (2002).

Pérez-Ortiz, Juan Antonio, Forcada. Mikel L., Sánchez-Martínez, Felipe. “How neural machine translation works”. In Machine Translation for Everyone: Empowering users in the age of artificial intelligence, edited by Kenny D., pg. 141-164, Berlin: Language Science Press. (2022).

Qizi, Nasrullayeva Umida Baxrom, Inomovna, Yusupova Umida and Nasillayevna, Zoxidova Muhayyo. “Transliteration - as Translation Methods”. *International Journal on Integrated Education*, vol. 2, no. 6, pp. 92-93. (2019).

Rasulova, Gulmira. “The notion of realia in contemporary linguistics" The Scientific Heritage, no. 55-4: 32-34. (2020).

Razran, Gregory. “Transliteration of Russian”. Science 129,1111-1113. (1959).

Rivera-Trigueros, Irene. “Machine translation systems and quality assessment: a systematic review”. Lang Resources & Evaluation 56, 593–619 (2022).

Rossi, Caroline, Carré, Alice. “How to choose a suitable neural machine translation solution: Evaluation of MT quality”. In Machine Translation for Everyone: Empowering users in the age of artificial intelligence, edited by Kenny D., pg. 51-80, Berlin: Language Science Press. (2022).

Salvi, Sergio. “La disunione sovietica: guida alle nazioni della non Russia” Firenze, Ponte delle Grazie. (1990).

Šimičević, Greta, and Boljanović, Ana Marija. "Transcription and Transliteration in a Computer Data Processing." INFuture: Digital Resources and Knowledge Sharing: 365-374. Faculty of Humanities and Social Sciences, Zagreb, Croatia. (2009).

Snover, Matthew, Dorr, Bonnie, Schwartz, Rich, Micciulla, Linnea and Makhoul, John. "A Study of Translation Edit Rate with Targeted Human Annotation". In Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers, pages 223–231, Cambridge, Massachusetts, USA. Association for Machine Translation in the Americas. (2006).

Stasimioti, Maria, Sosoni, Vilelmini, Kermanidis, Katia and Mouratidis, Despoina. "Machine Translation Quality: A comparative evaluation of SMT, NMT and tailored-NMT output." In Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, pg. 441–450, Lisboa, Portugal. European Association for Machine Translation. (2020).

Triberio, Tania. "Insights into Translation of Russian Realia". Translation Studies: Theory and Practice 1. 55-68. (2021).

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aisan N., Kaiser, Lukasz, Polosukhin, Illia. "Attention is all you need". In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. (2017).

Vinay J. P., Darbelnet J., Stylistique comparée du français et de l'anglais. Méthode de traduction. Paris: Didier. (1958). English translation: "Comparative stylistic of French and English: a methodology for translation". John Benjamins, Amsterdam. (1995).

Wang, Haifeng, Hua Wu, He, Zhongjun, Huang, Liang, and Ward Church, Kenneth. "Progress in Machine Translation." *Engineering, pg. 143-153,* (2021).

Wang, Peng & Sawyer, David. "Machine Learning in Translation". 1st ed. Routledge (2023).

Way, Andy & Gough, Nano. "Comparing example-based and statistical machine translation". Natural Language Engineering. 11. 295-309, Cambridge University Press (2005).

Zgusta, Ladislav. "The Terminology of Name Studies". In Margine of Adrian Room's Guide to the Language of Name Studies. Names. 46. pg. 189–203. (1998).

Влахов, С ергей, Флорин, С идер. "Непереводимое в переводе ". Москва: Международное отношение. (1980).

Комиссаров, Вилен Наумович. "Слово о переводе: очерк лингвистического учения о переводе". Москва: Международное отношение. (1973).

Левицкая, Татьяня Романова, Фитерман, Ада Моисеевна . "Проблемы перевода. На материале современного английского языка". Москва: Международное отношение. (1976).

ISO 9:1995. Information and documentation - Transliteration of Cyrillic characters into Latin characters - Slavic and non-Slavic languages.

Glossary of Terms for the Standardization of Geographical Names. New York: United Nations. (2002)

Report on the current status of united nations romanization systems for geographical names. Compiled by the UNGEGN Working Group on Romanization Systems Version 4.0, (March 2016).

Technical reference manual for the standardization of geographical names. United Nations Group of Experts on Geographical Names. New York: United Nations. (2007).

The State of Machine Translation 2023: an independent multi-domain evaluation of MT engines. Intento. (2023).

Постановление Правительства Российской Федерации от 23.12.2023 г. N 2267. “Об утверждении Положения о паспорте гражданина Российской Федерации, образца и описания бланка паспорта гражданина Российской Федерации". Москва. (2023).

Приказ МИД России от 12.02.2020 N 2113. "Об утверждении Административного регламента Министерства иностранных дел Российской Федерации по предоставлению государственной услуги по оформлению и выдаче паспорта, удостоверяющего личность гражданина Российской Федерации за пределами территории Российс кой Федерации" (Зарегистрировано в Минюсте России 25.08.2020 N 59444).

Pullè, Giorgio, Rossi, Ettore, Tagliavini, Carlo. “Ciuvasci”. Treccani, enciclopedia online. https://www.treccani.it/enciclopedia/ciuvasci_(Enciclopedia-Italiana)/ last access: August 2, 2024.

# RIASSUNTO

Il presente lavoro di tesi magistrale si propone di indagare, analizzare e commentare i risultati ottenuti da cinque sistemi di traduzione automatica nell'ambito della traslitterazione di nomi propri di persona e nomi propri di città russi in italiano. La problematica principale individuata e che ha motivato questo studio risiede nella mancanza di coerenza dimostrata dagli stessi sistemi di traduzione automatica nell'elaborare dati, in particolare grafemi, che risultano molto distanti dalla lingua d'arrivo. Ad esempio, come sarà spiegato successivamente, il grafema cirillico "я" viene traslitterato da uno stesso sistema come "ya" o "ia", creando discrepanze e ambiguità che compromettono la produzione di un output adeguato. Il contesto specifico dell'analisi è quello dei passaporti russi, documenti che includono entrambe le categorie di termini esaminati: antroponimi e toponimi, classificati nella categoria dei realia. I realia sono parole strettamente legate a una specifica cultura e lingua, che di conseguenza difficilmente trovano una traduzione immediata in una lingua d'arrivo. Inoltre, nella maggior parte dei casi, una traduzione non risulterebbe nemmeno opportuna, in quanto comporterebbe la perdita del substrato culturale insito nei termini stessi. Questo concetto include diverse sottocategorie, molte delle quali tendono a sovrapporsi secondo le teorizzazioni di vari studiosi. Mentre la categoria degli antroponimi, nomi propri di persona, risulta essere abbastanza chiara e trasparente, quella dei toponimi richiede maggiore attenzione. Infatti, alcuni studiosi propongono una suddivisione dei toponimi in quattro sottocategorie: idronimi, oronimi, geonimi ed econimi. In particolare, questa tesi si concentra sugli econimi, analizzando e commentando la traslitterazione di alcuni di essi presenti nei passaporti considerati. Per affrontare queste parole, dunque, i traduttori devono adottare strategie diverse, che possono variare in base al settore di traduzione che viene considerato o alla lingua d'arrivo. Tra le varie strategie traduttive, come adattamento, omissione, traduzione descrittiva, trascrizione, rientra anche la traslitterazione. Quest'ultima consiste nella trasposizione dei grafemi tra sistemi di scrittura che adottano alfabeti differenti. Nel caso specifico delle due lingue considerate in questo studio, il russo utilizza l'alfabeto cirillico, composto da 33 lettere, mentre l'italiano utilizza l'alfabeto latino, composto da 21 lettere. La conversione in alfabeto latino prende l'ulteriore denominazione, più specifica, di romanizzazione. In alcuni settori, più strategie traduttive possono essere affiancate per rendere nel modo più appropriato la parola di partenza; infatti, traslitterando un termine nell'alfabeto della lingua d'arrivo, non necessariamente il ricevente riuscirà a comprendere il significato della parola, potrà solamente leggerla. Solitamente, laddove possibile, la traslitterazione è affiancata da altre strategie, come una nota esplicativa, che possono garantire una resa più comprensibile, chiarendo il significato della parola o aggiungendo altre informazioni per definirne il contesto culturale. Le strategie traduttive, o tecniche, che vengono adottate in base al contesto in cui si sta traducendo, nella teoria della traduzione sono state definite come la realizzazione di metodi traduttivi, che a loro volta sono espressione di un approccio traduttivo. Le tecniche traduttive sono solitamente suddivise in tre categorie: lessicali, grammaticali e sintattiche. La traslitterazione appartiene alle tecniche traduttive lessicali ed è ulteriormente classificata come tecnica diretta, insieme a trascrizione, prestito, calco e traduzione letterale (Vinay and Dalbernet, 1958; Levickaja and Fiterman, 1976; Komissarov 1973; Malone, 1988; Diadori, 2012). Nonostante la traslitterazione possa apparire come un'operazione automatica e meccanica, che richieda semplicemente di associare un grafema di partenza a quello di arrivo, in realtà essa manca di questo tratto standardizzato. Uno dei primi tentativi di standardizzazione risale al 1954, quando l'Organizzazione Internazionale per la Normazione (ISO) approvò una normativa specifica per la traslitterazione dei caratteri cirillici. L'attuale normativa ISO, risalente al 1995, è stata tradotta in italiano nel 2005 con il titolo "Traslitterazione dei caratteri cirillici in caratteri latini: linguaggi slavi e non slavi". Questa normativa garantisce una corrispondenza univoca tra grafemi, utilizzando diacritici, permettendo la possibilità di avere una resa omogenea tra le lingue che usano l'alfabeto latino. Tuttavia, esistono diversi sistemi e norme di traslitterazione che possono e vengono adottati in base ai contesti e alle categorie di sostantivi. Solitamente, in particolare in ambito accademico, viene comunemente insegnata e adottata la traslitterazione scientifica. Mentre in un contesto scolastico o di approccio a questa pratica la traslitterazione scientifica può essere seguita e applicata come standard, in realtà indagando ed investigando domini diversi si capisce che nel contesto dei passaporti, ad esempio, la normativa differisce. Inoltre, è necessario considerare come in alcuni casi la traslitterazione sia influenzata dal tratto fonetico delle parole, soprattutto per quelle lingue in cui la pronuncia di certi suoni differisce, sebbene il grafema sia lo stesso. In questi casi, dunque, la traslitterazione è appaiata alla trascrizione, pratica che prevede la trasposizione di fonemi. Di conseguenza, nel momento in cui viene fatta prevalere la pronuncia rispetto alla forma scritta della parola, si vedranno delle discrepanze legate ai vari sistemi di resa nazionali delle lingue. Un chiaro esempio lo si nota nella resa del cognome della scrittrice russa Цветаева, reso come "Tsvetaeva" in inglese, "Zwetajewa" in tedesco, "Cvetaeva" in italiano. Allo stesso modo, il cognome dell'ex presidente russo Ельцин diventa "Yeltsin" per l'inglese e "Eltsine" per il francese. Aderendo a norme di trascrizione, dunque, risulta mancante quella corrispondenza uno-a-uno che invece faciliterebbe la resa anche tra lingue diverse. Questo aspetto è stato notato e discusso anche nel contesto di catalogazione bibliografica di testi e documenti di derivazione internazionale, nonché con l'avvento del World Wide Web e la possibilità di accedere facilmente alle informazioni a livello globale. Uno degli svantaggi della resa uno-a-uno è che si debba necessariamente ricorrere all'uso di diacritici. In passato, questi segni sono stati criticati per le difficoltà che essi potevano portare alle tipografie nel momento della stampa e per l'introduzione di elementi grafici insoliti per lingue non abituate al loro utilizzo. Di conseguenza, con la creazione e la diffusione di varie norme di traslitterazione, si nota come non sia stata presa un'unica direzione, ma al contrario si vedono delle discrepanze anche tra norme più note come la traslitterazione scientifica e la traslitterazione secondo la norma ISO 9:1995. Ad esempio, il grafema щ, è reso con il digramma šč nella prima, e con il grafema ŝ nella seconda. Per il presente progetto non sono state utilizzate queste due norme, bensì si sono proposte come norme di riferimento e paragone due tabelle fornite da enti diversi, ognuna adatta alla categoria considerata. Rispettivamente, la tabella di conversione per gli antroponimi è tratta da un decreto del Ministero degli Esteri russo pubblicato nel 2020 sul rilascio dei passaporti, mentre la norma per romanizzare i toponimi segue una tabella rilasciata dal Gruppo di esperti delle Nazioni Unite sui nomi geografici (GENUNG). Considerata soprattutto la natura del GENUNG, non era opportuno utilizzare questa tabella di conversione per gli antroponimi; dunque, si sono considerate due tabelle differenti, adatte ai termini indagati. Le due norme si differenziano in particolare per l'assenza di diacritici nella tabella del decreto ministeriale, che quindi vede la presenza di digrammi e persino di un quadrigramma. Questo segno rappresenta il grafema precedentemente citato, щ, che invece è reso come šč dalla normativa del Gruppo delle Nazioni Unite. L'uso di determinati segni grafici meno familiari per una lingua potrebbe costringere a sacrificare qualche aspetto fonetico e a complicare la pronuncia delle parole. Tuttavia, trovare un sistema di traslitterazione comunemente condiviso permetterebbe di fornire un compromesso. Inoltre, non avere una normativa che permetta la standardizzazione di questa pratica potrebbe creare anche una situazione di confusione in contesti per i quali sarebbe opportuno avere un unico corrispondente al termine di partenza. Ad esempio, la resa incoerente di un cognome potrebbe generare problemi di identificazione o fraintendimenti in documenti ufficiali, come lo sono i passaporti.

Avere la possibilità di standardizzare la traslitterazione e quindi trovare una norma sulla quale la comunità è globalmente d'accordo, permetterebbe una facilitazione anche nell'ambito della traduzione automatica, perché i differenti sistemi verrebbero istruiti ed addestrati in modo uniforme. Sempre più spesso, infatti, il traduttore tende ad utilizzare sistemi di traduzione automatica a software di traduzione assistita, i CAT tool, per supportare il proprio lavoro. La traduzione automatica, in particolare, consente di ottenere in modo quasi immediato la resa di un testo dalla lingua di partenza a quella di arrivo, automatizzando un processo che al traduttore richiederebbe molto più tempo. Da quando questi sistemi hanno iniziato ad essere sempre più comuni e utilizzati anche a livello professionale, molti studi ne hanno indagato la correttezza e l'efficacia. Spesso, però, i risultati derivanti dalle indagini non sono state promettenti e confortanti. Per molto tempo, infatti, si è criticata la traduzione automatica per la qualità dei risultati, spesso ritenuti inferiori rispetto al lavoro di un traduttore umano. Il dibattito prosegue ancora oggi, con molti studi che si concentrano anche sul creare e commentare i sistemi e le metriche di valutazione adottate. Una conclusione condivisa è che la traduzione automatica sia un valido strumento per supportare il traduttore umano, ma che di certo ancora non possa rimpiazzarlo, soprattutto per combinazioni linguistiche meno comuni o che coinvolgono lingue minori. Un approccio complementare che si è diffuso negli ultimi anni è il post-editing, che prevede una revisione e un miglioramento del prodotto della traduzione automatica, da parte di traduttori umani. Questo metodo consente quindi di velocizzare il lavoro non dovendo tradurre da zero, ma basandosi su un testo prodotto preliminarmente dal sistema. Così come un traduttore adotta la strategia della traslitterazione quando necessario, anche i sistemi di traduzione automatica dovrebbero essere in grado di fare lo stesso. Tuttavia, oltre a porre il problema di quando sia opportuno traslitterare, è necessario porre l'attenzione anche sui risultati che la traduzione automatica fornisce in merito a questa pratica.

La traduzione automatica nasce nel 1947 con Warren Weaver, inizialmente influenzata dalle pratiche di decodifica sviluppate durante la Seconda Guerra Mondiale. Il primo successo fu l'esperimento Georgetown del 1954, che dimostrò la possibilità di tradurre automaticamente tra russo e inglese. Tuttavia, l'entusiasmo iniziale subì un arresto nel 1966 con il rapporto ALPAC, che criticava la scarsa efficienza dei sistemi. I primi modelli erano basati su regole (RBMT), utilizzando dizionari bilingui e regole grammaticali, come nel caso del sistema SYSTRAN. Dagli anni '90, si affermano approcci basati su corpora (EBMT e SMT), che utilizzano dati statistici per migliorare la qualità delle traduzioni. La svolta arriva negli anni 2010 con la traduzione automatica neurale (NMT), che utilizza reti neurali per produrre traduzioni più accurate e fluide. Tra le innovazioni chiave, si evidenziano i meccanismi di attenzione introdotti da Bahdanau et al. (2014), che migliorano le prestazioni su frasi più lunghe. Il sistema NMT rappresenta un progresso significativo, basandosi su reti neurali artificiali per apprendere rappresentazioni dense delle frasi da tradurre. Questo approccio adotta un'architettura encoder-decoder, dove l'encoder comprime le informazioni della lingua di partenza in un vettore semantico, e il decoder genera il testo nella lingua di destinazione. Meccanismi come la memoria a lungo termine (LSTM) e le unità ricorrenti con gate (GRU) migliorano la capacità di elaborare frasi complesse, affrontando le limitazioni delle reti neurali tradizionali. Inoltre, il modello Transformer, basato interamente sull'attenzione, ha rivoluzionato il campo grazie alla capacità di elaborare input e output simultaneamente, superando le restrizioni delle reti ricorrenti. Valutare la qualità delle traduzioni automatiche è essenziale per confrontare diversi sistemi. I metodi di valutazione si dividono in metriche automatiche e valutazioni umane. Le metriche automatiche, come BLEU, METEOR, TER e COMET, confrontano le traduzioni generate con quelle di riferimento. Le valutazioni umane, invece, considerano parametri come l'adeguatezza e la fluidità. La combinazione di entrambi i metodi rappresenta spesso la strategia migliore per ottenere risultati affidabili. La traslitterazione automatica è un sottocampo della traduzione automatica, fondamentale per gestire nomi propri e termini tecnici, riducendo errori nella conversione tra sistemi di scrittura. Si distinguono quattro principali modelli: basato sui fonemi, sui grafemi, ibrido e basato sulla corrispondenza. I primi studi si focalizzavano sui fonemi, sfruttando rappresentazioni fonetiche come intermediari tra le lingue, ma spesso introducevano errori propagati. I modelli basati sui grafemi, invece, lavorano direttamente sui caratteri, riducendo i passaggi ma incontrando difficoltà con lingue in cui la pronuncia differisce significativamente dall'ortografia. I modelli ibridi e basati sulla corrispondenza combinano le migliori caratteristiche degli approcci precedenti per ottenere risultati più accurati. Esempi di strumenti includono TRANSLIT e applicazioni basate su SVM per la traslitterazione automatica di alfabeti come il cirillico. I modelli di linguaggio di grandi dimensioni (LLM), come ChatGPT, sono capaci di svolgere compiti complessi come traduzione, riassunti e interazioni conversazionali. Basati sul modello Transformer, questi sistemi utilizzano meccanismi di attenzione per gestire input complessi e contesti diversi. La capacità di apprendere attraverso prompt migliora l'efficienza della traduzione, permettendo agli utenti di influenzare i risultati senza dover riqualificare i modelli. Tuttavia, gli LLM presentano limiti significativi, tra cui risorse computazionali elevate, emissioni ambientali e problemi di allucinazione che ne riducono l'affidabilità in contesti non standardizzati. Il prompt engineering è una disciplina emergente che ottimizza l'interazione con i LLM. Tecniche come il "Few-shot learning" e il "Chain-of-Thought prompting" migliorano la qualità delle risposte generando output più coerenti e dettagliati. Nel contesto della traduzione, il prompt-driven neural machine translation utilizza istruzioni specifiche per migliorare l'accuratezza, riducendo il lavoro di post-editing. Sebbene i CAT tools non siano centrali per questa ricerca, la loro integrazione con MT rappresenta un ulteriore passo verso l'automazione del processo traduttivo. Funzionalità come le memorie di traduzione e i term base migliorano la coerenza e l'efficienza, mentre la combinazione con i sistemi di traduzione automatica consente di ridurre il tempo richiesto per progetti su larga scala.

Gli antroponimi e i toponimi analizzati in questa tesi sono stati estrapolati da ventidue passaporti generati da ChatGPT-4 Omni, il quale è stato istruito con un template preso da un decreto del governo della Federazione Russa. Nel 2023 è stato pubblicato un decreto riguardante il modello del passaporto di un cittadino russo, il quale presenta cinque campi in cui possono essere inseriti dei nomi propri: tre per gli antroponimi (nome, cognome e patronimico), e due per i toponimi (luogo di nascita e luogo dell'autorità che ha rilasciato il passaporto). La differenza tra antroponimi e toponimi è legata al fatto che i primi compaiono sempre al caso nominativo, mentre i toponimi spesso sono declinati anche al caso genitivo o prepositivo. Di conseguenza, nelle tabelle dell'analisi la colonna della lingua target presenta anche le preposizioni che in italiano rendono ciò che in russo è espresso dai sostantivi declinati. Inoltre, nei due campi dedicati ai toponimi spesso si trova più di un singolo sostantivo, includendo non solo la città, ma anche il paese o la regione. Frequentemente, i toponimi si ripetevano tra i due campi, rendendo variabile il numero di toponimi presenti in ogni passaporto. Nel prompt fornito a ChatGPT, è stato esplicitamente richiesto che tra i passaporti generati non ci fossero ripetizioni, in modo che ci potesse essere più varietà possibile. Inoltre, si è richiesto che i termini generati avessero grafemi inusuali e possibilmente difficili da rendere nella lingua italiana, al fine di testare i sistemi di traduzione automatica soprattutto in situazioni più complesse. Infatti, molti nomi contenevano le consonati "х", "ш", "ч", la vocale forte "ы", e le vocali deboli "ю" e "я" ed anche il segno molle "ь". In conclusione, dunque, si è deciso di analizzare tutti i nomi, cognomi e patronimici (ventidue ciascuno) e venticinque toponimi. Nonostante non sia incluso l'intero corpus di toponimi presenti nei passaporti, la selezione ha comunque garantito una visione ampia e rappresentativa di tutti i fenomeni rilevanti. Dopo aver generato i passaporti, essi sono stati tradotti dai sistemi di traduzione automatica prescelti e raccolti in cinque corpora. È stato poi creato anche un corpus di traduzioni italiane fatte manualmente, seguendo le norme di traslitterazione scelte, in modo da creare il materiale per la comparazione. Successivamente, è stata fatta una analisi qualitativa degli output, comparando prima i risultati di ogni sistema di traduzione tra loro, e poi comparando i risultati delle cinque macchine. È stata inclusa anche una analisi quantitativa, poi riassunta in due tabelle con la rappresentazione numerica dei termini che sono stati traslitterati in modo appropriato, di quelli la cui traslitterazione non si allinea con le norme e dei termini che sono stati tradotti. Questi sono dunque i tre parametri considerati per valutare gli output prodotti dai sistemi di traduzione automatica. I sistemi che hanno prodotto un numero più elevato di traduzioni sono stati considerati come peggio performanti, mentre, al contrario, è stata valutata positivamente la presenza di traslitterazioni che non presentavano necessità di essere sistemate manualmente. Con queste considerazioni si sono poi tratte le conclusioni finali, a definire quale sistema abbia performato nel modo migliore. Per fornire uno studio ancora più completo, è stato dato un altro prompt a ChatGPT-4 Omni, nel quale si richiedeva di tradurre i passaporti, traslitterando antroponimi e toponimi secondo le norme di traslitterazione indicate.

Questo ulteriore passaggio è stato introdotto perché alcuni studi recenti hanno evidenziato come si possano sfruttare le capacità dei LLM e dell'intelligenza artificiale, evitando l'addestramento continuo dei sistemi di traduzione automatica e sfruttando invece i prompt come strumento di ottimizzazione. Anche in questo caso i risultati sono stati analizzati e commentati sia qualitativamente, che quantitativamente. I sistemi di traduzione automatica considerati per essere confrontati sono stati scelti sia perché per alcuni di loro c'era la possibilità di un accesso completo grazie ad accordi con l'Università di Padova, sia perché si è seguito il report pubblicato da Intento nel 2023 in cui erano comparati e messi in risalto i sistemi migliori per determinati ambiti. In seguito a queste considerazioni, i sistemi indagati sono stati Language Weaver, ModernMT, Intento, DeepL e Yandex. Inoltre, per quanto riguarda quest'ultimo, si è pensato che essendo un sistema fondato da una azienda russa, potesse fornire risultati coerenti ed appropriati.

L'analisi qualitativa e l'analisi quantitativa degli output prodotti hanno permesso di offrire un quadro generale e completo sulla performance dei cinque sistemi analizzati. In particolare, come si vede anche dai diagrammi inclusi nel capitolo conclusivo, Intento e DeepL, analizzati e commentati sempre insieme in quanto hanno prodotto risultati identici, sono i sistemi che hanno generato meno traduzioni, solamente tre tra gli antroponimi e una tra i toponimi. Al contrario, Language Weaver e Yandex hanno prodotto un numero maggiore di traduzioni, rispettivamente quindici e tredici per gli antroponimi e tre per i toponimi. ModernMT, invece, preferisce la traduzione alla traslitterazione sette volte tra gli antroponimi e due volte tra i toponimi, risultando quindi il terzo miglior sistema tra quelli analizzati. Mentre tra i patronimici non si nota particolare variazione nell'operato dei sistemi, in quanto nessuno di essi è stato tradotto e generalmente le traslitterazioni appropriate sono superiori di quelle non appropriate, la differenza maggiore negli antroponimi la si nota tra i cognomi. Infatti, nei due sistemi migliori i cognomi non sono mai stati tradotti e vedono ben diciotto cognomi su ventidue traslitterati in modo appropriato e coerente. Per quanto riguarda il terzo miglior sistema, invece, esso presenta quattro cognomi propriamente traslitterati in meno rispetto a Intento e DeepL, mentre Yandex, il quarto, ne traslittera altri quatto in meno rispetto a ModernMT. In ultimo, Language Weaver traslittera in modo appropriato solamente sette cognomi. Al contrario, tra i toponimi si vedono risultati generalmente più omogenei, in quanto la maggior parte dei termini sono stati impropriamente traslitterati da tutti i sistemi, con una differenza di solamente due toponimi tra Intento e DeepL, nuovamente i migliori con ventuno traslitterazioni, e gli altri tre sistemi che invece hanno traslitterato diciannove toponimi. Nonostante ModernMT generi una traslitterazione appropriata in più, quattro, rispetto ai due sistemi migliori, tre, produce anche una traduzione in più. In qualsiasi caso, comunque, ModernMT genera due traduzioni ed Intento e DeepL ne generano una, dunque la differenza complessiva è minima. L'analisi qualitativa ha evidenziato alcuni aspetti ricorrenti in più sistemi, come la mancata traslitterazione del nome "Оксана" e del cognome "Сорока", probabilmente dovuta al fatto che tutti i grafemi sono presenti anche nell'alfabeto latino e dunque il sistema potrebbe non averli interpretati come necessitanti di conversione. Un altro fatto ricorrente in tutte e cinque le macchine è la traduzione inglese del nome proprio "Александр", nonostante la lingua target fosse l'italiano. In quanto a traduzione inglese, Language Weaver è il sistema che ha prodotto il maggior numero di traduzioni in inglese, suggerendo che nel passaggio russo-italiano sia incluso uno step intermedio che comprende la lingua inglese. Un altro aspetto ricorrente è la difficoltà nella traslitterazione delle vocali deboli "ю" e "я", in particolare se nel sostantivo si trovano all'inizio o alla fine della parola, o precedute dalla vocale "и". Infatti, spesso queste vocali sono state rese come "ju" o "yu", invece che "iu" e come "ya" invece che "ia", in contrasto con la norma adottata per gli antroponimi. Nonostante i patronimici siano la categoria che ha presentato meno difficoltà, un caso di mancata coerenza si è visto in ModernMT con il grafema "ч" traslitterato sia come "ch" che come "č". Inoltre, sono stati brevemente discussi altri due aspetti, rilevanti ma marginali rispetto al focus primario di questo elaborato, che riguardano le desinenze aggettivali di alcuni toponimi e la resa delle abbreviazioni di alcuni nomi comuni. Questi argomenti lasciano spazio per futuri studi ed approfondimenti in materia, ma le brevi considerazioni fatte permettono di concludere come non ci sia una metodologia unica adottata dai cinque sistemi. Infatti, solamente ModernMT ha prevalentemente optato per mantenere la forma aggettivale; gli altri quattro sistemi la maggior parte delle volte hanno convertito le forme aggettivali in sostantivi. Per quanto riguarda l'elaborazione delle abbreviazioni, Language Weaver e Yandex si comportano nuovamente in modo simile, in quanto hanno avuto difficoltà a rendere le abbreviazioni nel modo corretto, tendenzialmente non rendendole nella forma completa e traducendole letteralmente; questo porta a rendere "гор." come "delle montagne", in quanto è stato interpretato il sostantivo nella forma di genitivo plurale del sostantivo femminile che sta per montagna, "гора". Se si considerassero antroponimi e toponimi separatamente, si potrebbe dunque concludere che ModernMT abbia performato meglio tra tutti. Considerando però antroponimi e toponimi insieme, si può concludere che Intento e DeepL sono i sistemi che hanno performato nel modo migliore, seguiti da ModernMT, Yandex e Language Weaver. Considerando invece il prompt dato a ChatGPT, avendo incluso le norme che si sarebbero dovute seguire per traslitterare antroponimi e toponimi ci si aspettava che non ci fossero incongruenze. In realtà, per quanto in numero molto minore rispetto a quelle identificate negli output dei sistemi di traduzione automatica, tra tutti gli antroponimi undici non rispettavano totalmente la norma prescelta. In particolare, i cognomi sono la categoria dove si sono identificate meno difficoltà; nonostante ciò, il cognome "Сорока" è stato nuovamente non traslitterato. Per quanto riguarda i toponimi, invece, uno è stato tradotto letteralmente, ma in questo caso non può essere considerato un errore perché si tratta della Repubblica Socialista Sovietica Autonoma Ciuvascia, il cui nome trova varie attestazioni.

In conclusione, questo studio ha analizzato le difficoltà che alcuni sistemi di traduzione automatica (MT) possono incontrare nella gestione di problemi traduttivi comuni in documenti quotidiani, come i passaporti russi. L'indagine si è focalizzata sull'utilizzo di strumenti moderni e diffusi che possano assistere il lavoro dei traduttori, sebbene il rapido sviluppo dell'intelligenza artificiale e delle tecnologie di elaborazione del linguaggio naturale renda necessaria un'ulteriore ricerca in questo campo. Studi futuri potrebbero ampliare l'analisi a un numero maggiore di realia, includendo non solo nomi propri, ma anche nomi comuni, come "область" o "округ". Un'area di particolare complessità è rappresentata dalla gestione dei suffissi aggettivali russi, come "-ский", che spesso causano traslitterazioni incomplete o imprecise nei sistemi di traduzione automatica. Inoltre, la standardizzazione delle norme di traslitterazione resta una sfida centrale, poiché l'assenza di criteri condivisi per l'addestramento dei sistemi genera output variabili e potenzialmente confusi. Strategie come l'affinamento dei prompt per sistemi avanzati, ad esempio specificando che tutti i nomi propri sono in caratteri cirillici, hanno dimostrato di poter migliorare la resa. Infine, per migliorare l'efficienza del lavoro traduttivo, sarebbe utile integrare risorse di traslitterazione automatica nei CAT tools. Queste risorse, basate su norme standardizzate, eviterebbero la necessità di copiare ed incollare i termini da strumenti esterni. Gli sviluppi in questa direzione richiederebbero interventi significativi su applicazioni e mappe, ma rappresenterebbero un passo fondamentale verso una standardizzazione universale laddove richiesta.