

# Machine Translation-Friendly Language for Improved Machine Translation Output in Academic Writing

Arif Bakla

0000-0001-5412-4330

Anakara Yıldırım Beyazıt University, Faculty of Humanities and Social Sciences, Division of Translation and Interpretation, Ankara, Türkiye ror.org/05ryemn72 arifbakla@aybu.edu.tr

Ahmet Çekiç

0000-0002-7292-3301 Sivas Cumhuriyet University, Faculty of Education, Division of English Language Teaching, Sivas, Türkiye ror.org/04f81fm77 acekic@cumhuriyet.edu.tr

# Abstract

Academic writing aims to disseminate scholarly knowledge to a target audience in concise language. During this process, translators sometimes translate academic texts into foreign languages, so that the output of scholarly research can be read by speakers of different languages. In such cases, translators are to translate the source text into the target language in a clear and comprehensible language. This paper discusses how the authors of scholarly texts can structure their writing properly, so that they can help translators compose clear and concise translation output. Although the knowledge and skills of the translator are a key consideration in ensuring the quality of translated texts, the source text itself is highly influential in the quality of the output. Furthermore, although they might partially differ from one language to another, academic writing has some well-established conventions. Therefore, translators might experience some problems when they translate scholarly texts. This study aims to pinpoint such issues and to offer viable strategies that could be used for solving them before the translation process. Commonly encountered problems that impair the quality of the translation output are as follows: wordiness, ambiguous expressions, failing to consider the discourse structure of the target language while organising the source text, using conjunctions carelessly, using the passive voice unnecessarily, using indirect language to express ideas, using synonymous words sequentially, typos, and other similar problems. Though experience in academic translation can help solve some of these issues, it might not be possible to produce good translation output if the source text is poorly written. Success in academic translation can be achieved through collaboration between the author of an academic text and its translator. For this reason, it is necessary to raise the awareness of writers and translators on this issue.

# Keywords

Translation and Interpretation; Academic Translation; Academic Writing; Social Sciences; Translation Friendly Language; Strategies; Readability

# Highlights

- Machine translation can help academics who are less proficient in English compose scholarly works.
- For machine translation to be helpful in academic writing, the source text should be preedited carefully.
- Authors of scholarly publications should be familiar with the conventions of academic writing both in the source language and target language.
- Authors of scholarly publications should possess machine translation literacy to derive substantial benefits from machine translation.
- The latest developments in machine translation have made it quite promising for scholarly writing.

#### Citation

Bakla, Arif – Çekiç, Ahmet. "Machine Translation-Friendly Language for Improved Machine Translation Output in Academic Writing". *Eskiyeni* 54 (September 2024), 1269-1287. https://doi.org/10.37697/eskiyeni.1472519

#### Article Information

Ethical Statement	* This article is the revised and developed ver entitled "Translation-Friendly Academic Wri Strategies", orally delivered and published in the b Symposium of Social Sciences.	ting in Social Sciences: Practical		
Date of submission	14 April 2024			
Date of acceptance	16 September 2024			
Date of publication	30 September 2024			
Reviewers	Two Internal & Two External			
Review	Double-blind			
Plagiarism checks	Yes - intihal.net			
Conflicts of Interest	The Author(s) declare(s) that there is no conflict of interest			
Grant Support	No funds, grants, or other support was received.			
Complaints	eskiyenidergi@gmail.com			
Author Contributions	Conceptualization (CRediT 1)	Author-1 (%60) - Author-2 (%40)		
	Data Curation (CRediT 2)	Author-1 (%60) - Author-2 (%40)		
	Investigation - Analysis - Validation (CRediT 3-4-6-11)	Author-1 (%50) - Author-2 (%50)		
	Writing (CRediT 12-13)	Author-1 (%50) - Author-2 (%50)		
	Writing – Review & Editing (CRediT 14)	Author-1 (%60) - Author-2 (%40)		
S. Development Goals	-			
License	CC BY-NC 4.0			



## Akademik Yazmada Daha İyi Makine Çevirisi Çıktısı için Çeviri Dostu Akademik Dil

Arif Bakla

0000-0001-5412-4330 Ankara Yıldırım Beyazıt Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, Mütercim ve Tercümanlık Anabilim Dalı, Ankara, Türkiye ror.org/05ryemn72 arifbakla@aybu.edu.tr

Ahmet Çekiç

0000-0002-7292-3301 Sivas Cumhuriyet Üniversitesi, Eğitim Fakültesi, İngiliz Dili Eğitimi Anabilim Dalı, Ankara, Türkiye ror.org/04f81fm77 acekic@cumhuriyet.edu.tr

# Öz

Akademik yazım, bilimsel bilginin özlü bir dille hedef kitleye yayılmasını amaclamaktadır. Bu süreçte çevirmenler bazen akademik metinleri yabancı dillere çevirirler. Böylece akademik araştırmaların çıktıları farklı dilleri konuşanlar tarafından okunabilir. Bu tür durumlarda çevirmenler kaynak metni açık ve anlaşılır bir dille hedef dile çevirmelidir. Bu makale cevirmenlerin acık ve öz ceviri cıktıları olusturmalarına yardımcı olabilmeleri icin bilimsel metin yazarlarının yazılarını düzgün bir sekilde nasıl yapılandırabileceklerini ele almaktadır. Çevirmenin bilgi ve becerileri çeviri metinlerin kalitesinin sağlanmasında önemli bir husus olmakla birlikte, bizzat kaynak metin de çevirinin kalitesi üzerinde oldukça etkilidir. Ayrıca, bir dilden diğerine kısmen farklılık gösterse de akademik yazmanın bazı yerleşik kuralları vardır. Bu nedenle, çevirmenler akademik metinleri çevirirken bazı sorunlarla karşılaşabilirler. Bu çalışma bu tür sorunları tespit etmeyi ve çeviri sürecinden önce bunları çözmek için kullanılabilecek uygulanabilir stratejiler ortaya koymayı amaçlamaktadır. Yaygın olarak karşılaşılan ve çeviri çıktısının kalitesini düşüren sorunlar şunlardır: lafı dolandırmak, muğlak ifadeler, kaynak metni düzenlerken erek dilin söylem yapısını dikkate almamak, bağlaçları gelişigüzel kullanmak, gereksiz yere edilgen çatı kullanmak, fikirleri ifade etmek için dolaylı bir dil kullanmak, eşanlamlı kelimeleri ardışık olarak kullanmak, yazım hataları vb. Akademik çeviride deneyim sahibi olmak bu sorunların bazılarının çözümüne yardımcı olsa da kaynak metin kötü yazılmışsa iyi bir çeviri çıktışı üretmek mümkün olmayabilir. Akademik çeviride başarı akademik metnin yazarı ve çevirmeni arasındaki iş birliği ile sağlanabilir. Bu nedenle yazarların ve çevirmenlerin bu konuda bilinçlendirilmesi gerekmektedir.

### Anahtar Kelimeler

Mütercim ve Tercümanlık; Akademik Çeviri; Akademik Yazma; Sosyal Bilimler; Çeviri Dostu Dil, Stratejiler; Okunabilirlik

# Öne Çıkanlar

- Makine çevirisi İngiliz dilinde pek yetkin olmayan akademisyenlerin akademik çalışmaları yazmasına yardımcı olabilir.
- Makine çevirisinin akademik yazmada işe yaraması için kaynak metnin dikkatli bir şekilde ön biçimlemeye tabi tutulması gerekir.
- Bilimsel yayınların yazarları hem kaynak dilde hem de hedef dilde akademik yazım kurallarına aşina olmalıdır.
- Bilimsel yayınların yazarları makine çevirisinden ciddi anlamda faydalanabilmek için makine çevirisi okuryazarlığına sahip olmalıdır.
- Makine çevirisindeki son gelişmeler makine çevirisini akademik yazma noktasında son derece umut verici hale getirmiştir.

## Atıf Bilgisi

Bakla, Arif – Çekiç, Ahmet. "Akademik Yazmada Daha İyi Makine Çevirisi Çıktısı için Çeviri Dostu Akademik Dil". *Eskiyeni* 54 (Eylül 2024), 1269-1287. https://doi.org/10.37697/eskiyeni.1472519

### Makale Bilgileri

Etik Beyan	* Bu makale, 2. Türkiye Sosyal Bilimler Sempozyumu'nda sözlü olarak sunulan ancak			
	tam metni yayımlanmayan "Sosyal Bilim	nlerde Çeviri Dostu A	kademik Yazma:	
	Uygulamaya Dönük Stratejiler" adlı tebliğin iç	eriği geliştirilerek ve kısı	nen değiştirilerek	
	üretilmiş hâlidir.		0,1	
Geliş Tarihi	14 Nisan 2024			
Kabul Tarihi	16 Eylül 2024			
Yayım Tarihi	30 Eylül 2024			
Hakem Sayısı	İki İç Hakem - İki Dış Hakem			
Değerlendirme	Çift Taraflı Kör Hakemlik			
Etik Beyan	Bu çalışmanın hazırlanma sürecinde etik ilkel	ere uyulmuştur.		
Benzerlik Taraması	Yapıldı - intihal.net			
Çıkar Çatışması	Çıkar çatışması beyan edilmemiştir.			
Finansman	Herhangi bir fon, hibe veya başka bir destek a	lınmamıştır.		
Etik Bildirim	eskiyenidergi@gmail.com			
Yazar Katkıları	Araştırmanın Tasarımı (CRediT 1)	Yazar-1 (%60) -	Yazar-2 (%40)	
	Veri Toplanması (CRediT 2)	Yazar-1 (%60) -	Yazar-2 (%40)	
	Araştırma - Veri Analizi- Doğrulama (CRediT 3-4-6-11) Yazar-1 (%50) - Yazar-2 (%50)			
	Makalenin Yazımı (CRediT 12-13)	Yazar-1 (%50) -	Yazar-2 (%50)	
	Metnin Tashihi ve Geliştirilmesi (CRediT 14)	Yazar-1 (%60) -	Yazar-2 (%40)	
S. Kalkınma Amaçları	-			
Lisans	CC BY-NC 4.0			

# Introduction

English is an international language for academic communication as well. As Bowker and Ciro state 94% of the users of English are non-native and machine translation (MT) can be of help for them as they need support in writing in English (Bowker - Ciro, 2019). Speakers of English as an Additional Language (EAL) wish to publish their research in English in journals indexed in prestigious indexes such as SCI, SSCI, AHCI to meet academic promotion criteria. They also wish to get international recognition and disseminate their findings. It is not always easy to find competent translators who have expertise in the specific field of study. Although some publishers (e.g., Elsevier, Taylor & Francis, Wiley) provide translation and language-editing services, Turkish to English choice is not always available and the price for approximately 8000 words ranges between 1185\$ to 1422\$ for translation from Spanish to English and even higher for translation from Chinese to English. Towards the end of 20<sup>th</sup> century, with a shift from rule-based MT to corpus-based statistical MT, where the machine uses a statistical model to decide how certain pieces of source text should be translated based on a large collection of previously translated chunks of language pairs, the quality MT outputs increased to a large extent, and the quality of MT outputs increased even more with the advent of neural MT, which is fortified with artificial neural network that can be 'trained' with machine learning techniques (Bowker, 2020). In this respect, MT can be a solution for language-related problems in getting published.

Using MT in academic writing or preparing manuscripts for MT is a very recent issue in the scholarly circles. O'Brien et al. mark that "MT is still not good enough to produce highquality output for all languages, all types of text, and in all contexts" (O'Brien et al., 2018, 238). Therefore, there is not a conclusive answer to the question of whether MT can help (Bowker – Ciro, 2019; O'Brien et al., 2018). However, investigating and discussing the ways to use MT more effectively is worth the effort. Furthermore, although previous research has focused on Chinese, Korean, English (Miyata – Fujita, 2021), there is no study on writing MT friendly academic Turkish source text to be translated into English. The aim of this literature review study is to reveal the ways scholarly writing can be improved for better MT translations of source texts. Although mechanically it seems to be an easy process of just copying the source text and pasting it in the MT box, using MT for scholarly writing is a complex process, which entails MT literacy (Bowker – Ciro, 2019). With this need in mind, the current study seeks to help academics understand challenges in using MT for academic writing, its limitations, language features that cause problems in MT translation, and thus ultimately write more MT-friendly academic texts.

The critical role of MT-literacy for academics can play in improving the quality of scholarly writing can be better appreciated upon seeing costly, time-consuming and sometimes frustrating methods non-native scholars turn to. Lillis and Curry devised the term "literacy brokers" to refer to any people who directly intervene in the creation of manuscripts such as translators, reviewers, editor or proofreaders (Lillis - Curry, 2010). However, it is mostly a challenging and costly process to find expert translators who can effectively translate academic texts in a variety of academic fields which require mastery of field-specific knowledge of terminology and an awareness of textual characteristics. O'Brien et al. argue that MT can enhance writing scholarly articles, which are then automatically

translated and self-post-edited. This can decrease the cognitive load of academic writing in English and save time and money (O'Brien et al., 2018).

As Bowker underscores, "the fact that MT quality has improved significantly, along with the fact that MT is both relatively easy to access and to use, means that MT has become an increasingly attractive option for scholars who need to write in English as an additional language (EAL)" (Bowker, 2020, 289). The lack of MT literacy can lead to low-quality MT output, thereby decreasing non-native speakers' probability of getting published. Training academics, graduate students or anyone involved in academic writing to have MT literacy will have great academic and economic payoffs for the nations as it can save time, effort and ensure that the voices of non-native speakers are heard as in scientific spheres over the globe. In this respect, MT can ensure social justice by empowering disadvantaged nonnative academic writers.

#### 1. Academic Writing

Academic writing or scholarly communication can refer to the process through which scholars disseminate their findings, which is reviewed for quality and conserved for further use, to a wider community of researchers and others. English has become the language of academic writing, particularly with the expansion of scientific knowledge and multidisciplinary fields of study and with the emergence of new terms coined for newly found concepts, inventions and methods, particularly during the World War II. These two waves of came along with some affordances and drawbacks for scholarly communication (Bowker - Ciro, 2019). Bennett (2013, 2014a, 2015/cited in O'Brien et al., 2018) argue that English is unquestionably the lingua franca of academic communication and gatekeepers of publication i.e. editors and reviewers disfavour non-native speakers. Academic writing entails careful planning, organization of ideas, sticking to conventions of scholarly discourse and rhetorical conventions in academic genre, making careful word choices, use of discipline-specific terminology as well as avoiding ambiguity, redundancy and achieving clarity and brevity. Hence, as Day (1998, 227) points out, "the writing of an accurate, understandable paper is just as important as the research itself". In a recent study, Kuscu Özbudak (2023) revealed that the main reason why some Turkish academics utilize MT for self-translation of their own academic texts is that they feel inadequate in writing academic texts directly in English. This indicates that academics rely on MT even if they have certain level proficiency in the target language and underscores the importance of developing MT competency for increasing the quality and quantity of academic text publishing of Turkish academics for international readership.

Academic writing follows a set of somewhat internationally established rhetorical conventions and citations for formatting and citation, all of which can vary slightly across journals, publishing bodies and authorities. For rhetorical conventions of academic writing in English, you can refer to several guidelines (e.g., Galvan – Pyrczak, 2024) and to the Publication Manual of the American Psychological Association (2020, 7<sup>th</sup> Ed.). Although organizational features, layout and citation conventions are indispensable for and inherent to academic writing, they are not directly related to facilitating the quality of MT output. Thus, in line with the purpose of this paper, we focus more on linguistic features of academic writing, which can help scholars to generate more MT-friendly source texts and in turn more publishable MT output.

Academic writing can be regarded as a part of English for Academic Purposes (EAP), which emerged in the mid-1970s as an extension of English for Specific Purposes (ESP). As Hyland (2014, 2) puts it EAP "...seeks to understand and engage learners in a critical understanding of the increasingly varied contexts and practices of academic communication." This indicates that scholarly communication practices are not constant along all academic fields but echo various ways of knowledge construction and establishment of field-specific conventions as the hegemony of English in scholarly communication has swept away culture-specific academic writing conventions in many local contexts all over the world (Hyland, 2014). In line with this, academic writers should be aware of and observe genre-specific rhetorical conventions which van vary along academic discipline. One of the best ways to do this is to carefully examine submission guidelines and sample articles in prestigious journals in their specific fields and then follow format and rhetoric of sample articles. Therefore, the likelihood of getting an MT output for an academic manuscript published does not solely rely on refining lexical choice and grammatical patterns but also discourse features of source text.

## 2. Machine Translation

In his review of the history of MT, Korkmaz (2019) states that there is not an agreed upon history of the development of MT, even if has a relatively short history. In the twentieth century, particularly during the post-war period, translators and computer programmers dreamt of fully automatising translation processes, which was conceptualised by Hutchins and Somers (1992) as human-aided MT (versus machine-aided human translation). The earliest reflection of this mindset appeared in the late 1940s when Warren Weaver published a memorandum, and this memorandum encouraged research into machine translation over the next decade (Bowker – Ciro, 2019). During the 1950s and early 1960s, there was a burgeoning interest in translating texts by using computer programs which initially used a limited number of grammatical rules and bilingual word lists. These early efforts proved partially successful.

In 1966, however, a report issued by the Automatic Language Processing Advisory Committee, which was established two years earlier, delivered a major blow to the MT industry as it considered full automatization a hopeless dream that supposedly led the efforts to go down the drain. Following the issuing of this report, major funding bodies (e.g., the US government or research agencies) began to cut the financial support they provided to research projects. The ALPAC report, while considering fully automatized MT unattainable, suggested that ways of speeding up human translation should be investigated (ALPAC, 1966), which later gave way to the development of computer-assisted translation tools (machine-aided human translation) approximately two decades later. Despite the impact of this ill-fated report (on the near future of MT), research projects were not totally abandoned; several successful projects were underway, particularly in Canada and the USSR (e.g., the MÉTÉO project was specifically developed in 1981 and remained in use for 20 years as a sublanguage used for translating weather forecasts from English into French and provided 97% accuracy after light post-editing) (Rothwell et al., 2023).

While the earlier types of MT were rule-based systems (in the early 1950s), the 1980s and 1990s brought about new types of MT (i.e., example-based and statistical machine

translation, respectively). The new types of MT were considered highly successful, but the introduction of neural MT was a groundbreaking development in the mid-2010s as it significantly improved the quality of MT output. Achieved with the help of neural MT, the major update provided by Google Translate in 2016 made MT output highly useful, not only for daily-life purposes but also for more technical ones. Finally, the use of artificial intelligence, machine learning and large language models in MT has yet been another breakthrough in MT. Today, various technologies are used in combination to improve MT quality, such as neural MT, artificial intelligence and large language models and machine learning. With such new technologies, the translation industry has made a giant leap towards full automatization of translation tasks. In line with such advancements in translation technology, Sırkıntı and Sırkıntı (2024) pointed out that human translators have assumed the new role of post-editors rather than translators. Likewise, Çetiner and İşisağ (2019) reported that undergraduate translation students benefited from MT post-editing (PE) training, which led not only to significantly better performance but also to more positive attitudes towards MT PE. Current translation technologies make it possible to use the output of MT with light post-editing in the case of MT for information, while it might be necessary to spare some effort to use MT for dissemination purposes. Globally considered, recent developments in MT allow academic and technical writers to use MT engines to eliminate language barriers against scholarly publishing. In this regard, software aimed at facilitating MT for specific language pairs has also been developed. In a recent study have Sel et al., (2021) have created a parallel corpora for better Turkish-English academic translation. Bowker and Ciro (2019) claim that by using MT-friendly language (by carefully pre-editing scholarly source texts), scholarly writers with poorer English skills can use MT to compose academic texts in English and get published. To optimize the performance of MT engines and output, however, they should not only know the fundamentals of academic writing but also be equipped with strategies for editing source texts to make them entirely appropriate for MT.

# 3. Pre-editing in Machine Translation for Academic Writing

Pre-editing can be defined as "the practise of checking and revising the source text to make it more appropriate for MT" (Bakla, 2023, 227–228). According to Kokanova et al. (2022), although the nature of MT continuously changes, various pre-editing strategies could help improve the quality of MT output and there are some mistakes that can be eliminated through pre-editing. MT can be of great help for academics who want to write a scholarly paper on their own on the condition that the source text is edited and revised carefully before MT. This is particularly because the more the source text is written accurately and without ambiguity, the higher the quality of the MT output gets.

Wallwork (2023) suggests various revisions to improve the quality of the source text in an MT task, including but not limited to, writing clearly (by avoiding ambiguity), conciseness of writing, using shorter elements (e.g., reducing paragraph or sentence length), checking the word order of the source text so as to make it suitable for the target language, using keywords consistently across the whole text, avoiding vagueness of expression, checking punctuation/spelling, using pronouns carefully. Wallwork also recommends that the target equivalents of some keywords should be directly inserted into the source text to ensure that the correct target equivalents are used (155–156). Similar suggestions are made by Zhivotova et al. (2020), who group tips given for pre-editing scientific or technical texts under three categories: (a) vocabulary (e.g., reducing the number of abbreviations, replacing complex words and phrases with simpler alternatives, avoiding informal language and metaphorical expressions, using terminology consistently, avoiding words with context-sensitive meaning), (b) grammatical structures (e.g., avoiding vague sentences, using simple structures, dividing longer sentences up into shorter ones, eliminating too much parenthetical information and using grammatically correct sentences); and (c) overall text structure (e.g., checking the text for spelling and punctuation errors, checking the text for undue line breaks, self-checking the readability of the text by reading it). Similar strategies are also suggested by others (e.g., Kokanova et al., 2022).

Such lists of tips or suggestions indicate that there are quite a few points that should be pre-edited in a source text if it is intended for MT. The pre-editing process becomes much more important if the output text is to be published (MT for dissemination). Therefore, the following sections address major issues to be addressed when pre-editing a source text is intended for MT for dissemination.

# 3.1 Clarity and Brevity

Clarity and brevity are perhaps the two most important features of academic genre as it mainly aims to report the results or findings of research, which can be explicit enough for readers to be referred to or replicated. As academic writers spend too much time with their own data and findings, they can become insensitive to the issue of clarity in their writing. They are the ones who have the most detailed information about their research findings and who are held responsible for conveying their research results clearly. Clarity and brevity are also important to note that readers are not all native speakers of English. In addition to discourse characteristic of academic writing, clarity and brevity of the source text will enhance the quality of MT output.

Academic writers should avoid abbreviations and acronyms for the sake of brevity. This is because MT cannot properly process them, and their meaning can vary across different academic disciplines. Instead, it is wise to spell them out or use abbreviations and acronyms together with their spelled-out versions in parentheses.

Authors should also avoid forming sentences that are too complex or too long (i.e. sentences more than 30 words). The source text should be composed of plainly structured sentences to enhance the quality of MT output. On the other hand, sentences should not be too short because MT can have difficulty in translating them as such sentences might fail to provide adequate contextual information for effective translation.

Another strategy to be employed to achieve clarity is to use pronouns sparingly in the source text. MT can use incorrect gender for the pronoun (Translation Centre for the Bodies of the European Union, 2021). For example, in Turkish, to refer to third person singular we use the pronoun "O" regardless of gender. MT output of a Turkish source text can include inappropriate gender references in English. Pre-editing source text or post-editing of the target text can be a solution.

Another problem related to clarity is sentences with implied subjects, which are possible and common in Turkish texts, but this can lead to incorrect MT translation in English. In

Turkish, we can write a sentence without a subject by assuming that readers can refer back to the previous sentence(s) to detect the subject for the sentence, but it is not possible for MT to do this. Furthermore, this can lead to ambiguity, even for human translators. Therefore, the source text in Turkish should be written as clearly as possible by avoiding sentences with an implied subject. Another issue academic writers should avoid is line breaks in titles, in headings, in table or figure captions, as well as in body paragraphs. As MT can process them as two pieces, the output can be incorrect.

The source text should also be checked against typos (spelling errors) and dictos (spelling errors made when dictating in a word processor). Dictos are often more problematic as they are not spelling errors but involve a correctly spelled but unintended word as the speech-to-text tool misunderstands writes a similar sounding word (e.g., You say "too" but the speech-to-text tool spells it as "two").

Clarity can also be ensured and enhanced with proper use of punctuation. It is essential to use punctuation marks (e.g., commas, colons, semicolons, hyphens etc.) appropriately to facilitate MT outputs. Academic writers should be aware of the meaning-changing roles of punctuation marks and make sure that they have used them accurately in the source text. Even if there are some differences between languages in terms of the function of punctuation marks, MT tools are making improvements in processing them. To have a better understanding of meaning-changing role of punctuation marks please consider Example 1 (Trask, 1997, 3). In this example, where you place the punctuation mark leads to subtle changes in meaning.

- (1a): We had one problem: only Adel knew we faced bankruptcy.
- (1b): We had one problem only: Adel knew we faced bankruptcy.
- (1c): We had one problem only, Adel knew: we faced bankruptcy.
- (1d): We had one problem only Adel knew we faced: bankruptcy.

Clarity in the source text can also be ensured by avoiding ambiguity. Ambiguity, which refers to the case of language with two or more possible meanings, can result from structural and lexical elements. Lexical ambiguity refers to a case in which a word or phrase has more than one meaning. For example, the word "bank" English-Turkish true cognate can refer to a place where people put and/or withdraw money, can have different meanings in different contexts in both languages. MT tools are trained to be more context-sensitive in their translations, but academic writers should also check words which can led to lexical ambiguity (i.e., if words with two or more different meanings are translated consistently throughout the MT output). On the other hand, structural ambiguity refers to a case when a sentence has two or more different meanings because of its word order (e.g., I gave her dog milk. "I gave her [dog milk]" or "I gave [her dog] milk"). The source text should ensure that their source texts are free from any type of ambiguity (lexical or structural ambiguity) as much as possible to obtain quality MT outputs.

#### 3.2 Avoiding Figurative Language

Humans can easily understand nuances in meaning and appreciate figurative meaning in spoken and written language. However, machines are different in this respect. Although they can easily understand literal meaning, nonliteral language, particularly idioms, pose a challenge for MT tools. This was empirically supported by Kokanova et al. (2022), who carried out a study which compared six news texts by using Google Translate, Amazon Translate and PROMPT Professional Neural. Kokanova et al. concluded that current (neural) MT systems fail to understand some idiomatic expressions, collocations and phrasal verbs, and they cannot successfully deal with metaphorical meaning. Adopting a positive outlook, Schocket (2018) notes that neural MT is gradually becoming more skilful at recognising and translating idiomatic language, yet we think there is still a long way to go before idiomaticity can entirely be removed from the agenda of problematic issues in MT. Thus, even if figurative language is not a common feature in academic genre, academics should be cautious against any figurative language in the pre-editing and post-editing process of academic texts so that they can avoid possibly incorrect translations.

# 3.3 Avoiding Top-Heavy Sentences

A sentence is said to be a top-heavy one if its subject is too long (Bakla, 2021). When we begin to read a sentence, we initially tend to recognize the subject and immediately look for the predicate. This is because the subject presents the topic, while the predicate provides the comment about it. The longer it takes for a human to receive the comment, the more challenging it gets for him/her to comprehend the message conveyed by the sentence. This equally goes for machines as they have difficulty understanding longer sentences or sentences with longer subjects. MT engines are more likely to make mistakes in translation in longer units of meaning than in shorter ones. Consider Example 3. In 3a and 3c, the subject of the sentence is unnecessarily lengthy. It is possible to rephrase this sentence to increase its readability. The main idea is provided right at the beginning in 3b and 3d.

(3a) What you are trying to say about the basic aspects of this problem which has not been solved by experts investigating it for more than three decades is highly critical.

(3b) You have critical ideas about the basic aspects of this problem which has not been solved by experts investigating it for more than three decades.

(3c) That Greek philosophy contributed to Christian theology and had a deep impact on it is stressed by most theologians.

(3d) Most theologians stress that Greek philosophy contributed to Christian theology and had a deep impact on it.

As Bakla (2021) suggests, it is sometimes wise to begin a sentence by using the false subject "it" (4b), rather than using a lengthy gerundive phrase (4a) or moving the long gerundive phrase (4c) to the end of the sentence (4d). It is useful to give the main message as soon as possible in a sentence. It is the very reason why we write a topic sentence at the beginning of a paragraph to provide the gist of it.

(4a) Considering your family budget when deciding to buy an expensive household item, particularly in times of economic crisis, is important.

(4b) It is important that you consider your family budget when deciding to buy an expensive household item, particularly in times of economic crisis.

(4c) Identifying the boundaries of religious fundamentalism by clearly describing what makes someone a religious fundamentalist is one of the topics in books that address this issue.

(4d) One of the popular topics in current religious books is identifying the boundaries of religious fundamentalism by clearly describing what makes someone a religious fundamentalist.

It is also very common for Turkish authors to write sentences that begin with the expression "... gerçeği" (the fact that...), and this leads to top-heavy sentences. This is because the nominalizer "that" turns a full sentence into a noun, which then needs to be followed by a verb phrase (Example 5). This naturally results in a sentence with an unnecessarily long subject (5b). However, using the active voice in the Turkish sentence can solve this problem (5c).

(5a) Turkish: Osmanlı Devleti'nin İslamı üç kıtada yaymaya çalıştığı ve gittiği yerlere İslam kültür ve sanatını götürdüğü gerçeği göz ardı edilmemelidir.

(5b) English Translation: The fact that the Ottoman Empire tried to spread Islam in three continents and brought Islamic culture and art to the places it visited should not be ignored. (DeepL, June 2024, no post-editing)

(5c) Turkish: Osmanlı Devleti'nin İslamı üç kıtada yaymaya çalıştığı ve gittiği yerlere İslam kültür ve sanatını götürdüğü gerçeğini göz ardı etmemeliyiz.

(5d) English Translation: We should not ignore the fact that the Ottoman Empire tried to spread Islam in three continents and took Islamic culture and art wherever it went. (DeepL, June 2024, no post-editing)

### 3.4 Avoiding Noun Plague

Noun plague is a term coined by Wilson Follett to refer to the piling up of nouns to modify other nouns. Having more than two nouns in a row generally lowers the readability of a sentence and this can also be a problem in MT output; it is also called "noun-upon-noun syndrome" (Garner, 2016, 635). In Example 6 (adopted from Garner, 2016, 635), ten nouns are connected to form a single noun phrase, which proves difficult to understand, not only for machines but also for humans.

(6a) "Consumers complained to their congressmen about the National Highway Traffic Safety Administration's automobile seat belt 'interlock' rule."

(6b) "Consumers complained to their congressmen about the 'interlock' rule applied to automotive seat belts by the National Highway Traffic Safety Administration."

### 3.5 Checking Terminology

As Montgomery (2004, 1335/cited in Bowker - Ciro, 2019) points out, exponential increase in scientific knowledge and introduction of more precise terms for new inventions, concepts and techniques have brought along some challenges for academic writing such as increasing the need for shared and much more fine-tuned terminology for sharing and disseminating of knowledge, particularly with the ever-increasing need for multidisciplinary studies. Academic writers should use terminology consistently. Normally, using synonyms can be good for achieving variety in text and for increasing readability, but deliberately avoid using synonyms for terminology as it can lead to inconsistencies in MT and may lead to incorrect or imprecise translation of the terminology (Translation Centre for the Bodies of the European Union, 2021), which is particularly important for scholarly communication of research reports. Another point to take into consideration is that academic writers should be cautious against previously accumulated body of incorrect translations of terms. For example, the term "relational survey model" was translated into Turkish as "ilişkisel tarama modeli", which was later translated back to English as "relational screening model". This confusion created by ineffective translation between

language pairs has been so widely adopted by Turkish translators that it has led to emergence of a new pseudo term "relational screening model", which is used only used by only Turkish researchers as a quick scholarly database search can reveal. This means it is not used or recognised by the global scientific community at all. Therefore, academics should be aware of such false translations, which might have been partially established in English academic writing by Turkish researchers. Moreover, it could be a good strategy to insert the target equivalents of some key terms so that MT engines can translate key terms without committing any errors in translation. In this sense, scholarly writers can use domain specific glossaries to identify the target equivalents of key terminology.

A recent development in MT is the incorporation of user-generated MT glossaries into MT engines to improve the quality of translation by ensuring consistency of terminology. While termbases are used by linguists to be able to translate terms accurately and consistently, usually in computer-assisted translation tools, MT glossaries are used in much the same way by MT engines (Phrase, n.d.). Linguists prepare a list of key terms that appear in the source text and upload a bilingual term list to computer-assisted translation tool or MT software. For instance, the MT tool offered by Phrase (n.d.) allows its users to upload a list of key terms to be used by the MT tool for more consistent and accurate translation. Depending on the type of subscription, Phrase users can upload up to 10 glossaries with 50,000 entries. Similarly, the paid versions of DeepL (n.d.) allows its users to create their own glossaries in CSV (comma separated values) format and use it in MT for better outputs. This MT feature is gradually getting more advanced with better results. Academic writers are recommended to support MT engines by uploading their own MT glossaries; they have the chance to check the MT translation after the uploaded glossaries are used by the MT engine.

# 3.6 Considering the Features of Written Discourse

Conventions of academic writing could vary across languages that belong to different language families. For instance, the use of passive voice is firmly discouraged in academic writing in English. In fact, the encouraging or discouraging of passive voice in English is deeply rooted in philosophical discussions about how research should be done and how it should be reported. On the one hand, the proponents of the positivist tradition encourage the use of passive voice on the grounds that the presence of the researcher should be minimised to ensure the objectivity of the research undertaken. On the other hand, more recent approaches which emerged as a reaction to the well-established positivist tradition (e.g., constructivist, interpretivist or post-modernist perspectives) highlight the role of the author in planning, conducting and reporting scholarly research (particularly in qualitative studies). They even suggest that authors should explain how their outlooks shape the processes of conducting and reporting the results. In this respect, as qualitative research is grounded in such perspectives, qualitative researchers often use of personal pronouns like "I" or "we" while reporting their research. However, it should be noted that these perspectives might differ a lot across different languages. Therefore, being aware of such philosophical outlooks and how they are put into practice in different languages could inform writers of scholarly texts about how to avoid pitfalls while composing the source text to be translated into a particular target language.

In Example 7, the voice of the main verb is in passive (7a). In the first version of its English translation (7b), the passive voice is retained, and the subject of the sentence becomes much longer than the original, thereby reducing the readability of the whole sentence. In 7c, however, the active voice was preferred. This move apparently boosts the readability of the translation (7d) as the subject is very short and the verb immediately follows it. In some cases, the use of passive voice in the source text (Turkish) could lead to top-heavy sentences in English. By using the active voice, this problem can easily be solved.

(7a) Turkish: Bu çalışmada inanç ve akıl arasındaki ilişkiye dair temel hususlar kolaylıkla anlaşılabilecek örneklerle ele alınmaktadır. (Passive Voice)

(7b) MT Translation (Version 1): In this study, the basic issues regarding the relationship between faith and reason are discussed with easily understandable examples. (Passive voice) (Translated by DeepL, June 2024, with no post editing).

(7c) Turkish: Bu çalışma inanç ve akıl arasındaki ilişkiye dair temel hususları kolaylıkla anlaşılabilecek örneklerle ele almaktadır. (Active Voice)

(7d) MT Translation (Version 2): This study deals with the basic issues regarding the relationship between faith and reason with easily understandable examples. (Active voice) (Translated by DeepL, June 2024, with no post editing).

There are several characteristic examples of how academic writers adopt different perspectives in terms of using some linguistic elements, besides the active or passive voice debate. For instance, Turkish-L1 academic writers mistakenly use the first-person plural pronoun "we" in Turkish even though only one person writes an article. This is not only because it is uncommon to use "I" in scholarly writing in Turkish but also because the firstperson plural pronoun might sound more polite. However, the use of the first-person plural pronoun is assumed to be an indication of at least two authors who wrote the article. Similarly, native speakers of Turkish are said to be less direct in conveying the main message of a paragraph, unlike English in which the gist of a paragraph is directly provided at the beginning of it in the form of a topic sentence. Therefore, even if a text is being composed in Turkish, the author could form a topic sentence for each paragraph so that the output text follows the rules of academic writing in English. Our experience from the field of academic translation also indicate that Turkish authors tend to use conjunctions without considering meaning relations between clauses and sentences. It might be a good strategy to use more frequently used conjunctions that are not ambiguous. For instance, "cünkü" (because) is clear in its meaning, while "nitekim" (as a matter of fact, thus, indeed, in fact etc.) might be translated differently by machines in different cases.

Still another issue is that the syntactic properties certain conjunctions should be taken into consideration not only in L1 but also in L2. For example, a clause that introduces a reason can stand on its own when it is introduced using "çünkü" (because) in Turkish, whereas a clause introduced using "because" does not express a complete thought, so it needs a main clause in English. Otherwise, it becomes a fragment as seen in Example 8. In this example, the Turkish sentences (8a) are perfectly acceptable, while the English translation includes a fragment (the underlined part in 8b) as a clause introduced using "because" cannot stand on its own in academic English. Therefore, it should be attached to a main clause. When the revised version (8c) is used in MT, DeepL translates the sentences without forming a fragment (8d). (8a) Turkish: Makine çevirisi araçları ile İngilizceye çevrilmek üzere akademik bir metin hazırlayan yazarlar dikkatli olmalıdır. Çünkü insanların aksine makineler deyimsel anlamları yakalamakta ciddi anlamda zorlanırlar.

(8b) English Translation: Authors preparing an academic text to be translated into English by machine translation tools should be careful. <u>Because unlike humans, machines have serious difficulty</u> <u>in capturing idiomatic meanings.</u> (Translated by DeepL, April 2024, with no post editing).

(8c) Revised Turkish version: İnsanların aksine makinelerin deyimsel anlamları yakalamakta ciddi anlamda zorlandıklarından makine çevirisi araçları ile İngilizceye çevrilmek üzere akademik bir metin hazırlayan yazarlar dikkatli olmalıdır.

(8d) Authors preparing an academic text to be translated into English using machine translation tools should be cautious, as machines, unlike humans, have considerable difficulty in capturing idiomatic meanings. (Translated by DeepL, April 2024, with no post editing).

Another example is the use of the future tense in the introductory sections of a paper (e.g., when the author outlines the structure of a paper in the abstract, introduction or literature review). Consider the Turkish sentence in Example 9. It is not uncommon to see Turkish authors compose a sentence like the one in Example 9a (the main verb is in the passive voice and in the future tense although the study was carried out in the past).

(9a) Turkish: Bu çalışmada akademik yazmada makine çevirisi kullanılarak çevrilecek kaynak metnin yazılmasına dair hususlar ele alınacaktır.

(9b) MT Translation: In this study, the issues related to the writing of the source text to be translated using machine translation in academic writing will be discussed. (Translated by DeepL, April 2024, with no post editing).

(9c) MT Translation: In this study, the issues related to the writing of the source text to be translated using machine translation in academic writing are discussed. (Translated by DeepL, April 2024, with no post editing).

Normally, the introduction of a study report written in English provides an outline of it in the present tense, and it describes the procedures used in the study (e.g., data collection and analysis) in the past tense. However, when the source text (in Turkish) is translated into English, the output includes the future tense as the main verb of the sentence due to the source text (9b). The use of the future tense can confuse the reader, leading them to mistakenly think that the study has not been carried out yet. In short, it is invaluable to conform to the well-established conventions of academic writing in the target language when preparing the source text. In addition, the type of variety of English language, i.e. American English or British English or any other variety, favoured by the editor should also be taken into consideration. As a matter of fact, a recent study (Durmus – Yaman, 2024) revealed that Google translate provided American English as compared to British English. Thus, researchers can check which varieties of English MT tools favour and match the MT tool and the preferred English variety as requested by the target journal or publisher.

### **Concluding Remarks**

Since its earliest times, the overriding aim of MT has been to eliminate or minimize human involvement in the translation process. In the last few decades, there have been significant contributions to the realisation of this aim, particularly with the help of statistical and neural MT and the use of artificial intelligence. Thanks to these developments, MT can be used for translating academic texts into English for dissemination

purposes. Although the use of MT in academic writing poses some challenges, it seems possible that it can be used to compose scholarly texts, provided that the source text is carefully composed and pre-edited for successful translation. Academics with poor English skills can use the pre-editing checklist provided in the appendix, which is a more succinct presentation of what is suggested throughout this paper. Besides, they have various options at their disposal to publish MT output following additional revision and editing (postediting).

Globally considered, such uses of MT contribute to the aim of eliminating language barrier against the dissemination of scientific knowledge. Although this study suggests that the pre-editing that promotes MT-friendly language could significantly improve the quality of MT for academic writing purposes, it might be unclear, as Miyata and Fujita note (2021), how pre-editing works with black-box NMT systems. Therefore, prospective research could investigate to what extent the pre-editing strategies offered in this study could help improve the quality of MT output in academic writing.

### Appendix | Ek A Suggested Checklist for Pre-Editing of Scholarly Writing for Machine Translation

Clarity	Yes	No
Is the source text free from ambiguity (lexical or structural) that might		
confuse MT engines?		
Have you replaced idioms with literal expressions?		
Is your text free from metaphorical language?		
Is your text free from slang or colloquial language?		
Is the source text free from unnecessary jargon that might be		
incomprehensible to machines?		
re the abbreviated forms (e.g., clipped words, acronyms etc.) explained the		
first time they are mentioned in the text?		
Are pronoun references clear and comprehensible?		
Brevity	Yes	No
Have you made sure that the length of the sentences across the whole text is		
appropriate (Consider dividing longer sentences, particularly if they lead to		
poor readability)?		
Have you eliminated redundant expressions in the source text?		
Is the subject of each sentence as close as possible to its predicate?		
Have you shortened the subject of each sentence to avoid top-heavy		
sentences?		
Terminology		No
Have you used the target equivalents of key terminology in the source text?		
Have you uploaded an MT glossary to the machine translation tool if		
applicable?		
Have you used the same terms across the whole source text instead of using		
their synonyms to ensure consistency of MT?		
Spelling and Punctuation		No
Have checked the source text for possible typos?		
Have checked the source text for possible dictos?		
Have you made sure that the conjunctions are appropriately punctuated?		
Have you checked if punctuation does not lead to ambiguity in meaning?		
Discourse Features		No
Have you avoided using "we" if the study involves only one author?		
Have you mostly used the active voice across the manuscript?		
Have you the past tense to report the methods and procedures used in the		
The you the past tense to report the methods and procedures used in the		
study?		

#### References | Kaynakça

- ALPAC, Automatic Language Processing Advisory Committee. Language and machines: Computers in translation and linguistics. Washington: National Academy of Sciences, National Research Council, 1966.
- APA, American Psychological Association. *Publication manual of the American psychological association*. Washington: American Psychological Association, 7th Edition, 2020.
- Bakla, Arif. Putting pen to paper: academic paragraph writing. Ankara. Hemdem Akademi, 2021.
- Bakla, Arif. "Using ChatGPT for Translation Pedagogy. *ChatGPT in Foreign Language Education and Translation Studies*. Arif Bakla İsmail Çakır. 224–239. Ankara: Anı Yayıncılık, 2023.
- Bowker, Lynne. "Chinese speakers' use of machine translation as an aid for scholarly writing in English: a review of the literature and a report on a pilot workshop on machine translation literacy". *Asia Pacific Translation and Intercultural Studies* 7/3 (2020), 288–298. https://doi.org/10.1080/23306343.2020.1805843
- Bowker, Lynne Ciro, Jairo Buitrago. Machine translation and global research: Towards improved machine translation literacy in the scholarly community. Bingley: Emerald Publishing, 2021.
- Çetiner, Caner İşisağ, Uluç Korkut. "Undergraduate Level Translation Students' Attitudes towards Machine Translation Post-Editing Training". *International Journal of Languages' Education and Teaching* 7/1 (2019), 110-120.
- DeepL. "Uploading glossaries". DeepL. March 10, 2024. Accessed April 10, 2024. https://support.deepl.com/hc/en-us/articles/360021664739-Uploading-glossaries
- Ducar, Cynthia Schocket, Deborah Houk. "Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate", *Foreign Language Annals*, 51/4 (2018), 779– 795. https://doi.org/10.1111/flan.12366
- Durmuş, Rabia İ. Yaman, İsmail. "Tracing the Footprint of World Englishes on Machine Translation Tools". Ondokuz Mayıs Üniversitesi İnsan Bilimleri Dergisi 5/1 (2024), 39-58.
- Galvan, Melisa C. Pyrczak Fred, Writing Empirical Research Reports: A Basic Guide for Students of the Social and Behavioral Sciences. New York: Routledge, 2024.
- Garner, Bryan Andrew. Garner's modern English usage. New York: Oxford University Press, 2016.
- Hutchins, John William Somers, Harold Lambert. An Introduction to Machine Translation. London: Academic Press, 1992.
- Hyland, Ken. English for Academic Purposes: An Advanced Resource Book. United Kingdom: Routledge, 2014.
- Kokanova, Elena S. et al. "Pre-Editing English News Texts for Machine Translation into Russian". Language Studies and Modern Humanities 4/1 (2022), 25–30. https://www.doi.org/10.33910/2686-830X-2022-4-1-25-30
- Korkmaz, İnönü. "Makine Çevirisinin Tarihçesi". Uluslararası Sosyal ve Beşeri Bilimler Araştırma Dergisi 6/32 (2019), 155-166.
- Kuşçu Özbudak, Seda. "Linguistic reasons and professional concerns behind Turkish academicians' self-translation of academic texts". *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi* 36 (2023), 1561-1575. https://doi.org/10.29000/rumelide.1372497
- Lillis, Theresa Marry Curry, Marry Jane. Academic writing in a global context: The politics and practices of publishing in English. New York, NY: Routledge, 2010.
- Miyata, Rei Fujita, Atsushi. "Understanding pre-editing for black-box neural machine translation". Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume The 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 1539-1550. Stroudsburg, PA: Association for Computational Linguistics (ACL), 2021. https://doi.org/10.18653/v1/2021.eacl-main.132

- O'Brien, Sharon et al. "Machine translation and self-post-editing for academic writing support: Quality explorations". *Translation quality assessment: From principles to practice*, Joss Murkens et al. 1/237-262. Cham: Springer, 2018. https://doi.org/10.1007/978-3-319-91241-7\_11
- Phrase. "MT Glossaries". Phrase. March 10, 2024. Accessed April 10, 2024. https://support.phrase.com/hc/en-us/articles/5709675486876-MT-Glossaries-TMS
- Rothwell, Andrew et al. *Translation tools and Technologies*. Oxon: Routledge, 2023. https://doi.org/10.4324/9781003160793-1
- Sel, İlhami et al. "Creating a Parallel Corpora for Turkish-English Academic Translations. Computer Science," IDAP-2021: 5th International Artificial Intelligence and Data Processing symposium (Special), (2021), 335-340. https://doi.org/10.53070/bbd.990959
- Sırkıntı, Halise Gülmüş Hulisi Alp Sırkıntı. "Translating Creative Texts through Machine Translation: Deepl vs. Google Translate". Navigating Tapestry of Translation Studies in Türkiye 6 (2024), 54-67.
- *Writing for machine translation.* Luxembourg: Translation Centre for the Bodies of the European Union, 2021. https://doi.org/10.2817/191981
- Trask, Robert Lawrence. The penguin guide to punctuation. London: Penguin, 1997.
- Wallwork, Adrian. *English for writing research papers*. New York: Springer, 1st Edition, 2011. https://doi.org/10.1007/978-3-031-31072-0\_1
- Zhivotova, Alena A. et al. "Improving the Quality of Scientific Articles Machine Translation while Writing Original Text". *The 2020 International Multi-Conference on Industrial Engineering and Modern Technologies*. 1783–1786. New York: Curran Associates, Inc., 2020. https://doi.org/10.1109/FarEastCon50210.2020.9271442