asobid
Amasya Üniversitesi Sosyal Bilimler Dergisi

ISSN: 2548-0480 (Print)/(Online) 2602-2567
Sayı/Issue 10 ● Aralık/December 2021
Sayfa/Page: 237-266

Geliş Tarihi / Received: 30.11.2021 *Kabul Tarihi / Accepted:* 20.12.2021

Yıldız, Mehmet (2021). "A Translation Quality Assessment Tool Proposed". *Amasya Üniversitesi Sosyal Bilimler Dergisi (ASOBİD)*. S. 10, s. 237-266.

A TRANSLATION QUALITY ASSESSMENT TOOL PROPOSED

BİR ÇEVİRİ KALİTESİ ÖLÇME ARACI ÖNERİSİ

Dr. Öğr. Üyesi Mehmet YILDIZ*
Çanakkale Onsekiz Mart Üniversitesi
mehmetyildiz@comu.edu.tr

Abstract
Translation quality assessment has always attracted a great deal of scholarly attention, which has resulted in several translation assessment tools/rubrics. Yet very few were observed to incorporate translation solutions as evaluative parameters and, to the author's best knowledge, none was identified to assess the quality of translations into Turkish. These two features help the tool presented herein stand out and make a substantial contribution to the related literature. The tool was originally available in the author's doctoral dissertation (Yildiz, 2016), but an improved version was proposed in this paper. It was built on translation errors and translation solutions. To be able to judge a solution's acceptability, 25 rich points (PACTE, 2009) were identified in the excerpted manual. The rich points are located in the first part of the tool, while the second portion was solely based on the erroneous translation segments. The tool incorporates two types of errors – i.e., mechanical and transfer errors. The paper also proposes a grading table, featuring solution- and error-based grades in exponential increments. The minor, major, and critical errors are penalized with (-2), (-4), and (-8) points, whereas partially acceptable and acceptable solutions are awarded (+2) and (+4) points, respectively. The grading table is accompanied by a

* ORCID: orcid.org/ 0000-0001-9482-4358

rubric to describe how the degrees of errors and solutions can be operationalized, which is believed to promote objectivity and inter-rater reliability. The tool notably contributes to the literature with these two components as well. It is thought to be usable by translation schools, translators associations, and translation companies.
Keywords: Translation quality assessment, assessment tool, rubric, grading, error, solution

Öz

Çeviri kalitesinin ölçülmesi her zaman ilgi gören akademik alanlardan biri olmuştur. Bunun sonucunda da birçok çeviri değerlendirme aracı/yönergesi ortaya çıkmıştır. Ama bunlardan pek azı çeviri çözümlerini bir değerlendirme ölçütü olarak görmüştür ve hiçbiri Türkçeye yapılan çevirileri değerlendirmek için kullanılmamıştır. Bu iki farklılık, bu çalışmada sunulan ölçme aracının öne çıkmasına ve ilgili alanyazına kayda değer bir katkı sağlamasına imkan sunmaktadır. Söz konusu araç, yazarın doktora tezi (Yildiz, 2016) kapsamında üretilmiş ve bu çalışma ile daha da geliştirilmiştir. Değerlendirme aracı hem çeviri hataları hem de çeviri çözümleri üzerine inşa edilmiştir. Çeviri çözümünü bir parametre olacak işlevselleştirmek için alıntılanan kullanım kılavuzunda katılımcılara sorun oluşturması beklenen 25 unsur belirlenmiş ve bu unsurlar ölçme aracının ilk kısmını oluşturacak şekilde çalışmaya dahil edilmiştir. İkinci kısım sadece çeviri hatalarına odaklanmaktadır. Araç, iki hata türünü irdelemektedir: mekanik hatalar ve aktarım hataları. Bu çalışma, ölçme aracı ile birlikte çeviri çözüm ve hatalarına verilen puanların katlanarak arttığı bir puanlama tablosu da sunmaktadır. Küçük, büyük ve kritik hatalara sırasıyla (-2), (-4) ve (-8) puan verilirken kısmen kabul edilebilir ve kabul edilebilir çözümler (+2) ve (+4) puan ile ödüllendirilmiştir. Çalışmada aynı zamanda puanlama tablosunun anlamlandırılmasını kolaylaştıracak bir puanlama yönergesi de önerilmiştir. Bu yönergenin aracın nesnelliğini ve değerlendiriciler arası güvenilirliği geliştireceğine inanılmaktadır. Önerilen ölçme aracı, bu iki bileşeni ile de alana katkı sağlamaktadır. Aracın; çeviri eğitimi veren kurumlar, çevirmen dernekleri ve çeviri şirketleri tarafından kullanılabileceği düşünülmektedir.
Anahtar Sözcükler: Çeviri kalitesinin ölçülmesi, ölçme aracı, yönerge, puanlama, hata, çözüm

Subjective Translation Quality (Assessment)

Translation quality has been among the major concerns of commissioners/clients, translators, and assessors. These three agents of the translation market have their own definitions of translation quality. For commissioners, a high-quality translated text is supposed to have achieved the goal(s) set by them. For translators, a high-quality translation, from a functionalist perspective, is a text they produce by the *best* means at their disposal and hope to meet the commissioned needs. Lastly, for assessors, it is the translated text which manages to satisfy their expectations based on several "accuracy and fluency" criteria (Koby et al., 2014: 414). Since translation is a process governed by subjective decisions (Reiss, 2014: 91) and judges of translation quality, be they laymen or professionals, deploy different assessment criteria, translation quality tends to be subjective. But while a layperson can label a translated text simply as 'good' or 'bad', it can be claimed that professional assessors are expected to adopt an objective viewpoint and to employ objectivity-promoting rubrics to refrain from "anecdotal" (House, 1997: 3), "impression-based" (Martinez-Melis and Albir, 2001: 283), and macroanalytical evaluation (Yildiz, 2020: 575), which does not suggest that they necessarily and always do so. Therefore, this paper proposes an assessment tool that assessors of translation (quality) can use to promote the objectivity of their evaluative processes.

For researchers, subjectivity is inevitable and inherent in translation quality assessment (TQA). According to Hutchins and Somers, "what counts as a 'good' translation [...] is an extremely difficult concept to define precisely. Much depends on the particular circumstances in which it is made and the particular recipient for whom it is intended. Fidelity, accuracy, intelligibility, appropriate style and register are all criteria which can be applied, but they remain subjective judgements" (Hutchins and Somers, 1992: 2). Likewise, Samuelsson-Brown too questions the yardsticks of quality assessment and claims that the issue originates from intangibility of translation. He claims that "the

quality of a tangible object such as a metal bolt can be checked against a well-defined standard and such an assessment can be fairly objective. A translation is, however, an intangible entity, quality can be very subjective in many cases [...]. There are certain guidelines that can be applied [...]. (Samuelsson-Brown, 2010: 104). For Kahl (1991), what makes a 'good' or 'bad' translation is the criteria of the commissioner. Kahl (1991: 120) expresses that "commercial translations cannot intrinsically be judged as 'good' or 'bad'. Translations are a service, so they are always target-specific. They must first and foremost satisfy the customer's requirements [...]. Therefore translation quality can only be judged by these standards". In addition, it seems to Lilova (2008: 17) that "avoiding subjectivity by using rigorous and obligatory norms would be an experience doomed to failure" because she believes that "such norms cannot entirely reflect all the errors, all the cases which are particular to a given translation [and] would have such an inclusive nature, they would be so abstract that their application would always be influenced by the subjective acceptance of the person in charge of its evaluation" (Lilova, 2008: 17).

The subjective nature of assessment is evident in the theoretical approaches to the phenomenon as well, for different theoretical perspectives on translation quality would be favoring one definition over the other. For instance, while vanguards of a hermeneutic approach to TQA could be very well satisfied with so-called creative translational actions resulting from a translator's subjective interpretation of the source text, a behaviouristic viewpoint would suggest that a "good translation" should arouse an equivalent response in the target reader (House, 2015: 10-11). Moreover, a functionalist might maintain that a quality translation should fulfill the purpose provided by the commissioner, and an assessor advocating a descriptive viewpoint may have to start with figuring out whether a given text is a translation or just treat it like a translated text unless otherwise is said or evidenced. Then, (s)he can question its conformity with target cultural norms (House, 2015: 11-12). If a member of the manipulative school, an assessor looking for an imposed political view needs to reveal

how and why a source text has been chosen for and manipulated in translation (House, 2015: 13). As for linguistic stance, translation evaluation is not a mere, one-to-one structural and semantic matching anymore but employs a broadened scope with the inclusion of speech acts, pragmatics, sociolinguistics, and discourse analysis (House, 2015: 14).

The true merit of objectivity is well appreciated when it comes to assessing several works because an evaluative procedure failing to pass consistent judgments would be faulted for being unreliable. Eyckmans and Anckaert (2017: 51) are of the view that TQA "should rely on human judgement (and therefore be subjective)" and that "when (multiple) translations need to be assessed", subjectivity turns into a problem that can be solved by having recourse to "a sound methodological basis for translation assessment practices in which the subjective process of judging translation quality is embraced and the measurement error that comes with it is calculated, expressed and controlled by means of a reliability coefficient". This suggests that absolute objectivity is unachievable in practice, yet subjectivity should not prevail in an entire assessment process. The upcoming title attempts to present a reasonable weighting between these two extremities.

Need for Less Subjective/More Objective TQA

From the above paragraphs, it can be concluded that the translation is subjective, translation assessment is subjective, and translation quality is subjective. In other words, subjectivity is inevitably inherent in TQA. Yet given that subjectivity in assessment is a major concern and objectivity is a mainstay of assessment, then it is crucial to administer an assessment process as objectively as possible.

For Wilss (1982: 221-222), "a taxonomy of translational-critical criteria guaranteeing a systematic description, explanation, and evaluation of [target language text]" is needed "to overcome the state of inherent methodological instability". It is very difficult to pass "any 'final judgment' on the quality of a translation that fulfills the demands of scientific objectivity" (House, 2001: 255)

since "the primary difficulty surrounding the issue of translation evaluation is its subjective nature" (Bowker, 2000: 183). Nevertheless, assessors should still pursue objectivity because subjectivity "does not invalidate the objective part of the assessment" (House, 2001: 256). Moreover, as Williams (2013: 420) discusses, "there is general agreement about the need for a translation to be good, satisfactory or acceptable", yet there is an "ongoing debate" over and a need for the means of determining to what degree a translation is good, satisfactory, or acceptable. Such a need is also observable "in developing and implementing a [TQA] model" (2013: 420) and "both the language industry and translation studies urgently need a method to measure translation quality as objectively as possible" (Koby et al., 2014: 415-416). Accordingly, the present paper attempts to contribute to the efforts to satisfy this need for a tool to assess translation quality 'as objectively as possible' by quantifying this phenomenon.

Some research is available as to what to operationalize/consider to assess objectively. Marcel Thelen (2008: 419) produces an umbrella list of criteria based on previous considerations on TQA criteria. "(1) (various types and grades of) equivalence between ST and TT as a yardstick for accuracy/fidelity in the areas of form and meaning, (2) compatibility with the locale of the TT's readers, and (3) correct and natural use of the language and terminology of the TT". It can be drawn from this list that a TQA tool is expected to observe different types and grades of equivalences, consider the sociocultural characteristics of target readership, and evaluate the accuracy and fluency of a target segment. For Martinez-Melis and Albir (2001: 283), (1) assessors should consider several criteria and the assessee should be informed of them, (2) the criteria should be in congruence with the context, function, and purpose of the assessment, (3) "the object of assessment […] must be clearly defined, as well as the level at which it is being carried out", and the assessor should be aware of how competently he/she can assess the intended object, and (4) an assessor should know of markers to judge whether an assessee has the intended skills and, if he/she does, to what degree.

It is seemingly unlikely to produce an overarching quality assessment tool, but there are some common properties applicable to objective TQA tools. The most prominent of these properties should be predetermined apt criteria with high validity and reliability, the lack of which, as Mueller (2004) argues, potentially leads to two major measurement errors, namely invalidity and unreliability. Therefore, an assessment tool is expected to offer accurate and consistent results each time it is administered. Yet validity is considered the more important characteristic (Mueller, 2004: 163) because an assessment tool should be able to measure what it is intended to measure. Otherwise, it would repetitively and reliably produce the same or similar invalid results. The tool proposed here is believed to help administers obtain relatively more valid and reliable results by presenting objectivity-promoting criteria, which are supposed to generate accurate results likely to recur in the same or similar evaluative settings.

Methodology

Source Text and Justification of Selection

Sampling a representative source text can be listed among the critical considerations in TQA (Martinez-Mateo et al., 2017). A randomly selected text would fall short of revealing the 'true' quality of a target text by being partially representative of the entire text. Therefore, the sampled source text should result from a conscious selection process.

Farahzad (1992: 275-276) discusses how a source text is operationalized for an evaluative activity. The text should be authentic and "self-contained for its translation to require no knowledge of the precise linguistic context". It should not be too challenging to exceed assessees' "level of command of both source and target languages as well as their level of translational competence". Lastly, assessees should be presented with brief information about the text (Farahzad, 1992: 275-276).

In this paper, the source text is a user's manual of a camera, written in English and consisting of 262. Factoring in the

abovelisted parameters, the manual was operationalized for the purpose of the study because it is an authentic, non-classwork, self-contained stand-alone textual material, whose suitability for the participating students was assessed in a pilot study (further explained below). The other reasons why this text type was chosen are that "the style of a non-literary text generally contains fewer or more controlled ambiguities, gaps and possibilities for the reader's engagement" (Boase-Beier, 2011: 76) and that denotational meaning is prioritized over connotational meaning (Newmark, 1995: 16). Thus, manuals are thought to lend themselves to a less speculative assessment.

Moreover, the participants were asked to preserve the text type as they translated, which made it relatively easy to judge whether their decisions conformed to the characteristics of manuals. Manuals' informativeness and instructionality allowed the author to less effortfully decide whether the translated units reflected the facts about the device. Because manuals' instructions are typically clear to understand, he ably questioned whether the text was capable of exerting the desired extratextual, pragmatic effects.

- Do not [1]**place the strap around the neck** of an infant or child**.** Placing the [2]**camera** [3]**strap** around the neck of an infant or child could result in strangulation.

- [4]**Observe proper precautions** when handling batteries. [5]**Batteries may leak** or explode if improperly handled. [6]**Observe** the following precautions when [7]**handling batteries** for use in this product.

  - Only use a Rechargeable Li-ion Battery EN-EL9 ([8]**supplied**). Use the Quick Charger
  - MH-23 (supplied) to charge the [9]**battery**.
  - Do not [10]**short** or disassemble the battery.
  - Be sure the product is off before [11]**replacing the battery**. If you are using an AC adapter, be sure it is [12]**unplugged**.
  - Do not attempt to insert the battery [13]**upside down** or [14]**backwards**.
  - Do not expose the battery to flame or to excessive [15]**heat**.
  - Do not [16]**immerse** in or expose to water.
  - [17]**Replace the terminal cover** when transporting the battery. Do not transport or store the battery with metal objects such as necklaces or

hairpins.

- [6]**Batteries are prone to leakage** [18]**when fully discharged**. [19]**To avoid damage** to the product, be sure to remove the battery [20]**when no charge remains**.
- When the battery is not in use, attach the [21]**terminal cover** and store in a cool, dry place.
- The battery may be hot [22]**immediately after use** or when the product has been used on battery power for an extended period. Before removing the battery turn the camera off and allow the battery to cool.
- [23]**Discontinue use immediately** should you notice any changes in the battery, such as [24]**discoloration** or [25]**deformation**.

Table 1. Source text and rich points

In a pilot study, ten translation students, who were then excluded from the main research, were asked to translate a user's manual into Turkish for the author to be able to find the working lexical, phrasal, collocational, and sentential rich points. PACTE (2009: 212) defines *rich points* as "specific source-text segments that contained translation problems". The underlined and enumerated source segments in Table 1 are the rich points. 25 rich points were identified in the manual.

Table 2 provides the by-domain distribution of the rich points. It reveals that 14 of the rich points occurred at the lexical level, whereas 11 at the syntactical level.

DOMAINS		#	RICH POINTS
A. Lexical Level	A1. Words and technical terms	1	Strap
		2	Camera
		3	Battery
		4	To observe
		5	(supplied)
		6	To short
		7	Unplugged
		8	Upside down
		9	Backwards

		10	Heat
		11	To immerse
		12	Terminal cover
		13	Discoloration
		14	Deformation
B. Syntactical Level	B¹. Phrasal Level	15	when fully discharged
		16	to replace the terminal cover
		17	when no charge remains
		18	to avoid damage
		19	immediately after use
		20	discontinue use immediately
	B². Collocational Level	21	to place the strap around the neck
		22	to observe proper precautions
		23	to handle batteries
		24	to replace batteries
		25	batteries may leak/batteries are prone to leakage

Table 2. Rich points and respective domains

These source segments which are likely to pose translation problems for the participating students were operationalized in the tool. Thanks to these rich points, which correspond to the first section of the assessment tool, the author was able to award positive points for the students' viable translation solutions. The following title offers further details about the production of this

first part and the second section of the tool and about the scoring/grading.

The Tool

TQA from a broader perspective

TQA tools consist of several components, including but not limited to awarded points, type, location, and seriousness of errors or solutions (depending on whether they are error- or solution-based), and guiding rubrics. While some incorporate more of these constituents to yield more valid results, others may produce partially valid outputs by incorporating few. This is why there arises a need for more comprehensive TQA tools relying on a wide array of components to consider translation quality characteristics from a broader viewpoint.

Research is available as to what and what not a TQA tool must incorporate. Williams (2009: 5) offers a list of vital issues to take into account when producing a valid and reliable TQA tool.

a. Evaluator

b. Level of target language rigor

c. Seriousness of errors of transfer

d. Sampling versus full-text analysis,

e. Quantification of quality

f. Levels of seriousness of error

g. Multiple levels of assessment

h. Purpose or function of TQA

To elaborate on this list, an assessor should have the linguistic and extralinguistic knowledge to be able to assess a target text, yet because an assessor is likely to hold different levels of sensitivity and tolerance to linguistic and transfer errors of varying criticality, they should meticulously observe a set of rubrics. Error-based assessment is time-saving and can be quantified with relatively little effort, but an assessor should ascertain that an excerpt contains diverse translation problems featuring varying

levels of difficulty and is representative of the entire text. Translation segments are produced in view of different intratextual and extratextual aspects, such as "accuracy, target language quality, format (appearance of text), register, situationality", thus these aspects should be deemed integral to TQA, the purpose or function of which should affect how the foregoing features should be prioritized (Williams, 2009: 6).

Error-based TQA practices built on number and seriousness of identified errors is a major concern among researchers because "they tend to ignore the macrotextual features of the target text" (Williams, 2013: 421) and "macrotextual issues of coherence and cohesion" Williams (2009: 8) and the fact that a translation with more errors than another may nonetheless may of better overall quality and meet the client's requirements more effectively" (Williams, 2013: 421). Another criticized property of error-driven assessment is its "the over-hyped objectivity claim", yet it seems that they tend to be subjective due to "different error typologies", "arbitrary error weightings and acceptability thresholds", "unreliability and inconsistency in decision-making as a result of the subjectivity", "the reductionist and atomistic nature of error analysis", and "applicability and practicality" concerns resulting from its "labor-intensive", "time-consuming"[1] , and "cognitively taxing" nature (Han, 2020: 260).

Still, it is evident that error detection is an indispensable property of TQA assessment (Bowker, 2000; Waddington, 2001; Thelen, 2008; Koby and Champe, 2013; Martinez-Mateo et al., 2017; Vandepitte, 2017; Nord, 1991), but the partial representativeness of error-based analysis, which overly relies on intertextual comparison, should be compensated for by taking other intra- and extratextual factors into account. According to Al-Qinai (2000: 499), "the assessment of a translated text seeks to measure the degree of efficiency of the text with regard to the syntactic, semantic and pragmatic function of ST within the cultural frame and expressive potentials of both source language and target language". This proposition suggests that assessing a

[1] This view is supported by Williams (2009).

target text requires considering intra- and extratextual parameters ranging from the adequate rendering of a morpheme to the perlocutionary effects of the entire target text or its segments. Thus, a translation decision should be evaluated against both its local and peripheral impact(s).

Williams (2009: 8) reports that macrotextual issues of coherence and cohesion are ignored in the assessment guidelines since "microtextual analysis of samples has been used extensively not only because it saves time but also because it provides error counts as a justification for a negative assessment" (2009: 6). Eyckmans and Anckaert (2017: 42) state that "analytical [criteria-based] methods of assessment came to replace the holistic and intuitive approaches" as of the 1980s, which "was motivated by the need to objectify the evaluation process". Now, there is reviving scholarly interest in holistic assessment but in one that is rubric-based and far less intuitive. Moreover, the literature of translation studies has recently come to incorporate a growing amount of research on the combination of microtextual and macrotextual analyses to produce more valid assessment tools because "quality in translation is a multifaceted reality [, so] a general comprehensive approach to evaluation may need to address multiple components of quality simultaneously" (Colina, 2009: 239).

Waddington (2001) presents a conciliative model of assessment to combine microtextual and microtextual analysis. In the study in which he compares error-based analysis with holistic assessment, his statistical results point to the benefits of combining error-based assessment and holistic evaluation (2001: 35). Similarly, Martinez-Mateo (2016: 42) claims that the combination of holistic qualitative and error-based microtextual quantitative assessment "is necessary and complementary in order to provide the full picture of translation quality". Martinez-Mateo et al. (2017), from a functionalist perspective, propose a "Modular Assessment Pack" comprising a qualitative and a quantitative module. The qualitative module incorporates four parts to pass holistic judgments about the adequacy of the target text at stake, namely "functional, pragmatic and textual adequacy",

"specialized lexical units and content adequacy", "non-specialized lexical units and content adequacy", and "normative and stylistic adequacy". The first part "measures the TT adequacy in relation to its aim, defined by the assignment specifications and the needs of the target audience". The second "refers to the TT's conveyance of specialized lexical units and content in an adequate and coherent manner", while the third and fourth to "the TT's conveyance of non-specialized lexical units and content in an adequate and coherent manner" and "the observance of grammar, spelling and punctuation rules in the TT and the use of an adequate style, bearing in mind the aim and the target audience" (Martinez-Mateo et al., 2017: 24-25). On the other hand, their quantitative module relies on (1) error typology, (2) integration of "the qualitative module dimensions and the quantitative module error typology", (3) functionalism-oriented textual comparison to identify errors (4) textual categorization according to "the aim and quality requirements of the TT", and (5) tagging the detected error.

The discussion above foregrounds several features of a valid assessment tool, which can be coarsely reduced to five textual parameters:

1. **Intratextual factor** refers to linguistic and non-linguistic internal features of a source and target text, such as content, subject matter, textual organization, and sub- and suprasentential features (for more, Nord, 1991).
2. **Intertextual coherence** "refers to a relation between translatum and source text, defined in terms of the skopos" (Vermeer, 2012: 193).
3. **Extratextual factor** refers to the external factors that occur outside the text but still make sense with the text and make the text itself meaningful, such as the text's real-life impact, its receivers and their perception, and its function (for more, Nord, 1991).
4. **Microtextual factor** refers to "phonemic, morphosyntactic, and lexical (subsentence/sentence) elements of discourse" (Williams, 2004: 162).

5. **Macrotextual factor** refers to "coherence and cohesion" (Williams, 2009: 8), therefore, to "structure[s] larger than the sentence, such as the paragraph, section, and chapter" (Williams, 2004: 162).

The TQA proposed in this paper is believed to satisfy these standards to a great extent because it simultaneously addresses the micro- and macrotextual factors while taking into account the intra-, inter-, and extratextual factors. The subsequent title presents the grading rubric for the readers to be able to envisage how different levels of translation errors and solutions were conceptualized and how they were applied to the translations of the lexical, syntactical, and textual elements in the proposed tool.

The Grading Table and the Rubric

As the above discussions of holistic assessments suggest, a translation tool should not be solely relied on atomistic analyses but on micro- and macrotextual factors. This is why an assessor should be simultaneously and interchangeably able to adopt micro- and macrotextual perspectives on the evaluated target texts, and such a broad-spectrum evaluative approach should be supported by a holistic assessment tool.

Table 3[2] presents two levels of translation solutions – to the identified rich points – and three levels of translation errors, arising from unacceptable renditions of the rich points and the remaining source segments. It is clear that the awarded points exponentially increase; in other words, the deduced points (-2), (-4), and (-8) refer to "minor", "major", and "critical" errors, whereas (+2) and (+4) to "partially acceptable" and "acceptable" solutions. The following are the reasons why such a grading was preferred to a one-point-increment scoring – e.g., minor: (-1), major: (-2), and critical: (-3):

[2] The error grading part of this table was adapted from the ATA examination's Flowchart for Error Point Decisions (Ver. 2009) (Koby and Champe, 2013) and the entire table was produced with the help of Dr. Geoffrey S. Koby, Kent State University, USA.

a) to highlight the criticality levels more discernibly,

b) to help assessors make their decisions more confidently because they would, for example, know what constitutes a critical error, how a minor error differs from a partially acceptable solution is, and what point to award them,

c) to promote the representativeness of the points and thus the validity of the tool because a one-point distance between two neighboring criticality levels, even between a minor and critical error, would fail to present the gravity of a translation error's impact.

d) to increase the credibility of the tool because this exponential grading allows to better appreciate the difference between the quality of a text infested with critical errors and of another permeated by major errors.

GRADING TABLE					
	CRITICAL	**MAJOR**	**MINOR**	**PARTIALLY ACCEPTABLE**	**ACCEPTABLE**
ERROR	**(-8)**	**(-4)**	**(-2)**		
SOLUTION				**(+2)**	**(+4)**

Table 3. The grading table

This exponential grading was operationalized as explained in Table 4. The table reveals that the degrees of errors and solutions are judged in view of their local or wider impacts on the target text. Minor errors denote the translations that are not the exact equivalents but make sense in the context, while major errors are the ones making no sense and have an impact at a supralexical level to cause a semantic disturbance at textual locales ranging from sentences to paragraphs. Critical errors exert negative extratextual effects besides making no sense in the context. In

other words, their impacts do not only resonate across the entirety of a target text but also could potentially lead to corporal injuries and/or mechanical damages as far as the source text herein is concerned.

The table provides two types of solutions, i.e., acceptable and partially acceptable solutions. The solutions were only graded in the presence of the rich points. The viable solutions were considered "acceptable", while the ones causing slight semantic disturbances were regarded as "partially acceptable". To specify the difference between a minor error and a partially acceptable solution, the former is concerned with the analyses of lexical equivalences, while the latter is assessed against naturalness/appropriateness. For instance, the use of "plastik koruma [plastic cover]" instead of "terminal kapağı" for "terminal cover" is a minor error, while "Piller sızma yapabilir [Batteries may make leakage]" suggested for "Batteries may leak" instead of "Piller sızdırabilir" sounds colloquial and thus is a partially acceptable solution. "Pilleri kesmeyiniz [Don't cut the batteries]", which is the suggested translation for "Don't short the batteries", could theoretically connote that a user must not cut batteries but can short them. This assumed practice might result in an explosion and thus lead to bodily injuries or mechanical damages, which makes the suggested target segment a critical error.

ERRORS		
DEGREE	**DESCRIPTION**	**EXAMPLE**
Minor	Errors that are not the exact equivalents but make sense in context.	*plastik koruma [plastic cover]* for *terminal cover* instead of *terminal kapağı*.
Major	Errors that make no sense in context.	*uç kısımlar [tips/parts at the tip]* for *terminal cover* instead of *terminal kapağı*.

Critical	Errors that make no sense in context and have negative extratextual effects.	*Pilleri kesmeyiniz [Don't cut the batteries]* for *Don't short the batteries* instead of *Pillere kısa devre yaptırmayınız*.
SOLUTIONS		
DEGREE	**DESCRIPTION**	**EXAMPLE**
Acceptable	Solutions that perfectly make sense in the context.	*Fotoğraf makinesi* for *Camera*.
Partially Acceptable	Solutions that make sense in the context but cause a slight semantic disturbance.	*Piller sızma yapabilir [Batteries may make leakage]* (*sızma yapabilir* sounds colloquial) for *Batteries may leak* instead of *Piller sızdırabilir*.

Table 4. The rubric of errors and solutions

How the exponential grading in Table 3 and the rubric in Table 4 were positioned and operationalized in the proposed tool is presented in the following section.

The Proposed TQA Tool

Table 5 presents the TQA tool, which was built to cover a wide range of assessment-related parameters. Vertically, it is composed of translation problems – i.e., rich points –, target context, translation errors, and translation solutions. The table also presents that translation problems, errors, and solutions are further ramified into subdomains, which are horizontally positioned in the tool and occur at lexical, subsentential, and sentential/suprasentential levels.

PROBLEMATIC	TARGET	ERROR	SOLUTION

DOMAIN	CONTEXT	CRI[1]	MAJ[2]	MIN[3]	PAC[4]	AC[5]
A. LEXICAL LEVEL						
A[1]. Words ve technical terms						
strap	askı kayışını yeni doğan bebeğinizin					4
camera	Makinenin askı kayışını					4
battery	Makinenize **bataryaları** yerleştitirken				2	
to observe	gerekli önlemleri alınız.					4
(supplied)	(ürünle birlikte tedarik edilen)					4
to short	Bataryayı kısa devre yaptırmayınız.					4
unplugged	**cihaza takılı olmadığından**		-4			
upside down	baş aşağı şekilde veya tersinden takmayın					4
backwards	baş aşağı şekilde veya **tersinden** takmayın		-4			
heat	aşırı ısıya					4
immerse	Suya					4

	batırmayınız					
terminal cover	**bağlantı ucunu** takın.		-4			
discoloration	**renk atması** veya deformasyon gibi				2	
deformation	renk atması veya **deformasyon** gibi				2	
B. SUBSENTENTIAL LEVEL						
B¹. Phrasal Level						
when fully discharged	Bataryalar tamamen **şarj edildiğinde**	-8				
replace the terminal cover	bağlantı **ucunu da değiştirin**.	-8				
when no charge remains	şarjı bittiğinde					4
to avoid damage	Makinenin zarar görmesini engellemek için					4
immediately after use	kullanıldıktan hemen sonra					4
discontinue use immediately	bataryayı kullanmayı derhal bırakın.					4
B².						

Collocational Level						
to place the strap around the neck	askısını bebeklerin boynuna asmayın.					4
to observe proper precautions	gerekli önlemleri alınız.					4
to handle batteries	Makinenize bataryaları **yerleştirirken**		-4			
to replace batteries	Bataryayı değiştirmeden önce					4
batteries may leak/batteries are prone to leakage	sızıntı yapabilir.					4
C.NON-RICH POINT ELEMENTS						
C[1]. Communicativeness						
Quantity	**yeni doğan** bebeğinizin ve çocuğunuzun			-2		
Quality	Bataryayı **baş aşağı şekilde veya tersinden** takmay**ın [ya çalışmayınız]**	-8				
Manner	Bataryayı **baş aşağı**			-2		

	şekilde veya tersinden takmayın					
	Makine **yüksek batarya performansında** kullanıldığında	-8				
	Bataryayı kullanmadığınız zaman **bağlantı ucuna takınız**.	-8				
Relation	**Uygunsuz** kullanımda piller akabilir.			-2		
C². Coherence						
Reference	**Bunun** takılı olmadığından emin olun.			-2		
Substitution	**Böyle bir işlem**, yeni doğan			-2		
Ellipsis	Pil üzerinde kısa devre yapmayın veya **[pili]** sökmeyin.			-2		
Conjunction	bebeğinizin **ve [veya]** çocuğunuzun			-2		
Lexical cohesion	**[Pili]** Suya sokmayın		-4			

Descriptions:	SUBTOTAL	-74	70
[1]Critical error [2]Major error [3]Minor error [4]Partially acceptable [5]Acceptable	TOTAL SCORE	-4	

Table 5. The Proposed TQA Tool

Moreover, the respective target segments are provided in the table. If unacceptable, the relevant segments are highlighted in bold. The sample grading shows that the student in Table 5 grossed -74 and +70 points in the 'errors' and 'solutions' sections, respectively, amounting to -4 points in total. Both the subscore of -74 and the total score of -4 suggest that the produced target text is unacceptable.

It can be realized that the tool roughly comprises two sections: the first section relies on rich points lending themselves to the assessment of solutions and errors, while the second features the segments containing minor, major, and critical errors. The second part is important because it allows assessors to consider the translation errors other than the ones suggested as solutions to the rich points. For a holistic evaluation, any transfer/translation or mechanic/linguistic error was considered. Because this part of the tool was not constructed on rich points, it solely took errors into account. The errors were presented under two categories: communicativeness and coherence. The former is related to the capacity of a statement to communicate a specific message/content, while the latter is concerned with textual properties. In the "communicativeness" section, Grice's four conversation maxims were operationalized – i.e., quantity, quality, manner, and relation – because they were observed to lend themselves to the problematization of the communicative characteristics of a manual. To be specific, Smith (2003: 2-32) lists the following stylistic features of a manual, which were categorized under the four Gricean Maxims:

a. Quantity

1. Keep sentences short and limit each sentence to one main idea or step.

2.Use no more than one subordinate clause per sentence.

b. Quality

3. describe the actions or procedures necessary to perform a task.

4. explain how a product works and applications for which it is used.

5. describe how the product may be misused.

6. warn consumers about hazards (safety information).

c. Manner

7. Write in the active voice.

8. Avoid noun strings

9. Avoid multiple negatives

10. Use everyday words that are familiar to and appropriate for your audience

11. Use specific, concrete words rather than ambiguous or abstract ones.

d. Relation/Relevance

12. Include only relevant information and details. This will minimize clutter and confusion and can draw consumers' attention to important information.

Since these excerpted 12 stylistic characteristics of a manual are perfectly applicable to Grice's four maxims[3] and they are not only related to translation/transfer but also mechanical errors, they were included in the tool. "Mechanical errors are [linguistic errors] that can routinely be fixed without reference to the source text", whereas transfer errors require referring back to the source text to understand the performance of "the segment to reflect the true meaning of the source" (Lacruz et al., 2014: 77). Therefore, the errors in the "rich points" section are transfer errors because these translations fail to establish the retrospective communicative links to the source text.

The "coherence" subsection, which is comprised of reference, substitution, ellipsis, conjunction, and lexical cohesion, is solely concerned with mechanical/linguistic errors. They are representative of a respective translator's knowledge of the target language. They are the five cohesive devices of Halliday and Hasan (Baker, 2011: 190). Reference is used "for the relationship which holds between a word and what it points to in the real world". In the case of substitution, "an item (or items) is replaced by another item" (190), while ellipsis "is a case of leaving something unsaid which is nevertheless understood" (196). Conjunction "involves the use of formal markers to relate sentences, clauses and paragraphs to each other" (200). Lastly, lexical cohesion is "the role played by the selection of vocabulary in organizing relations within a text" (210). Since these are cohesive devices to form a "network of lexical, grammatical and other relations which provide links between various parts of a text" (Baker, 2011: 301), they can also be operationalized to assess the textual quality of a target text.

Conclusion

The present paper proposes a translation quality assessment tool, a grading table, and a rubric to contribute to the related

[3] For more, see Baker (2011: 237).

domain of translation studies, namely translation evaluation. The author believes that the tool can be used by translation schools for formative and summative assessment, translators associations in certification exams, and translation companies to hire translators.

The proposed TQA tool relies on not only translation errors but also translation solutions. To be able to judge a solution's acceptability, 25 rich points were identified in the excerpted manual, of which 14 were lexical and 11 were supralexical. The rich points were positioned in the first part of the tool, while the second part was solely built on erroneous translation segments not to miss any flawed target unit that would potentially degrade the quality of a translated text. The tool is also believed to prove beneficial in identifying two types of errors, i.e., mechanical and transfer errors.

Moreover, the grading table of the proposed tool is thought to be a notable contribution to the domain of translation evaluation. The table features solution- and error-based grades in exponential increments. The grades are determined in terms of three parameters: solutions, errors, and their respective levels of (un)acceptability. To elaborate, minor, major, and critical errors are penalized with (-2), (-4), and (-8) points, whereas partially acceptable and acceptable solutions are awarded (+2) and (+4) points, respectively. Such grading allows for better discernment between the levels of translation errors and solutions. Along with the grading table, a rubric was presented to describe how the degrees of errors and solutions could be operationalized, which is considered to promote objectivity and inter-rater reliability.

This TQA tool was administered to judge the qualities of non-literary target texts. Hence, future research can discuss its applicability to literary translations. The tool, the grading table, and the rubric were tailored for the purpose of this research study; thus, researchers may adapt this tool to their respective research settings. Besides, they can also modify the tool to include some new assessment parameters, such as typographical errors. Because the tool was administered to translation students, it can be used to assess the quality levels of translations rendered by professionals.

References

Al-Qinai, J. (2000). Translation Quality Assessment. Strategies, Parametres and Procedures. *Meta: journal des traducteurs/Meta:Translators' Journal*, Vol. 45, Issue 3, pp. 497–519. https://doi.org/10.7202/00187.

Baker, M. (2011). *In Other Words: A Coursebook on Translation (2nd ed.)*. London/Newyork: Routledge.

Boase-Beier, J. (2011) Stylistics and Translation. *The Oxford Handbook of Translation Studies* (Eds. K. Malmkjaer and K. Windle). Oxford/New York: Oxford University Press. pp. 71–82.

Bowker, L. (2000). A Corpus-Based Approach to Evaluating Student Translations. *The Translator*, Vol. 6, Issue 2, pp. 183-210. https://doi.org/10.1080/13556509.2000.10799065.

Colina, S. (2009). Further Evidence for a Functionalist Approach to Translation Quality Evaluation. *Target*, Vol. 21, Issue 2, pp. 235–264.

Eyckmans, J., and Anckaert, P. (2017). Item-Based Assessment of Translation Competence: Chimera of Objectivity versus Prospect of Reliable Measurement. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, Issue 16, pp. 40–56.

Farahzad, F. (1992). Testing Achievement in Translation Classes. *Teaching Translation and Interpreting: Training, Talent and Experience* (Eds. C. Dollerup and A. Loddegaard). Amsterdam: John Benjamins. pp. 271-78.

Han, C. (2020). Translation quality assessment: A methodological review. The Translator, Vol. 26, Issue 3, pp. 257–273. https://doi.org/10.1080/13556509.2020.1834751.

House, J. (1997). *Translation Quality Assessment: A Model Revisited*. Tübingen: Gunter Narr Verlag.

House, J. (2001). Translation Quality Assessment: Linguistic Description versus Social Evaluation. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, Vol. 46, Issue 2, pp. 243-257. doi:10.7202/003141ar.

House, J. (2015). *Translation Quality Assessment: Past and Present*. London/New York: Routledge.

Hutchins, W. J. and Somers, H. L. (1992). *An Introduction to Machine Translation*, London/San Diego: Academic Press.

Kahl, P. (1991). Translation Quality - How Can We Tell It's Good Enough?. *Proceedings of Translating and the Computer 12: Applying Technology to the Translation Process* (Ed. C. Picken), London: Aslib. pp. 149-158.

Koby, G. S. and Champe, G. G. (2013). Welcome to the Real World: Professional-Level Translator Certification. *The International Journal for Translation and Interpreting Research*, Vol. 5, Issue 1, pp. 156-173. doi:ti.105201.2013.a09.

Koby, G. S., Fields, P., Hague, D. R., Lommel, A., and Melby, A. (2014). Defining Translation Quality. *Revista Tradumàtica: Tecnologies de la Traducció*, Vol. 12, pp. 413-420.

Lacruz, I., Denkowski, M., and Lavie, A. (2014). Cognitive Demand and Cognitive Effort in Post-Editing. *Proceedings of AMTA 2014 Third Workshop on Post-Editing Technology and Practice* (Eds. S. O'Brien, M. Simard, and L. Specia), pp. 73 - 84.

Lilova, A. (2008). The Perfect Translation – Ideal and Reality (Trans. by M. J. Stern). *Translation Excellence: Assessment, Achievement, Maintenance* (Ed. M. G. Rose). Amsterdam: Benjamins. pp. 9-18

Martínez-Mateo, R. (2016). Aligning Qualitative and Quantitative Approaches in Professional Translation Quality Assessment. *Encuentro*, Issue 25, pp. 36–44.

Martínez-Mateo, R., Montero Martínez, S., and Moya Guijarro, A. J. (2017). The Modular Assessment Pack: A New Approach to Translation Quality Assessment at the Directorate General for Translation. *Perspectives*, Vol. 25, Issue 1, pp. 18-48. doi:10.1080/0907676X.2016.1167923.

Martínez-Melis, N. and Hurtado Albir, A. (2001). Assessment in Translation Studies: Research Needs. *Meta:Journal des Traducteurs/Meta: Translators' Journal*, Vol. 46, Issue 2, pp. 272–287. https://doi.org/10.7202/003624ar.

Mueller, C. W. (2004). Conceptualization, Operationalization, and Measurement. *The SAGE Encyclopedia Social Science Research Methods* (Eds. M. Lewis-Beck, A. Bryman, and T. F. Liao). California: SAGE. pp. 161-165.

Newmark, P. (1995). *A Textbook of Translation*, London: Longman.

Nord, C. (1991). *Text Analysis in Translation: Theory, Methodology and Didactic Application of a Model for Translation-Oriented Text Analysis*. Amsterdam/Atlanta: Rodopi.

PACTE (2009). Results of the Validation of the PACTE Translation Competence Model: Acceptability and Decision Making. *Across Languages and Cultures*, Vol. 10, Issue 2, pp. 207-230.
https://akjournals.com/view/journals/084/10/2/article-p207.xml

Reiss, K. (2014). *Translation Criticism – The Potentials and Limitations: Categories and Criteria for Translation Quality Assessment (Trans. by E. F. Rhodes)*. London/New York: Routledge.

Samuelsson-Brown, G. (2010). *A Practical Guide for Translators (5th ed.).* Bristol/Buffalo/Toronto: Multilingual Matters.

Smith, T. P. (ed.) (2003). *Manufacturer's Guide to Developing Consumer Product Instructions*, Washington.

Thelen, M. (2008). Translation Quality Assessment or Quality Management and Quality Control of Translation?. *Translation and Meaning – Part 8* (Eds. B. Lewandowska-Tomaszczyk and M. Thelen). Maastricht: Hogeschool Zuyd. pp. 411-424

Waddington, C. (2001). Should Translations Be Assessed Holistically or Through Error Analysis?. *Hermes: Journal of Linguistics*, Issue 26, pp. 15-38.

Williams, M. (2004). *Translation Quality Assessment: An Argumentation-Centred Approach*. Ottawa/Ontario: University of Ottawa Press.

Williams, M. (2009). Translation Quality Assessment. *Mutatis Mutandis*, Vol. 2, Issue 1, pp. 3-23. http://aprendeenlinea.udea.edu.co/revistas/index.php/mutatismutandis/article/view/1825/1609.

Williams, M. (2013). A Holistic-Componential Model for Assessing Translation Student Performance and Competency. *Mutatis Mutandis*, Vol. 6, Issue 2, pp. 419-443.

Wilss, W. (1982), *The Science of Translation: Problems and Methods*. Tübingen: Gunter Narr Verlag.

Vandepitte, S. (2017). Translation Product Quality: A Conceptual Analysis. *Quality Aspects in Institutional Translation* (Eds. T. Svoboda, Ł. Biel, and K. Łoboda). Berlin: Language Science Press. pp. 15–29.

Vermeer, H. J. (2012). Skopos and Commission in Translational Action (Trans. by A. Chesterman). *The Translation Studies Reader (3rd ed.)* (Ed. L. Venuti). London and New York: Routledge. pp. 191–202

Yildiz, M. (2016). *Mütercim-Tercümanlık Öğrencilerinin Özel Alan Çevirisi Kapsamında Yazılı Çeviri Edinçlerinin Ölçülmesi*. Unpublished Doctoral Dissertation, Istanbul: Istanbul University Institute of Social Sciences.

Yildiz, M. (2020). A Critical Perspective on Translation Quality Assessments of Five Translators' Organizations: ATA, CTTIC, ITI, NAATI, and SATI. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, Issue 18, pp. 568-589. doi:10.29000/rumelide.706390.