Araştırma Makalesi
Research Article

ISSN 2548-0502

söylem
FİLOLOJİ DERGİSİ

Mart **2025:** 819-826
Çeviribilim Özel Sayısı

Evaluating Large Language Models in Translation: A Theoretical and Practical Analysis Based on Skopos Theory

ARŞ. GÖR. **DİLARA BAL*** **-** PROF. DR. **ŞABAN KÖKTÜRK****

Abstract

The aim of this study is to analyse the translation competences of large language models, such as GPT-4, through various theoretical lenses within translation studies. This study can be considered unique as its evaluation of these models' translation performance is based on skopos theory. This research assesses how well large language models align with these theoretical frameworks and their effectiveness in producing contextually appropriate and culturally sensitive translations. The research explores the architecture and operational principles of large language models, explaining their application in translation. Methodologically, the study employs a comparative analysis of translations generated by large language models across language pairs amongst Turkish, English and Spanish. The analysis focuses on key theoretical aspects, such as the purpose and functionality of translations. Additionally, the study examines the cultural and contextual appropriateness of translations generated by large language models, evaluating their ability to maintain cultural nuances and meet the expectations set by the respective translation theories. The findings reveal the strengths and limitations of large language models in adhering to theoretical principles, providing insights into their potential to enhance or challenge traditional translation practices. This research advances the theoretical understanding of machine translation and offers practical recommendations for improving the translation capabilities of large language models. By integrating theoretical analysis with practical applications, the study aims to provide insight into future developments in translation technologies and their role in the future of translation studies.

Keywords: Large language models, translation theories, neural machine translation, skopos theory, translation technologies.

Çeviride Büyük Dil Modellerini Araştırmak: Skopos Kuramı Üzerinden Kuramsal ve Uygulamaya Dayalı Bir Analiz

Öz

Bu çalışmanın amacı, GPT-4 gibi büyük dil modellerinin çeviri edinçlerini çeviribilim alanındaki kuramlar aracılığıyla incelemektir. Bu çalışma, büyük dil modellerini skopos kuramına göre incelemesi bakımından özgün olarak değerlendirilebilir. Bu araştırmada, büyük dil

* * Sakarya University, PhD student, Social Sciences Institute, Translation Studies (Sakarya, Türkiye), e-mail: dilarabal@sakarya.edu.tr, ORCID: 0000-0002-3934-0681

** Sakarya University, Translation Studies (Sakarya, Türkiye), e-mail: skokturk@sakarya.edu.tr, ORCID: 0000-0002-2575-0137

Gönderilme Tarihi: 15 Aralık 2024 Kabul Tarihi: 13 Mart 2025

modellerinin bu kuramsal çerçevelerle ne kadar uyum sağladığı ve bağlamsal bütünlüğü ve kültürel hassasiyeti yansıtmada ne kadar başarılı olabileceği değerlendirilmiştir. Araştırmada, büyük dil modellerinin yapısı ve işleyiş ilkeleri incelenmiş ve çeviride uygulamalarına bakılmıştır. Çalışmanın yönteminde Türkçe, İngilizce ve İspanyolca ele alınarak dil çiftleri arasında büyük dil modelleri tarafından üretilen çevirilere bakılmış ve karşılaştırmalı analiz kullanılmıştır. Bu analizde, çevirilerin amacı ve işlevselliği gibi temel kuramsal yönlere odaklanılmıştır. Ayrıca, çalışmada büyük dil modelleri tarafından üretilen çevirilerin kültürel ve bağlamsal uygunluğunu incelenmiş ve kültürel nüanslar da dahil olmak üzere çeviri kuramlarının belirlediği standartlar çerçevesinde çeviri edinçleri değerlendirilmiştir. Bulgularla birlikte büyük dil modellerinin kuramsal ilkelere bağlı kalmadaki güçlü ve zayıf yönleri gösterilerek geleneksel çeviri uygulamaları tartışılmıştır. Bu araştırma yalnızca makine çevirisinin kuramsal yönlerine vurgu yapmakla kalmamış, aynı zamanda büyük dil modellerinin çeviri edinçlerini geliştirmek için öneriler sunmuştur. Kuram ve uygulamayı bütünleştirerek çeviri teknolojilerindeki gelişmelere ve çeviribilimin geleceğindeki rollere ilişkin de yorumlar yapılmıştır.

Anahtar Kelimeler: Büyük dil modelleri, çeviri kuramları, nöral makine çevirisi, skopos kuramı, çeviri teknolojileri

INTRODUCTION

Translation is a crucial medium for cross-cultural communication, historically reliant on human translators' nuanced understanding of language and culture. The advent of Neural Machine Translation (NMT) and, more recently, Large Language Models (LLMs), such as GPT-4, has transformed the translation landscape, raising questions about their ability to fulfil functional and cultural requirements. This study evaluates the translation competence, which can also be linked to transfer competence (Kirsten, 2019), of LLMs using Skopos theory, which prioritises the purpose of the translation over linguistic equivalence. By examining their outputs, this research investigates whether these models can achieve the goals set for translations in different cultural and linguistic contexts.

Recent studies, such as Castilho et al. (2018), have highlighted the growing integration of NMT into translation workflows. While these advancements offer increased efficiency, they often fall short in addressing cultural adaptation and context, central principles in Skopos theory. Similarly, Koehn (2020) notes that although NMT achieves higher fluency compared to statistical models, challenges persist in low-resource languages and culturally sensitive texts, requiring further refinement.

Large language models have introduced a new paradigm in translation technologies, offering unprecedented scalability and speed. However, their potential to address the nuances of cultural and contextual translation remains underexplored. By situating this study within Skopos theory, it aims to evaluate whether these models can function effectively within the parameters set by translation studies. The study is particularly significant given the global reliance on automated tools for communication in a highly interconnected world.

1. Translation Technologies

We currently rely on technology for most tasks to do with translation (Şahin, 2013 & 2023). We are at a point where technology has reached vital tasks in life as well. Here are some examples of how many tools might exist within translation technologies.

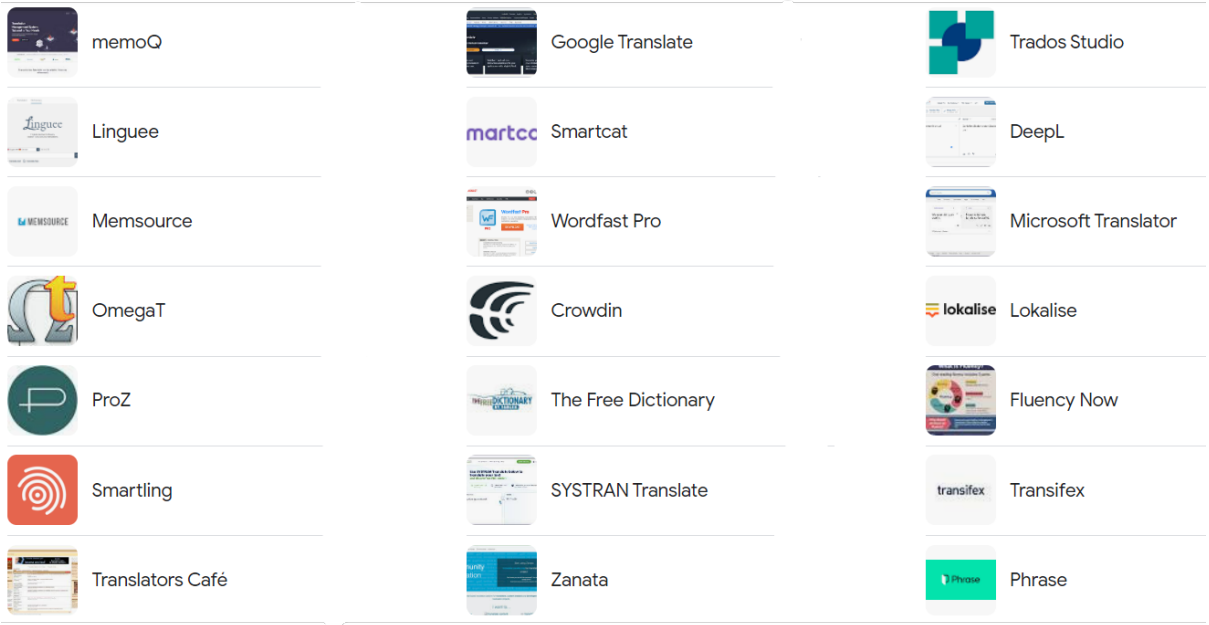


Image 1: AI-based tools for translation (Source: Google)

Some of the tools which are commonly used in translation are listed in the image above. Examples of these are Trados Studio, Smartcat, and memoQ, which have been used as translation technologies for many years. However, these tools have upgraded themselves as technology as a general field has improved. Thus, they include a part of neural machine translation and artificial intelligence as well. As artificial intelligence has become increasingly popular, even more online tools have emerged.

1. Skopos Theory

Developed by Hans J. Vermeer in the late 1970s, Skopos theory shifts the focus from linguistic fidelity to the function of the target text. A successful translation meets the intended goals of the audience, making functionality and cultural appropriateness central concerns. This study investigates whether LLM-generated translations fulfil their functional purposes and effectively address cultural nuances. The theory's emphasis on purpose is particularly relevant in evaluating machine-generated translations, which often prioritize literal accuracy over functional equivalence. Skopos theory underscores the translator's agency in shaping the text according to its intended purpose. In the context of LLMs, this theory provides a critical lens to evaluate whether these tools can mimic such agency or if they require human intervention to bridge the gap between linguistic and cultural expectations. Jiménez-Crespo (2017) highlights the importance of functional equivalence in translation studies, particularly in the context of crowdsourcing and collaborative translations. These principles align closely with Skopos theory, which prioritizes the needs and expectations of the target audience. Applying this framework to LLMs allows us to critically assess their ability to deliver purpose-driven translations rather than purely linguistic fidelity.

METHODOLOGY

A comparative analysis was conducted across three language pairs: Turkish-English, English-Spanish, and Turkish-Spanish to evaluate LLM translation performance. Key methodological steps include:

i. **Text Selection**: Representative texts covering literary, technical, and colloquial content were chosen to ensure a broad evaluation of translation outputs.

ii. **Translation Process**: Translations were generated using GPT-4 with standardised prompts, ensuring consistency in input conditions.

iii. **Analysis Criteria**:

a. Fidelity to the source text's purpose.

b. Contextual appropriateness.

c. Cultural sensitivity.

d. Accuracy in linguistic structure.

iv. **Validation**: Native speakers, translators and online grammar tools (e.g., Grammarly and LanguageTool) evaluated the outputs to identify strengths and limitations.

FINDINGS

The findings reveal varied performance levels across language pairs:

- **Spanish to English**: High accuracy due to linguistic similarities and extensive training data for both languages.
- **English to Spanish**: Slightly lower performance; challenges with grammatical gender and verb conjugation.
- **Turkish to English**: Strong results but affected by structural differences and idiomatic expressions.
- **English to Turkish**: Struggles with agglutinative structures, cultural nuances, and limited training data.
- **Spanish to Turkish**: Moderate accuracy, constrained by grammatical complexity and limited data.
- **Turkish to Spanish**: Least accurate, reflecting significant linguistic and cultural disparities.

On a quiet afternoon, María found an old, dusty key hidden in her grandmother's attic. Curiosity led her to an ancient wooden chest tucked away in a corner. As she turned the key, the chest creaked as it opened, revealing a package wrapped in cloth. Inside the package, there were handwritten letters and a black-and-white photo of her grandmother with someone she had never seen before. María felt that she was about to uncover a secret that had been hidden for generations.

Image 2: *An example of the type of output received.*

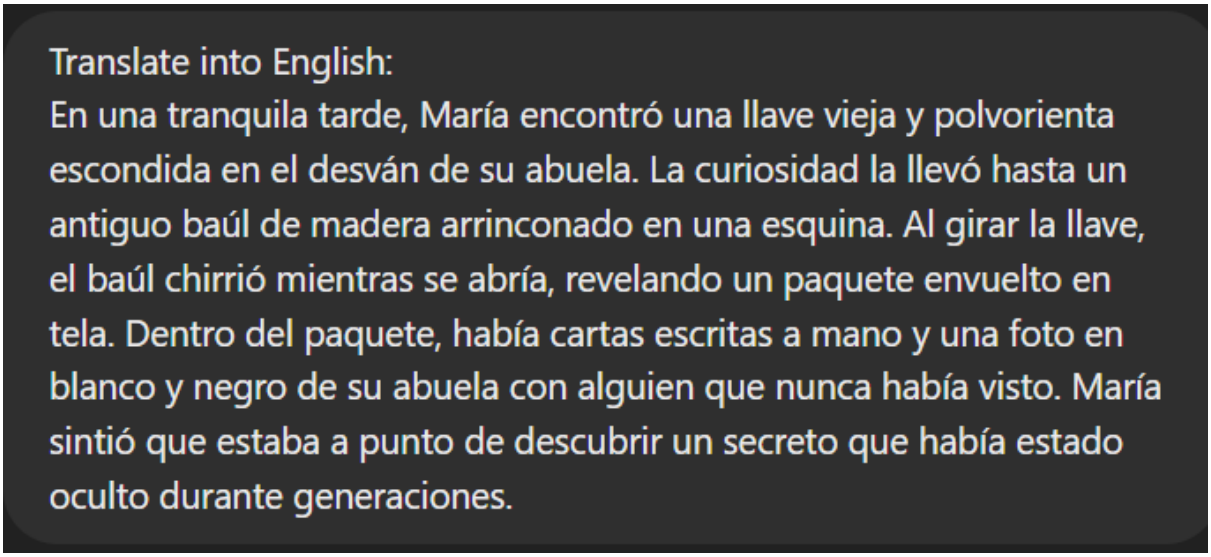


Image 3: *An example of the type of output received.*

Error Type	Turkish-English	English-Turkish	Spanish-English	Turkish-Spanish
Spelling/Grammar	10%	15%	5%	20%
Contextual Errors	15%	25%	10%	30%
Cultural Adaptation	20%	30%	15%	40%
Meaning Loss	25%	20%	10%	35%

Table 1: *Error Distributions amongst language pairs.*

DISCUSSION

Castilho et al. (2018) argue that while integrating NMT into workflows improves efficiency, issues surrounding cultural adaptation persist. These findings are reinforced by Koehn (2020), who emphasizes that challenges in domain-specific and low-resource translations remain significant. The results align with the principles of Skopos theory, highlighting that LLMs often meet the functional goals of translations in high-resource languages but falter in maintaining cultural nuances in less-resourced pairs. For instance:

- **Turkish-English**: LLMs perform well due to structural predictability.

• **Turkish-Spanish**: LLMs encounter difficulty with idiomatic expressions and culturally specific content, which aligns with findings by Jiménez-Crespo (2017).

Language Pair	AI/LLM (GPT) Translation Correctness (%)
Spanish to English	90%
English to Spanish	88%
Turkish to English	85%
English to Turkish	80%
Spanish to Turkish	75%
Turkish to Spanish	70%

• **Table 2:** *Percentages of correction*.

Supervised learning algorithms have enhanced the models' ability to produce fluent translations (Wang, 2023). However, significant gaps remain in addressing cultural subtleties and maintaining the intended function of the target text. Ethical considerations, such as biases in training data and the potential devaluation of linguistic expertise, also require attention. Certain recommendations are as follows:

• **Hybrid Models**: Incorporate human oversight to address cultural and contextual gaps, ensuring higher-quality outputs.

• **Low-Resource Languages**: Invest in curated datasets for underrepresented languages, drawing from initiatives like OpenNMT (Toral et al., 2018).

• **User Feedback Integration**: Enable dynamic learning from user corrections to improve real-time performance.

• **Ethical Transparency**: Promote accountability by making training datasets and algorithms more transparent.

CONCLUSION

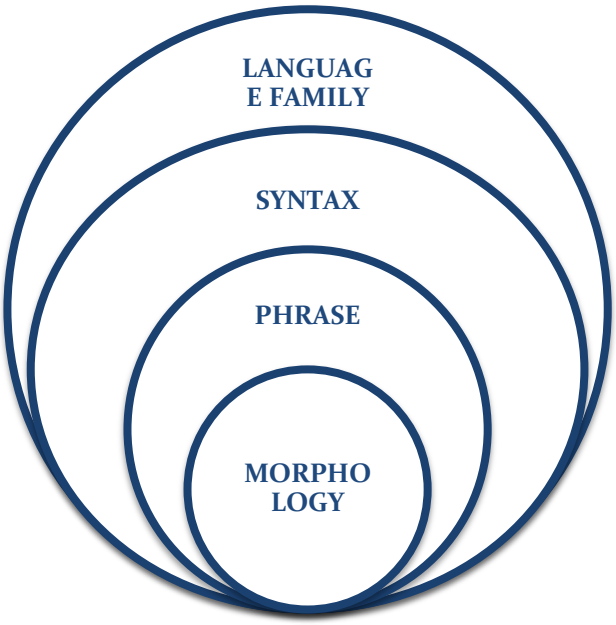


Figure 1: Reasons for different outcomes amongst languages

We could argue that the percentages turned out the way they did partly due to the languages belonging to different language families, with variations in syntax, phrase structure, and morphological complexity. Prior studies have demonstrated that linguistic typology significantly impacts the quality of neural machine translation (NMT) outputs (Koehn & Knowles, 2017; Aharoni, Johnson, & Firat, 2019). For instance, languages with rich morphology, such as Turkish and Finnish, pose greater challenges for NMT systems than more analytically structured languages like English or Chinese. Research on low-resource languages (Östling & Tiedemann, 2017) has further highlighted the disparities in translation quality when working with underrepresented linguistic data, as current models are disproportionately trained on high-resource languages.

Beyond linguistic differences, cultural and contextual sensitivity is another critical factor affecting translation quality. Large language models, while capable of generating fluent and grammatically correct translations, often struggle to capture cultural nuances, idiomatic expressions, and pragmatic meanings (Fan et al., 2021). This limitation has been widely discussed in studies examining biases in NMT outputs, where translations tend to reflect the dominant cultural perspectives encoded in training data (Bender et al., 2021). The issue of bias extends to ethical concerns, particularly regarding the transparency of training datasets and the socio-political implications of automated translation in marginalized communities.

Given these challenges, recent studies advocate for hybrid approaches that combine machine efficiency with human expertise (Toral & Way, 2018). While NMT has significantly reduced the time and effort required for translation, human post-editing remains essential in ensuring accuracy, particularly in legal, medical, and literary translation domains. Post-editing studies suggest that professional translators can improve machine-generated texts through targeted interventions, refining semantic accuracy, stylistic appropriateness, and cultural relevance. Additionally, broader evaluations across diverse language pairs are necessary to assess how NMT models perform across typologically distinct languages and whether machine-generated outputs remain consistent in quality.

Future research should further investigate the ethical implications of AI-driven translation, emphasizing training data transparency, fairness in representation, and equitable access to language technology. The development of more inclusive multilingual models should consider the needs of underrepresented languages, incorporating interdisciplinary insights from computational linguistics, sociolinguistics, and translation studies. As technology evolves, the role of human translators will remain vital, particularly in areas requiring cultural sensitivity and ethical judgment, reinforcing the idea that while AI can augment translation, it cannot fully replace the nuanced decision-making abilities of human professionals.

REFERENCES

Aharoni, Roee, Johnson, Melvin, & Firat, Orhan. (2019). Massively multilingual neural machine translation. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 3874–3884.

Bender, Emily M., Gebru, Timnit, McMillan-Major, Angelina, & Shmitchell, Shmargaret. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.

Fan, Angela, Bhosale, Shruti, Schwenk, Holger, Ma, Xiaoqing, El-Kishky, Ahmed, Goyal, Naman, ... & Edunov, Sergey. (2021). Beyond English-centric multilingual machine translation. Journal of Machine Learning Research, 22(107), 1–48.

Jiménez-Crespo, Miguel Ángel (2017). *Crowdsourcing and Online Collaborative Translations: Expanding the Limits of Translation Studies*. John Benjamins Publishing.

Kocmi, Tom, & Federmann, Christian (2023). *Large language models are state-of-the-art evaluators of translation quality*.

Koehn, Philipp (2020). *Neural Machine Translation*. Cambridge University Press.

Koehn, Philipp, & Knowles, Rebecca. (2017). Six challenges for neural machine translation. Proceedings of the First Workshop on Neural Machine Translation, 28–39.

Malmkjær, Kirsten (2009) . What is translation competence? Revue française de linguistique appliquée, Vol. XIV(1), 121-134. https://shs.cairn.info/journal-revue-francaise-de-linguistique-appliquee-2009-1-page-121?lang=en.

Marzena Karpinska and Mohit Iyyer (2023). Large Language Models Effectively Leverage Document-level Context for Literary Translation, but Critical Errors Persist. In *Proceedings of the Eighth Conference on Machine Translation.* pp. 419–451, Singapore. Association for Computational Linguistics.

Östling, Robert, & Tiedemann, Jörg. (2017). Neural machine translation for low-resource languages. Machine Translation, 31(1–2), 187–207.

Şahin, Mehmet (2013). *Çeviri ve Teknoloji*. İzmir Ekonomi Üniversitesi Yayınları.

Şahin, Mehmet (2023). *Yapay Çeviri*. Çeviribilim Yayınları.

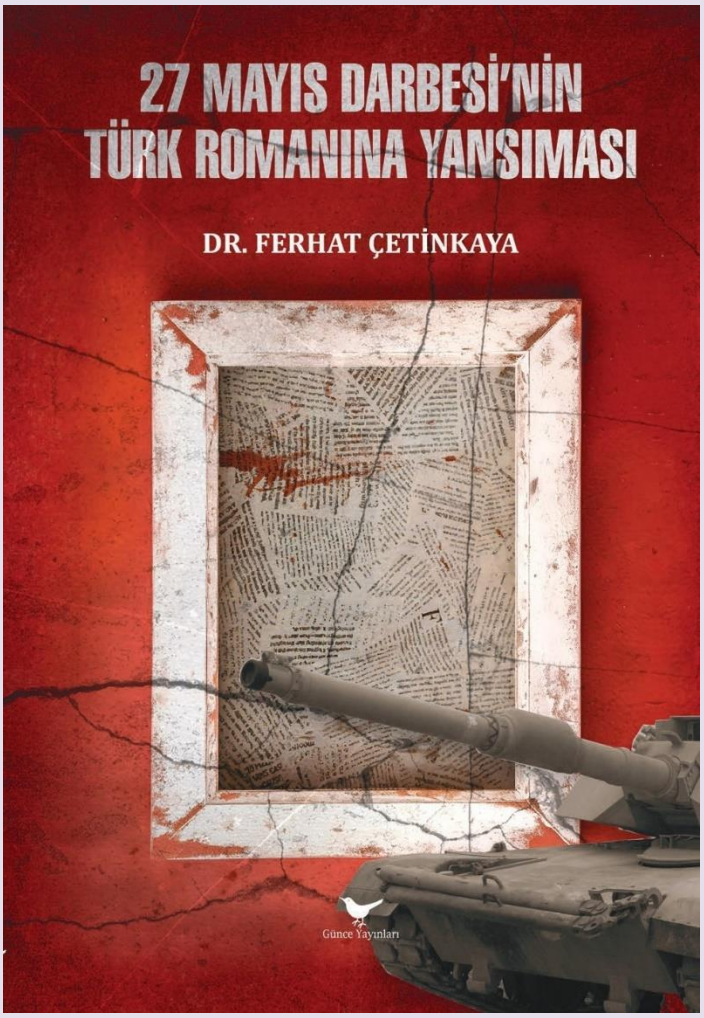
Toral, Antonio, & Way, Andy. (2018). What level of post-editing is worthwhile? Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers), 192–208.Castilho, Sheila, et al. (2018). Evaluating MT for massive open online courses: A multifaceted comparison between PBSMT and NMT systems. *Machine Translation, 32*(4), 275–298.

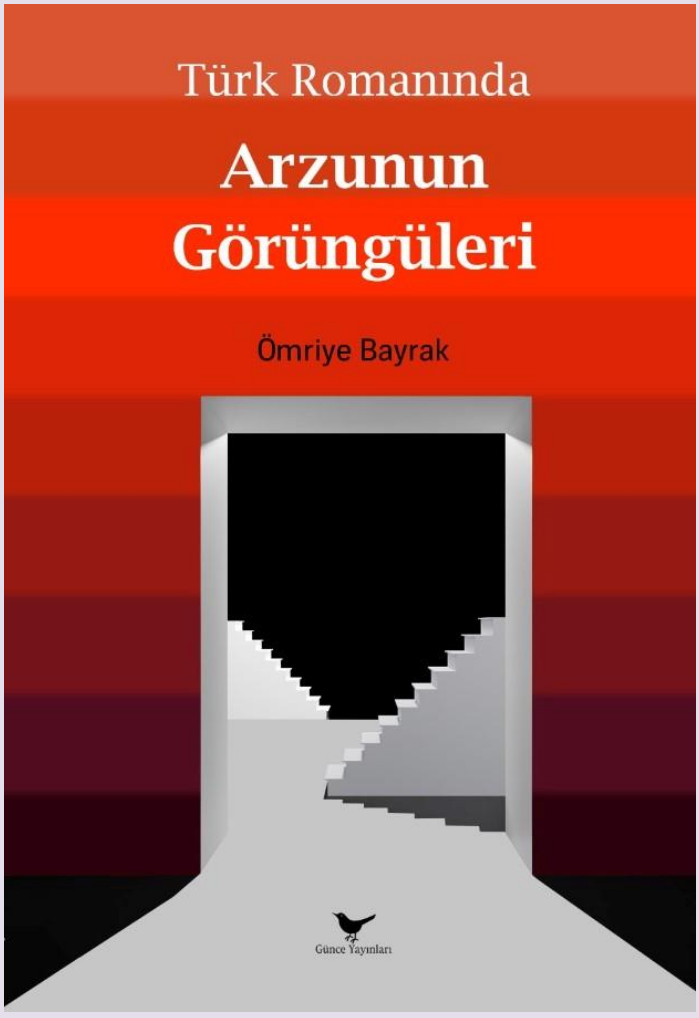
Toral, Antonio, et al. (2018). Post-editing Effort of a Novel with Statistical and Neural Machine Translation. *Computational Linguistics, 44*(3).

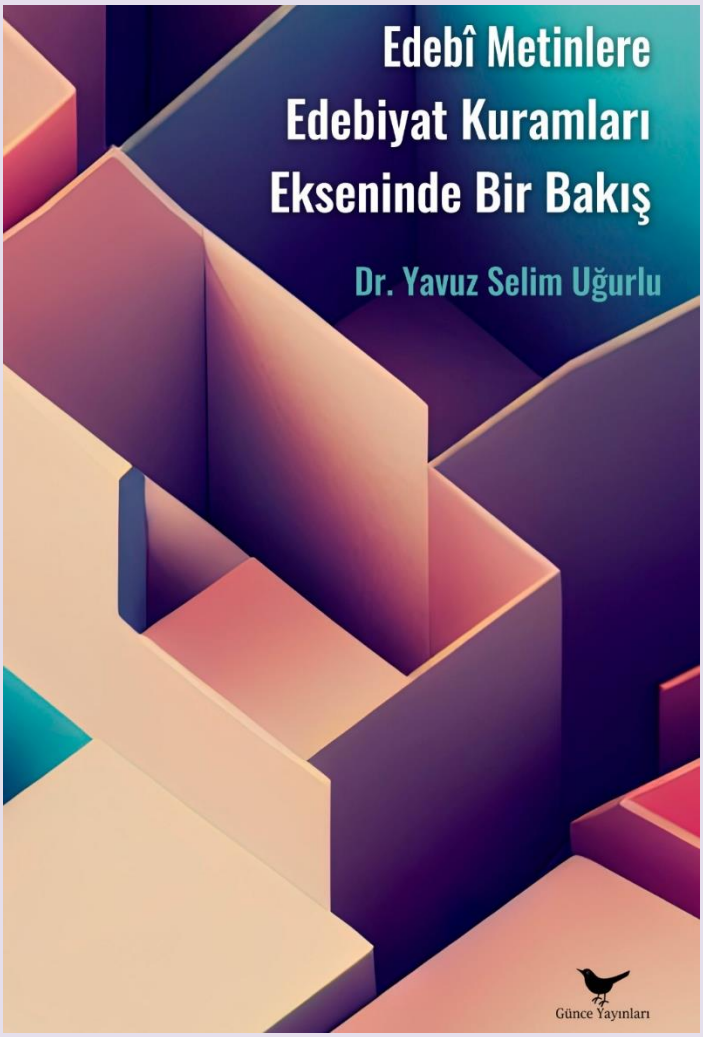
Wang, Longyue, et al. (2023). *Document-Level Machine Translation with Large Language Models.*

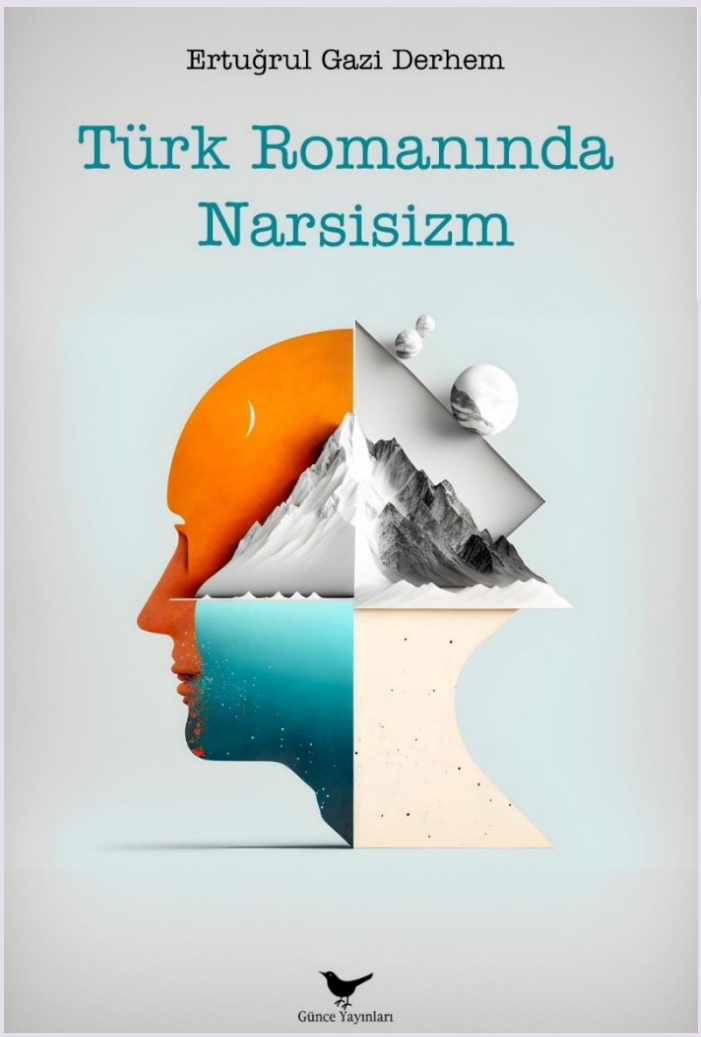
Wolf, Michaela, & Fukari, Alexandra. (Eds.) (2007). *Constructing a Sociology of Translation*. John Benjamins Publishing.

R


27 MAYIS DARBESİ'NİN
TÜRK ROMANINA YANSIMASI
DR. FERHAT ÇETİNKAYA
Günce Yayınları


Türk Romanında
Arzunun
Görüngüleri
Ömriye Bayrak
Günce Yayınları


Edebî Metinlere
Edebiyat Kuramları
Ekseninde Bir Bakış
Dr. Yavuz Selim Uğurlu
Günce Yayınları


Ertuğrul Gazi Derhem
Türk Romanında
Narsisizm
Günce Yayınları