International Journal of Language and Translation Studies, 5/2 (2025), 281-307

İnsan Çevirmen ve Makine Çevirisi. Türkçe-İtalyanca Sözlükbiliminde Bir Vaka Çalışması *

Ellen PATAT [1]

Geliş Tarihi: 20.06.2025
Kabul Tarihi: 01.12.2025
Yayım Tarihi: 30.12.2025
Değerlendirme: İki Dış Hakem / Çift Taraflı Körleme
Makale Türü: Araştırma Makalesi

Atıf Bilgisi:

Patat, Ellen (2025). İnsan Çevirmen ve Makine Çevirisi. Türkçe-İtalyanca Sözlükbiliminde Bir Vaka Çalışması. *International Journal of Language and Translation Studies,* 5/2, 281-307.

Benzerlik Taraması: Yapıldı – iThenticate

Etik Bildirim: lotusjournal@selcuk.edu.tr

Çıkar Çatışması: Çıkar çatışması beyan edilmemiştir.

Finansman: Bu araştırmayı desteklemek için dış fon kullanılmamıştır.

Öz

Bu makale, Türkçe-İtalyanca sözlükçülük alanında insan çevirisi ile otomatik çeviri çıktıları arasında karşılaştırmalı bir analiz sunmaktadır. Gelişmekte olan çok dilli çevrimiçi sözlük Turklex'ten alınan örnek cümlelerden oluşturulan bir derlem kullanılarak, çeviriler profesyonel bir çevirmen tarafından manuel olarak yapılmış, ardından dört farklı makine çeviri aracı – Google Translate, DeepL, Gemini ve ChatGPT-4 – aracılığıyla otomatik olarak üretilmiştir. Sözlükçülükte doğruluğun kritik bir unsuru olan madde başlıklarının aktarımına özel önem verilmiştir. Çeviriler arasındaki farklılıklar ve örtüşmeler incelenerek, çalışma güncel NMT (Sinirsel Makine Çevirisi) sistemleri ile yapay zeka destekli sohbet robotlarının, daha az incelenmiş dil çiftlerine uygulandığında gösterdiği güçlü ve zayıf yönleri değerlendirmektedir. Bildiğimiz kadarıyla, bu çalışma, sözlükçülük bağlamında Türkçe-İtalyanca çeviri üzerine yapılmış ilk odaklı araştırmadır. Elde edilen bulgular, hem makine çeviri araçlarının değerlendirilmesine ışık tutmayı hem de bu araçların dil kaynakları geliştirmeye entegrasyonu üzerine geniş çaplı tartışmalara katkı sağlamayı amaçlamaktadır. Araştırma, ChatGPT'nin diğer araçlardan kıyasla bir miktar daha iyi performans sergilemesine rağmen, yapay zeka destekli sohbet robotları ile NMT sistemlerinin bağlamsal olarak incelikli ve anlamsal açıdan doğru çeviriler üretme konusunda hâlen insan çevirmenlerin gerisinde kaldığını ortaya koymuştur.

Anahtar Kelimeler: Makine Çevirisi, Türkçe-İtalyanca, Sözlükbilim, Çevrimiçi Türkçe-İtalyanca Sözlük, Sinirsel Makine Çevirisi (NMT), Yapay Zeka Destekli Chatbot

* Etik Beyan: *Bu çalışmanın hazırlanma sürecinde bilimsel ve etik ilkelere uyulduğu ve yararlanılan tüm çalışmaların kaynakçada belirtildiği beyan olunur.

[1] Öğr. Gör. Dr., İstanbul Üniversitesi, Edebiyat Fakültesi, İtalyan Dili ve Edebiyat, ellen.patat@istanbul.edu.tr.

Human Translator vs Machine Translation. A Case Study in Turkish-Italian Lexicography *

Ellen PATAT [2]

Date of Submission: 20.06.2025
Date of Acceptance: 01.12.2025
Date of Publication: 30.12.2025
Review: Double-blind peer review
Article Type: Research Article

Citation:

Patat, Ellen (2025). Human Translator vs Machine Translation. A Case Study in Turkish-Italian Lexicography. *International Journal of Language and Translation Studies*, 5/2, 281-307.

Plagiarism Check: Yes - iThenticate

Complaints: lotusjournal@selcuk.edu.tr

Conflict of Interest: The author(s) has no conflict of interest to declare.

Grant Support: The author(s) acknowledges that they received no external funding to support this research.

Abstract

This paper presents a comparative analysis between human translation and automated outputs within the field of Turkish-Italian lexicography. Using a corpus of example sentences extracted from *Turklex*—a developing online multilingual dictionary—translations were carried out manually by a professional translator and subsequently generated by four different machine translation tools: Google Translate, DeepL, Gemini, and ChatGPT-4. Particular attention is paid to the rendering of headwords, a critical element in lexicographic accuracy. By examining the divergences and overlaps across these translations, the study highlights the strengths and limitations of current NMTs and AI-driven chatbots when applied to a less commonly examined language pair. To the best of our knowledge, this is the first focused investigation into Turkish-Italian translation within a lexicographical framework. The findings aim to inform both the evaluation of machine translation tools and the broader discussion on their integration into language resource development. The research found that while ChatGPT slightly outperformed the other tools, AI chatbots and NMTs still fall short of human translators in delivering translations that are both contextually nuanced and semantically accurate.

Keywords:
Machine translation, Turkish-Italian, Lexicography, Online TR-IT dictionary, NMT, AI-driven Chatbot

* Ethical Statement: * It is declared that scientific and ethical principles have been followed while carrying out and writing this study and that all the sources used have been properly cited.

[2] Instructor, PhD. Istanbul University, Faculty of Letters, Department of Italian Language and Literature, ellen.patat@istanbul.edu.tr.

Introduction

At Asialex 2023, Rundell reflected on current practices in dictionary compilation and shared his preliminary observations regarding the use of ChatGPT for lexicographic tasks. Among the exploratory questions posed, one stands out: "Can ChatGPT generate good dictionaries with minimal human input (thereby making lexicographers redundant)?" (Rundell, 2023, p. 15). His response is unequivocal: "This time, the answer is a straightforward 'no'. […] ChatGPT can produce plausible-looking dictionary text, at least for headwords at the simpler end of the spectrum. But closer examination almost always reveals problems, whether of omission, invention, or inauthenticity" (Rundell, 2023, p. 16). However, this position has been challenged by scholars who argue that ChatGPT already performs remarkably well in several phases of the dictionary-making process, often delivering outputs that require little or no further revision (de Schryver, 2023, p. 8). Rundell's question serves as a starting point for the present research from the perspective of the translator, who plays a multifaceted role in dictionary making, particularly in the compilation of bilingual dictionaries. Translators contribute by identifying accurate equivalents between source language (SL) and target language (TL), taking into account not only direct meanings but also nuances, idiomatic usage, and cultural context. They help ensure that headwords, definitions, and usage examples reflect authentic language use. Their linguistic expertise enriches both the lexicographical accuracy and usability of the dictionary.

The role of translators has become increasingly interconnected with the development of Large Language Models (LLMs) for machine translation (MT)—a rapidly evolving field, as researchers and developers continuously seek to enhance translation quality, fluency, and adaptability. The integration of Artificial Intelligence (AI) into the translation process has transformed the discipline, prompting both experts and educators to reassess traditional approaches and explore innovative methodologies. As LLMs are inherently more focused on language processing than any previous technologies, it is unsurprising that they have generated exceptionally high expectations within the lexicographic community (de Schryver, 2023, p. 4). The advent of Neural Machine Translation (NMT) has marked a significant revolution. Neural networks demonstrate exceptional performance in tasks related to natural language processing and other applications, owing to their capacity to model complex, non-linear relationships within data. AI-driven tools leverage advanced machine learning and natural language processing techniques to produce rapid and contextually relevant translations. As a transformative technology, NMT models translate larger segments of text rather than isolated fragments, effectively incorporating contextual elements and preserving the nuanced meanings of the source text. Undoubtedly, the integration of AI into translation practices offers several advantages, including improved speed, cost-effectiveness, and consistency. While it is widely acknowledged that this AI-powered approach has enhanced some aspects of the translation process, AI translation tools could still face significant limitations, particularly in accurately conveying cultural nuances, idiomatic expressions, and context-sensitive content. Hence, the expertise of human translators and linguists seems to be still indispensable.

The rise of AI in translation has also sparked important ethical discussions, such as concerns over data privacy, transparency in algorithmic decision-making, and the potential for over-reliance on technology. The debate surrounding these issues emphasizes the need for a balanced approach that combines AI efficiency with human expertise, fostering a future where translation is both accurate and culturally resonant. As the landscape of translation continues to evolve, the collaboration between AI technologies and human translators is set to shape the future of the industry. The ongoing advancements in AI are likely to enhance translation quality and accessibility, while the indispensable insights provided by human translators will ensure that the complexities of language and culture are effectively navigated. These preliminary reflections lay the groundwork for this study's examination of the ongoing evolution of this collaboration, which has been attracting significant scholarly attention, as evidenced by the growing body of literature on the subject. Nevertheless, this study seeks to address the lack of research, to the best of our knowledge, within the Turkish-Italian context, with particular reference to the field of lexicography (see 2.2). It is considered important to verify the accuracy of the following claim: "If you need a high-quality translation between languages, look no further than ChatGPT, one of the most cutting-edge AI-powered language models available" (Khoshafah, 2023, p. 2).

Research Objectives

- To compare the choices made by AI-powered tools with the reference standard, namely the translator's rendering, of headwords in context.
- To compare the performances of AI language models: NMTs vs AI chatbots.

Research Questions

1. Do the translator's choices match the translations provided by AI language models?
2. Which AI language models displays a better performance within the given research context?

Hypotheses:

1- AI chatbots are closer to the reference standard.

2- AI chatbots are expected to perform better than NMTs.

The objective of this comparative analysis is to evaluate the relative performance of the AI-driven translation tools in relation to the human translator's output. The corpus for this preliminary study is drawn from TURKLEX, a new, work-in-progress online multilingual dictionary (see 2.2.1). The project for a Turkish (TR)-Italian (IT) / IT-TR dictionary is necessary considering that the lack of a printed or online learners' dictionary for Turkish as a foreign language represents a notable gap in educational resources (Gürlek, 2023, p. 629). An online TR-IT/IT-TR dictionary would not only aid language learning but also facilitate accurate translation by providing deeper insights into cultural contexts, enhancing the learning experience and bridging the gap for students and professionals alike. Therefore, this study contributes to the ongoing debate on the efficacy of current machine translation technologies.

1. Historical development: From human to machine translation.

The evolution of translation methodologies spans from manual approaches, relying on personal knowledge and bilingual dictionaries, to later significant advancements, such as standardized dictionaries, which made translations more systematic and accessible. The late 20^{th} and early 21^{st} centuries marked a transformative leap with artificial intelligence (AI), transitioning from previous models to neural machine translation (NMT) and generative models. These innovations enabled tools to better handle complexities like idiomatic expressions and cultural references. An AI translator is a cutting edge tool that employs artificial intelligence to perform bidirectional language processing, facilitating both speech-to-text and text-to-speech transformations across multiple languages. Utilizing deep learning architectures and large-scale multilingual corpora, it dynamically adapts to linguistic variability and improves its performance through continuous model optimization. By harnessing cutting-edge machine learning algorithms and sophisticated natural language processing (NLP) techniques, they deliver translations that are increasingly accurate and contextually nuanced.

Historically, the progression from rigid rule-based systems to the nuanced, context-aware capabilities of modern AI chatbots seems to have significantly improved translation accuracy, fluency, and accessibility. The earliest form of machine translation, Rule-Based Machine Translation (RBMT), relied on predefined linguistic rules, grammar, and bilingual dictionaries to convert text from one language to another. These systems operated strictly on syntactic and morphological patterns, requiring extensive human input to develop the rules for each language pair. RBTM struggled with idiomatic expressions, contextual meaning, and scalability for new language pairs.

Emerging in the 1990s, Statistical Machine Translation (SMT) shifted the focus to data-driven approaches. These systems, such as early versions of Google Translate, used large bilingual corpora to calculate the probabilistic likelihood of word or phrase alignments between languages. SMT modeled translations based on statistical probabilities, enabling systems to handle larger-scale translation tasks but produced fragmented sentences, struggling with long-range dependencies and idiomatic nuances.

An enhancement of SMT, Phrase-Based Statistical Machine Translation (PBSMT) focused on translating phrases rather than individual words, allowing systems to better capture contextual relationships within short text segments but still lacked a deep understanding of sentence-level context and cultural subtleties. To address the limitations of both RBMT and SMT, hybrid systems combined the two approaches. These systems used rule-based techniques to handle linguistic nuances while leveraging statistical models for broader data-driven translation.

Neural Machine Translation (NMT) marked a paradigm shift in translation technology. Using deep learning techniques, particularly sequence-to-sequence models and attention mechanisms, NMT processes entire sentences as contextual units. Google Translate or DeepL, for instance, significantly reduced the gap between human and machine translation quality, enhancing fluency and accuracy. The integration of translation technology into conversational AI systems has led to significant changes in practices, strategies, and methodologies within the field. AI-powered chatbots combine NMT with natural language processing (NLP) to facilitate real-time, multilingual communication in a conversational context. Open AI's ChatGPT or Meta's Seamless M4T provide context-aware translations that adapt to ongoing dialogue and are able to manage idiomatic expressions and cultural subtleties during interactions. Several other AI translation tools have gained popularity in recent years, each offering unique features, such as Microsoft Translator, iTranslate, Amazon Translate, or Taia.

	NMT	AI chatbot
Purpose	Designed to perform translations of text or speech from one language to another.	Designed to simulate human conversation and provide interactive responses to user queries.
Functionality	Focused on linguistic conversion, providing direct translations for text or spoken input.	Based on uses NLP to understand user input, generate responses, and engage in dialogue.
Context understanding	Provides translations without a conversational context or awareness of user intent	Context-aware, using dialogue history and user intent to generate meaningful and interactive responses.
Integration	Embedded into other tools or platforms for translation support	Integrated into customer service systems, websites, or messaging platforms to assist users in real-time conversations.

Table 1: Features of NMT vs AI Chatbot

To date, translation processes can be broadly categorized based on the degree of human involvement. Quite intuitively, Human Translation (HT) is carried out entirely by human translators, whereas MT, machine translation, describes the automation of the whole translation process. The latter refers to computerized systems responsible for the production of translations with or without human assistance (Xiu and Xeauyin, 2018, p. 17). It includes: Computer-Assisted Translation (CAT), where human translators make use of digital tools and software to support their work; and Human-Assisted or ‘humanaided’ Machine Translation (HAMT), in which machines produce the initial translation and human intervention is limited to reviewing or post-editing the output (Xiu and Xeauyin, 2018, p. 17; Son and Kim, 2023, p. 2).

In a steadily expanding body of literature, a wide range of study designs and theoretical frameworks have been employed to investigate NMT systems and AI-driven chatbots (Kembaren, Hasibuan, and Natasya, 2023; Sahari, Al-Kadi and Ali, 20023; Son and Kim, 2023; Yilmaz, Naumovska, and Aggarwal, 2023). Research has assessed Google Translate (GT) through experimental designs (Karnal & Pereira, 2015; Habeeb, 2020) and has increasingly focused on ChatGPT, evaluating its capabilities both in translation and in broader domains (Deng and Lin, 2022; Fuchs, 2023; Fan and Gong, 2023; Kocmi and Federmann, 2023; Liu et al., 2023; Li et al., 2024). In the specific context of translation, ChatGPT has shown performance comparable to commercial tools when translating high-resource European languages; however, its effectiveness declines notably with low-resource or linguistically distant languages (Jiao et al., 2023, p. 9; Liu et al., 2023, p. 15). Dedicated studies on ChatGPT’s role in translation are emerging (Siu, 2023), including comparative research involving human translators and AI across diverse language pairs—such as Arabic-English (Al Rousan, Jaradat, and Malkawi, 2025) and English with languages that use gender-neutral pronouns like Bengali, Farsi, Malay, Tagalog, Thai, and Turkish (Ghosh and Caliskan, 2023). In addition, recent scholarship has also addressed cross-platform comparisons in multilingual contexts. These concentrate on the translation of: an English-language AI questionnaire into 33 languages using Google Translate and GPT-3.5 (Kunst and Bierwiaczonek, 2023); Indonesian short stories through Google Translate and DeepL (Agung, Budiartha, and Suryani, 2024); patient education materials in Spanish and Chinese (Khoong, Steinbrook, and Brown, 2019). Similarly, in regards to Turkish, Çetin and Duran (2024) compared the outputs of human translators, Google MT, DeepL, and ChatGPT in the domains of education, healthcare, and law.

1.1 The Interdependence of Translation and Practical Lexicography: A Systematic Approach to Language Transfer

The relationship between translation and practical lexicography (or lexicographic practice)—i.e. the planning and compilation of concrete dictionaries (Bergenholtz and Gouws, 2012, p. 39)—is grounded in their mutual focus on language and meaning. In dictionary-making, the team is organized into several areas of expertise to ensure comprehensive and accurate results. Linguistic experts—beginning with lexicographers, who create the initial entries—followed by translators, who provide cross-linguistic equivalents, are responsible for compiling lexical data and ensuring terminological accuracy, particularly in specialized domains, often in consultation with subject-matter experts. Lexicography serves as a foundational resource for translation by providing precise definitions, contextual examples, and usage patterns of words. Translation, in turn, relies on lexicographical resources to ensure accurate equivalents, facilitating the transfer of meaning, connotation, and cultural context across languages.

In the practical domain, translators are tasked with translating headwords and the related content from a SL to a TL, ensuring that not only direct equivalents but also the subtleties and cultural nuances are captured. Consequently, translation informs lexicography by illustrating the real-world application of language, while lexicography aids translation by ensuring that the appropriate headword and its meaning are faithfully conveyed, preserving the integrity and richness of both languages involved. Example sentences (i.e. illustrative examples) in dictionaries enhance dictionary quality by providing context that reflects authentic language use. Chen (2016) highlights representativeness as a key evaluation criterion, while Han (2008) notes that such examples strengthen the explanatory role of headwords, making entries clearer and more practical (as cited in Tan, Long, and Bamigbade, 2023). Inter- or multidisciplinary research has shown that generative pre-trained transformers are capable of producing high-quality dictionary entries (de Schryver, 2023; Lew, 2023, 2024).

2. Study context

2.1. Machine translators

Advancements in artificial intelligence have profoundly transformed the field of machine translation, resulting in the development of highly sophisticated and widely accessible tools. In the present study, four prominent AI-driven translation systems—Google Translate, DeepL, Gemini, and ChatGPT—were selected as points of comparison, as each exemplifies a distinct yet complementary approach to automated language processing and multilingual communication. While built on distinct design principles and optimized for varying functionalities, these translation tools, collectively contribute to a more robust and versatile translation ecosystem. Google Translate, developed by Google, is a prominent example NMT system that employs a sequence-to-sequence model architecture, typically structured around an encoder-decoder framework augmented with attention mechanisms to ensure semantic fidelity and contextual accuracy in target-language generation. In parallel, Gemini, introduced by Google DeepMind in December 2023, represents a cutting-edge general-purpose multimodal model (GPMM) founded on a large-scale transformer architecture—an infrastructural backbone shared by numerous advanced models in NLP and translation domains. Another major player in the field, DeepL Translator, was released in 2017 by DeepL GmbH, a Cologne-based company also known for the Linguee lexical database; this NMT system integrates a user-centric interface and leverages extensive bilingual corpora and iterative user feedback to refine its translation performance. ChatGPT by OpenAI is a generative language model built on the GPT (Generative Pre-trained Transformer) architecture, primarily optimized for conversational engagement but also capable of performing translation tasks. Its dialogic functionality allows for interactive disambiguation and nuanced contextual interpretation, thereby enhancing its applicability in multilingual communication scenarios.

Google Translate and DeepL are both dedicated NMT systems focused primarily on providing quick and accurate translations across multiple languages, with Google Translate excelling in coverage and accessibility, and DeepL emphasizing nuance and stylistic quality. Gemini, as a multimodal model, extends translation capabilities beyond text to include inputs like images and audio, enabling cross-modal comprehension. Meanwhile, ChatGPT, though not a dedicated translation tool, complements these systems through its interactive capabilities, allowing users to clarify ambiguities, receive paraphrased interpretations, or explore linguistic context in dialogue form. Together, these tools cater to diverse user needs—from casual translation to professional linguistic tasks—making their functions different in focus but complementary in overall utility.

2.2 TR-IT/IT-TR dictionaries

Lexicography plays a critical role in the development of linguistic and cultural understanding between languages. In the context of IT-TR/TR-IT lexicography, the field has evolved significantly over the past century, driven by academic, diplomatic, and commercial exchanges between Italy and Turkey. Bilingual dictionaries serve not only as linguistic tools but also as bridges that facilitate intercultural communication, language acquisition, and translation practices. Historically, early Italian-Turkish dictionaries were often compiled by diplomats, missionaries, or language instructors, and were primarily focused on basic vocabulary and phraseological expressions intended for practical communication (see Rocchi's works). Over time, these resources became more sophisticated, incorporating idiomatic usage, morphological analysis, and contextual examples. In recent decades, there has been a growing emphasis on corpus-based lexicography, enabling more accurate representations of contemporary language use, register variation, and collocational patterns, which is nevertheless scarce in this specific language combination.

In the Italian tradition of linguistic and lexicographic studies, the interdependence between translation and lexicography has long been recognized as a cornerstone of language mediation. As De Mauro (1999) and Serianni (2005) emphasize, bilingual dictionaries not only mirror linguistic equivalences but also embody cultural and pragmatic asymmetries that the translator must negotiate. Dardano and Trifone (2003) further underline the descriptive and pedagogical functions of lexicography, highlighting the importance of contextual precision and semantic adequacy in cross-linguistic meaning transfer. Within this framework, bilingual lexicography emerges as a key arena where linguistic competence and interpretative awareness converge.

Italian scholarship has also explored the broader implications of bilingual lexicography from both a descriptive and didactic standpoint. Works by D'Achille (2010) and Lo Cascio (2003) provide insights into the structural and functional characteristics of contemporary Italian, offering a basis for understanding how lexicographic choices influence translation strategies. Marello's seminal studies (1989, 2015) and Rossi's (2005) reflections on equivalence and metalexicography reveal how bilingual dictionaries operate not as neutral repositories of words but as dynamic interpretative tools that encode cross-cultural meanings.

In more recent years, Bernardini and Zanettin (2004, 2020) have expanded this discussion by integrating corpus-based translation studies into the lexicographic domain, showing how computational tools and digital corpora can enhance both the consistency and explanatory potential of bilingual entries. In this perspective, bilingual lexicography is not merely a static record of equivalences but a dynamic interface where translation, data-driven analysis, and linguistic description intersect — an approach particularly relevant when assessing how neural machine translation systems perform across less-resourced language pairs such as Turkish–Italian.

Modern IT-TR (for instance, Demiryan, 2011) and TR-IT dictionaries vary widely in scope and quality. Digital lexicography has opened new possibilities for bilingual dictionary development. Online platforms and mobile applications now allow for real-time updates, integration of user feedback, and the inclusion of multimedia resources, enhancing the user experience and linguistic accuracy. However, the field still faces underrepresentation in terms of comprehensive, high-quality resources, particularly in the Turkish-Italian direction. Collaborative efforts between academic institutions and lexicographic publishers remain essential to enrich this area and support both language learning and cross-cultural communication.

The languages involved exhibit significant differences that influence translation strategies. Differences lie in grammatical aspects, including word order, morphology, gender and number agreement, tense and aspect, syntax, adjectives placement. Some studies focus on some types of differences, for instance, quantity expressions (Güneş, 2018); letters and sounds (Kaya, 2020); Turkish lexicon of Italian origin (Özkan, 2020; Manzelli, 2021) or Turkish-Ottoman elements in the Italian vocabulary (Rocchi, 2020). In comparing the syntax of Turkish and Italian, several fundamental differences emerge that are crucial for translation studies. As to morphosyntactic typology, Turkish, an agglutinative language, predominantly follows an SOV (Subject-Object-Verb) word order, with the verb typically at the end of the sentence, hence, influencing the processing of sentence structure and syntactic relationships. In contrast, Italian adheres to a SVO (Subject-Verb-Object) order, where the verb precedes the object. This structural divergence necessitates careful attention when translating, as it affects the syntactic flow and the placement of focus within the sentence. Additionally, Turkish employs case marking through suffixes to indicate grammatical roles, directly affixed to the noun, whereas Italian relies on prepositions to express similar grammatical relationships. The agreement system in Turkish is simpler, with verbs agreeing with the subject in person and number, without gender distinctions, while Italian verbs must agree in both person and number, and adjectives must concord in gender and number with the noun they modify. Furthermore, Turkish does not use definite or indefinite articles, instead relying on context or demonstratives to convey definiteness, whereas Italian obligatory articles (definite and indefinite) accompany nouns. When negating sentences, Turkish uses verb suffixes to form negation, whereas Italian places “non” before the verb. The structure of relative clauses also differs: Turkish relative clauses follow the noun they modify and are introduced by a relative pronoun with suffixes, while Italian relative clauses are introduced by pronouns like *che* (that, which), *cui* (whom, which), and *il quale* (the one who). These syntactic disparities highlight the complexities translators face when transferring meaning across these languages, requiring a nuanced understanding of word order, morphological features, and syntactic structures to ensure accurate and contextually appropriate translations.

2.2.1 TURKLEX: A new online dictionary

Turklex is a new, large-scale, general lexicographic project launched in 2023, coordinated by Assoc. Prof. Mehmet Gürlek from the Department of General Linguistics at Istanbul University. The project brings together a multidisciplinary team of linguists, lexicographers, and language experts. It forms part of a broader initiative to enhance digital language resources and support academic research, language education, and cultural preservation. The project’s structure includes one comprehensive monolingual dictionary and a set of seven bilingual dictionaries, each focusing on Turkish in relation to different target languages. By employing both traditional fieldwork methods and advanced corpus-linguistic tools, Turklex seeks to contribute significantly to the development of modern Turkish lexicography and to facilitate comparative linguistic studies with other languages, including Italian.

3. Methodology

3.1 Design, population, and data collection

The corpus for this preliminary study consists of 100 entries—source headwords (SHs), encompassing all word classes—along with their corresponding example sentences, extracted from the TURKLEX project (see 2.2.1). The objective was to compile a sample sufficiently large to yield meaningful insights while remaining manageable for detailed anlaysis. Initially, 142 entries were examined, and a set of exclusion criteria was applied. The selection process began with entries listed under the letter “A.” Entries lacking example sentences were excluded, as minimal contextual information is essential for accurate translation. Consequently, sentences retained were required to have complete syntactic structures; isolated predicates, fragments, or collocations were omitted to ensure analytical clarity and comparability across translation. The resulting corpus includes sentences varying in length, syntactic complexity, and lexical diversity, representing a range of word classes.

In accordance with the principles of translational plurality and complementarity (Deng 2024: 4), the human translator determines the most appropriate strategy for each sentence, considering both the source context and the specific conditions of the translation process. Each sentence was first translated from TR into IT by a professional human translator (HR). Between February 10 and 15, 2025, these same sentences were translated by four machine translation (MT) systems: Google Translate (GT), DeepL (DL), Gemini (GE), and ChatGPT-4 (CH). To ensure methodological consistency, each MT system was used in a single-query mode, without initiative prompting, and the same software versions were applied throughout the experiment. As such, no iterative prompting was employed, maintaining a controlled experimental design. The resulting translations were then systematically analyzed in comparison to the human rendition, with particular attention to convergences and divergences in the translated headwords (THs) and in the contextual adequacy of examples (Fig. 1).

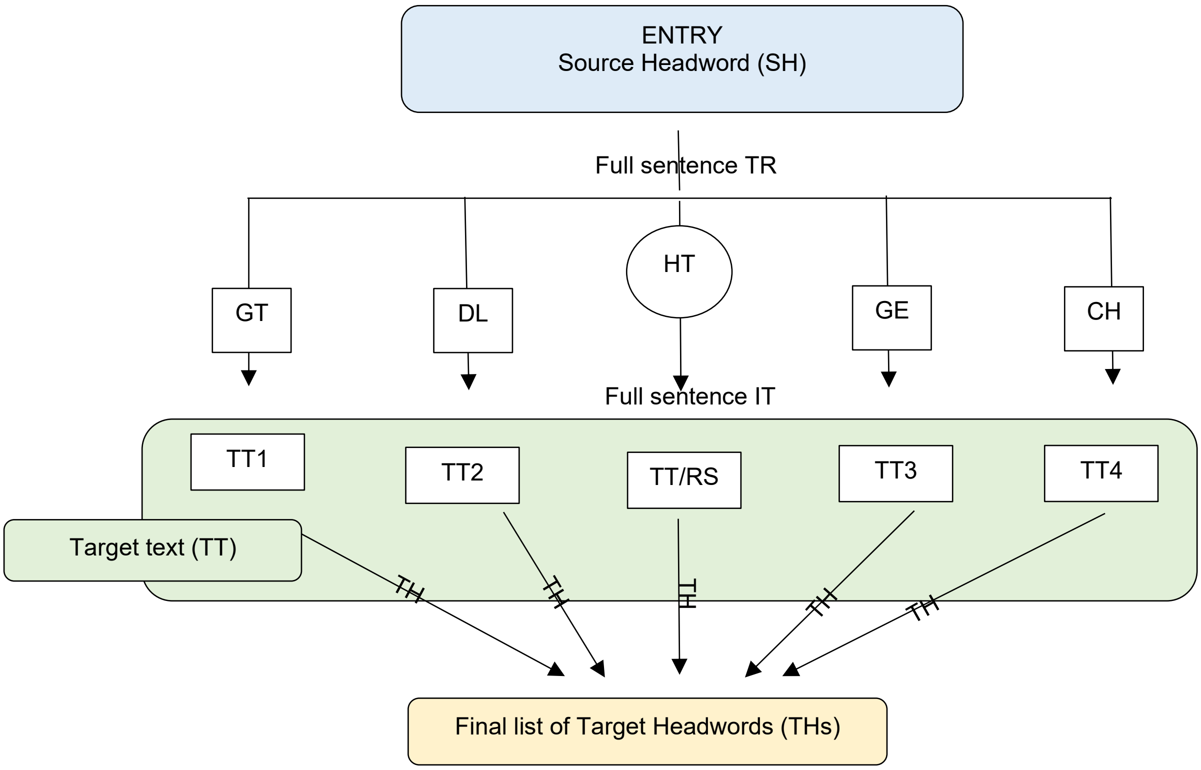

Fig. 1 The process of determining the final list of Target Headwords for comparison

The focus was on multiple dimension of equivalence: lexical, i.e., whether the translated headwords preserved meaning and sense of the source; syntactic, whether sentence structures were appropriately maintained; semantic, whether the meaning of the source sentence was fully conveyed; naturalness, whether the translation was idiomatic and coherent in Italian.

Differences and similarities between human and machine translations were recorded using structured tables, highlighting patterns of convergence and divergence in the translation of HW and example sentences. This approach allows for a transparent and reproducible analysis, providing both qualitative insights and quantitative summaries of translation performance.

The design intentionally focused on TR-IT translation as less-resourced language pair, highlighting the challeges faced by current NMT systems while establishing a clear framework for future comparative studies.

3.2 Data analysis

Assessing translation quality plays a vital role in evaluating the effectiveness of machine translation (MT) systems. While automatic evaluation metrics—such as METEOR (Banerjee and Lavie, 2005) and TER (Snover et al., 2006)—offer objective and scalable assessments, they often fail to capture nuances that align with human perception (Callison-Burch et al., 2006). Given these limitations, our study relies on human translator evaluation, a reliable method for assessing translation quality, as it is still challenging to create contextually accurate translations (Naveen and Trojovsky, 2024). This approach ensures that the assessment reflects a more authentic measure of translation performance, grounded in real-world expectations and professional standards.

As to this study, the data were quantitatively analyzed using statistical methods, including independent t-tests and ANOVA, to evaluate the performance differences between RS and systems.

4. Results and Discussion

Table 2 presents the minimum and maximum values related to the evaluation results of the human translation used as the reference standard. In addition, Skewness and Kurtosis values are also provided. These values are used to determine whether parametric or non-parametric analyses should be applied.

	N	Minimum	Maximum	Skewness	Kurtosis
GT	100	,00	1,00	,000	-2,041
DL	100	,00	1,00	,122	-2,026
GE	100	,00	1,00	,041	-2,040
CH	100	,00	1,00	,122	-2,026

Table 2. Descriptive Statistics

As shown in Table 1, the minimum value is "0" while the maximum value is "1." When examining the skewness and kurtosis values, Kline (2011, p. 63) states that for the assumption of normality to be met, skewness and kurtosis values should be less than 3. In this study, it is assumed that the quantitative data follow a normal distribution. Therefore, parametric methods are employed in the analyses. An independent samples t-test is used for the analysis of RQ 1 and Hypothesis 2. For the analysis of RQ 2, a one-way analysis of variance (ANOVA) is used.

Table 3 presents the results of the evaluation based on the human translation used as the reference standard. In the table, the mean scores of the responses provided by the four tools are shown, along with the standard deviation of their scores.

	GT	DL	GE	CH
Ave.	,5000	,4700	,4900	,5300
SD.	,50252	,50161	,50242	,50161

Table 3. Comparative evaluation of translations by translation systems

Table 3 presents the means and standard deviations of the translations produced by GT, DL, GE, and CH. The mean score for GT 0.5000, with a standard deviation of 0.50252. The mean score for DL is 0.4700, with a standard deviation of 0.50161. The mean score for GE is 0.4900, with a standard deviation of 0.50242. The mean score for CH is 0.5300, with a standard deviation of 0.50161. CH demonstrates the highest mean performance among the tools, suggesting a higher level of accuracy in comparison to the other AI-driven and NMT systems. GT follows closely, while DL and GE show slightly lower mean scores, respectively. Despite the relatively small differences, the standard deviations indicate a similar level of variability in performance across all tools, with DL and CH showing slightly higher consistency in their results.

When compared to this reference standard, CH correctly translated 53 tasks, GT 50 tasks, GE 49 tasks, and DL 47 tasks. As can be seen, CH has the highest number of correct translations, while DL has the lowest. This suggests that, in this study, AI-driven models like CH demonstrate a closer alignment with the RS compared to NMT systems like DL.

Hypothesis 1 investigates whether “AI chatbots are closer to the reference standard.” AI chatbots demonstrate a slightly superior alignment with the reference standard compared to NMT systems. With a total of 102 correctly translated tasks, the combined performance of CH and GE exceeds the 97 correct translations produced by GT and DL. This difference, though not necessarily statistically significant, indicates that collectively AI-driven chatbots are generally more effective in providing translations that closely match the reference standard. This could be attributed to the advanced capabilities of AI chatbots, which may handle nuances and context more effectively than traditional NMT systems. Nevertheless, the statistical proximity of GE’s score to GT’s may indicate that traditional NMT models still hold substantial merit in producing high-quality, contextually sound translations. GE, with its multimodal capabilities and more diverse corpora, leverages a broader context but has yet to fully capitalize on these advantages within translation tasks.

Within the scope of this study, Hypothesis 2 evaluates whether “AI chatbots are expected to perform better than NMT systems.” To test this hypothesis, an independent samples t-test is conducted.

	Ave.	SD	t-value	p-value
NMTs	,4850	,50103	-0,499	0,618
AI chatbots	,5100	,50115		

Table 4. Difference Tests Between NMT and AI Chatbots

As demonstrated in Table 4, while the performance of AI chatbots does not show a statistically significant difference when compared to NMT systems ($t = -4.099$, $p = 0.618 > 0.05$), the mean performance score of the AI chatbots ($\overline{X} = 0.5100$) is slightly higher than that of the NMT systems ($\overline{X} = 0.4850$). This suggests that, although the difference is not statistically substantial, AI chatbots may offer a marginal advantage in translating tasks compared to NMT systems. However, these findings highlight the complexity of machine translation performance, where even slight improvements may not always reach statistical significance, raising important considerations for future research on human translator versus machine translator performance. In particular, the small but consistent difference could point to the evolving capabilities of AI chatbots, which may offer a more nuanced approach to translation, even in the absence of significant statistical evidence.

As to RQ 1, an independent t-test is used to analyze the differences between the reference translation and the translations provided by GT, DL, GE, and CH. Two groups are formed for the analysis: human translation and AI language model translations.

Systems	N	Mean	Std. Deviation	t value	p value
AI Language Models	400	,4975	,50062	-10,030	0,000*
Human Translator	100	1,0000	,00000		

Tablo 5. Difference Test between RS and AI Language Models

As shown in Table 5, the mean of the human translation is 1, with a standard deviation of 0. The mean for the translations provided by the tools is 0.4975, with a standard deviation of 0.50062. The translator's choices differ statistically significantly from the translations provided by the AI language models ($t = -10.030$, $p = 0.000 < 0.05$). In other words, the translator's choices do not match the translations provided by the AI language models.

Regarding RQ 2, an ANOVA analysis method is used to examine the differences in the performance of the AI language models. Each of the four models in the study is assigned to separate groups for the analysis.

System	N	Mean	Std. Deviation	F value	p value	Difference
GT	100	0,5000	0,50252	25,278	0,000*	1,2,3,4<5
DL	100	0,4700	0,50161			
GE	100	0,4900	0,50242			
CH	100	0,5300	0,50161			
HT	100	1000	0,00000			

Table 6. Difference test for models

As shown in Table 6, overall, the models perform similarly in the context of the study. The HT's choices differ statistically significantly from the translations provided by AI language models ($F=25.278$, $p=0.000<0.05$). A Tukey post hoc analysis was conducted to determine between which groups the differences occurred. According to the results of the Tukey analysis, AI language models are statistically significantly lower than human translation.

4.1 Illustrative Examples

While a comprehensive analysis of individual examples falls beyond the scope of the present study, a few representative instances are provided. Given the corpus size (100 entries) and the variety of linguistic phenomena it contains, the selected instances illustrate typical patterns, highlighting linguistic issues, and allowing the methodological rationale to be clearly conveyed without overwhelming detail.

4.1.1 Context-sensitive verb translation

Example #1 - <abandı>

SS: Çocuğun üzerine <u>abandı</u>.

HR: <u>Si gettò</u> sul bambino.

GT: <u>Si premette</u> sul bambino.
DL: Il bambino era <u>prepotente.</u>
GE: <u>Si è gettato</u> sul bambino
CH: <u>Si è accasciato</u> sopra il bambino

All translations capture the general notion of physical action, but significant differences emerge in lexical choice and semantic interpretation. The suitable, contextually appropriate translation should emphasize the sudden physical movement: GT renders it literally, which conveys pressure rather than a sudden jump; DL misinterprets the subject entirely, producing a sentence about the child's behavior rather than the action performed on him; and CH selects a more descriptive, idiomatic expression, which captures the action's physicality but slightly differs in nuance. GE closely mirrors the HT in structure and meaning, although the HT's use of the past historic versus GE's rendering in the present perfect differ primarily in their temporal and aspectual implications.

4.1.2 Context-sensitive noun / adjective translation

Example #2 - <acı>

SS: Yemeklerde ve salatada <u>acı</u>yı severim.

HR: Mi piace <u>il piccante</u> nelle pietanze e nell'insalata.

GT: Mi piacciono i cibi <u>piccanti</u> e le insalate.

DL: Mi piacciono i cibi <u>piccanti</u> e le insalate.

GE: Mi piacciono i cibi <u>piccanti</u>, sia nei piatti che nelle insalate.

CH: Mi piace il piccante nei piatti e nelle insalate.

All translations convey the general idea, but subtle differences emerge in lexical and syntactic interpretation. HT clearly separates the concept of <acı>, "picante", as a quality applied to both "pietanze" (yemek) and "insalata" (salata), reflecting the original sentence's structure. GT and DL, on the other hand, shift the focus slightly, producing "i cibi piccanti e le insalate," which can be read as two separate entities, losing the attribute's shared application. GE restores clarity by explicitly linking "piccante" to both dishes and salad; CH also preserves the semantic relationship, phrasing it as "nei piatti e nelle insalate," but slightly alters the sentence rhythm.

Example #3 - <acılı>

SS: Çok acılı bir aile, kazada akrabaları vefat etmiş.

HT: È una famiglia molto addolorata, i parenti sono morti in un incidente.

GT: Una famiglia molto addolorata: i loro parenti sono morti nell'incidente.

DL: Una famiglia molto triste, i suoi parenti sono morti in un incidente.

GE: Una famiglia molto sfortunata ha perso i suoi parenti in un incidente.

CH: Una famiglia molto addolorata, i loro parenti sono morti in un incidente.

In Example #3 as well, all translations convey the general meaning, but differences appear in lexical choice and nuance. The HT uses "addolorata", capturing both emotional intensity and social context, and includes the copula "è" to form a complete, grammatically correct sentence, which is absent in MT outputs. GT preserves the lexical meaning but slightly alters punctuation and emphasis with a colon, where as DL translates <acılı> as "triste", which conveys emotion but may slightly understate the intensity or cultural nuance of mourning. GE interprets the sentence more broadly with "sfortunata", emphasizing misfortune rather than emotional grief. Finally, CH aligns closely with HR, though it retains i loro parenti, which is slightly redundant in Italian but maintains clarity. This example demonstrates how MT systems vary in rendering culturally and emotionally loaded adjectives, highlighting the importance of contextual sensitivity in noun/adjective translation.

4.1.3 Context-sensitive adverb translation

Example #4 - <acilen>

SS: Acilen eve gel.

HT: Torna a casa immediatamente.

GT: Torna subito a casa.

DL: Torna subito a casa.

GE: Torna a casa immediatamente.

CH: Torna a casa urgentemente.

All translations convey the basic imperative meaning, but subtle differences appear in lexical choice and register: “immediatamente”, the HT’s choice, emphasizes a formal, clear sense of urgency. GT and DL provide a slightly more colloquial equivalent, which is natural in spoken Italian (and was included in the synonyms); GE aligns with HR, preserving the formal urgency. CH’s rendering is correct but slightly marked and less idiomatic in everyday Italian imperatives. This highlights how machine translation systems vary in rendering context-sensitive adverbs of urgency, affecting tone, register, and idiomaticity.

4.1.4 Context-sensitive discourse marker translation

Example #5 - <acaba>

SS: İlginç bir kitap; yazarı kim acaba?

HT: È un libro interessante, chissà chi è l’autore?

GT: Un libro interessante; mi chiedo chi sia l’autore?

DL: Un libro interessante; chi è l’autore?

GE: È un libro interessante; chissà chi è l’autore?

CH: È un libro interessante; mi chiedo chi sia l’autore.

While all translations convey the general meaning, differences appear in rendering the discourse marker <acaba>. “Chissà” naturally conveys the speaker’s speculative curiosity in Italian. GT and CH opt for a grammatically correct but slightly more formal and reflective, less idiomatic, form in conversational Italian. DL omits the equivalent, rendering a neutral question, losing the nuance of speculative uncertainty. GE aligns closely with HR, preserving the idiomatic sense of curiosity. Example #5 illustrates how machine translation systems handle context-sensitive discourse markers differently, affecting the nuance of speaker attitude and conversational tone.

Overall, despite the lack of a statistically significant difference within models, it is noteworthy that CH achieves a slightly higher performance, while DL demonstrates the lowest performance. This suggests that, although the overall performance does not reach a significant level of differentiation, there are still observable variations between the models, with CH outperforming DL, which may indicate differences in the models' ability to handle translation tasks effectively. This highlights the nuances in model performance that can be attributed to various factors related to their design, training, and intended purpose. CH is a general-purpose language model designed for a wide range of tasks, which are not solely focused on translation. This broad focus means that its translation abilities are secondary to its primary function, potentially affecting its accuracy in comparison to systems specifically built for translation tasks. CH outperforming DL, a translation-specific system with years of refinement and focus on parallel corpora for translation, is due to the continuous improvement of AI-driven models, which may have developed more sophisticated translation capabilities in recent iterations, benefitting from more advanced language understanding techniques. The present data highlight the evolving nature of AI-driven chatbots, which, while originally designed for broader language tasks, appears to be edging closer to specialized NMT systems in terms of translation accuracy. This shift may reflect both advancements in AI technology and the expanding capabilities of models trained on diverse, high-quality data sets.

5. Conclusion

In conclusion, this study confirms and extends previous research findings, particularly in relation to the significant advantages of human translations in terms of accuracy and contextual appropriateness (Çetin and Duran 2023; Siu 2023: 29). The results align with the broader literature, reaffirming that while AI-driven systems such as ChatGPT and Gemini are powerful tools for translation tasks, they still fall short in comparison to human translators for more context-sensitive yet generic translation work. The careful selection of corpus in this study, which prioritized syntactically complete units, ensured a well-defined context for evaluating translation accuracy, further highlighting the unique strength of human translators in handling linguistic subtleties and cultural nuances. ChatGPT slightly outperformed the other tools in the context of this study. However, as the study also reveals, AI chatbots and NTMs are not yet on par with human translators in terms of delivering translations that are contextually rich and semantically precise. While chatbots multimodal capabilities provide some edge in translating beyond word-for-word accuracy, they still struggle to fully replicate the deep cultural competence and subject-specific knowledge that professional human translators possess.

Moreover, the study acknowledges the evolving synergy between human translators and large language models (LLMs). As LLMs continue to improve, they can support human translators by providing fast draft translations and multilingual resources, thus enhancing productivity and efficiency. However, this interaction is complementary rather than substitutive—humans remain essential for refining the output of AI systems and ensuring the quality and cultural relevance of translations. Ultimately, the integration of AI tools into the translation workflow offers potential, but the indispensable role of professional translators is crucial for maintaining the accuracy, depth, and cultural sensitivity required in high-quality translation work.

6. Limitations

This study presents several methodological and technical limitations that should be acknowledged. First, the use of a personal account for accessing and interacting with machine translation platforms, which may introduce potential bias related to personalized algorithmic behavior. Second, the non-deterministic nature of the implemented translation systems poses significant challenges in maintaining consistency across translations. Although efforts were made to use the same model across all translation instances within a very limited span of time, the internal model may still have undergone updates not visible to the user. Third, human translation served as the reference standard in this study, with subjective choices potentially affecting the evaluation of what constitutes an accurate or preferred rendering. While expert judgment offers a high level of linguistic reliability, it also brings inherent interpretative bias. Moreover, no iterative prompting—i.e., the process of refining translations through successive user inputs—was applied in this study. This choice leaves open the question of whether subsequent interactions could lead to significant improvements in translation quality. Finally, the probabilistic nature of AI models, tends to favor high-frequency lexical items commonly used in the target language. This frequency-weighted prediction strategy often leads to generic or overly common word choices, which may not align with the lexicographical precision required in dictionary-based translation tasks.

References

Agung, I G.A. M., Putu G. B., & Nil Wayan, S. (2024). Translation performance of Google Translate and DeepL in translating Indonesian short stories into English. *In Proceedings: Linguistics, Literature, Culture and Arts International Seminar (LITERATES*), 178-185.

Al R, Rafat, R. J., & Malkawi, M. (2025). ChatGPT translation vs. human translation: an examination of a literary text. *Cogent Social Sciences*, 11 (1), 1-21. DOI: 10.1080/23311886.2025.2472916.

Banerjee, S., & Lavie, A. (2025). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *In Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, 65-72.

Bergenholtz, H., & Gouws, R. H. (2012). What is Lexicography?, *Lexikos*, 22, 31-42.

Bernardini, S., & Zanettin, F. (2020). When corpora meet translation: A resource for teaching, research and practice. In C. O'Brien & M. Winters (Eds.), *The Routledge Handbook of Translation and Technology* (pp. 383–398). London: Routledge.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. In D. McCarthy, S. Wintener (Eds.), *11th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Cimputational Linguistics, 249-256.

ChatGPT. Available from: https://chatgpt.com (February 2025).

Çetin, Ö., & Duran, A. (2024). A Comparative Analysis Of The Performances Of Chatgpt, Deepl, Google Translate And A Human Translator In Community Based Settings, *Amasya Üniversitesi Sosyal Bilimler Dergisi (ASOBİD)*, 9 (15), 120-173.

de Schryver, G.M. (2023). Generative AI and Lexicography: The Current State of the Art Using ChatGPT, *International Journal of Lexicography*, XX, 1-33, https://doi.org/10.1093/ijl/ecad02.

DeepL Translator. Available from: https://www.deepl.com/it/translator (February 2025).

D'Achille, P. (2010). *L'italiano contemporaneo*. Bologna: Il Mulino.

Dardano, M., & Trifone, P. (2003). *Grammatica italiana: Con nozioni di linguistica*. Bologna: Zanichelli.

De Mauro, T. (1999). *Grande dizionario italiano dell'uso (GRADIT)*. Torino: UTET.

Demiryan, R. (2011). *Dizionario. Italiano-Turco*, İnkılap Kitabevi.

Deng, J., & Lin, Y. (2022). The Benefits and Challenges of ChatGPT: An Overview, *Frontiers in Computing and Intelligent Systems*, 2(2), 81-83, http://dx.doi.org/10.54097/fcis.v2i2.4465.

Deng, Y. (2024). ChatGPT-based Research on the Influence of AI on English Translation, *Applied Mathematics and Nonlinear Sciences*, 9(1), 1-14, DOI:10.2478/amns-2024-2044.

Fan, P., & Gong, X.H. (2023). The Application of ChatGPT in Translation Teaching: Changes, Challenges, and Responses. *International Journal of Education and Humanities* 11(2). 49-52 DOI:10.54097/ijeh.v11i2.13530.

Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: Is Chat GPT a blessing or a curse?, *Frontiers in Education*, 8. https://doi.org/10.3389/feduc.2023.1166682.

Gemini AI. Available from https://gemini.google.com/app?hl=tr (February 2025).

Ghosh, S., & Caliskan, A. (2023). ChatGPT Perpetuates Gender Bias in Machine Translation and Ignores Non-Gendered Pronouns: Findings across Bengali and Five other Low-Resource Languages, *ArXiv*, arXiv:2305.10510.

Google Translate. Available from https://translate.google.com.tr (February 2025).

Güneş, S. (2018). İtalyanca ve Türkçe Belirsiz Nicelik İfadelerinde Adların Kullanımı Üzerine Bir İnceleme, *RumeliDE Journal of Language and Literature Studies*, 214-231.

Gürlek, M. (2023). Türkçe Öğrenici Sözlüğü için Bir Maddeyi Hazırlamak, *Türkiyat Mecmuasi,* 33(2), 625-640, DOI: 10.26650/iuturkiyat.1306809.

Habeeb, L. S. (2020). Investigating the Effectiveness of Google Translate among Iraqui Students, *International Journal of Innovation, Creativity and Change*, 12(12), 104-105.

Jiao, W., Wang, H., Wenxuan, J., Wang, X., Shi, S., & Tu, Z. (2023). Is Chat GPT A Good Translator? Yes With GPT-4As the Engine, *arXiv*, 1-10, https://arxiv.org/pdf/2301.08745.

Karnal, A.R., Vanmacher Pereira, V. (2015). Reading Strategies in a L2: A Study on Machine Translation, *The Reading Matrix: An International Online Journal*, 15 (2), 1-11.

Kaya, R. (2020). İtalyanlarin Türkçe Öğretimi: Harfler Ve Ses Karşiliklari, *Hacettepe Üniversitesi Yabancı Dil Olarak Türkçe Araştırmaları Dergisi*, 6, 51-75.

Kembaren, F. R. W., Hasibuan, A. K., & Natasya, A. (2023). Technology Trends in Translation: A Comparative Analysis of Machine and Human Translation, *Absorbent Mind. Journal of Psychology and Child Development*, 3 (2), 168-183, 10.37680/absorbent_mind.v3i2.4486.

Khoong, E. C., Steinbrook, E., & Brown, C. (2019). Assessing the Use of google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions, *JAMA Intern Med.*, 179 (4), 580-582. doi:10.1001/jamainternmed.2018.7653.

Khoshafah, F. (2023). ChatGPT for Arabic-English Translation: Evaluating Accuracy, *Research Square,* 1-18, https://doi.org/10.21203/rs.3.rs-2814154/v2.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. London: Guilford publications.

Kocmi, T., & Federman, C. (2023). Large Language Models Are State-of-the-Art Evaluators of Translation Quality, *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, European Association for Machine Translation, 193-203.

Lew, R. (2023). ChatGPT as a COBUILD lexicographer, *Humanities and Social Communications*, 10, 704. https://doi.org/10.1057/s41599-023-02119-6.

Lew, R. (2024). Dictionaries and lexicography in the AI era, *Humanit Soc Sci Commun*, 11, 426. https://doi.org/10.1057/s41599-024-02889-7.

Li, J., Dada, A., Puladi, B., Kleesiek, J., & Egger. J. (2024). ChatGPT in healthcare: A taxonomy and systematic review, *Computer Methods and Programs in Biomedicine*, 245. https://doi.org/10.1016/j.cmpb.2024.108013.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). *Summary of ChatGPT/GPT-4 research and perspective towards the future of large language models*. arXiv:2304.01852. https://doi.org/10.1016/j.metrad.2023.100017.

Lo Cascio, V. (2003). *Grammatica dell'italiano per stranieri*. Firenze: Le Monnier.

Naveen, P., & Trojovsky, P. (2024). Overview and Challenges of Machine Translation for Contextually Appropriate Translations, iScience, 27 (10), 1-25, DOI:10.1016/j.isci.2024.110878.

Manzelli, G. (2020). Il contributo dell'italiano al lessico ludico del turco di Turchia. *Italiano LinguaDue*, 2, 82-97.

Marello, C. (1989). *Dizionari bilingui: Studi di metalessicografia.* Alessandria: Edizioni dell'Orso.

Marello, C. (2015). La lessicografia bilingue in Italia oggi. In R. Simone (Ed.), *Enciclopedia dell'Italiano* (pp. 1–6). Roma: Istituto dell'Enciclopedia Italiana Treccani.

Özkan, N. (2020). L'italiano nel turco, *Italiano LinguaDue*, 2, 98-106.

Rocchi, L. (2020). Gli elementi turco-ottomani nel *Vocabolario italiano*, e *kurdo* di Maurizio Garzoni (1887) e la loro importanza storico-documentaria, *Rivista internazionale di Tecnica della Traduzione / International Journal of Translation*, 22, 231-249.

Rossi, F. (2005). *Lessicografia e traduzione: Problemi di equivalenza tra italiano e lingue europee*. Torino: UTET.

Rundell, M. (2024). Automating the Creation of Dictionaries: Are We Nearly There?, *Proceedings of the 16th International Conference of the Asian Association for Lexicography:"Lexicography, Artificial Intelligence, and Dictionary Users"*. Seoul: Yonsei University, 1–9.

Sahari, Y., Al-Kadi Talib, A.M., & Ali, J. K. M. (2023). A Cross Sectional Study of ChatGPT in Translation: Magnitude of Use, Attitudes, and Uncertainties, *J Psycholinguist Res*, 52, 2937–2954, https://doi.org/10.1007/s10936-023-10031-y.

Siu, S. C. (2023). ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation, *SSRN*, 1-36, https://dx.doi.org/10.2139/ssrn.4448091.

Serianni, L. (2005). *Italiano: Grammatica, sintassi, dubbi*. Milano: Garzanti.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation, In *Proceedings of Association for Machine Translation in the Americas (AMTA)*, 223-231.

Son, J., & Kim, B., (2023). Translation Performance from the User's Perspective of Large Language Models and Neural Machine Translation Systems, *Information*, 14 (10), 574, https://doi.org/10.3390/info14100574.

Sungran, K. (2023). An Analysis of ChatGPT's Language Translation Based on the Korean Film Minari, *Journal of English Teaching through Movies and Media*, 24 (4), 1-14 doi: https://doi.org/10.16875/stem.2023.24.4.1.

Tan, R., Long, X., & Bamigbade, O. O. (2023). Comparative Research on Machine Translation and Human Translation of Examples in Dictionary from the Perspective of Skops Theory, *IEIT 2023, AHSSEH* 10, 342–353, https://doi.org/10.2991/978-94-6463-230-9_41.

Yilmaz, E. D., Naumovska, I., & Aggarwal, A. V. (2023). AI-Driven Labor Substitution: Evidence from Google Translate and ChatGPT. *INSEAD*. http://ssrn.com/abstract=4400516.

Xiu, P., & Xeauyin, L. (2018). Human translation vs machine translation: The practitioner phenomenology, *Linguistics and Culture Review*, 2(1), 13-23, https://doi.org/10.37028/lingcure.v2n1.8.

Zanettin, F., Bernardini, S., & Stewart, D. (2023). *Corpora in Translator Education*. London: Routledge.