DATA &
METADATA

Check for
updates

SYSTEMATIC REVIEW

Translation as a linguistic act in the context of artificial intelligence: the impact of technological changes on traditional approaches

La traducción como acto lingüístico en el contexto de la inteligencia artificial: el impacto de los cambios tecnológicos en los enfoques tradicionales

Nataliia Yuhan[1] ✉, Yuliia Herasymenko[2] ✉, Oleksandra Deichakivska[3] ✉, Anzhelika Solodka[4] ✉, Yevhen Kozlov[5] ✉

[1]State Institution «Luhansk Taras Shevchenko National University», Department of Oriental Philology and Translation, Educational and Scientific Institute of Philology and Translation. Poltava, Ukraine.
[2]Berdyansk State Pedagogical University, Department of Foreign Languages and Teaching Methods. Berdyansk, Ukraine.
[3]Ivan Franko National University of Lviv, Department of English Philology, Faculty of Foreign Languages. Lviv, Ukraine.
[4]Admiral Makarov National University of Shipbuilding, Department of German Philology, Faculty of Philology. Mykolaiv, Ukraine.
[5]National Technical University «Kharkiv Polytechnic Institute», Department of Business Foreign Language and Translation, Educational and Scientific Institute of Social and Humanitarian Technologies. Kharkiv, Ukraine.

ABSTRACT

The purpose of this article is to study translation as a human speech act in the context of artificial intelligence. Using the method of analysing the related literature, the article focuses on the impact of technological changes on traditional approaches and explores the links between these concepts and their emergence in linguistics and automatic language processing methods. The results show that the main methods include stochastic, rule-based, and methods based on finite automata or expressions. Studies have shown that stochastic methods are used for text labelling and resolving ambiguities in the definition of word categories, while contextual rules are used as auxiliary methods. It is also necessary to consider the various factors affecting automatic language processing and combine statistical and linguistic methods to achieve better translation results. Conclusions - In order to improve the performance and efficiency of translation systems, it is important to use a comprehensive approach that combines various techniques and machine learning methods. The research confirms the importance of automated language processing in the fields of AI and linguistics, where statistical methods play a significant role in achieving better results.

Keywords: Technological Changes; Linguistics; Innovations; Language Technologies; Automatic Translation.

RESUMEN

El propósito de este estudio es explorar la traducción como acto lingüístico humano en el contexto de la inteligencia artificial. A través del análisis de literatura pertinente, se enfoca en cómo los cambios tecnológicos alteran los métodos tradicionales y estudia las conexiones entre estos y su emergencia en la lingüística y el procesamiento automático del lenguaje. Se identifican como principales técnicas los métodos estocásticos, basados en reglas y los automatizados mediante autómatas o expresiones finitas. Las investigaciones indican que los métodos estocásticos son útiles para el etiquetado de textos y resolver ambigüedades en la categorización de palabras, mientras que las reglas contextuales funcionan como apoyo. Es crucial considerar los distintos factores que influyen en el procesamiento del lenguaje y emplear una mezcla de técnicas estadísticas y lingüísticas para optimizar los resultados de la traducción. Las conclusiones

subrayan la necesidad de un enfoque integral que integre múltiples técnicas y métodos de aprendizaje automático para mejorar el rendimiento de los sistemas de traducción. Este análisis refuerza la relevancia del procesamiento automático del lenguaje en la intersección de la IA y la lingüística, donde los enfoques estadísticos son clave para obtener resultados superiores.

Palabras clave: Cambios Tecnológicos; Lingüística; Innovaciones; Tecnologías Del Lenguaje; Traducción Automática.

INTRODUCTION

Introduce the Problem

The last decade has seen significant changes in the technology sector. This has contributed to the rapid development of the fields of machine translation (MT) and natural language processing (NLP). These areas remain relevant in the field of translation studies.[39] Significant progress in research on text and speech translation automation, including machine translation, speech synthesis, and recognition, as well as the use of evaluation metrics, combined with the development of artificial intelligence (AI), is actively expanding the scope of translation tools.[5] These technological innovations, along with the democratisation of neural networks, are leading to a rethinking of the way translation professions work and are organised.

This trend is supported by the rapid growth of the language industry and companies' active investment in translation technologies to integrate them into various business processes, interfaces, platforms, and applications.[21] In recent years, the implications of these developments have gone beyond scientific research. The status of a professional translator is undergoing changes, as is the general idea of the essence of translation.[28] Thus, two key transformations are taking place. The first significant change is the key role of data in the translation process. Collecting, cleaning, annotating, and structuring big data from corpora are critical steps for effective training of translation algorithms.[25] This data can be considered a valuable resource or even part of the common good.[34] All of these aspects need to be carefully assessed, put into perspective, and possibly regulated according to the specifics of translation and the needs of society as a whole.

A second important transformation is the shift from a content-focused approach to one that emphasises the use of the big data, voluminous documents that are now being created, distributed, evaluated, and recycled online. This calls into question the distinction between texts "for information", "purely functional" or "literary texts".[1] Moreover, flexible content management, user engagement metrics, and texts generated automatically by large language models such as the GPT-4 model[50] and its related models imply an interaction between human linguistic production and machine translation that requires further study and analysis.[20]

The pace and scale of technological, economic, and societal change raises a number of questions. The enthusiasm for neural machine translation contrasts with numerous industries where language processing technologies, although advanced, are still not fully developed. In this context, we can observe a certain gap between languages that have sufficient digital resources and those for which such resources are insufficient.[5] This can also mean a gap in access to machine translation interfaces, limitations in the development of training corpora, and unclear translations.

Explore importance of the problem

The main issue that stands out in the discussion is the uncertainty surrounding the role, status, and prospects of translators, as well as the sustainability of the traditional translation model.[46] In the face of these significant changes, translator training is at a crossroads as it needs to find ways to align these new trends with more traditional skills and approaches. Of course, it's hard to refuse the possibilities of progress, but in the field of translation, AI programs still have a large number of shortcomings. Attempts to solve the problem by increasing capacity have led to an increase in the number of unresolved issues, which exacerbates the problem. For example, Bohatyrets V.[6] describes errors in machine translation that are often associated with incorrect recognition of adjective functions, including attributive and predicative. For example, when an adjective is used before a noun or when an adjective is used as part of a predicate and indicates a feature or state of the subject of a sentence. In such cases, the adjective is placed after the verb "to be" (are, is) and indicates a feature of the subject. The author notes that such mistakes in machine translation are the least of them. The machine can confuse these functions, which leads to translation errors, and requires additional efforts to correctly determine the context and use adjectives correctly in translation.

In this aspect, scientists identify the criteria for NLP software reliability,[41] which can be described in three aspects (figure 1):
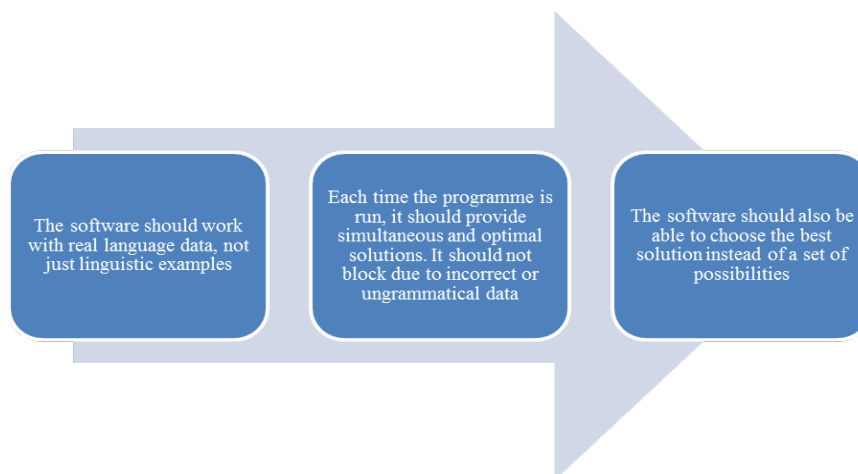
Figure 1. Reliability criteria for language software

The problem that the article highlights is the need for software to process real language data instead of being limited to linguistic examples. A key aspect is the need for the software to be able to provide reliable and optimal solutions every time it is run, to avoid blocking due to incorrect or ungrammatical data, and to select the best solution among different possibilities. It is important that the NLP software is reliable and efficient, and its performance should be evaluated against the user's needs and expectations to select the best option.

Describe relevant scholarship

In recent years, the idea of adapting machine translation to literary texts for probabilistic automatic translation systems has emerged. The main problems arise in the context of using neural systems. Neural networks have several advantages, such as less literal and more natural translation, efficiency in dealing with texts containing a large number of words, and fewer errors.[36] This leads to improved results in the translation of literary works. The architecture of neural networks with attention mechanisms has further improved the quality of translations.[30] Automatic machine translation models allow capturing linguistic dependencies on a large scale and have become a topic of discussion in translation studies, attracting more and more attention from scholars.[14] The use of these models is at the heart of the GPT-2 and GPT-3 language models, which are widely used in the press and social media.

One of the features of the neural approach is its need for a large amount of data and the need to have training corpora from a specific industry to achieve the highest performance.[7] Adapting neural networks to specific languages or industries where there are limited resources is one of the key challenges today. Many methods have been developed to adapt neural networks to new domains in order to expand the capabilities of machine translation.[33] The ideal scenario is to train the system only on adapted data,[11] but this scenario is not always realistic in practice and is a central challenge in the field of automatic machine translation of literature.

In both educational and linguistic fields, the push towards digital solutions in response to immediate and substantial needs (be it a global pandemic or the requirements of global digital communication) underscores a shared trajectory towards more integrated, technologically reliant methodologies. The findings from[17] research underscore the importance of flexibility, rapid adaptability, and the ongoing refinement of technological applications in real-world scenarios, principles that are equally relevant to the advancement of translation technologies.

In any case, the creation and provision of industry-specific translation engines has become quite commonplace and is one of the main selling points of service providers today. This conclusion is supported by Iskakova M.[19] study, which points to the need for machine translation systems to adapt to literary texts. With the emergence of artificial intelligence, there is a growing interest in creating specialised machine translation systems for literature.[44] Additional studies evaluating the effectiveness of publicly available systems such as Google Translate or DeepL allow us to assess the quality of the translations they produce, but it is important to remember that these systems are not specifically designed for literary translation, so it may be unfair to dismiss the possibility of automatic literary translation based on them. Work on automatic literary translation, such as,[24] shows that successful results can be obtained using machine translation systems trained on literary data.

State hypotheses and their correspondence to research design

The research hypothesis is that advances in neural machine translation and natural language processing have revolutionised the field of translation studies, making machine translation tools more widely applicable

in professional environments. The integration of specialised text corpora has allowed for adaptability to specific fields, expanding the use of this technology in translation and education. However, there is still a gap in the understanding and acceptance of machine translation among translators and linguists, leading to ongoing discussions and debates. The purpose of this study is to investigate the relationship between linguistic theories, computational methods, and the practical application of automatic translation tools. Along with the hypothesis, the research questions are outlined: What is the current state of neural machine translation and natural language processing technologies in the field of translation studies? How do specialised corpora affect the adaptability of machine translation tools for specific fields? What challenges and limitations do translators and linguists face when implementing machine translation technologies?

METHOD

The method used in this study is literature analysis. This study is a systematic review focusing on articles, publications, and studies related to neural machine translation and translation studies. It uses data from specialized text corpora to assess the adaptability of machine translation tools to specific industries. The study also included descriptive statistics, comparative analysis, and content analysis of scientific sources to identify trends, differences, and similarities in approaches to automatic translation. The participant (subject) characteristics involve evaluating the reliability and accuracy of machine translation software for practical use and exploring the challenges that translators and linguists face in implementing these technologies.

Identify subsections

Descriptive statistics were also used to analyses the data to identify the main trends in the industry. Comparative analysis and content analysis of scientific sources were also conducted to identify the main differences and similarities in approaches to automatic translation.

Participant (subject) characteristics

These methods made it possible to conduct a study and evaluate the reliability and accuracy of machine translation software for practical use. The content analysis of the related literature confirmed the generalized opinion of scholars that neural machine translation technologies have improved significantly and become more widely used in the professional activities and education of translators. However, there are challenges that translators and linguists face in the process of implementing these technologies. The sample for this study consisted of various texts and literature related to machine translation technologies, particularly focusing on neural machine translation. The information was collected through a systematic review of academic papers, books, articles, and other sources discussing the advancements and challenges of machine translation technologies.

The selection process involved identifying relevant literature using specific search terms and databases such as Google Scholar, JSTOR, and other academic repositories. The selected texts were then reviewed and analyzed to extract key findings and insights about the reliability and accuracy of machine translation software. The information was processed by conducting a content analysis of the literature, which involved categorizing and summarizing the key themes, trends, and challenges identified. The analysis focused on evaluating the current state of machine translation technologies, the improvements made in neural machine translation, and the practical implications for translators and linguists. To reproduce the study, researchers can follow a similar methodology by conducting a systematic search for literature on machine translation technologies, selecting relevant texts, and conducting a content analysis to evaluate the reliability and accuracy of the software. Researchers can also replicate the study by using different sources or databases to gather information and by applying similar criteria for selecting relevant texts for analysis.

Sampling procedures

The integration of computational methods and linguistic theories can help improve the effectiveness of automatic translation tools. The use of these methods has proven that this integration can lead to improved translations and reduced errors.

Sample size, power, and precision

One of the main challenges faced by translators and linguists is the persistent gap in understanding and acceptance of machine translation. This can lead to discussions and debates about its effectiveness and reliability.

Measures and covariates

The study showed that in recent years, neural machine translation technologies have improved significantly, allowing them to become more widely used in the professional activities and education of translators.

The integration of specialised text corpora has also contributed to the adaptability of machine translation technologies to various industries.

Research design

Thus, the study was aimed at analysing the current state of neural machine translation technologies and their impact on the field of translation studies. The challenges faced by translators and linguists in the process of implementing new technologies were investigated. The possibilities of integrating computational methods and linguistic theories to improve the efficiency of automatic translation were also explored.

RESULTS AND DISCUSSION
Effective methods of natural language processing

In general, there are two main methods of building an architecture in the field of translation using artificial intelligence systems: stochastic and rule-based methods. There is also a third type of method based on finite automata or expressions, which can be seen as an intermediate between stochastic and theoretical methods of natural language processing.[3]

Stochastic methods rely on statistical models and probabilistic approaches to generate translations based on large amounts of data. Some practical examples are demonstrated below (table 1):

Table 1. Examples of the main methods of building an architecture in the field of translation	
Method	Example
Statistical Machine Translation (SMT)	Research introduced the IBM Models for machine translation, which used statistical methods to align words and phrases in parallel corpora. These models computed translation probabilities based on observed data and were foundational in the development of SMT systems.
Phrase-Based Machine Translation:	The phrase-based model, which breaks sentences into smaller phrases and translates them based on statistical patterns learned from training data. This approach improved translation accuracy by considering larger contextual units than word-based models.
Neural Machine Translation (NMT)	Subsequent advances in NMT architecture represent a shift towards deep learning methods in machine translation. NMT models use neural networks to learn mappings from source to target language sequences, achieving state-of-the-art performance by capturing complex dependencies and long-range dependencies.
Rule-Based Methods	Early rule-based systems like SYSTRAN, developed in the 1960s, used linguistic rules and dictionaries to perform translation. These systems relied on grammatical rules, syntactic structures, and dictionaries of word translations curated by experts.

In the context of translation as a linguistic act in the context of artificial intelligence, let's take a closer look at stochastic methods, as they have the greatest impact on traditional methods. Stochastic methods are rule-based methods used, for example, for text labelling, i.e., for assigning a category or part of speech to each word in a text.[8] When working with lexicons, translators often encounter ambiguities: for example, the word "book" can be a noun, verb, or adjective, and "pen" can be a plural or singular noun.[31]

Rules that help resolve ambiguity in determining the possible categories for a word are called contextual rules because they use words that come before or after it. Usually, these rules are based on the immediate context, i.e. words that are close by, one, two, or more words away. These rules can be word-specific or more general, targeting categories. For example:
- After the article, the word is a noun (not a verb).
- After the verb "to have", the word was the past participle of the verb "to be".
- After "a", the word "book" is a masculine noun.

Obviously, these rules are not a description of the language grammar, but they help to resolve situations of ambiguity in defining categories for words.

In this context, it can be noted that these are ad hoc rules. Their main purpose is to be integrated into computer processing, and in most cases, their use is "procedural", i.e. it is important which rule is applied.[35] For example, for labelling, we stop the procedure as soon as a word is assigned a unique label.

The rule-based approaches vary in their applicability and effectiveness across different languages, particularly considering the complexities and resources available for each.

Stochastic Methods, such as Statistical Machine Translation (SMT) and Neural Machine Translation (NMT), rely heavily on large amounts of data and statistical models. They are generally effective for languages with:
- *Abundant Data:* Languages like English, Chinese, Spanish, and French have extensive parallel

corpora and resources, making them suitable for training robust statistical models.
- *Complex Syntax and Semantics:* Stochastic methods excel in capturing nuanced relationships between words and phrases, which is beneficial for languages with complex grammatical structures.

However, stochastic methods may face challenges with:
- Low-Resource Languages: Languages with limited digital resources, such as parallel corpora and linguistic annotations, pose difficulties for training effective statistical models.
- Idiosyncratic Language Features: Languages with irregularities or unique syntactic structures may not fit well into statistical frameworks, leading to lower translation accuracy.

Rule-based approaches leverage linguistic rules and expert knowledge to generate translations. They are particularly effective for structured languages as languages with clear grammatical rules and predictable syntactic structures benefit from rule-based approaches. For instance, languages like German or Arabic, which have well-defined grammatical genders and case systems, can be handled effectively with rule-based systems. The Domain-Specific Translation or the rule based methods allow for precise customisation based on specific domains (e.g., legal, medical), where terminology and syntax follow strict guidelines.

However, rule-based methods may struggle with:
- *Ambiguity and Context:* Languages with extensive homonymy (multiple meanings for the same word) or complex contextual dependencies may challenge rule-based systems, which often rely on explicit rules that may not capture all contextual nuances.
- *Maintenance and Adaptability:* Developing and maintaining comprehensive rule sets can be labor-intensive and may require continuous updates as languages evolve or new linguistic phenomena are discovered.

While stochastic and rule-based methods each have their strengths and challenges, their suitability varies significantly depending on the linguistic characteristics and resources available for a particular language. Advances in AI and machine learning continue to shape the landscape of machine translation, aiming to address the complexities inherent in diverse language pairs and contexts.

Rules are most often made "by hand", by the intuition of the person creating them, based on the knowledge of grammar and the intuition of the speaker, who may have relevant knowledge of the language. However, there is usually a development phase of rule-based systems during which the effect of the selected rules is tested and adjustments are made. In such cases, it is possible to change the order in which the rules are applied, remove some rules, add other words, using the trial and error method. This allows the system to find the best translation options.

The rules and their order of application may be determined automatically by learning or computing from the corpus to determine the optimal rules and the sequence of their application. However, after that, further calculations are no longer related to the application of the rules.[43] This method is known as the "transformation method" and can be seen as a compromise between stochastic and rule-based methods.

Thus, methods based on regular expressions or finite state machines allow for a broader and more complex understanding of the context that is taken into account in the action of the rules. Many studies aim to take into account the syntactic analysis of utterances, but not to perform it completely.[32] To avoid complications when trying to analyse utterances with incorrect grammar or, conversely, to avoid being overwhelmed by a large number of options, it is better to limit yourself to obtaining fragments of the analysis. This is the main goal of fragment parsing. In this context, Shakun N.[40] argues that linguistic AI programs are not designed to produce a linguistic description but to produce a result. Grammar is seen as a programming language for recognisers, with the aim of writing patterns that are reliable indicators of bits of syntactic structure, even if these bits of the structure are "boundaries" or "kernels" rather than traditional word combinations.

The dilemma faced by developers of language processing systems is to choose between models that better match the representation of language but cannot provide implementations that follow acceptable practices, or to prioritise linguistic expressiveness. This dilemma is clearly articulated by those that propose a compromise solution in the form of building an approximate model that provides efficiency but preserves linguistic expressiveness. Systems rely on finite state language models. However, these models are inadequate for linguistic interpretation, as they cannot express the relevant syntactic and semantic patterns. Advanced Phrase Structure Formalisms (APSGs), such as unification grammars, can express many of these patterns, but they are less suitable for modelling language because of the intrinsic cost of state transitions in APSG parsers.

Recruitment
Stochastic methods are probabilistic methods based on corpus-based statistics. We first describe the general methodology borrowed from Bednarek & Carr, which we then illustrate through a precise explanation of two

treatment examples, and we conclude with some remarks on this type of method.

General methodology

Based on the literature analysis, we can describe a general methodology that is applicable to all types of automatic translation tasks. This methodology consists of three stages (table 2):

Table 2. General methodology for building an AI translation system architecture		
Stage	Procedure	Result
Determining the phase	After the task is defined, the phase in which the problem is modelled, identifying the probabilities of certain events, creates the need to make "simplified assumptions"	It is assumed that the appearance of a word does not depend only on the two previous words, and the correct choice of hypotheses is a key factor in the quality of the resulting model
Building training data	The annotated corpora are used to estimate the values of elementary probabilities determined at the modelling stage	When new data is received, the elementary probabilities calculated in the previous step are applied to it, which allows solving the problem
Calibration	The body of the natural language processing tool is used for calibration	The housing is central to the development of this tool

However, researchers highlight the difficulties in implementing these methods: the cost of the programs, in terms of execution time and memory space, which require overly simplified models, the lack of certain data that would be useful - we do not always have an adequate corpus to solve the problem we want to deal with - and the very imperfect estimation of low probabilities.[2]

One of the key issues in the machine translation process is spell-checking. This method of spell-checking focuses on individual words, regardless of context. A lexicon, or dictionary, contains correctly spelled words, which allows detecting "nonwords" in a text.[51] Typically, an error can be caused by the insertion, deletion, substitution, or transposition of a single letter. Errors (insertion parasites) involving more than two letters are not counted.

- hat → hot
- potion → passion
- kidney → kindly
- horses → horsnes

Another example concerns the detection of the nonword acress in an English text. According to the above hypothesis, this nonword can be formed from six correctly spelt words in seven different ways:

- cress → acress (inserting parasite A in the first position)
- actress → acress (delete t)
- caress → acress (inversion of letters C and A)
- access → acress (replacing the second occurrence of C with R)
- across → acress (replace o with e)
- acres → acress (insertion of the parasite s in the penultimate position)
- acres → acress (insertion of a parasite in the same position)
- acres → acress(insert parasite s in the last position)
- acres → acress (inserting the parasite s in the last position)

This is one of two ways to get the word "acress", which no longer exists in English and is just an example of letter substitution to study the impact of errors on text comprehension.

Statistics and Data Analysis

To evaluate these six options, they are assigned scores based on spelling errors in the corpus. However, automatic processing of these errors would be difficult due to the statistical significance of certain errors, such as the replacement of "ac" with "ca". To simplify this process, letter replacements and insertions can be treated separately, taking into account only one letter of the context before the replacement. This makes the calculations simpler and avoids a large amount of training data. In any case, this can only be processed manually, which once again confirms the vital position of humans in machine translation and the inability to abandon traditional methods.

Most importantly, however, it should be noted that the method mixes all possible causes of calculation errors, such as the proximity of keys on the keyboard, phonological and spelling uncertainties, etc.

Ancillary Analyses

You can predict the next word in a sequence of words to determine the probability of its occurrence in a certain context using the n-gram method. For example, the word "eat" can be most often followed by the words "on", "some", "lunch", "a", "Indian", "Thai", or "British". This reflects the linguistic peculiarities of the language, not the actual information about the availability of restaurants with Indian or British cuisine in a particular region.

Participant Flow

The impact of technological change on traditional translation approaches can be demonstrated by the ability of AI to generate probabilistic approaches based on statistics. They face two opposing requirements. These methods attempt to determine the probability of an event E occurring based on a number of observed values i1, i2...p p observed in reality, to determine the probability of event E. To do this, in a set of authenticated and recorded data - a corpus in the case of language data processing - we look for situations in which the indices p appears, and among these situations, we count those in which the event E actually occurred. At the same time, the more situations where i1, i2...p are confirmed, the more significant the statistical result. However, as the number of indices has increased, so has the number of situations where all of these indices have been tested at a low level.[16]

That is why it is necessary to limit the number of indices that are considered: for example, n-grams that become 1-grams, the context of a spelling error that is reduced to a maximum of one character, etc. Thus, the degree of similarity between the situation under study and the situations considered in the probability calculation necessarily decreases.

Intervention or Manipulation of Fidelity

The models that are defined in this context are approximations: it is decided to consider certain aspects of technological realities and to take into account human performance. The question then is whether there are linguistic or cognitive assumptions underlying the choice of artificial intelligence. It can be said that such hypotheses do exist, but that they are very general, very imprecise: for example, in this context,[12] emphasise that it is possible that an AI spelling error may be related to a minimum number of letters but not to the machine's cognitive processes.

In this paper, we agree with this opinion of scientists. After all, in any case, spelling mistakes in machine translation are most likely related to the position of words or letters in the corpora. Also, the numerical results often reflect certain properties of languages. For example, there is a high probability that a transitive verb may be followed by a determiner, which can be interpreted as taking into account rewriting rules:

- SV → V [trans] SN
- SN → Det (Adj) N

But most likely, AI models ignore general linguistic rules, the system considers words rather than syllables. Linguistic theory is based on a hypothesis that considers linguistic works as linear sequences of words. In this sense, even if models exist, they fall short of traditional and long-standing ideas about language syntax and translation.

In line with this, these AI translation applications do not focus on separating the level of an autonomous syntactic system but rather combine different linguistic levels and cognitive properties to provide the best service to users.

Baseline Data

Natural language processing (NLP) work, including stochastic NLP work, is quite satisfactory, but their translations do not always accurately convey the meaning of the information. Some indications show that it is important not to separate NLP research from linguistics, as this may lead to a halt in TAL work or even to a foundation on stochastic NLP research in linguistics. Therefore, it is important to address the question of the place of NLP in the context of linguistics.

An analysis of the scientific literature has shown that this topic is a subject of much debate. The issue of terminology remains important: from "probabilistic processing"[42] to "corpus linguistics"[13] and then to "probabilistic natural language modelling",[48] although the essence of the process remains unchanged.

According to a study by Mishra & Kumar,[29] there are two main techniques in the field of AI translation and automatic language processing: those based on more or less advanced parsing and statistical and numerical methods that detect associations in corpora. Therefore, it is important to use a combination of statistical and linguistic approaches to achieve optimal results. According to[38] automatic terminology building requires the

use of both statistical and linguistic methods, which can cooperate or conflict with each other. To achieve optimal performance, it is necessary to integrate rather detailed human language skills, which remains a relevant and open topic for future research.

According to the analysis, scholars distinguish between the goals of artificial intelligence in translation and the goals of linguistics.[37] Some give corpus linguistics a dual and partially contradictory characterisation. However, there are those who believe that NLP is related to applied linguistics.[45] In fact, scientists do not clearly distinguish between NLP and corpus linguistics, which causes confusion. Nevertheless, the study results show that regardless of the paradigm of using AI or corpora, all processes involve automated language processing.

The emergence of AI introduces a new concept among linguists - probabilistic natural language "modelling".[23] However, scientists still refer to the concept of "language models", referring to the fact that these models play a key role in the functioning of natural language processing systems that face real-world problems.[47] The challenges mentioned include speech recognition, machine translation, and information retrieval. Wallis S.[49] adds that it is important that the training of the models used is automated. These tasks include speech recognition, machine translation, and information retrieval. It is necessary that the training of the models is automated, and although the models are not models for representing language knowledge, but rather models for automatic processing, they still require human intervention.

The use of n-grams is considered obvious, but the question arises as to the probability of unobserved n-grams. In the context of the study[18] point out that unobserved n-grams can indeed improve machine language knowledge. Nowadays, digital approaches such as stochastic models or probabilities in grammars are generally accepted.[9]

According to,[15] automatic language processing specialists, based on their practical achievements, are eager to explore linguistics. They believe that the study of linguistics always opens up two main areas of linguistic phenomena: observation of texts, which makes us look at them, and introspection, which allows us to draw conclusions about language production. The researchers believe that all possible phenomena can be found in existing texts, which helps us to better understand the language we hear every day.

Today, computer language processing is not yet able to accurately capture human language abilities,[27] but some computational linguists are exploring the computational properties of language to remedy this.[22] One of the interesting and challenging questions about translation as a linguistic act in the context of artificial intelligence is how technological changes affect traditional approaches and how humans can address these issues using standard approaches to unconstrained text processing. Statistical methods are considered the most promising at the moment, so many scientists are actively interested in them in the context of the cognitive capabilities of computer linguistics.

So, we can draw several conclusions based on the literature analysis. First of all, there are two main methods of building an architecture for translation: stochastic and rule-based methods. There is also a third type of method, based on finite automata or expressions, which can be seen as an intermediate between stochastic and theoretical methods of natural language processing.

Stochastic methods are used to label texts and resolve ambiguities in the categorisation of words. The rules that help with this are called contextual rules, which are based on the immediate context.

In addition, it is important to consider various factors that affect automatic language processing, such as spelling errors, which can be detected and corrected using special methods. Statistical and data analysis, as well as other additional studies, can help improve the accuracy and efficiency of translation systems.

Ultimately, to achieve optimal results in AI translation, it is important to balance the use of statistical and linguistic methods and combine them to achieve better performance. Such an integrated approach can help improve translation results and ensure better text understanding.

Thus, these studies confirm the importance of automated language processing in the field of artificial intelligence and linguistics. Probabilistic modelling of natural language is becoming an increasingly important aspect of this research. Scientists seek to combine computer language processing with linguistic principles to achieve better results in machine translation and other tasks. Despite the complexity of the tasks, statistical methods are considered a promising area of research that can solve a number of problems in this field.

CONCLUSION

In conclusion, the combination of statistical and rule-based methods in automated language processing is essential for improving the accuracy and efficiency of translation systems. By integrating these approaches, researchers can achieve better results in machine translation and text understanding. Additionally, automated language processing plays a crucial role in the advancement of artificial intelligence and linguistics, as evidenced by the growing interest in probabilistic modelling of natural language. Further research in this area will continue to push the boundaries of machine translation and other language processing tasks.

REFERENCES

1. Al Ismail YA. The Evolution of Empirical Research in Translation Studies: From Cognitive Insights to AI-Enhanced Horizons. Int J Linguist Lit Transl. 2023;6(12):61-5. Available from: https://doi.org/10.32996/ijllt.2023.6.12.8

2. Amaar A, Aljedaani W, Rustam F, Ullah S, Rupapara V, Ludi S. Detection of fake job postings by utilizing machine learning and natural language processing approaches. Neural Process Lett. 2022;54(3):2219-47. Available from: https://doi.org/10.1007/s11063-021-10727-z

3. Arzhevitin S, Bortnikov G, Bublyk Y, Lyubich O. Impact of martial state on the performance of the Ukrainian banking sector. Financ Credit Act Probl Theory Pract. 2023;1(48):23-41. Available from: https://doi.org/10.55643/fcaptp.1.48.2023.3966

4. Bednarek M, Carr G. Computer-assisted digital text analysis for journalism and communications research: introducing corpus linguistic techniques that do not require programming. Media Int Aust. 2021;181(1):131-51. Available from: https://doi.org/10.1177/1329878x20947124

5. Benmansour M, Hdouch Y. The Role of the Latest Technologies in the Translation Industry. Emirati J Educ Lit. 2023;1(2):31-6. Available from: https://www.emiratesscholar.com/system/publish/061223101243907.pdf

6. Bohatyrets V. AI and Machine Translation Post-editing: Advances and Challenges (Insights for Students of International Studies). Media Forum: Analytics, Forecasts, Inf Manage. 2023;13:195-209. Available from: https://doi.org/10.31861/mediaforum.2023.13.198-209

7. Bushman I. Education in the 21st century: philosophical foundations and principles. Future Philos. 2022;1(2):4-15. Available from: https://doi.org/10.57125/FP.2022.06.30.01

8. Cherniaieva O, Orlenko O, Ashcheulova O. The infrastructure of the Internet services market of the future: analysis of formation problems. Future Econ Law. 2023;3(1):4-16. Available from: https://doi.org/10.57125/FEL.2023.03.25.01

9. De Sutter G, Lefer MA. On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. Perspect. 2020;28(1):1-23. Available from: https://doi.org/10.1080/0907676X.2019.1611891

10. Ding J. Corpus-based Translation Studies: Examining Media Language through a Linguistic Lens. In: SHS Web of Conferences. Vol. 185. EDP Sciences; 2024. p. 01012. Available from: https://doi.org/10.1051/shsconf/202418501012

11. Fan K, Chunlei W. Translation Studies in the Era of AI: Characteristics, Fields and Significance. Int J Transl Interpr Stud. 2023;3(4):58-67. Available from: https://doi.org/10.32996/ijtis.2023.3.4.7x

12. Fatima R, Samad Shaikh N, Riaz A, Ahmad S, El-Affendi MA, Alyamani KAZ, Nabeel M, Ali Khan J, Yasin A, Latif RMA. A Natural Language Processing (NLP) evaluation on COVID-19 rumour dataset using Deep Learning techniques. Comput Intell Neurosci. 2022;2022:1-17. Available from: https://doi.org/10.1155/2022/6561622

13. Feder A, Keith KA, Manzoor E, Pryzant R, Sridhar D, Wood-Doughty Z, et al. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. Trans Assoc Comput Linguist. 2022;10:1138-58. Available from: https://doi.org/10.1162/tacl_a_00511

14. Grego K. From the cognitive turn to AI: Reflections on recent trends in translation (studies). In: Proceedings of the Scenari Multimediali e Didattica della Traduzione; 2023 Dec 14-16; Milan. Retrieved from https://air.unimi.it/handle/2434/1020932

15. Gries ST, Jansegers M, Miglio VG. Quantitative methods for corpus-based contrastive linguistics. In: New Approaches to Contrastive Linguistics. Berlin: De Gruyter; 2020. p. 53-84. Available from: https://doi.org/10.1515/9783110682588-003

16. Harrington K, Ronan P. Demystifying corpus linguistics for English language teaching. In: Demystifying

Corpus Linguistics for English Language Teaching. Cham: Springer International Publishing; 2023. p. 1-17. Available from: https://link.springer.com/chapter/10.1007/978-3-031-11220-1_1

17.   Huda O. Use of the Moodle Platform in Higher Education Institutions During Training Masters: Experience Under Martial Law. ELIJ [Internet]. 2023 Jun 25 [cited 2024 Jun 27];1(2):4-20. Available from: https://www.el-journal.org/index.php/journal/article/view/2

18.   Hundt M, Rautionaho P, Strobl C. Progressive or simple? A corpus-based study of aspect in World Englishes. Corpora. 2020;15(1):77–106. Available from: https://doi.org/10.3366/cor.2020.0186

19.   Iskakova M. Electronic Technologies to Ensure Individual Learning of Education Seekers with Special Needs. Future Soc Sci. 2023;1(1):4-20. Available from: https://doi.org/10.57125/FS.2023.03.20.01

20.   Jalilbayli OB. Philosophy of linguistic culture and new perspectives in modern Azerbaijani linguistics. Future Philos. 2022;1(4):53-65. Available from: https://doi.org/10.57125/FP.2022.12.30.05

21.   Jiang K, Lu X. Integrating machine translation with human translation in the age of artificial intelligence: Challenges and opportunities. In: Advances in Intelligent Systems and Computing. Springer Singapore; 2021. p. 1397–1405. Available from: https://link.springer.com/chapter/10.1007/978-981-33-4572-0_202

22.   Kang Y, Cai Z, Tan C-W, Huang Q, Liu H. Natural language processing (NLP) in management research: A literature review. J Manag Analytics. 2020;7(2):139–172. Available from: https://doi.org/10.1080/23270012.2020.1756939

23.   Khan AI, Al-Badi A. Open source machine learning frameworks for industrial internet of things. Procedia Computer Science. 2020;170:571–577. Available from: https://doi.org/10.1016/j.procs.2020.03.127

24.   Khasawneh MAS, Al-Amrat MGR. Evaluating the Role of Artificial Intelligence in Advancing Translation Studies: Insights from Experts. Migration Lett. 2023;20(S2):932-943. Available from: https://migrationletters.com/index.php/ml/article/view/3745/2734

25.   Khasawneh MAS. The Potential of Ai in Facilitating Cross-Cultural Communication Through Translation. J Namib Stud: Hist Politics Culture. 2023;37:107-130. Available from: https://namibian-studies.com/index.php/JNS/article/view/4654

26.   Koka NA, Akan MF, Kana'n BHI, Khan MR, Zulfiquar F, Jan N. Impact of artificial intelligence (ai) on translation quality: assessment and evaluation. J Southwest Jiaotong Univ. 2023;58(4):907-919. Available from: http://jsju.org/index.php/journal/article/view/1787

27.   Mandera P, Keuleers E, Brysbaert M. How useful are corpus-based methods for extrapolating psycholinguistic variables? Q J Exp Psychol. 2015;68(8):1623-42. Available from: https://doi.org/10.1080/17470218.2014.988735

28.   Maraieva U. On the formation of a new information worldview of the future (literature review). Future Philos. 2022;1(1):18-29. Available from: https://doi.org/10.57125/FP.2022.03.30.02

29.   Mishra BK, Kumar R, editors. Natural Language Processing in Artificial Intelligence. 1st ed. Apple Academic Press; 2020. Available from: https://doi.org/10.1201/9780367808495

30.   Mohamed YA, Khanan A, Bashir M, Mohamed AHHM, Adiel MAE, Elsadig MA. The impact of artificial intelligence on language translation: A review. IEEE Access: Practical Innovations, Open Solutions. 2024;12:25553-79. Available from: https://doi.org/10.1109/access.2024.3366802

31.   Munawaroh S, Hastami Y, Suwandono A, Probandari AN, Hartono, Wiyono N, Ghozali DA, Herawati F, Afifah UM, Hanifah AANN. Reading Holy Quran to Improve Episodic Memory in Elderly. Future Med. 2023;2(3):4-11. Available from: https://doi.org/10.57125/FEM.2023.09.30.01

32.   Nikolenko K. Artificial Intelligence and Society: Pros and Cons of the Present, Future Prospects. Future Philos. 2022;1(2):54-67. Available from: https://doi.org/10.57125/FP.2022.06.30.05

33.   Petchenko M, Fomina T, Balazyuk O, Smirnova N, Lugova O. Analysis of Trends in the Introduction of Digitalization in Accounting: The Ukrainian Case. Financ Credit Act Probl Theory Pract. 2023;1(48):105–113. Available from: https://doi.org/10.55643/fcaptp.1.48.2023.3951

34.   Poplavskyi M, Rybinska Y, Ponochovna-Rysak T. The specificity of synesthesia in contemporary American and English poetry and its impact on the reader. Cogito. 2020;12(3):297-315. Available from: https://www.ceeol.com/search/article-detail?id=1034449

35.   Rakhimov T. Research on moral issues related to the use of artificial intelligence in modern society. Future Philos. 2023;2(2):30-43. Available from: https://doi.org/10.57125/FP.2023.06.30.03

36.   Ramakrishnan R. CSR and Sustainable Development interrelations. Law Bus Sustain Herald. 2022;2(1):40-48. Available from: https://lbsherald.org/index.php/journal/article/view/33

37.   Salloum S, Gaber T, Vadera S, Shaalan K. A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access. 2022;10:65703-27. Available from: https://ieeexplore.ieee.org/abstract/document/9795286.

38.   Salloum S, Gaber T, Vadera S, Shaalan K. Phishing email detection using natural language processing techniques: a literature survey. Procedia Computer Science. 2021;189:19-28. Available from: https://doi.org/10.1016/j.procs.2021.05.077

39. Schmitt PA. Translation 4.0-evolution, revolution, innovation or disruption? Lebende Sprachen. 2019;64(2):193-229. Available from: https://doi.org/10.1515/les-2019-0013

40.   Shakun N. Anthropological dilemmas of information society development modern stage in the context of globalisation challenges. Future Philos. 2022;1(3):52-63. Available from: https://doi.org/10.57125/FP.2022.09.30.04

41.   Shrivastava R, Jain M, Vishwakarma SK, Bhagyalakshmi L, Tiwari R. Cross-cultural translation studies in the context of artificial intelligence: Challenges and strategies. In: Advances in Cognitive Science and Communications. Springer Nature Singapore; 2023. p. 91–98. Available from: https://link.springer.com/chapter/10.1007/978-981-19-8086-2_9

42.   Smith N, Hoffmann S, Rayson P. Corpus tools and methods, today and tomorrow: Incorporating linguists' manual annotations. Literary Linguistic Computing. 2007;23(2):163–180. Available from: https://doi.org/10.1093/llc/fqn004

43. Sofilkanych M. The formation of a new information culture of the future: the socio-philosophical content. Future Philos. 2022;1(1):56-67. Available from: https://doi.org/10.57125/FP.2022.03.30.05

44.   Soysal F. Çeviribiliminin Yapay Zeka (YZ) ile Geliştirilmesi: Zorluklar, İmkânlar ve Öneriler. Karamanoğlu Mehmetbey Üniv Uluslararası Filoloji ve Çeviribilim Derg. 2023;5(2):177–191. Available from: https://doi.org/10.55036/ufced.1402649

45.   Tan S, Joty S, Baxter K, Taeihagh A, Bennett GA, Kan MY. Reliability testing for natural language processing systems. arXiv preprint. 2021. Available from: https://doi.org/10.48550/arXiv.2105.02590

46.   Tarasova V, Romanchuk S, Kapitan T, Demeshko I, Leleka T. New Slang Expressions-Neologisms to Denote the Phenomena of War: A Translation Aspect of the Neglect. World J English Lang. 2023;13(8):558. Available from: https://www.sciedupress.com/journal/index.php/wjel/issue/view/1223

47.   Tejedor-García C, Escudero-Mancebo D, Cámara-Arenas E, González-Ferreras C, Cardeñoso-Payo V. Assessing pronunciation improvement in English learners using a controlled computer-assisted pronunciation tool. IEEE Trans Learn Technol. 2020;13(2):269-282. Available from: https://doi.org/10.1109/TLT.2020.2980261

48.   Verma S, Paul A, Kariyannavar SS, Katarya R. Understanding the applications of natural language processing on COVID-19 data. In: 2020 4th International Conference on Electronics, Communication and

Aerospace Technology (ICECA); 2020 Nov; pp. 1157-62. IEEE. Available from: https://ieeexplore.ieee.org/abstract/document/9297490.

49.   Wallis S. Statistics in corpus linguistics research: A new approach. Routledge; 2020. Available from: https://doi.org/10.4324/9780429491696

50.   Wang L. The Impacts and Challenges of Artificial Intelligence Translation Tool on Translation Professionals. SHS Web of Conferences. 2023;163:02021. Available from: https://doi.org/10.1051/shsconf/202316302021

51.   Wu Y. Discussion of the Use of Concordance Programmes in the EFL Classroom. J Linguist Commun Stud. 2023;2(4):79-84. Available from: https://www.pioneerpublisher.com/JLCS/article/view/523.

FINANCING

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS
Conceptualization: Nataliia Yuhan and Yuliia Herasymenko.
Methodology: Oleksandra Deichakivska.
Software: Yevhen Kozlov.
Validation: Nataliia Yuhan, Yuliia Herasymenko, and Oleksandra Deichakivska.
Formal analysis: Anzhelika Solodka.
Supervision: Nataliia Yuhan.
Project administration: Yuliia Herasymenko.
Investigation: Yevhen Kozlov.
Resources: Oleksandra Deichakivska.
Data curation: Anzhelika Solodka.
Writing—original draft: Nataliia Yuhan.
Writing—review and editing: Yuliia Herasymenko.