

Leena SALMI (Turku) /

Marja KIVILEHTO (Tampere)

Translation Quality Assessment: Proposals for developing the Authorised Translator's Examination in Finland

Abstract

The Finnish system for authorising translators to produce legally valid translations was reformed in 2008, from a test measuring language skills into an examination containing translation assignments. The examination consists of two translation assignments and a test on the examinees' knowledge of the authorised translator's professional practices. In the assessment of the translation products, a predefined, two-dimensional assessment system is used in which translations are marked for both content and language quality. In this paper, we discuss the Finnish assessment system and compare it with the assessment systems used on examinations by the American Translators' Association (ATA), the Canadian Translators, Terminologists and Interpreters Council (CTTIC), and the translation quality evaluation models SAE J2450 and MQM-DQF used within the translation industry. Drawing on a previous analysis of the use of the assessment system in the Finnish examinations and a pilot survey among the assessors in the Finnish system, we propose a new, simplified assessment model.

Keywords: translator certification, translation assessment, translation of official documents, legal translation, translators' professional practices

1. Introduction

The systems of authorising or certifying translators vary between countries. In some countries the examinees have to take translation tests, while in others, such as Canada, it is possible to receive a certification by mentorship (CTTIC 2016a). In Finland the system of authorising translators was reformed in 2008. It now consists of two translation assignments and a test on the examinees' knowledge of the professional practices in an authorised translator's work. In this article, we discuss the assessment criteria used on this examination and on other similar examinations and assessment systems.

In addition to the systems as such, the terminology related to certification also varies: certification, accreditation and authorisation are all used, with slightly

different meanings (see Melby 2013; Hlavac 2013; Pym et al. 2012: 15–16 and 25–26, for discussions on terminology). We use the terms *authorisation* and *authorised translator*, as the certification here relates to the production of legally valid translations only, and as these are the terms used in the Finnish system (see FNBE 2012).

Translation quality assessment is a vast area that is of interest in various contexts, such as translator training, machine translation, technical communication, or the practice of quality assessment in professional translation service provision. Our perspective in this article is on assessing the translation product in an examination context designed to determine whether a candidate demonstrates a level of competence required for producing legally valid translations. This is why we limit our discussion on the assessment of translations. This can be based on either an analysis of errors or on a comparison of the translation with predefined criteria describing what the translation should be like or what translation skills it should demonstrate (Angelelli 2009: 40–41; Turner et al. 2010). Error analysis has been regarded as reliable, and that is why it is used on many certification examinations (cf. Hale et al. 2012: 58). Nevertheless, research shows that criterion-based assessment can be as trustworthy as the assessment based on error analysis (Turner et al. 2010).

On the following pages, we present the assessment system in the Finnish Authorised Translator's Examination, compare it with some other assessment systems used in similar examinations and within the translation industry, and discuss possible modification of the Finnish system.

2. The assessment system used on the Authorised Translator's Examination in Finland

The Finnish Authorised Translator's Examination consists of two translation assignments: one in the field of law and administration and the other in another specialist field chosen by the examinee. The four specialist fields available are business and economy, medicine, technology and education. The length of the examination texts is defined in terms of characters and should be approximately

2,000 characters (with spaces) or equivalent for languages which use non-Latin alphabets. The examinees have 2 hours and 45 minutes per translation assignment at their disposal. They are allowed to use different translation aids, with the exception of machine translation, translation memories and direct contact with other persons. After the examination, the assessors begin their work. Two assessors mark the translations independently, but discuss their assessments afterwards (FNBE 2015, FNBE 2016a). The goal is that the assessments are as consistent and transparent as possible regardless of language combinations and directions. Language combinations and directions are chosen by the examinee, provided that there are examiners available for the combination in question, and either the source or the target language in the combination must be Finnish, Swedish or Sami.

2.1. The assessment system in general

In Finland, the assessment of translation assignments in the Authorised Translator's Examination is based on an error analysis. In their work, the assessors look at the correspondence of the source and target texts, on one hand, and at how acceptable the translations are as target texts, on the other hand. The assessors use a scoring chart (see Appendix 1) with two error categories: errors of equivalence of content ('C' errors) and errors of acceptability and readability ('A' errors). In the first error category, the meaning of the source text is not transferred into the translation, and in the second category, there are problems in the target language (style, register, grammar, punctuation, etc.). Both categories currently include seven¹ error types of different degrees of severity. The severity of an error is expressed in terms of fault points varying from 1 to 9. One 9-point error is enough to fail the translation. However, it is not only the number of fault points that counts on the examination. Other factors, such as the function of the translation, the difficulty of the source text and the time constraint are taken into account as well. Even translators' notes can affect the assessment: if the examinee shows in a note that he or she knows the proper professional practice, an inadequate translation solution can be accepted. All these factors are considered in the overall

¹ From 2012–2014 there were eight 'C' errors and seven 'A' errors on the scoring chart.

statement written by the assessors where they express their views about the translation (FNBE 2015: 4–5).

The scoring chart is based on a proposal presented by Andrew Chesterman (2001) in an article in *Kääntäjä – Översättaren*, the journal of the Finnish Association of Translators and Interpreters, where he proposed dividing the errors into these two categories. For defining the severity of an error, Chesterman (2001) proposed using a severity scale with two levels; the severity could be defined as either severe or not severe, or given either 4 or 1 fault points, respectively. The chart was drafted by a working group preparing the implementation of the examination in 2008.

2.2. Assessment in the Finnish Authorised Translator's Examination in 2012–2014

In order to see how the scoring chart is used in the Finnish Authorised Translator's Examinations, we conducted a small-scale study of translations produced in the examinations between 2012 and 2014. The results of this study are reported in more detail in Kivilehto (2016), Kivilehto and Salmi (forthcoming) and Kivilehto (2017). The data contained 56 translations from and into English (“the English setting”), 48 from and into Swedish (“the Swedish setting”), and 28 from and into German (“the German setting”). We counted all errors marked in the translations and analysed which error types had been used by the assessors. For the error types in use, see the scoring chart in Appendix 1. Table 1 shows the numbers of translations in the data, as well as the main results of the study.

Table 1: Error types used by the assessors on the Finnish Authorised Translator's Examination (2012–2014)

	English	Swedish	German
Number of evaluated translations	56	48	28
Most common error type	C7 (41%)	A6 (23%)	C7 (42%)
Second most common error type	A5 (26%)	C7 (20%)	
Error types seldom used	C3 (0%) C4 (0.1%)	C4 (0.4%) C3 (0.6%)	

As can be seen from Table 1, the most common error type both for language pairs including English and German was a 'C' error, more precisely C7, “an individual word/term that is imprecise, unsuitable or irrelevant or an omission or an addition not essentially affecting the meaning of the text”. This type accounted for over 40% of all errors in both settings. In language pairs including Swedish, however, the most common error type was the 'A' error A6, “individual style errors and unidiomatic expressions”, accounting for 23% of all errors.

The second most common error type in the English setting was the acceptability error A5, “a structural error that does not cause misinterpretation”, which accounted for about 26% of all errors, and in the Swedish setting, the C7 with 20%. In the German setting, the rest of the errors were divided fairly evenly, so that in addition to the C7 with 42%, no other type accounted for more than 10%, which is why they are not shown in Table 1. In contrast to these frequently employed error types, error types seldom used were C3, “the translation function is disregarded, leading to inadequate results”, C4, “unfounded alternative translation equivalents”, and C8, “incomplete or erroneous equivalents for the cultural and social context of the source language”.

The percentages show that C7 was frequent and C3 infrequent in all language settings. The frequency of C7 may depend on the description of this error type. There were actually three different kinds of errors included in the description: incorrect term, omission and addition. We also noticed that 'A' errors were used in the Swedish setting more often than 'C' errors. One reason for 'C' errors occurring more seldom in the Swedish setting might be that the examinees have a good knowledge of Swedish, it being one of the other official languages of Finland. However, information on the linguistic background of the examinees is not available in the system.

In sum, the results of our small-scale study show that some of the error types are used by the assessors quite often, while others occur quite seldom. This points to the fact that the scoring chart might need some adjustments.

2.3. Survey among assessors

In November 2016, we conducted a pilot survey on the assessment system among the assessors of the Finnish examination. The survey was distributed in a seminar organised shortly after the examination to allow the assessors to discuss the assessment parameters (cf. Salmi / Kinnunen 2015: 235).

We received 24 responses in the seminar. There are over 60 assessors on the register, but not all of them have translations to assess every year. This is why we decided to only distribute the pilot questionnaire on paper only to those present at the seminar, as they would be the ones actively doing the assessment. We asked them to evaluate the functionality and ease of use of the current error categories (equivalence of content and acceptability of target language) on a scale of 1 to 5, with four different features expressed with adjectives (see Table 2). We also asked whether they considered there to be too few or too many error types. In addition, the questionnaire contained three open-ended questions (comments on the general evaluation, comments on the numbers of error types, and suggestions for developing the current system). The averages of responses regarding the general evaluation of the current categories are presented in Table 2:

Table 2: Evaluation of current error categories, scale from 1 to 5

	Average	Respondents
Functionality (criteria are functional / not functional)	4.1	21
Logic (criteria are logical / illogical)	3.8	21
Comprehensibility (criteria are easy / difficult to grasp)	3.6	22
Ease of use (criteria are easy / difficult to use)	3.5	21

As can be seen from Table 2, the current division into categories was found to be functional (average 4.1 on a scale of 1 to 5), but their comprehensibility and ease of use were less highly rated, though above scale midpoint (averages 3.5 and 3.6). Seven respondents explicitly stated in their comments that it is sometimes difficult to decide which error type to choose, and two mentioned that some of the error types overlap. On the other hand, five gave positive comments

on the current system, and two specified that once you get used to the categories, the system works well.

For the response to the question on the number of error types (1=too few, 2=ok, 3=too many), the average was 2.3 (23 respondents). One respondent checked both “1” and “3”, which was interpreted as “1” for the analysis. Table 3 shows the distribution of the responses:

Table 3: Responses to question related to number of error types

The number of error types is...	
...too large	9
...as it should be	12
...too small	2

Table 3 shows that half of the respondents (12) were satisfied with the number of error types. However, the number of respondents (9) who considered the number to be too large was clearly greater than that of those considering it too small (2).

3. Assessment systems elsewhere

As mentioned above, the terms used in the field of authorised translation vary. In some countries certified translators are called ‘sworn translators’ or ‘official translators’ (Vigier et al. 2013: 40). In Finland, such translators are ‘authorised translators’. In the United States there exist no such appellations as ‘sworn translator’ or ‘official translator’, whereas in Canada, the title is ‘certified translator’. In addition to the appellations, there is also variation in the certification and authorisation procedures, not to mention the assessment systems. In this section, we discuss, as examples, the assessment systems in two certification examinations, ATA (the United States) and CTTIC (Canada), as well as in two translation quality evaluation models used within the translation industry, the SAE J2450 and the MQM-DQF harmonised system. ATA and CTTIC are examples

of certification examinations, and the SAE J2450 and the MQM-DQF are examples of error categorisation.

3.1. Assessment in certification examinations – ATA and CTTIC

The American Translators Association (ATA) arranges translation examinations for translators who want to become ATA certified. The examination consists of three translation assignments: A, B and C. Assignment A is compulsory for all examinees, while they can choose between B or C. A is a general text assignment, and B and C are somewhat specialised texts. B is technical, scientific or medical, and C is financial, business or legal in nature. The texts contain about 225 to 275 words each and the examinees have three hours to translate the texts. Dictionaries, glossaries and non-interactive Internet resources are allowed, but CAT tools and interactive Internet resources are prohibited (ATA 2016a, ATA 2016b).

The assessment system of the ATA translation examination is a combination of error- and criterion-based assessment. Errors fall into 23 error types listed alphabetically from *addition* (A) to *word form* (WF) (see ATA 2016d). The translations are assessed from the point of view of four areas of performance: usefulness/transfer, terminology/style, idiomatic writing and target mechanics. Each of these areas contains a description of performance with five error levels: standard, strong, acceptable, deficient and minimal – standard being the ideal performance. Errors are marked using a five-level fault point system (1, 2, 4, 8 and 16 fault points), with the fail threshold being 18 fault points (ATA 2011, 2016c, ATA 2016d).

The Canadian Translators, Terminologists and Interpreters Council (CTTIC) organises translation examinations for professionals who want to use the appellation ‘certified translator’. The examination consists of two translation assignments, one of which is a general text and compulsory, and the other a somewhat specialised text assignment and selected from among two texts, either scientific/technical/medical or economic/administrative. The texts contain approximately 175–185 words each and the examinees have three hours to translate the

texts. Dictionaries and reference works are allowed, but all technological aids are prohibited (CTTIC 2016a, CTTIC 2016b).

The assessment system of the CTTIC translation examination is error-based. Errors fall into two categories: translation errors and errors of expression. Translation errors are failures to render the meaning of the source text, i.e., misinterpretations and omissions of phrases and larger units. Errors of expression are errors of the target language, including violation of syntax, grammar, vocabulary, spelling and typographical rules. Errors are rated using fault points ranging from 3 to 10 points, 3–5 points standing for minor errors and 10 points for major errors. To pass the examination, the examinee must gain 70% for each assignment (CTTIC 2016c).

In sum, both ATA and CTTIC rely on detecting and counting errors. What is different, though, is that the ATA system is more comprehensive, combining both error- and criterion-based assessment.

3.2. Assessment in Translation Quality Models – SAE and DQF-MQM

Different systems to evaluate translation quality have been developed within companies providing translation services (for a fairly recent overview, see O'Brien 2012). We present here two such systems where error-based assessment is used: the typology described in the SAE J2450 standard, as well as the DQF-MQM, which is a recently formulated comparison of two existing frameworks, the Multidimensional Quality Metrics (MQM) and the Dynamic Quality Framework (DQF) system formulated by TAUS, a resource centre for translation industries.

The SAE J2450 translation quality metric is a standard created by the Society of Automotive Engineers for quality measurement in the automotive industry. It has seven error categories: wrong term, syntactic error, omission, word structure or agreement error, misspelling, punctuation error and miscellaneous error. The classification intentionally does not contain a category for style, register, tone, etc., as it is “designed only for the evaluation of translations of service information” (SAE International 2016: 5). A severity level is defined for each error,

with two severity levels ('serious' and 'minor'). The error type and the severity level are converted into a numeric value using a table with a predefined scale (e.g., syntactic error: serious=4 points, minor=2 points), and the fault points are added up. The sum is then divided with the number of words in the source text document, giving a weighted score (SAE International 2016).

The SAE model takes some account of subjectivity, admitting that the severity ranking is "always a judgment call by the evaluator, and necessarily so" (SAE International 2016: 5). It provides a specific rule for how to deal with hesitation: when hesitating between error types, the rule is to choose the error type appearing first in the list, and when between severity levels, to mark the error as 'serious'. This is mechanical and arbitrary, but intended to create consistency, as long as all evaluators follow the same rule (SAE International 2016).

The Multidimensional Quality Metrics (MQM) was developed to provide a framework for describing and defining quality metrics used to assess the quality of translated texts and to identify issues in those texts. It does not provide one single metric, but rather a vocabulary for describing metrics and creating them with different types of foci. It "does not specify what *must* be checked, but rather provides ways to describe what *is* checked" (Lommel et al. 2015: 23, emphasis in the original).

Simultaneously with the development of the MQM, the Dynamic Quality Framework (DQF) system had been formulated by TAUS, a resource centre for translation industries. These two were compared for harmonisation in an EU-funded project, and Lommel et al. (2015) present, as a result of the process, an MQM-compatible version of the DQF typology (Lommel et al. 2015: 12–13). There are seven main error categories: accuracy, fluency, terminology, style, design, locale conventions and verity, and a varying number (up to seven) of sub-categories within each of them. *Accuracy* relates to correspondence with the source text content (addition, omission, ambiguities etc.), and *fluency* to the target language grammar, punctuation and other conventions; *terminology* and *style* have their own categories. *Design* is related to the layout and formatting of the translation and *locale conventions* to how issues like addresses or measurements are expressed (usually, they should follow the target culture conventions). *Verity* refers to rendering culture-specific items. The DQF error typology is detailed in

the TAUS Quality Dashboard document (TAUS 2016) with examples. The examples given reflect a functionalist approach and a domesticating strategy in that using the Fahrenheit scale in a text translated into French, for example, is considered an error (TAUS 2016: 23).

Both systems have been implemented in various CAT tools, allowing the quality assessment of both human and machine translation. For example, a translation memory system can compare the output in each segment to pre-established guidelines or glossary entries, during or after the translation process. This is partly why the DQF-MQM comparison differentiates between “issues” and “errors”; issues are “potential problems” that an automatic system may notice but that are not necessarily errors (for example, the target language equivalent for a term could be present in a glossary but missing from a translated segment, yet if the translator has just replaced the term with a pronoun, this is not an error) (Lommel et al. 2015: 7). Another example of an issue would be an explanatory addition related to a culture-specific item that the target audience would not be familiar with.

In addition to classifying the error type, the severity level of each error is defined. The MQM system distinguishes between two levels, ‘major’ and ‘minor’. The DQF system has five levels, of which Severity 1, 2 and 3 are errors that are ‘critical’, ‘major’ or ‘minor’, respectively; Severity 4 is used to mark other information that is not an error, and kudos can be used to praise for exceptional achievement (TAUS 2016: 23). In other words, MQM can be considered to have two and DQF three severity levels for errors.

The harmonised system also contains the idea of weight, meaning that some issue type may be given more importance than others. For example, terminology might be weighted at 2.0 and style at 1.0 (Lommel et al. 2015: 8).

4. Proposals for a simplified assessment model

In developing the assessment used in the Finnish Authorised Translator’s Examination, we need to remember the nature of the examination. It is meant to sift out the translators capable of producing legally valid translations. In this type of

translating, documentary translation, rather than instrumental translation, is usually needed, as a legally valid translation is meant to have the same evidentiary weight as the original and it needs to represent the original document “as is”. Accuracy is essential and nothing should be omitted, unless agreed beforehand with the authority requesting the translation. All culture-specific items should reflect the source-text culture, rather than be localised using target-culture items, and be explained in a translator’s note, if necessary. For example, educational or occupational qualifications should not be rendered with the nearest target language equivalent, as systems vary and as drawing equivalents is a task for the authorities of the target culture (FNBE 2016b; Nord 1997: 47; Taibi / Ozolins 2016: 91).

The current system relies on error analysis, and the severity of an error is indicated by fault points. Depending on the error type, the points given can range from 1 to 2, 2 to 6 or 6 to 9, but the exact score for each error is decided by the assessor. There are currently 14 different error types, which was felt by 9 of our 23 respondents to be too many. As the data analysis discussed in 2.2. showed, some types are used frequently and others hardly at all. There would therefore be room for rethinking the system. In the following, we present our suggestions for simplifying it.

The current division into errors of equivalence of content ('C' errors) and errors of acceptability and readability ('A' errors) is the same as the division in the CTTIC system into Translation and Language errors. Our first suggestion for simplifying the current system is to retain this division ('C' vs. 'A' errors), but without subcategories.

An alternative to this would be to modify the MQM-DQF error typology to correspond to the needs of a system for testing examinees in producing legally valid translations. The error types ‘accuracy’ (correspondence with the source text), ‘fluency’ (adherence to the target language conventions regarding grammar, spelling, punctuation etc.), ‘terminology’ and ‘style’ could be used without modification. As the analysis of the examination assessments showed, there is strong evidence from the assessments that use of terminology can form an error type of its own. ‘Design’ could here refer to how the conventions of producing legally valid translations are taken into account (for example, providing a hea-

ding for the translation or adding a translator's confirmation at the end). 'Locale conventions' could relate to how elements like addresses or measurements are dealt with in the translations. 'Verity' could refer to the manner of rendering culture-specific items and taking the target audience into account (as defined in the instructions given to examinees).

In the current Finnish system, fault points are given as in the SAE system, where the number of fault points is related to the error type. In the CTTIC system, fault points are given according to the severity of the error (3, 5 or 10), the total fault score is subtracted from 100, and 70 points are needed to pass (CTTIC 2016c). In fact, the use of severity levels instead of fault points given by the assessor would simplify the rating of the errors. The systems reviewed here have 2–5 severity levels, and a division into three ('minor', 'major' and 'critical') might be sufficient. In the SAE system, a weighted score is reached by dividing the total fault score by the number of words in the original. As mentioned in section 2, in the Finnish system, the length of the examination texts is approximately 2,000 characters (or equivalent). A similar score could therefore be used (with the fault points subtracted from a given figure or divided by the number of characters in the source text). However, if fault points are used, whether in the form of severity levels or numbers, a pass rate could also be defined and used consistently, as in the CTTIC and ATA systems. This is our first suggestion: errors classified into either two or seven error types, and the severity of the errors defined and summed up as a single figure which is then used with a pass rate.

A pass rate expressed as a number or as a percentage presents some problems, however. A legally valid translation may not contain severe errors, and this is reflected in the current scoring chart by the inclusion of error types that are rated at 9 points but lead to a failed examination (e.g., complete misinterpretation of an idea or the omission of a sentence). One critical error can therefore be enough to fail an otherwise acceptable translation. On the other hand, a translation may contain plenty of minor errors but still be acceptable, as long as it renders the content of the original as required for the translation task in question. A strict pass rate might not do justice to such different cases.

Another problem with the use of pass rates is discussed in the CTTIC marker's guide (CTTIC 2016c). The document advises the assessors to try to distance

themselves from the 70% pass mark as far as possible and, for translations that are “borderline cases”, determine their overall acceptability and adjust their marking accordingly. This is also recommended to avoid “complaints from unhappy candidates” (CTTIC 2016c: 4–5).

If fault points are not used to define passing or failing, the assessors should give an overall assessment on the translation, taking a stand on whether a) the translation is suitable to and usable in its intended purpose, as defined in the translation instructions given to the examinee; and b) the translation has conveyed the meaning of the source text. This could be developed along the lines of the ATA system, where a criterion-based assessment is used (as described in 3.1.) for assessing translations from the point of view of four different areas of performance and defining an error level (from 1 to 5) for each of them. An acceptable translation could be defined to reach, say, at least the second highest level in all four areas. In this second suggestion, the assessment would combine error-based and criterion-based assessment. In fact, this brings us back to the proposal by Andrew Chesterman in 2001, in which he proposed rating the translation on a scale from 0 to 3 and accepting only the translations rated at 3.

As these are just initial proposals for a new scoring system, testing with a sample set of examination translations would be needed next, in order to gain data on the usability of the proposed error types and to refine them. A more thorough survey of the assessors' opinions could also be administered in a digital form to all the current and former assessors.

In any case, marking the errors in each translation is necessary and useful for the legal protection of the examinees, as they have the right to receive information on why they have failed and how they could improve their skills. The CTTIC instructions for candidates clearly state that CTTIC is “not a teaching institution” and that the “exam is only a tool used to assess candidates' knowledge and capabilities, not a development tool” (CTTIC 2016b). However, this is not quite the stand taken by the Finnish Authorised Translators' Examination Board, as “examinees are informed of the way in which assessment criteria are applied to their performance in the examination” (FNBE 2012: 9), and this is currently done by sending failed examinees a copy of their assessed translation.

5. Conclusions

In this article, we have given an overview of the assessment system used in the Finnish Authorised Translator's Examination. We focussed on how assessors assess the translations and gave an account of what they think about the assessment system. We also discussed the assessment systems in two certification examinations – ATA and CTTIC – and two quality assessment models – SAE and MQM-DQF.

As our overview points out, the assessment systems of the Finnish examination and of the ATA and CTTIC examinations resemble each other. The CTTIC system has the same division into categories as the Finnish system (errors in rendering the meaning of the source texts vs. errors of expression in the target language). In each examination, error types are marked and fault points are counted. There are, however, differences in how detailed the systems are. The ATA system is the most detailed with its 23 error types, whereas the Canadian system seems more holistic. The SAE and MQM-DQF systems fall in between these, having fewer than ten error types.

The analysis of the translation assignments in the Finnish examination, as well as our pilot survey among the assessors, provides grounds for a rethinking of the assessment system. The analysis showed that not all the error types are used, while some are used very often. Therefore, it might be useful to introduce a more holistic system and to reduce the number of error types. We would like to propose either a simplified version of the current error type division, corresponding to the CTTIC system with two error categories, or a new error division system based on the MQM-DQF classification with six or seven error types. The translations could be rated as pass / fail by using either fault points with a determined pass threshold, as in the SAE or CTTIC systems, or using descriptors as in the ATA system. In any case, our next step will be to test the different rating systems with a sample set of examination translations in order to study their usability, and to inquire the assessors' opinions with a more comprehensive survey than the pilot one reported in this paper.

References

- Angelelli, Claudia V. (2009): Using a rubric to assess translation ability. Defining the construct. In: Angelelli, Claudia / Jacobson, Holly E. [eds.] (2009): *Testing and Assessment in Translation and Interpreting Studies. A call for dialogue between research and practice*. Amsterdam / Philadelphia: John Benjamins. 13–47.
- ATA (2011) = American Translators Association: Rubric for Grading. ATA CERTIFICATION PROGRAM Version 2011. http://www.atanet.org/certification/aboutexams_rubic.pdf [19.6.2017].
- ATA (2016a) = American Translators Association: Certification Examination Overview. https://www.atanet.org/certification/aboutexams_overview.php [14.12.2016].
- ATA (2016b) = American Translators Association: Computerized Certification Exam. https://www.atanet.org/certification/aboutexams_computerized.php [14.12.2016].
- ATA (2016c) = American Translators Association: Flowchart for Error Grading. https://www.atanet.org/certification/aboutexams_flowchart.pdf [14.12.2016].
- ATA (2016d) = American Translators Association: Framework for Standardized Error Marking. https://www.atanet.org/certification/aboutexams_error.php [14.12.2016].
- Chesterman, Andrew (2001): Polemiikka virallisen kääntäjän kokeesta [Debate over the official translator's examination]. In: *Kääntäjä – Översättaren* 9, 10.
- CTTIC (2016a) = Canadian Translators, Terminologists and Interpreters Council: Certification. <http://www.cttic.org/certification.asp> [21.11.2016].
- CTTIC (2016b) = Canadian Translators, Terminologists and Interpreters Council: The Candidate's Guide for the CTTIC Standard Certification Examination in Translation. Revised February 2011. http://www.cttic.org/examDocs/CandidateGuide_En1011.pdf [13.12.2016].

- CTTIC (2016c) = Canadian Translators, Terminologists and Interpreters Council: CTTIC Standard Certification Translation Examination Marker's Guide. Revised March 2005. <http://www.cttic.org/examDocs/guide.markersE.pdf> [14.12.2016].
- FNBE (2012) = Finnish National Board of Education: Qualification requirements for authorised translators' examinations 2012. Regulations and Guidelines 22/011/2012.
- FNBE (2015) = Opetushallitus: Käsikirja 2015. Auktorisoidun kääntäjän tutkinto. Toimintaohjeet tutkintotehtävien arvioijalle [Finnish National Board of Education: Handbook 2015. Authorised translator's examination. Instructions for assessors of translation assignments].
- FNBE (2016a) = Opetushallitus: Ohjeita tutkintoon 12.11.2016 osallistuvalla [Finnish National Board of Education: Instructions to those participating in the examination on 12 Nov 2016]. http://www.oph.fi/koulutus_ja_tutkinnot/auktorisoidut_kaantajat/tutkintoon_osallistuminen/tutkinnon_suorittamisesta [14.10.2016].
- FNBE (2016b) = Opetushallitus: Yleistä auktorisoidusta kääntämisestä [Finnish National Board of Education: About authorised translating in general]. http://oph.fi/koulutus_ja_tutkinnot/auktorisoidut_kaantajat/usein_kysyttya/yleista [27.12.2016].
- Hale, Sandra B. / Garcia, Ignacio / Hlavac, Jim / Kim, Mira / Lai, Miranda / Turner, Barry / Slatyer, Helen (2012): Improvements to NAATI Testing. Development of a Conceptual Overview for a New Model for NAATI Standards, Testing and Assessment. The National Accreditation Authority for Translators and Interpreters (NAATI). <https://www.naati.com.au/media/1062/intfinalreport.pdf> [18.11.2016].
- Hlavac, Jim (2013): A Cross-National Overview of Translator and Interpreter Certification Procedures. In: *Translation & Interpreting. The International Journal for Translation & Interpreting Research* 5 (2), 32–65.
- Kivilehto, Marja (2016): *Käännösfunktion huomiotta jättäminen, joka johtaa epätäsmälliseen lopputulokseen*. Auktorisoidun kääntäjän tutkinnon käännöstehtävien arvioinnista [*The translation function is disregarded, leading*

- to an inadequate result. On assessing translation assignments in the authorised translator's examination]. In: Hirvonen, Pia / Rellstab, Daniel / Siponkoski, Nestori [eds.] (2016): *Text and textuality*. Vaasa: University of Vaasa. (= VAKKI Publications. 7). 391–401.*
- Kivilehto, Marja (2017): Miten auktorisoidun kääntäjän tutkinnon käännöstehtävät vastaavat tutkinnon tavoitteita erikoisalojen kääntämisen näkökulmasta? [How do the translation assignments in the authorised translator's examination meet the requirements of the examination from the point of view of specialised translation?]. In: Hartama-Heinonen, Ritva / Kivilehto, Marja / Laukkanen, Liisa / Ruokonen, Minna [eds.] (2017): *MikaEL. Electronic Journal of the KäTu Symposium on Translation and Interpreting Studies, Vol. 10*. <https://www.sktl.fi/liitto/seminaarit/mikael-verkkojulkaisu/> [15.7.2017]. 136–149.
- Kivilehto, Marja / Salmi, Leena (forthcoming): Translation assessment system in translator authorisation in Finland. In: Koby, Geoffrey / Lacruz, Isabel [eds.]: *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 14.
- Lommel, Arle / Görög, Attila / Melby, Alan / Uszkoreit, Hans / Burchardt, Aljoscha / Popović, Maja (2015): Harmonised Metric. Project Report, QT21 project (funded by the European Union's Horizon 2020 program for ICT). <https://www.taus.net/evaluate/dqf-tools#error-typology> [17.12.2016].
- Melby, Alan K. (2013): Introduction. In: *Translation & Interpreting. The International Journal for Translation & Interpreting Research* 5 (1), 1–12.
- Nord, Christiane (1997): *Translating as a purposeful activity: Functionalist approaches explained*. Manchester: St. Jerome.
- O'Brien, Sharon (2012): Towards a Dynamic Quality Evaluation Model for Translation. In: *The Journal of Specialised Translation* 17, 55–77.
- Pym, Anthony / Grin, François / Sfreddo, Claudio / Chan, Andy L. J. (2012): The Status of the Translation Profession in the European Union. Studies on Translation and Multilingualism, DGT/2011/TST. Brussels: European Commission. http://ec.europa.eu/dgs/translation/publications/studies/index_en.htm [23.10.2016].

- SAE International (2016): Surface Vehicle Recommended Practice. Translation Quality Metric. SAE J2450, August 2016.
- Salmi, Leena / Kinnunen, Tuija (2015): Training translators for accreditation in Finland. In: *Interpreter and Translator Trainer* 9 (2), 229–242.
- Taibi, Mustapha / Ozolins, Uldis (2016): *Community Translation*. London / Oxford / New York / New Delhi / Sydney: Bloomsbury.
- TAUS (2016) = Translation Automation User Society: Harmonized DQF-MQM Error Typology. <https://www.taus.net/evaluate/qt21-project#harmonized-error-typology> [17.12.2016].
- Turner, Barry /Lai, Miranda / Huang, Neng (2010): Error Deduction and Descriptors – A Comparison of Two Methods of Translation Test Assessment. In: *Translation & Interpreting. The International Journal for Translation & Interpreting Research* 2 (1), 11–23.
- Vigier, Francisco / Klein, Perla / Festinger, Nancy (2013): Certified Translators in Europe and the Americas: Accreditation Practices and Challenges. In: Borja Albi, Anabel / Prieto Ramos, Fernando [eds.] (2013): *New trends in Translation Studies, vol. 4: Legal Translation in Context: Professional Issues and Prospects*. Oxford: Peter Lang. 27–51.

Appendix 1

Scoring chart used by the assessors of the Authorised Translator's Examination in 2012–2014

Error category	Error type		Points
Equivalence of content (C); precise and faultless use of special terminology.	An idea is completely misinterpreted An entire sentence is left out	C1	9 p. leads to a failed examination
	A wrong term leading to the misinterpretation of the translation	C2	9 p. leads to a failed examination
	The translation function is disregarded, leading to an inadequate result	C3	6 p. may lead to a failed examination
	Unfounded alternative translation equivalents, i.e. the choice is left to the evaluator	C4	6 p. may lead to a failed examination
	An omission or an addition essentially affecting the meaning of the text, e.g. a general and crucial abbreviation is not translated	C5	6–4 p. depending on the severity of the omission; 6 p. may lead to a failed examination
	Misinterpreted structure	C6	6–2 p.
	An individual word/term that is imprecise, unsuitable or irrelevant for the content or culture but does not necessarily lead to the misinterpretation of the translation An omission or an addition not essentially affecting the meaning of the text	C7	4–2 p.
	Incomplete or erroneous equivalents for the cultural and social context of the source language	C8	2 p.
A structural error that is likely to cause misinterpretation	A1	6–4 p. depending on the severity of the error; 6 p. may lead to a failed examination	
Acceptability and readability of text (A). General acceptability and readability of text; usage according to orthographical, morphological and syntactic norms; register and style correspond to the text function and the intended use of the translation.	Inconsistent terminology or style	A2	6–2 p.
	A spelling mistake that affects the interpretation of the text section	A3	4–2 p.
	Inadequate translation in terms of the information structure of the text	A4	2 p.
	A structural error that does not cause misinterpretation	A5	2 p.
	Individual style errors and unidiomatic expressions	A6	2–1 p.
	A spelling mistake that does not affect the interpretation of the text section	A7	1 p.