



**International
Standard**

ISO 5060

**Translation services — Evaluation
of translation output — General
guidance**

*Services de traduction — Évaluation des résultats de traduction
— Recommandations générales*

**First edition
2024-02**

STANDARDSISO.COM : Click to view the full PDF of ISO 5060:2024

STANDARDSISO.COM : Click to view the full PDF of ISO 5060:2024



COPYRIGHT PROTECTED DOCUMENT

© ISO 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
3.1 General concepts	1
3.2 Concepts related to the people involved in the evaluation process	3
3.3 Concepts related to the evaluation process	3
3.4 Concepts related to errors	4
3.5 Concepts related to penalty points	5
4 Human resources	6
4.1 General	6
4.2 Professional competences of evaluators	6
4.3 Qualifications of evaluators	6
5 Pre-evaluation phase	7
5.1 Evaluation approaches	7
5.2 Implementing the evaluation of translation output	7
5.3 Specifications	7
5.4 Error typology	7
5.5 Error type weights	9
5.6 Severity levels	9
5.7 Critical errors	9
5.8 Counting repeated errors	9
6 Evaluation phase	9
6.1 General	9
6.2 Basic evaluation principles	10
6.3 Sampling the translation output to be evaluated	10
6.4 Translation evaluation scorecard	10
6.5 Quality rating	11
7 Post-evaluation phase	11
7.1 Feedback	11
7.2 Dispute resolution	11
Annex A (informative) Translation use case, evaluation purpose and constraints	12
Annex B (informative) Developing a translation evaluation strategy	13
Annex C (informative) Developing a translation evaluation system	14
Annex D (informative) Sampling	15
Annex E (informative) Cohesion and coherence	16
Annex F (informative) Examples of scorecards for evaluating translation output	17
Bibliography	21

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 5, *Translation, interpreting and related technology*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Globalization, the accompanying growth of the translation market, the professionalization of translation services and the growing availability of online machine translation tools have driven the demand for standardized and harmonized principles and procedures for the evaluation of translation output.

In an industrial world in which products have to comply with legal regulations and client specifications, translation output has to be regarded as a product that has to fulfil certain requirements. The systematic comparison of a target language content with the source language content giving regard to the translation project specifications provides the means to arrive at a valuable and objective judgement of achievement.

STANDARDSISO.COM : Click to view the full PDF of ISO 5060:2024

[STANDARDSISO.COM](https://standardsiso.com) : Click to view the full PDF of ISO 5060:2024

Translation services — Evaluation of translation output — General guidance

1 Scope

This document gives guidance on the evaluation of human translation output, post-edited machine translation output, and unedited machine translation output. It also provides guidance on the qualifications and competences of evaluators. The role of sampling is also discussed in this document.

This document focuses on an analytic translation evaluation approach using error types and penalty points configured to produce an error score and a quality rating.

A further focus is the human evaluation of translation output only. This document follows an approach designed to reflect minimum complexity. The rationale behind this approach is to keep this document applicable for as many users as possible in the translation sector.

The guidance provided in this document can also support the evaluation of source texts intended for translation.

This document is applicable to translation service providers (TSPs), including individual translators, translation companies or in-house translation services, their clients and other interested parties in the translation sector, such as translator education and training institutions.

This document does not apply to related elements such as the processes of assuring the quality of translation output and corrective actions.

This document does not apply to interpreting.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 20539, *Translation, interpreting and related technology — Vocabulary*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 20539 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1 General concepts

3.1.1

translate

render *source language content* (3.1.6) into *target language content* (3.1.7) in written form or signed language

[SOURCE: ISO 20539:2023, 3.1.8]

3.1.2

translation

set of processes to render *source language content* (3.1.6) into *target language content* (3.1.7) in written form

Note 1 to entry: A translation can refer to formats other than text-based formats (e.g. an audio file, image, etc.).

[SOURCE: ISO 17100:2015, 2.1.2, modified — “may” has been changed to “can” in Note 1 to entry.]

3.1.3

revision

bilingual editing

examination of *target language content* (3.1.7) against *source language content* (3.1.6) for its suitability for the agreed purpose

[SOURCE: ISO 17100:2015, 2.2.6, modified — “bilingual examination” has been replaced by “examination” in the definition and Note 1 to entry has been changed into an admitted term.]

3.1.4

evaluation of translation output

evaluation

bilingual examination of *target language content* (3.1.7) against *source language content* (3.1.6) while classifying any *errors* (3.4.1) with respect to *translation evaluation specifications* (3.3.5) and for the purpose of reaching a *quality rating* (3.3.10)

Note 1 to entry: The term “assessment” is frequently used as a synonym for evaluation. While both involve inspection, “evaluation” is the preferred term, since formally there are some differences. Evaluation is product-oriented and judgmental (to accept or reject a product), whereas assessment is process-oriented and diagnostic (to provide feedback on how to improve the process used to produce the product). The preference for “evaluation” is also to avoid the confusion that the abbreviated term “QA” gives rise to, mostly referring to “quality assurance”.

3.1.5

corrective action

action taken to eliminate the cause of nonconformities or *errors* (3.4.1) during *translation* (3.1.2) or in *target language content* (3.1.7)

Note 1 to entry: Corrective action involves an investigation to identify what went wrong and what action can be taken to ensure that it does not happen in the same way again.

[SOURCE: ISO 17100:2015, 2.5.5, modified — Singular “nonconformity” has been changed to plural “nonconformities” and “in the translation process” has been changed to “during translation”.]

3.1.6

source language content

language content to be *translated* (3.1.1)

[SOURCE: ISO 17100:2015, 2.3.2]

3.1.7

target language content

language content *translated* (3.1.1) from *source language content* (3.1.6)

[SOURCE: ISO 17100:2015, 2.3.3]

3.1.8

translation output

result of *translation* (3.1.2)

[SOURCE: ISO 20539:2023, 3.3.3]

3.1.9

client-TSP agreement

arrangement between a client and a translation service provider (TSP) which specifies the conditions of a *translation service* (3.1.10)

Note 1 to entry: This can take the form of a contract, memorandum of understanding, oral agreement, email, etc. Oral agreements can be documented in writing.

3.1.10

translation service

production and delivery of *target language content* (3.1.7) according to a *client-TSP agreement* (3.1.9)

[SOURCE: ISO 20539:2023, 3.3.4, modified — “specifications issued by a client” has been changed to “a client-TSP agreement”.]

3.1.11

segment

unit of text produced for a computer application to facilitate *translation* (3.1.2)

Note 1 to entry: A segment can be a sentence, heading or other unit of text, such as phrase, word or a single character.

[SOURCE: ISO 18587:2017, 3.2.9]

3.2 Concepts related to the people involved in the evaluation process

3.2.1

evaluator

person who performs *evaluation of translation output* (3.1.4) in accordance with the *translation evaluation system* (3.3.4) of the *evaluating entity* (3.2.2)

3.2.2

evaluating entity

person or group of people who determines a *translation evaluation strategy* (3.3.3)

3.3 Concepts related to the evaluation process

3.3.1

requirement

need or expectation that is stated, generally implied or obligatory

[SOURCE: ISO 9000:2015, 3.6.4, modified — Notes to entry have been deleted.]

3.3.2

translation project specifications

set of agreed upon and defined *requirements* (3.3.1) for producing *translation output* (3.1.8)

Note 1 to entry: ISO 17100:2015, Annex B, lists a set of sample project specifications.

3.3.3

translation evaluation strategy

organization-wide policy implemented based on the specific purpose for the *evaluation of translation output* (3.1.4)

Note 1 to entry: The purpose for the evaluation of translation output can vary from project to project. Therefore, it can be necessary to apply various translation evaluation strategies.

3.3.4

translation evaluation system

set of steps and actions performed during the *evaluation of translation output* (3.1.4) implemented based on the *translation evaluation strategy* (3.3.3)

3.3.5

translation evaluation specifications

set of principles for the *evaluation of translation output* (3.1.4), based on the *translation project specifications* (3.3.2)

3.3.6

translation evaluation sample

sample

amount of content taken from *target language content* (3.1.7) to be submitted to an *evaluation of translation output* (3.1.4)

3.3.7

sampling

procedure whereby a part of the *translation output* (3.1.8) is selected to provide a *translation evaluation sample* (3.3.6)

3.3.8

translation evaluation scorecard

translation evaluation scoresheet

table containing all relevant *error types* (3.4.3) with their assigned *penalty points* (3.5.1), for the purpose of producing an *error score* (3.3.9) and the *quality rating* (3.3.10) for the evaluated *translation output* (3.1.8), and for documenting the evaluation

3.3.9

error score

evaluation penalty point total (3.5.3) in relation to the number of characters, number of words or number of lines evaluated in the *translation output* (3.1.8)

3.3.10

quality rating

classification of evaluated *translation output* (3.1.8), based on comparing the *error score* (3.3.9) with an applicable *error score threshold* (3.5.6) and rating system

3.3.11

risk assessment

process of identifying potential risk based on the likelihood that an *error* (3.4.1) will occur and analysing and evaluating its risk level according to the perceived severity of its impact

3.4 Concepts related to errors

3.4.1

error

failure to adhere to *translation project specifications* (3.3.2)

3.4.2

error typology

classification system of *errors* (3.4.1)

3.4.3

error type

class of *translation error* (3.4.1) identified by name, definition and position in an *error typology* (3.4.2)

3.4.4

main error type

superordinate *error type* (3.4.3)

3.4.5

error sub-type

subordinate *error type* (3.4.3)

3.4.6

severity level

rating of the impact of an *error* (3.4.1) on the *target language content* (3.1.7) and its suitability for its intended purpose

Note 1 to entry: Example of severity levels: minor, major and critical.

3.4.7

neutral item

annotation pointing out an element that is not an *error* (3.4.1) but is flagged for further attention

3.4.8

minor error

error (3.4.1) that does not prevent the *translation output* (3.1.8) from fulfilling its intended purpose

3.4.9

major error

error (3.4.1) that can prevent the *translation output* (3.1.8) from fulfilling its intended purpose

3.4.10

critical error

error (3.4.1) that does prevent the *translation output* (3.1.8) from fulfilling its intended purpose

3.4.11

repeated error

error (3.4.1), recurring within a text, that is annotated, and that can be but is not always assigned *penalty points* (3.5.1)

3.4.12

severity multiplier value

numerical value reflecting *severity levels* (3.4.6) of a given *error* (3.4.1) multiplied with *error counts* (3.4.13)

3.4.13

error count

number of *errors* (3.4.1) found during an *evaluation of translation output* (3.1.4)

3.5 Concepts related to penalty points

3.5.1

penalty point

numerical value calculated for each *error type* (3.4.3) based on the *severity multiplier value* (3.4.12)

3.5.2

error type penalty point total

sum of *penalty points* (3.5.1) for a given *error type* (3.4.3)

3.5.3

evaluation penalty point total

sum of *error type penalty point totals* (3.5.2) of an *evaluation of translation output* (3.1.4)

3.5.4

error type weight

numerical value used to modify the unweighted *error type penalty point total* (3.5.2)

3.5.5

weight, verb

apply *error type weights* (3.5.4)

3.5.6

error score threshold

target value above or below which an *evaluation of translation output* (3.1.4) is assigned a passing rating

4 Human resources

4.1 General

The translation service provider (TSP) should have a process in place to ensure that people performing evaluation of translation output have the competences and qualifications outlined in [4.2](#) and [4.3](#).

4.2 Professional competences of evaluators

Evaluators should have the following competences:

- a) Translation competence: The ability to translate content, including the ability to address the problems of language content comprehension and language content production and the ability to render the target language content in accordance with the translation project specifications.
- b) Revision competence: The ability to objectively compare target language content against source language content and the ability to identify potential errors and suggest or make corrections in order to ensure that the revised target language content meets the translation project specifications.
- c) Evaluation competence: This competence requires attention to detail, the ability to follow instructions precisely and the ability to communicate feedback constructively. This competence includes the ability to objectively classify and annotate any errors found during the evaluation process, and assign them to their relevant error type and severity level in order to establish an error score by filling out the translation evaluation scorecard.
- d) Linguistic and textual competence in the source language and the target language: The ability to understand the source language, fluency in the target language, and general or specialized knowledge of text-type conventions. This linguistic and textual competence includes the ability to apply this knowledge when producing translation output or other target language content.
- e) Competence in information acquisition and processing: The ability to understand the processes required for the translation quality evaluation and to apply translation project specifications and translation evaluation specifications efficiently. In some cases, research competence can also be required to verify decisions taken and to develop suitable strategies for efficiently accomplishing the tasks required.
- f) Cultural competence: The ability to make use of information on the behavioural standards, up-to-date terminology, value systems and locale that characterize both source and target language cultures.
- g) Technical competence: The knowledge, abilities and skills required to perform the technical tasks required for the translation evaluation process by employing technical resources including the tools and IT systems that support the entire evaluation process.
- h) Domain competence: The ability to understand domain-specific source language content and to evaluate adequate correspondence in the target language by verifying the appropriate terminology, style and register.

4.3 Qualifications of evaluators

The TSP should determine the evaluator's qualifications by obtaining documented evidence that the evaluator fulfils at least one of the following criteria.

- a) The evaluator has obtained a degree in translation, linguistics, language studies or an equivalent degree from a recognized institution of higher education. The degree should include significant training in translation and revision.
- b) The evaluator has obtained a degree in any other field from a recognized institution of higher education and has the equivalent of two years of documented experience in translation and revision.
- c) The evaluator has the equivalent of five years of documented experience in translation and revision.

NOTE Higher education refers to universities and other institutions of education at the third (tertiary) level. Degree refers to the certificate of qualification from one of said institutions.

5 Pre-evaluation phase

5.1 Evaluation approaches

There are two evaluation approaches common in the translation sector, holistic and analytic.

- Although the holistic evaluation approach relies on both quantitative and qualitative evaluation, its focus is on the analysis of the quality of the overall translation output rather than on applying an error typology to individual segments. This approach can be helpful when evaluating some types of content, such as for transcreation or marketing purposes or for quick decision-making.
- The analytic evaluation approach involves a segment-based comparison of target language content against source language content taking into account translation project specifications. Nonconformities are regarded as errors and are annotated with respect to error type and severity level. Penalty points are assigned to each error type and severity level, leading to an error type penalty point total and an evaluation penalty point total.

This document covers the analytic evaluation approach only.

5.2 Implementing the evaluation of translation output

This subclause shows the relevant evaluation factors, the translation evaluation strategy and the translation evaluation system that should be considered before starting with the evaluation of translation output. The evaluation process should also be aligned with the translation project specifications.

- a) Relevant evaluation factors: Translation use cases, purposes and any constraints pertaining to the evaluation of translation output should be identified before establishing any translation evaluation strategy. For a list of questions, see [Annex A](#).
- b) Evaluation strategy: A translation evaluation strategy should be developed based on the relevant translation evaluation factors. For a list of questions, see [Annex B](#).
- c) Evaluation system: An evaluation system should be aligned with a translation evaluation strategy. For a list of questions, see [Annex C](#).

5.3 Specifications

The evaluation of translation output should be based on the translation evaluation specifications and the underlying documented and agreed-upon translation project specifications.

Conformance with specifications should be measurable, and evaluators should be able to quickly decide whether the evaluated translation output conforms or does not conform with the specifications.

5.4 Error typology

In accordance with the specifications, the evaluating entity can freely choose the main error types and error sub-types which conform to the specifications.

The following error typology consists of 7 main error types and 34 error sub-types.

The order in which the main error types and error sub-types occur in this error typology is not intended as a hierarchy of error types.

- a) Terminology:
 - inconsistent with terminological resource;

- inconsistent use of terminology;
 - wrong term.
- b) Accuracy:
- mistranslation;
 - addition;
 - omission;
 - do not translate (DNT);
 - untranslated.
- c) Linguistic conventions:
- grammar;
 - punctuation;
 - spelling;
 - unintelligible;
 - character encoding.
- d) Style:
- organization style;
 - third-party style;
 - inconsistent with external references;
 - register;
 - awkward style;
 - unidiomatic style;
 - inconsistent style.
- e) Locale conventions:
- number format;
 - currency format;
 - measurement format;
 - time format;
 - date format;
 - address format;
 - telephone format;
 - shortcut key.
- f) Audience appropriateness:
- culture-specific reference.

g) Design and markup:

- character formatting;
- layout;
- markup tag;
- truncation/text expansion;
- links/cross-reference.

Each error type should be defined and contain some examples of errors for these error types.

For the impact of cohesion and coherence, see [Annex E](#).

5.5 Error type weights

The translation evaluation strategy can provide for the use of error type weights. Assignment of these numerical values represents the importance of a particular error type in relation to other error types (e.g. a mistranslation error can be assigned more weight than a punctuation error). A typical weight has the numerical value of 1, and can be adjusted up or down in order to weight certain error types if desired for a given content type or set of specifications.

5.6 Severity levels

If severity levels are used, multiplier values can be assigned to each severity level for the purpose of calculating the applicable penalty points. The severity level can, for example, be minor, major or critical, depending on the impact the error has on the usability of the target language content for its intended purpose.

Depending on how the error affects the purpose of the translation output, different severity levels can be assigned to the same type of error by counting them as either neutral items, minor errors, major errors or critical errors.

5.7 Critical errors

Depending on the purpose of the evaluation, the evaluating entity should determine a severity level for critical errors. Critical errors pose serious danger to the end user of translation output and can, among other consequences, cause bodily harm or loss of life, financial losses or damages, loss of reputation, render the target language content incomprehensible, and prevent the intended use of the translation output.

If critical errors are identified during an evaluation of translation output, even an evaluation of translation output with an error score that would otherwise pass should automatically produce a “fail” evaluation result. Evaluators may decide not to complete the evaluation.

5.8 Counting repeated errors

The translation evaluation strategy should specify whether every occurrence of an error is counted individually or as a repetition.

6 Evaluation phase

6.1 General

The evaluation process depends on the translation evaluation strategy and the evaluating entity. The evaluator can use quality control and quality evaluation tools for evaluation support. The evaluation strategy can vary based on the translation project.

6.2 Basic evaluation principles

Errors are annotated according to the relevant translation project specifications and translation evaluation specifications.

Evaluators should evaluate the translation output objectively and neutrally. They should adhere to the client specifications, such as style guides, and avoid personal stylistic or other preferences.

If evaluators wish to flag the translator's or evaluating entity's attention to an issue not covered by the translation specifications, they may do so using neutral comments.

Evaluators should have been thoroughly trained in the evaluation specifications and procedures that apply for the translation evaluation project at hand (see [4.2](#) and [4.3](#)).

Evaluators should be able to justify their evaluation results.

The evaluation should be reproducible independently of the evaluator or the evaluation context. In order to achieve this, proper training and guidance should be provided and updated regularly, based on how the evaluation system performs.

6.3 Sampling the translation output to be evaluated

Translation evaluation specifications should specify whether the entire target language content or parts of it are to be evaluated. Sampling is a common method to determine the extent of translation output to be evaluated, but it bears the risk that errors exist in the unevaluated target language content and that an evaluation of other samples from the same translation output produces different results.

The decision on whether and how to conduct sampling depends on the translation evaluation strategy. Factors can include risk assessment of the document, time, budget and resources available.

For a list of questions see [Annexes A, B and C](#). [Annex D](#) shows possible criteria for sampling.

6.4 Translation evaluation scorecard

A tool for translation evaluation should include the following data and metadata based on the translation evaluation, which may exist in the form of a scorecard or in the form of data in a database, from which a translation scorecard can be generated, or in the form of separate scorecard documents:

- identification (ID) or name of translation project;
- date of translation project;
- ID of evaluator;
- ID of TSP;
- number of characters/words/lines/pages evaluated;
- error types;
- error type weights;
- number of errors per error type;
- evaluation penalty point total;
- severity levels;
- severity multiplier values;
- error score of translation output;
- error score threshold;

- quality rating of the translation output;
- threshold for quality rating;
- pass or fail rating.

All error types and severity levels used in the translation evaluation scorecard should be assigned numerical values reflecting their effect on the translation output.

Every error that is recognized according to the specifications should be assignable to an error type and a severity level. If an error is not assignable to an error sub-type, the selection of error sub-types should be revisited (e.g. by assigning the error to a main error type or by re-aligning the error sub-type definition). If an error can be assigned to more than one category, it should be dealt with in accordance with the translation project specifications and translation evaluation strategy.

The evaluation penalty point total is calculated on the basis of the numerical values of error types and severity levels.

In order to calculate an error score, the evaluation penalty point total should be divided by the total volume of text (words or characters) evaluated in the source language content or target language content, based on the translation evaluation strategy and the translation project specifications.

The translation evaluation strategy can specify that an error score threshold should be defined to determine whether the translation output passes or fails in relation to the error score. The value of the error score threshold depends on the purpose of the evaluation.

[Annex F](#) shows examples of scorecards.

6.5 Quality rating

The error score can be further used to determine a quality rating. A quality rating can be qualitative (e.g. good, acceptable, poor) or quantitative (e.g. points or percentages).

Often, a quality rating is pass or fail, and the error score is the result of dividing the evaluation penalty point total by the evaluation word count.

Quality ratings can also be based on thresholds that are ranges (e.g. 0,001 to 0,004 being good, 0,005 to 0,009 being average, > 0,01 being insufficient) rather than a single number in the case of pass/fail judgement.

The translation evaluation strategy can specify that a threshold should be defined to determine whether the translation output receives a pass or fail quality rating.

7 Post-evaluation phase

7.1 Feedback

Evaluators should provide the TSP with clear, comprehensible feedback on the significance of the error score and any associated information (e.g. error score, filled-out translation evaluation scorecard, detailed report of errors and quality rating).

The evaluating entity should provide an opportunity for the TSP to give counter-feedback on the evaluation.

7.2 Dispute resolution

In the case of differences of opinion regarding the evaluation, the evaluators and the TSP should engage in constructive dispute resolution.

Annex A (informative)

Translation use case, evaluation purpose and constraints

The following questions can be used to help users of this document to analyse their need for evaluation. This list is not exhaustive.

- a) Translation use case:
 - What type of source language content will be dealt with?
 - What risk level is involved?
 - What kind of translation project specifications apply?
- b) Purpose of the evaluation:
 - Will translation output be ready for delivery?
 - Will translation output be subject to approval by the client?
 - Whose work will be evaluated: a TSP, a translator, a reviser, or a post-editor?
- c) Constraints:
 - Will there be enough budget for an evaluation?
 - Will there be enough time for any evaluation to be conducted?
 - If so, will there be enough time to evaluate the full translation output?
 - Will there be enough human resources for an evaluation?
 - Are the evaluators trained (see [4.2](#) and [4.3](#))?

Annex B (informative)

Developing a translation evaluation strategy

The following questions can be used to help users of this document to develop their own translation evaluation strategies. This list is not exhaustive. The following questions result from the answers in [Annex A](#).

a) Follow-up questions linked to the use case:

- What kind of translation evaluation scorecard should be used?
- What percentage of the translated volume, projects or documents will be evaluated?
- Can the same translation evaluation strategy apply to other use cases?

b) Follow-up questions linked to the purpose:

- At which stage in the process will the evaluation take place?
- How often will evaluations be conducted?
- Will locked segments be taken into account?
- How will repeated errors be handled?

c) Follow-up questions linked to the constraints:

- Will sampling be necessary?
- Will sampling be appropriate?
- Will the evaluation be conducted in-house or outsourced?
- Which language combinations will be evaluated?

Annex C (informative)

Developing a translation evaluation system

The following questions can be used to help users of this document to develop their own translation evaluation systems. This list is not exhaustive. The following questions result from the answers in [Annex B](#).

a) Sampling:

- What portion of the translated volume, projects, documents will be sampled?

NOTE For detailed recommendations on sampling, see [Annex D](#).

b) Scorecard:

- Which error types will be used for evaluation?
- Which severity levels (e.g. minor, major or critical) will be used for evaluation?
- How many penalty points will be assigned to the pertinent severity levels?
- How will the results of the evaluation be used?

STANDARDSISO.COM : Click to view the full PDF of ISO 5060:2024

Annex D **(informative)**

Sampling

This document takes into account the trend in documentation practice that more and more content is produced in fragments only or in small information units using content management systems (CMS). As a consequence, evaluation of translation output can involve the consideration of only fragments or small information units.

When not evaluating the entire target language content or sampling randomly, the following factors can help the TSP to decide which parts of the translation output to choose for evaluation:

- The kind of matches that will be included (e.g. no-matches, fuzzy matches, 100 % matches, machine translation segments).
- The percentage of no-matches, fuzzy matches or 100 % matches: Target language content with a higher percentage of no-matches reflects the quality of new translation output better than target language content with a high percentage of 100 % matches.
- Signal words and important, related text passages: Errors in target language content containing warning notices such as “DANGER”, “WARNING” or “CAUTION” can have greater impact on the usability of the translation output than errors in other parts of target language content that do not contain these signal words (see ISO 3864-2:2016, 4.3).
- Some parts of the translation output that can be more important for the evaluation, such as sections relevant to usability of the translation output (e.g. sections containing numbers, instructions or legal implications).
- Whether whole paragraphs or only single sentences from a document are evaluated.

Annex E (informative)

Cohesion and coherence

Cohesion and coherence are universal textual characteristics that can manifest across various error types. Each instance of a cohesion or coherence error should therefore be annotated under one of the seven main error types depending on how it manifests and the consequences of the error.

Cohesion is a surface feature that provides linkage between parts of a text connecting actual words and expressions. Cohesion errors can be annotated as:

- mistranslation under “accuracy” [see [5.4 b](#))] if they distort the intended meaning;
- mistranslation under “linguistic conventions” [see [5.4 c](#))] if the translation output is ungrammatical;
- mistranslation under “style” [see [5.4 d](#))] if they result in awkward or unidiomatic translations; or
- mistranslation under “terminology” [see [5.4 a](#))] for errors relating to terminology.

Coherence is the logical progression of the intended underlying message of the text. It is difficult to identify coherence-related errors unless they are tied to one or more cohesion markers. The smaller the translation evaluation sample, the more difficult it becomes to detect and annotate coherence-related errors.