

Translation Quality Metric

Foreword—In the past, quality measurement of language translation in the automotive industry was somewhat subjective, if such measurement was undertaken at all. Cost and effort go into obtaining translations from suppliers, the same as for any manufactured product. It is necessary to have a tangible method for measuring the quality of translation deliverables as precisely as for any manufactured product.

The objective of the proposed quality metric is to establish a consistent standard against which the quality of translation of automotive service information can be objectively measured regardless of the source language, regardless of the target language, and regardless of how the translation is performed (i.e., human translation or machine translation).

The metric allows an evaluator to tag errors in a translation and compute a weighted, numeric score that represents the quality of the translation. The metric consists of four parts:

- Seven error categories
- Two sub-categories
- Two meta-rules to help decide ambiguities on the assignment of an error to the categories and sub-categories
- Numeric weights

Consistent use of a metric across the industry will allow tighter control of translation quality.

The risks of low-quality translations of service information include erosion of customer confidence, higher warranty costs, and (at an extreme) damage to vehicles or injury to people.

SAE Technical Standards Board Rules provide that: "This report is published by SAE to advance the state of technical and engineering sciences. The use of this report is entirely voluntary, and its applicability and suitability for any particular use, including any patent infringement arising therefrom, is the sole responsibility of the user."

SAE reviews each technical report at least every five years at which time it may be reaffirmed, revised, or cancelled. SAE invites your written comments and suggestions.

TO PLACE A DOCUMENT ORDER: +1 (724) 776-4970 FAX: +1 (724) 776-0790
SAE WEB ADDRESS <http://www.sae.org>

TABLE OF CONTENTS

1.	Scope.....	2
2.	References	2
3.	Definitions (How To Use This Document).....	2
4.	Technical Requirements (Content Of The Metric)	2
4.1	General Comments.....	2
4.	Guidelines for Evaluators	4
4.3	Error Categories	5
4.3.1	Wrong Term (WT)	5
4.3.2	Syntactic Error	6
4.3.3	Omission.....	7
4.3.4	Word Structure or Agreement Error.....	7
4.3.5	Misspelling	8
4.3.6	Punctuation Error.....	8
4.3.7	Miscellaneous Error	8
	Appendix A Translation Metric Score Sheet	10
	Appendix B Quick Reference Guide	11

1. **Scope**—This SAE Recommended Practice is applicable to translations of automotive service information into any target language. The metric may be applied regardless of the source language or the method of translation (i.e., human translation, computer assisted translation or machine translation).

Note that the current version of the metric does not measure errors in style, making it unsuitable for evaluations of material in which style is important (e.g., owner's manuals or marketing literature). The metric can be expanded to accommodate style and other requirements of particular new media.

2. **References**—There are no referenced publications specified herein.
3. **How to Use this Document**—The *General Comments* section introduces the correct use of the metric. *Guidelines for Evaluators* indicates scoring practices. The error categories are defined in the next section.

4. **Technical Requirements (content of the metric)**

- 4.1 **General Comments**—The SAE J2450 metric should be regarded as only one element in a total Quality Assurance process, albeit an important one.

This document defines SAE J2450 and should be regarded as a reference document, not a training document. The task force was chartered to create a metric, and it is up to users of that metric to determine how to train quality evaluators in the use of the metric. We recommend that a training course be established based on this definition document before the implementation of any quality assessment based on SAE J2450.

In using SAE J2450 each error found by the evaluator should be marked in two ways. First, it should be classified into one of the seven categories described below, e.g., 'wrong term'. After its primary category has been identified, the evaluator should indicate if it is a 'serious' or a 'minor' error, according to how severe the evaluator considers the error. Both the primary (category) and the secondary (serious/minor) classification are judgment calls by the evaluator.

While reasonably complete definitions are given below to help the evaluator identify the major categories with some consistency, there may yet be room for ambiguity and the final category selection may, in fact, be arbitrary. For example, the English word 'replace' may be translated into French as either 'replacer' (to put back into place) or 'remplacer' (to replace with another item), according to the context of use. That is, the English verb 'replace' is ambiguous with respect to French. Suppose that an English text describes a situation where a seat belt must be replaced in the event of an accident because the belt may be weakened by the first accident and fail in a subsequent accident. If a French translation then uses the word 'replacer', an evaluator should mark this verb as an error. However, is it a 'wrong term' error, a 'misspelling' error or a 'miscellaneous' error? Without talking to the original translator, it is fundamentally impossible for an evaluator to know if the translator mistakenly thought that 'replacer' and not 'remplacer' as the correct verb, or whether the translator simply mistyped and accidentally omitted the 'm'.

In such cases where the primary category classification is ambiguous and there is no evidence as to which category an error belongs, then an evaluator should place it into the category that appears the earliest in the following list. This is completely arbitrary, but if all evaluators follow this simple rule, then greater consistency of classification across evaluators will result.

The secondary classification of an error as either 'serious' or 'minor' is always a judgment call by the evaluator, and necessarily so. While it is virtually impossible to define the notion of a serious or minor error, as a general guideline if an error is clearly serious in its consequences for a technician or its effect on the meaning of the translation, then it should be marked as 'serious'. If not, then it should be sub-classified as 'minor'. It is a natural and expected consequence that there will be cases when an evaluator hesitates in this choice. Whenever an evaluator hesitates, let us take it again as a general guideline to sub-classify the error as 'serious'. This again is completely arbitrary, but it seems the safer alternative since safety may be an issue in a service text.

Once the primary and secondary classifications have been assigned, then a numeric weight between 1 and 5 is looked up in the definition and assigned to the error. 5 corresponds to the most severe error, while a 1 indicates an error with relatively small consequences. This assignment of weights is a mechanical process and there is no room for judgment. While sometimes this assignment of numeric weights will over-value the severity of an error, it will under-value it at other times. The underlying assumption of SAE J2450 is that these deviations will tend to cancel each other, i.e., "regression to the mean" will normalize these deviations on a large evaluation text or with a large number of evaluation texts.

The SAE J2450 categories are meant to be used to tag errors that are linguistic in nature. Thus, a translation deliverable that is free of any SAE J2450 errors may still be unacceptable to the client due to other problems, such as formatting errors.

Also, SAE J2450 is designed only for the evaluation of translations of service information where the style, tone, register etc., of the target language deliverable is not as important to the client as it may be for owners manuals or marketing literature. The target customer of the translation is the service technician. Therefore, SAE J2450 deliberately avoids tagging "errors" of style.

The definitions as follows are generally syntactic in nature. That is, they depend upon the surface form of the translation deliverable and are generally divorced from the meaning. However, meaning is accommodated in the notion of a 'serious' versus a 'minor' occurrence of an error type.

The general concept is that once an error has been identified by category, the evaluator decides whether the effect on the translation of that error is severe or not. If severe, then the error is classified as a 'serious' occurrence of that error type, and otherwise it is regarded as a 'minor' occurrence. This is necessarily a judgment call by the evaluator.

A simple example will clarify. If the English term 'accelerator' is translated into a target language by the phrasal equivalent of 'gas pedal', the technician reading the translation is not likely to be confused. This error would be marked 'wrong term, minor'. However, if 'accelerator' were translated as 'brake' the translation should clearly be marked as 'wrong term, serious'.

Care should be taken by any evaluator in the application of these SAE J2450 categories NOT to confuse an error with the cause of that error. That is, a translation deliverable may contain a token of a wrong term error (or an error of any other type) because the translator made a mistake or because there is an error or ambiguity in the source language text. Errors in the target text that are caused by source language errors or ambiguities are still errors in the target text. Any contractual or working relationships between clients and suppliers should recognize this distinction, and not penalize suppliers for errors that are caused by the client's source documents, even though there is a SAE J2450 error in the target text. Again: Do not confuse the existence of an error with its cause. SAE J2450 is only concerned with the existence of an error, regardless of the reason for that error. Causal analysis should be performed on all errors, according to the needs of the client.

Finally, although some translators and evaluators may be accustomed to working with a category called 'mistranslation', the SAE J2450 metric has no such category. Any error in a translation is regarded as a mistranslation, potentially affecting the meaning of the target text. The first six categories below attempt to identify which kind of a mistranslation an error is, and the seventh ('miscellaneous') is a catch-all for when the evaluator knows something is an error, yet cannot categorize it easily as any of the first six types.

The remainder of this document will define SAE J2450, but the reader should keep in mind that the metric itself consists of four elements. They are:

- a. The seven primary error categories,
- b. The two secondary subcategories (i.e., serious and minor),
- c. The two meta-rules:
 1. When an error is ambiguous, always choose the earliest primary category.
 2. When in doubt, always choose 'serious' over 'minor.'
- d. The numeric weights.

4.2 Guidelines for Evaluators—An evaluator should be thoroughly familiar with the definitions of the seven categories that follow, as well as the general comments in the preceding section. SAE J2450 offers no guidelines as to how to select texts for evaluation. Texts may be evaluated in part or in whole, according to the desires of the client organization. If in part, texts may be selected at random or in some predetermined manner. Similarly, SAE J2450 does not suggest how to use the results of an evaluation. This includes the acceptance threshold for a numeric score associated with an evaluation. Scores will be normalized by dividing by the number of words in a sample, but SAE J2450 does not attempt to define what any client finds acceptable. There will be as many methods to employ SAE J2450 as there are client-supplier contracts for producing translations.

However, SAE J2450 does offer guidelines as to how a text that is selected for evaluation is actually tagged for errors. We assume that the evaluator has access to the source text as well as the target text, as well as any reference material such as client terminology lists that may be required.

The task is procedurally simple. When the evaluator locates a non-stylistic error, then he or she should take five actions. Note that if the same error occurs one or more times in a text, it should be marked each time it occurs. The five actions are:

- a. Mark the location of the error in the target text. This should be done by circling the error in the target language text. An empty circle indicates an omission. If markups are performed electronically, then an equivalent, overt form of the markup should be devised.

- b. Indicate the primary category of the error. That is, mark whether it is a 'wrong term', an 'omission', etc. This should be done by writing the category abbreviation (see Table 1) above the circled error. Meta-Rule 1: If more than one category is possible for a given error, select the category that appears earliest in the following table.
- c. Indicate the sub-classification of the error as either 'serious' or 'minor'. Write out the subclassification abbreviations immediately following the primary category. For example, to indicate a serious syntactic error, write 'SE - s' above the circled text. Meta-Rule 2: If you are not sure which sub-classification to use, then choose 'serious'.
- d. Look up the numeric value of the error. These are listed at the end of each definition and in Table 1.
- e. Compute the normalized score. For the sample text, sum up the individual numeric weights and divide by the number of words in the source language text.

As a quick reference guide, the evaluator may refer to the following tables:

TABLE 1—ERROR CATEGORIES, CLASSIFICATIONS, AND WEIGHTS

Category Name: (abbreviation)	Sub-Classification: (abbreviation)	Weight: serious/minor
a. Wrong Term (WT)	serious (s)	5/2
b. Syntactic Error (SE)	minor (m)	4/2
c. Omission (OM)		4/2
d. Word Structure or Agreement Error (SA)		4/2
e. Misspelling (SP)		3/1
f. Punctuation Error (PE)		2/1
g. Miscellaneous Error (ME)		3/1

4.3 Error Categories

4.3.1 WRONG TERM (WT)—This category cannot be defined until the concept of 'term' has been defined. We define a term as any

- a. Single word,
- b. Multi-word phrase used as a single, lexical constituent (i.e. part of speech),
- c. Abbreviation,
- d. Acronym,
- e. Number or numeral,
- f. or proper name, including trade names, brand names, registered trademarks, place names, and personal names.

Thus, 'automotive' is a single word; 'powertrain control module' is a multi-word phrase; 'Hz' is an abbreviation, 'ABS' is an acronym, '3.14' is a numeral, 'three' is a number, and 'Society of Automotive Engineers' is a name. All are terms, as are function words such as prepositions, conjunctions, determiners and articles.

Given the definition of term, we define a 'wrong term' to be any target language term that

- a. Violates a client term glossary;
- b. Is in clear conflict with de facto standard translation(s) of the source language term in the automotive field;
- c. Is inconsistent with other translations of the source language term in the same document or type of document unless the context for the source language term justifies the use of a different target language term, for example due to ambiguity of the source language term;
- d. Denotes a concept in the target language that is clearly and significantly different from the concept denoted by the source language term.

An example of (b) would be 'Parking Brake' translated as "Parkbremse" (aircraft term) in German rather than "Handbremse" (automotive domain). In Dutch, the equivalent error would be the use of 'parkeerrem' (aircraft) rather than 'handrem' (automotive).

Clause (c) is important because it captures the error of inconsistency of terminology, which is a frequent error. An example of (c) would be a translation of English 'accelerator' into French that sometimes uses 'accélérateur' and sometimes 'pédale des gaz'. In German, an inconsistency might be to translate 'odometer' sometimes as "Odometer" or sometimes as "Kilometerzähler".

An example of (d) would be "black interior" translated into Spanish as "interior blanco" (white interior).

A wrong term could appear in a translation if the translator is not a native speaker of the target language. For example, if English 'glow plug' were translated into Dutch in a literal fashion, then the wrong term would be 'gloeiplug' rather than the correct 'bougie'.

Serious weight: 5

Minor weight: 2

4.3.2 SYNTACTIC ERROR—A syntactic error comprises the following cases:

- a. A source term is assigned the wrong part of speech in its target language counterpart.
- b. The target text contains an incorrect phrase structure, e.g., a relative clause when a verb phrase is needed.
- c. The target language words are correct, but in the wrong linear order according to the syntactic rules of the target language.

To show an example of a clause (a) error, the English sentence 'The throttle valve connects to the accelerator pedal' could be mistranslated into French as 'Le papillon des gaz connecte à la pédale d'accélérateur.' In this example, the French verb 'connecter' is transitive and should take a direct object, not a prepositional phrase complement.

A clause (a) error may or may not appear as an incorrect inflection on the target term, depending upon the language, but a part of speech error should be assigned to this category nonetheless.

Clause (c) could occur, for example, in Romance languages if an adjective that was supposed to follow its noun were instead placed in front of the noun and this placement was not driven by reasons of style or emphasis. For example, the English "Check gauge lamp" could be mistranslated into Portuguese as "Luz dos indicadores de verificação" instead of "Luz de verificação dos indicadores".

An English to German example of this error would be "This car was built in Germany" translated into German as "Dieses Auto wurde gebaut in Deutschland."

Serious weight: 4

Minor weight: 2

4.3.3 OMISSION—An error of omission has occurred if:

- a. A continuous block of text in the source language has no counterpart in the target language text and, as a result, the semantics of the source text is absent in the translation;
- b. A graphic which contains source language text has been deleted from the target language deliverable.

By a 'continuous block of text' is meant anything from a term to a sentence to a paragraph or to any larger block of text. The corresponding missing text in the target language may or may not need to be continuous. For example, a verb in English may translate to a so-called separable verb in German (where the verb is supposed to appear first, followed later in the sentence by a preposition) and that preposition is then missing in the German. For example, the German 'ablesen' is such a verb that is a translation of the English 'to read off (instruments)', where the verb 'lesen' would normally appear, then be followed by the direct object, and finally by the preposition 'ab'.

In Spanish, it would be an omission to translate "Remove the wheelnuts and the wheel" as "Desmonte la rueda" (Remove the wheel).

Notice that the previous definition implies that even if a large section of text were absent from the target language text, that would constitute only one occurrence of this error, albeit a very serious one in its implications.

Part (b) of the definition implies that if the missing graphic does NOT contain any source language text, then there is NO occurrence of an error with respect to SAE J2450. This would still be, of course, an error and a problem that needs to be rectified by the translation supplier. But the point is that SAE J2450 is limited to linguistic errors, and does not therefore encompass all potential problems with a translation deliverable.

Note that *omission* does not imply that the source and target language words must be in a 1:1 correspondence.

Serious weight: 4

Minor weight: 2

4.3.4 WORD STRUCTURE OR AGREEMENT ERROR

- a. An error of incorrect word structure has occurred if an otherwise correct target language word (or term) is expressed in an incorrect morphological form, e.g., case, gender, number, tense, prefix, suffix, infix, or any other inflection.
- b. An error of agreement has occurred when two or more target language words disagree in any form of inflection as would be required by the grammatical rules of that language.

Note that clause (a) implies that an article of the wrong gender/number/case would be tagged in this category, and not as a Wrong Term error. An example of this could be the use of the wrong Dutch article 'De (rather than: het) gekozen voertuig wordt niet ondersteund'.

An example of (a) would be English 'The vehicle selected is not supported' translated to 'Gewählte Fahrzeug wird nicht unterstützt' rather than the correct 'Gewähltes Fahrzeug wird nicht unterstützt' in German.

An example of (b) would be German 'Der Motor läuft' translated to English "The motors is running" or "The motor are running" rather than "The motor is running." In Spanish, "fog lights" should be translated as "luces antiniebla" rather than "luces antinieblas".

Another example of (b) would be English 'for a new combustion engine' translated into German as 'für ein neues Verbrennungsmotor' rather than "für einen neuen Verbrennungsmotor."

The intent of part (b) of this rule is to assure that even though more than one word may be wrong, there is only one occurrence of the error, rather than one for each word that has the incorrect inflection. This is to recognize that the error is a single error in the mind of the translator and that this error simply manifests itself in several words. However, if the same error occurs in a separate phrase elsewhere in the document, it will again count as a new occurrence of the error.

Serious weight: 4

Minor weight: 2

4.3.5 MISSPELLING—A misspelling has occurred if a target language term:

- a. Violates the spelling as stated in a client glossary,
- b. Violates the accepted norms for spelling in the target language,
- c. Is written in an incorrect or inappropriate writing system for the target language.

Clause (a) of course assumes that the client glossary is correct.

An example in German of a misspelling could be in the compound “Stromkreis Kraftstofffreigabe Bereich” for “Fuel Enable Circuit Range” where the sequence of 3 f’s should have been reduced to just 2 consecutive f’s. The English “Screw on” could be misspelled in Portuguese as “Atarrachar” instead of “Atarraxar”.

Clause (c) is intended to flag as an error words in target languages such as Japanese that use more than one writing system, and the word in question was written with the wrong writing system, even though it may have been spelled correctly in that (wrong) system. For example in Japanese if a foreign term, normally written in Katakana, were instead written in Hiragana, albeit with the correct syllabic symbols in Hiragana for that word’s pronunciation.

Serious weight: 3

Minor weight: 1

4.3.6 PUNCTUATION ERROR—The target language text contains an error according to the punctuation rules for that language.

For example, in German “Am 2 Zylinder” for “On the 2nd cylinder” there should be a period following the ‘2’ to indicate an ordinal number. A serious punctuation error could involve a missing decimal point or comma according to the target language, e.g., “116” instead of “11.6” or “11,6”.

In Dutch and other languages, there could be a missing concatenation mark for implied compounds, such as the translation of “particle separators and filters” as “partikelafscheiders en filters” instead of “partikelafscheiders en -filters”.

Serious weight: 2

Minor weight: 1

4.3.7 MISCELLANEOUS ERROR—Any linguistic error related to the target language text which is not clearly attributable to the other categories listed previously should be classified as a miscellaneous error.

This rule recognizes that not all errors fall neatly into the other six error categories, and that some evaluators may proofread a translation and recognize it as clearly wrong, yet not be able to say precisely what it is that is wrong.

This category should NOT be used to tag stylistic problems as miscellaneous errors. Remember that stylistic errors are ignored by SAE J2450.

Examples of miscellaneous errors include (but are not limited to):

- a. A literal translation of an idiom, e.g., the English idiom 'to push the envelope' meaning 'to perform beyond existing boundaries.'
- b. A lexically correct, but culturally offensive honorific (as in Asian and African languages),
- c. A superfluous or 'extra' block of text that has no relevance to the source text,
- d. A mistranslation that reverses the meaning of the original, e.g., taking English as the target language, 'Do not engage the brake' instead of 'Engage the brake.'

An example of an error not easily categorized in any preceding category could be the French l'usage "the usage" appearing as l' usage. Here a space appears after the apostrophe, which could have occurred as a typo, for instance.

'Acorn nut' is an idiom that could be accidentally translated literally. An example of an incorrect honorific would be 'you' translated as /tšài/ rather than /khun/ in Thai.

Superfluous text could occur where the translator includes two synonyms in a text, perhaps intending to check with the client for the preferred term, only to forget to do so, leaving both terms in the delivered translation. Thus, in a translation to French the English term 'relay box' may show up as "bôte de relais / bôte à relais".

A Spanish example would be to translate "Remove defective part" as "Remueva la pieza defectuosa o dañada" (Remove the defective or damaged part).

Generally the process of translation is performed as part of the process of localization, which makes a product culturally appropriate for the target audience as well as linguistically correct. With this in mind, a translator may be perfectly justified in adding text to a translation in order to explain some cultural reference that may not be familiar to the target audience of the translation. Such additions should not be regarded as superfluous, and no error should be marked.

Serious weight: 3

Minor weight: 1

PREPARED BY THE SAE VEHICLE E/E SYSTEMS DIAGNOSTIC STANDARDS COMMITTEE