UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Programa de Doctorado en Territorio y Sociedad: Evolución histórica de un espacio tricontinental (África, América y Europa)

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Methodologies applied to translation quality management and evaluation in professional environments

Metodologías aplicadas en la gestión y evaluación de la calidad de la traducción en entornos profesionales

Tesis presentada por

Jennifer Vela Valido

Para la obtención del grado de Doctora por la Universidad de Las Palmas de Gran Canaria con Mención de Doctorado Internacional

Directora: Dra. Laura Cruz García

Codirector: Dr. Jorge Díaz Cintas

LAS PALMAS DE GRAN CANARIA, 2021

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Dª Amelia Rodríguez Rodríguez, COORDINADORA DEL PROGRAMA DE DOCTORADO 6022 - Programa de Doctorado en Territorio y Sociedad: Evolución histórica de un espacio tricontinental (África, América y Europa) (DOCTESO) DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA,

INFORMA,

Que la Comisión Académica del Programa de Doctorado, en su sesión de fecha .... tomó el acuerdo de dar el consentimiento para su tramitación, a la tesis doctoral titulada "Methodologies applied to translation quality management and evaluation in professional environments" presentada por la doctoranda Dª Jennifer Vela Valido, dirigida por la Doctora Laura Cruz García y codirigida por el Doctor Jorge Díaz Cintas.

Y para que así conste, y a efectos de lo previsto en el Artº 11 del Reglamento de Estudios de Doctorado (BOULPGC 7/10/2016) de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a…….... de... de dos mil............

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

ESCUELA DE DOCTORADO

Programa de doctorado *Territorio y Sociedad: Evolución histórica de un espacio tricontinental (África, América y Europa) (DOCTESO)*

Título de la Tesis

Methodologies applied to translation quality management and evaluation in professional environments

Tesis Doctoral presentada por Dª Jennifer Vela Valido

Dirigida por la Dra. Dª. Laura Cruz García

Codirigida por el Dr. D. Jorge Díaz Cintas

La Directora   **El Codirector**   **La Doctoranda**

Las Palmas de Gran Canaria, a 17 de diciembre de 2021

Resumen

Introducción

Hoy en día, las diferentes industrias profesionales cuentan con diversas normativas y estándares de control de calidad que regulan los procesos que deben seguir las empresas para gestionar la calidad de sus productos y servicios en su ámbito específico. La industria de la traducción no es una excepción, por lo que existen diversos estudios académicos, iniciativas profesionales y esfuerzos normativos que pretenden definir y estandarizar los procesos y las prácticas que deben seguirse para garantizar la calidad de una traducción. Sin embargo, a pesar de que las iniciatives en estos tres ámbitos (académico, profesional y normativo) tienen el mismo objetivo común, en ocasiones parecen seguir enfoques paralelos y, a veces, contradictorios.

La mayoría de los investigadores y profesionales del ámbito de la traducción coinciden en que el panorama de la traducción profesional ha evolucionado considerablemente desde 1990 (Thelen, 2008; Drugan, 2013; Esselink, 2019). La llegada de Internet, la llamada "revolución digital" y la globalización trajeron consigo un aumento de la demanda de traducciones y el desarrollo de nuevas y mejores herramientas electrónicas que ayudaran a la industria de la traducción a satisfacer este aumento de la demanda. Estos acontecimientos también trajeron consigo una mayor concienciación sobre el papel de la industria de la traducción como facilitadora de otras industrias, así como la necesidad de definir qué es realmente una " traducción de calidad" y cómo medir esta calidad.

Otro de los factores clave de la importancia de esta área de la traducción ha sido el desarrollo de estándares, directrices y normativas sobre este tema. No es de extrañar que el sector de la traducción, como muchos otros sectores profesionales, haya aumentado gradualmente su presencia en organismos reguladores internacionales como el Comité Europeo de Normalización (CERN), la Organización Internacional de Normalización (ISO) o la Sociedad Americana de Pruebas y Materiales (ASTM). Todos estos organismos cuentan hoy en día con comités específicos que establecen directrices y normas que abarcan diferentes aspectos de la industria de la traducción, incluida la gestión, el control y la evaluación de su calidad.

Justificación

Aunque la calidad de una traducción es un aspecto esencial en los estudios académicos de traducción, las publicaciones de la industria y los esfuerzos de estandarización, son escasos los estudios interdisciplinarios que conectan estos tres ámbitos. Con algunas excepciones recientes, la bibliografía académica sobre este tema tiende a centrarse en lo que debería ser la calidad de la traducción en un contexto ideal sin tener en cuenta las limitaciones o prioridades de la industria de la traducción, ni los conceptos y procesos presentados en las normas y directrices internacionales sobre traducción y calidad de la traducción.

Una de las razones de esta falta de contextos reales en los estudios académicos sobre calidad de traducción es probablemente la dificultad para poder acceder a la propia industria y a fuentes de información propiedad de empresas privadas, con algunas excepciones notables gracias a la colaboración de organismos públicos como las Naciones Unidas (Cao & Zhao, 2008) y la Comisión Europea (Koskinen, 2014). Esta limitación, así como la necesidad de realizar estudios de traducción descriptivos, han sido reconocidas por varios investigadores como Holmes (1988), que ya en 1988 recomendaba intentar tener un contacto lo más estrecho posible con los fenómenos empíricos objeto de estudio.

Otra de las lagunas identificadas más adelante, a principios del siglo XXI, por investigadores como Chesterman y Williams en su publicación *The Map. A beginner's Guide to Doing Research in Translation Studies* (Chesterman & Williams, 2002) es que, a pesar de que la tecnología forma parte de la profesión de la traducción desde hace un par de décadas, no se han investigado suficientemente muchos aspectos de la tecnología de la traducción, los flujos de trabajo asociados y los mecanismos de gestión, evaluación y control de calidad.

Por otro lado, si examinamos las publicaciones profesionales del sector de la traducción, vemos que suelen tener un tamaño y un alcance bastante limitados, ya que tienden a describir las prácticas y limitaciones de un entorno concreto (un tipo de empresa, un tipo de proveedor o un tipo de traducción). Además, dada la preeminencia de los enfoques puramente descriptivos, estas publicaciones no ofrecen marcos comparativos con otros casos de uso o modelos teóricos o normativos. Es decir, se centran en la definición de las prácticas de gestión, evaluación y control de calidad en un entorno concreto y no ofrecen orientaciones ni modelos con base científica.

En cuanto a la regulación y estandarización de los procesos de gestión de calidad de la traducción, en la última década se han publicado varias normas relacionadas con los servicios de traducción y la calidad de la traducción, siendo ISO y ASTM las organizaciones más activas en este sentido. Sin embargo, es ciertamente difícil encontrar estudios empíricos sobre el impacto de estas normas en el mundo académico y profesional. Estas normas también parecen tener un limitado impacto, ya que ni los compradores de servicios de traducción ni los proveedores de servicios de traducción están formalmente obligados a cumplir con estas normas internacionales (con la excepción de ciertos organismos internacionales que sí exigen una certificación a los proveedores de servicios de traducción que desean trabajar con ellos).

La desconexión y las limitaciones en el alcance de los enfoques académico, profesional y normativo plantean varios retos para el avance del campo de la gestión de la calidad de la traducción y su aplicación a la industria de la traducción. ¿Pueden utilizarse los modelos teóricos en los enfoques prácticos de la gestión de la calidad de la traducción? ¿Qué pueden aprender los profesionales del sector a partir de un estudio descriptivo de las prácticas actuales de gestión de la calidad de la traducción llevadas a cabo por los proveedores de servicios de traducción y las empresas con equipos internos de gestión de la calidad de la traducción? ¿Cuál es el impacto real y las oportunidades futuras de las diferentes normas y directrices del sector?

Objetivos

El objetivo principal de esta tesis es dar respuesta a las preguntas planteadas con anterioridad a través de una descripción del estado de la cuestión en cada uno de los tres enfoques (académico, industrial y normativo) y de una investigación empírica sobre las prácticas actuales de gestión y evaluación de la calidad en la industria de la traducción.

Para alcanzar este objetivo global, esta tesis propone seis preguntas de investigación:

1. ¿Cuál es el perfil demográfico y profesional de las empresas y los participantes que realizan tareas de gestión y evaluación de la calidad de la traducción?
2. ¿Cuál es el nivel de madurez en términos de procesos y análisis de la traducción?
3. ¿Cuál es el nivel de madurez de los procesos de gestión de la calidad de la traducción?
3.1. ¿Cuál es el nivel de madurez global en la gestión de la calidad de la traducción?

Metodología

La revisión de los enfoques académico, profesional y nos proporciona tres ángulos diferentes sobre cómo debe gestionarse y evaluarse la calidad de la traducción en contextos profesionales. Sin embargo, cada uno de ellos presenta diferentes limitaciones en cuanto a su alcance y aplicabilidad, y es difícil encontrar ejemplos de su aplicación en la industria de la traducción. Por ello, esta tesis pretende complementar esta investigación descriptiva con una investigación empírica destinada a describir las prácticas actuales de gestión y evaluación de la calidad en la industria de la traducción y analizar la adopción, el impacto y la idoneidad de los enfoques propuestos por los organismos académicos, profesionales y normativos.

Tras revisar la bibliografía disponible sobre prácticas de gestión y evaluación de la calidad de la traducción, no se pudo identificar ningún marco o metodología clara que pudiera servir de guía para esta investigación empírica. Sin embargo, sí se identificaron diferentes métodos

específicos que podían ser adecuados para esta investigación, por lo que se decidió utilizar una combinación de métodos cualitativos y cuantitativos, también llamado enfoque de métodos mixtos (Creswell et al., 2003).

Los datos empíricos de este estudio se han recogido utilizando los métodos de encuesta y entrevista. Ambos métodos contaron con un cuestionario maestro que abarcaba las seis preguntas de investigación mencionadas anteriormente, con cuatro variaciones para adaptarse a los distintos grupos de informantes: empresas compradoras de servicios de traducción, empresas proveedoras de servicios de traducción, expertos en gestión y evaluación de la calidad de la traducción y empresas proveedoras de tecnologías de la traducción.

Los cuatro tipos de cuestionarios se publicaron juntos a través de una herramienta de encuestas en línea y comenzaban con un conjunto de preguntas demográficas compartidas. Tras este primer conjunto de preguntas, cada cuestionario tenía diferentes preguntas en función del perfil profesional de los participantes. El objetivo de esta primera sección demográfica compartida era relacionar la demografía del perfil de los encuestados y las empresas con las respuestas dadas a las preguntas de la investigación.

Las preguntas se diseñaron basándose en fuentes como la literatura académica revisada en la sección 2.2, las metodologías profesionales presentadas en la sección 2.3 y los enfoques normativos descritos en la sección 2.4. También se consultaron varias encuestas del sector de la traducción realizadas por organizaciones de investigación del sector, como Nimdzi (2021) y la Encuesta Europea del Sector de los Idiomas ELIS (ELIA et al., 2021), coorganizada por EUATC, ELIA, FIT Europe, GALA y la red universitaria EMT. Estas dos encuestas fueron especialmente relevantes en el diseño de la demografía del estudio y proporcionaron algunas orientaciones para ciertas preguntas relacionadas con las normas y certificaciones, la composición del personal, los principales avances, las prácticas operativas, la tecnología y las tendencias.

Por último, esta investigación contó con diferentes métodos de recogida y análisis de datos, que se realizaron por separado. Los resultados de cada uno de los grupos de datos recogidos se fusionaron posteriormente en la fase de análisis y discusión.

Conclusiones

En total, se recogieron más de 4.000 puntos de medición de los 68 participantes que formaron parte del estudio.

Los resultados de la encuesta realizada para el estudio empírico sugieren que, en general, las empresas internacionales con grandes volúmenes de traducción y varios idiomas tienden a tener una necesidad más importante de programas de gestión y evaluación de la calidad y que el tipo de recursos humanos más comúnmente implicados en la gestión de la calidad de la traducción son los gestores o directores de traducción/localización, los gestores de calidad de traducción/localización y los gestores de programas de traducción/localización.

Los datos recopilados también indican que más del 50% de los compradores de servicios de traducción no tienen en cuenta ninguna norma sobre servicios de traducción o gestión de la calidad para elaborar sus programas de gestión de la calidad de las traducciones. Este porcentaje se eleva al 70% en el caso de los proveedores de servicios de traducción, al 87% en el caso de los proveedores de tecnología de la traducción y al 100% en el caso de los expertos en gestión de la calidad de la traducción.

Se han observado resultados similares en cuanto al uso de métricas estandarizadas, ya que menos del 50% de los compradores de traducciones utilizan algún modelo existente para diseñar sus métricas, y el 23% de ellos no utiliza ninguna métrica. Por otro lado, casi el 50% de los proveedores de servicios de traducción utilizan métricas estandarizadas como MQM-DQF o LISA QA, y el 21% no utiliza ninguna métrica. En cambio, el 77% de los expertos en gestión de la calidad de la traducción están familiarizados con estas normas (en particular las normas ISO) y las métricas de evaluación como MQM-DQF o LISA QA, y el 75% de las soluciones tecnológicas de traducción están diseñadas para ser compatibles con las métricas de evaluación estandarizadas mencionadas anteriormente.

Los resultados de la encuesta realizada también sugieren que los compradores de servicios de traducción tienden a tener un sistema de calificación con dos niveles (aprobado/suspendido), umbrales definidos y un sistema de puntuación numérica. Además, suelen utilizar el modelo MQM-DQF como base para su metodología, que incluye cuatro categorías principales de error (Precisión, Fluidez, Terminología y Estilo) y tres niveles de error (Menor, Mayor, Crítico). Al igual que los compradores de servicios de traducción, los proveedores de servicios de traducción utilizan los modelos MQM-DQF o LISA QA como

base para su metodología, que incluye cuatro categorías principales de error (Precisión, Fluidez, Terminología y Estilo) y tres niveles de error (Menor, Mayor, Crítico). Sin embargo, su sistema de calificación es más complejo porque utiliza entre tres y cinco niveles, umbrales definidos de aprobado/suspendido y un sistema de puntuación numérica.

Según los datos analizados, los tipos de funciones más habituales en los programas de gestión de la calidad de la traducción de los compradores de servicios de traducción son principalmente los revisores y correctores, seguidos de los traductores, los gestores de calidad y los responsables lingüísticos. Solo el 38,7% de los participantes mencionó tener evaluadores ejerciendo esta función específica. En cuanto a la formación y las cualificaciones, los compradores de servicios de traducción prefieren evaluadores con experiencia como traductores y revisores, y consideran que los gestores de la calidad de la traducción deben tener competencias culturales y lingüísticas, experiencia o certificación en estudios lingüísticos y experiencia en programas de gestión de la calidad de la traducción. Los proveedores de servicios de traducción siguen un planteamiento similar, y solo el 35,7% de estos participantes mencionó tener evaluadores ejerciendo esta función específica. En cuanto a la formación y las cualificaciones, los proveedores de servicios de traducción suelen ofrecer formación específica para los evaluadores. Los proveedores de servicios de traducción también consideran que los gestores de la calidad de la traducción deben tener titulación académica o experiencia como traductores, capacidad analítica demostrada y experiencia en programas de gestión de la calidad de la traducción.

Los datos recopilados en este estudio sugieren que la mayoría de los compradores y proveedores de servicios de traducción utilizan los módulos disponibles en las herramientas comerciales CAT o TMS, tanto para la gestión de la calidad como para la evaluación. Sin embargo, los mayores proveedores de servicios de traducción tienden a utilizar herramientas independientes de gestión y evaluación de la calidad de la traducción o a desarrollar herramientas propias, mientras que los proveedores de servicios de traducción más pequeños no utilizan ninguna herramienta o confían en hojas de cálculo para realizar estas tareas. Esta práctica es confirmada por la mayoría de los expertos en calidad de la traducción que participaron en el estudio. Un número importante de expertos también recomienda utilizar herramientas propias de gestión de la calidad de la traducción (siempre que sea posible) o herramientas específicas de calidad de la traducción como ContentQuo, QA Distiller, Verifika, Xbench o TQAuditor.

Los principales problemas en la gestión de la calidad de la traducción para los compradores de servicios de traducción son la falta de adecuación de las herramientas utilizadas para la gestión y evaluación de la calidad de la traducción, la falta de recursos humanos dedicados y la necesidad de establecer métricas y umbrales de evaluación claros, escalables y personalizables que puedan utilizarse para evaluaciones analíticas y holísticas. Para los proveedores de servicios de traducción, los principales problemas son, de nuevo, la falta de adecuación de las herramientas utilizadas para la gestión y evaluación de la calidad de la traducción, la necesidad de establecer metodologías de gestión de la calidad de la traducción rentables y flexibles y las limitaciones de tiempo que encuentran para realizar las tareas de gestión y evaluación de la calidad de la traducción. Desde el punto de vista de los expertos en gestión de la calidad de la traducción, los principales puntos débiles que han observado son la necesidad de establecer metodologías de gestión de la calidad de la traducción rentables y flexibles y el hecho de que los modelos de evaluación actuales son lentos, caros y rígidos.

Las principales soluciones adoptadas por los compradores de servicios de traducción para superar los retos de gestión de la calidad de la traducción mencionados son el uso de herramientas más avanzadas (desarrolladas interna o externamente), la adopción de un enfoque interfuncional y entre equipos para la calidad de la traducción, y conocer los últimos avances en este campo de mano de los proveedores de servicios de traducción o de los proveedores de tecnología de traducción. Los proveedores de servicios de traducción proponen el aumento de los recursos humanos especializados en la gestión de la calidad de la traducción (interna o externamente), la inversión en herramientas de automatización y la búsqueda de nuevas herramientas para realizar la gestión o la evaluación de la calidad de la traducción. Entre las soluciones apuntadas por los expertos en calidad de la traducción se encuentran el desarrollo de metodologías holísticas de gestión de la calidad de la traducción, el uso de tecnologías para realizar evaluaciones asistidas por ordenador y la publicación de reglamentos o normas para contratar y formar a evaluadores cualificados.

Acknowledgements

The path toward this dissertation has been bewildering, and its completion is thanks in large part to several people who have challenged, supported and inspired me along the way.

First and foremost, I would like to thank my supervisors Laura Cruz García and Jorge Díaz Cintas for their incredible support and guidance, as well as their generous mentorship over the past few years.

Many thanks to Tabea De Ville and Alejandro Bolaños García Escribano, for their feedback and help during the last stages of this dissertation.

Special thanks go to my friends, my whole family and my parents, Ana Estrella Valido Cruz and Angel Luis Vela Marín, who have always encouraged me and emboldened me to pursue my education and never give up.

I would also like to thank my partner, Tiago Rocha, who has been my rock, proofreader, sounding board and lighthouse along this process.

Finally, I want to thank all those that have agreed to be interviewed, fill questionnaires, provide feedback, and share insights into the different realities of the translation industry. I am enormously grateful for your time and generosity and I hope this research inspire many more studies on this topic.

Table of contents

List of tables

List of figures

xxi

1 Introduction

1.1 Introduction and overview

Nowadays, every industry has several regulations and quality control standards that regulate the processes that companies must follow to manage the quality of their products and services in their specific field. The translation industry is not an exception, and there are several academic studies, industry approaches, and regulatory efforts that aim to define, standardise and guide the processes and best practices that should be followed in order to ensure the quality of a translation (both as a process and as a product). Even though these three perspectives have the same common goal, they seem to follow parallel and, sometimes, contradictory approaches.

Most researchers and professionals in the field of translation agree that the professional translation landscape has evolved significantly since the 1990s (Thelen, 2008; Drugan, 2013; Esselink, 2019). The advent of the Internet, the "digital revolution", and globalization brought an increase in the demand for translations and the development of new and better electronic tools that would allow the translation industry to meet this surge in demand. These developments also brought an increased awareness of the role of the translation industry as an enabler to other industries and the need to define and agree on what "translation quality" really means and to whom.

The other key driver of this focus on translation quality in the industry has been the development of standards, guidelines, and norms covering this topic. It is not a surprise that the translation industry, as many other mature industries, is increasing its presence and weight in international regulatory bodies such as the European Committee for Standardization (CERN), the International Organization for Standardization (ISO), or the American Society for Testing and Materials (ASTM). These bodies count on specific committees that establish objective and standardised guidelines and norms encompassing different aspects of the translation industry, including translation quality. This increase in the number of norms and standards covering the topic of translation quality runs parallel to the increase in academic translation studies and industry publications looking into the same topic but from different perspectives.

1.2 Justification

Even though translation quality is an essential aspect in academic translation studies, industry publications, and standardisation efforts, interdisciplinary studies connecting these three angles are rare. With a couple of recent exceptions, literature tends to focus its efforts on what translation quality "should be" without considering the translation industry's limitations or priorities, or the concepts and processes presented in international norms and guidelines on translation and translation quality. This lack of real-world contexts in translation studies is probably influenced by the difficulty for academic researchers to be able to access the industry itself and to sources of information owned by private companies, with a few notable exceptions thanks to the collaboration of public organisations such as the United Nations (Cao & Zhao, 2008), the European Commission (Koskinen, 2014) and the European Parliament (Drugan et al., 2018). This limitation and the need for descriptive translation studies have been recognised by several researchers such as Holmes, as they are focused on maintaining "the closest contact with the empirical phenomena under study" (Holmes, 1988:176).

Other gaps identified by researchers such as Chesterman and Williams in their publication *The Map. A beginner's Guide to Doing Research in Translation Studies* (Chesterman & Williams, 2002) is that even though technology has been an integral part of the translation profession for a couple of decades, there has not been enough research on many aspects of the technology, as well as workflows, translation processes and "mechanisms of quality control" (Chesterman & Williams, 2002:15)

If we examine industry publications, they are often quite limited in size and scope, as they tend to describe the practices and limitations of a particular environment (a type of company, a type of provider, or a type of translation). Moreover, given the pre-eminence of purely descriptive approaches, these publications do not offer comparative frameworks with other use cases or theoretical or regulatory models. Therefore, they focus on what translation quality "is" in a particular environment and do not offer guidance or scientific-based models.

Additionally, if we take a look at industry surveys conducted by industry research organisations such as Slator (Slator, 2021), CSA Research (CSA Research, 2021b), Nimdzi (Nimdzi, 2021) and professional associations like EUATC, ELIA, FIT Europe, GALA (ELIA et al., 2021), we can see that they cover topics such as certifications, staff composition, major developments or services lines. However, none of them provides specific

information about the current practices on translation quality management and evaluation in the industry.

In terms of regulation and standardisation, quality management as a discipline had a significant impulse in the late 1980s with the publication of the ISO 9000 series, which covered a broad range of industries, including the translation industry. In 2000, the certification of this standard was established, giving way to a significant rise in the number of companies (primarily translation service providers) that went through the certification processes. In the last decade, many more standards relating to translation services and translation quality have been published, being ISO and ASTM the most active standardisation organisations in this regard. However, these standards seem to have certain limitations, and empirical studies on their impact on the academic and the professional world are pretty difficult to find. The primary limitations observed are their limited impact on the end-clients (also called "buyers), the fact that providers are not formally bound to these international standards (with the exception of those international bodies that do require a certification from those translation service providers that wish to work with them, or to participate in international tenders), and the need for relevancy and accuracy with simplicity and usability.

The disconnection and limitations in the scope of the academic, professional, and regulatory approaches bring several challenges to the advancement of the translation quality management field and its application to the translation industry. Can theoretical models be used in practical approaches to translation quality management? What can practitioners and professionals learn from a descriptive study of the current practices in translation quality management carried out by translation service providers and companies with internal translation quality management teams? What are the real impact and future opportunities of the different norms and guidelines within the industry?

1.3 Research questions

The primary purpose of this dissertation is to provide an answer to the questions posed above by describing three approaches on translation quality management and evaluation (academic, industry, and regulatory) and carrying out an empirical research on the current quality management and evaluation practices in the translation industry. With the above goal in mind, this study proposes six research questions:

1. What is the demographic and professional profile of the companies and participants that carry out translation quality management and evaluation tasks?
2. What is the level of maturity in terms of translation processes and analytics?
3. What is the level of maturity in terms of translation quality management processes?
   3.1. What is the overall level of maturity in translation quality management?
   3.2. What is the overall level of knowledge and use of translation quality management norms and standards?
   3.3. What are the documented processes and human resources used in translation quality management programs?
      3.3.1. How does the translation quality management program work?
      3.3.2. What are the translation quality evaluation methodologies used?
      3.3.3. How does the translation quality evaluation process work?
      3.3.4. What kind of human resources are involved in translation quality management and evaluation processes?
   3.4. What are the main tools and technologies used in translation quality management and evaluation processes?
4. What are the main pain points and limitations in translation quality management and evaluation processes?
5. What solutions and workarounds are applied in the industry to overcome the current pain points and limitations?
6. What are the future challenges and trends in translation quality management, and what kind of initiatives would help the advancement of the translation quality management practices in professional environments?

1.4 Thesis structure

This dissertation will first offer a literature review divided into four main sections.

The first section presents the key definitions of the main concepts related to translation quality as a product and as a process, while the second section offers a description of the most relevant academic approaches and theoretical models to translation quality. The third section focuses on the current professional approaches to translation quality, the similarities and differences between academic and professional approaches, and the tools and technologies used by the industry Finally, the fourth section reviews the main norms, guidelines, and standardisation efforts in translation quality management and evaluation.

After the literature review, the methodology used to design and carry out an empirical study of the translation industry's current quality management and evaluation practices will be presented.

The next section of this dissertation will cover the language data and the analysis of the research's quantitative and qualitative results. This research covers four different angles: end-clients of translation services (also called "buyers"), providers of translation services (or TSPs), translation technology companies, and subject matter experts and researchers specialised in translation quality management methodologies.

After the presentation of the results of the study, an attempt will be made to categorise the main trends and insights gained with this study. These conclusions will give way to further recommendations on future lines of research that will widen the scope of this study, confirm or qualify the results presented, and overcome its inherent limitations.

The translation field comprises a great number of different tasks that sometimes are considered a hybrid between translation and other types of content creation or modification, such as transcreation, respeaking, or voice to text translation. Since these modalities have particular characteristics and limitations that set them apart from the traditional translation approaches, this dissertation does not attempt to cover these "hybrid" tasks but focus on the ones that can be treated in a more homogeneous way. Furthermore, the scope of this dissertation is the translation of written texts into written texts without any audiovisual elements. Therefore, this study does not include audiovisual translation scenarios (such as subtitling, voice-over, or dubbing), but localization ones (such as web translation or software translation) are considered here. Finally, this dissertation will be covering management and evaluation methodologies performed by humans on both human translations and machine translations. These variances will be briefly described whenever the methodologies vary depending on whether the translation has been done by a human translator or by a translation engine.

The approach followed in this dissertation is two-fold: The first aims to offer a descriptive study of the current approaches followed in academic, professional, and regulatory fields with the assumption that this effort will provide a compilation of valuable insights from these three different angles. The second approach is an empirical study based on the data compiled

through a series of surveys and interviews conducted to understand the current views and experiences on translation quality management and evaluation in the industry.

Ultimately, this dissertation will also offer detailed empirical information aimed to bridge the "academy-industry" divide by bringing theory and practice closer together with the ultimate purpose of contributing to the advance of translation quality management studies and their application to the professional methodologies.

2 Literature review

One of the main challenges that characterise the current panorama of the translation quality management processes is the multiplicity of approaches, standards, theories, norms, and terminology used in three parallel yet complementary levels: the academic field, the business field, and the regulatory field. For this reason, it is essential to establish a common ground among these three levels to be able to understand and use the most standardised terms and approaches in the translation quality management field. In order to achieve this goal, this dissertation will start by reviewing the efforts made to define the concept of "translation quality" and discussing some of the main challenges posed by the scarcity of standardisation and consistency in the application of the different steps associated with translation quality management systems. We will also provide a summary of the key terminology used in this doctoral dissertation, according to prescriptive bodies of knowledge, such as the ISO norms, the ASTM standards, and the DQF-MQM framework. In the second part of this chapter, we will proceed to review the most relevant literature available in the three levels mentioned above: the academic level (theoretical studies and academic research initiatives), the industry level (publications and research initiatives conducted by professional bodies and organisations) and, finally, the regulatory field (norms and standards published by regulatory bodies).

2.1 Defining translation quality

Even though the question of how to define quality in translation has been broadly researched and discussed from the second half of the 20th century onwards (House, 1981; Horguelin & Brunette, 1998; Horton, 1998; Hatim & Mason, 2014), there is still quite a lot of controversy on this topic, especially from the theoretical and academic viewpoint. This controversy is probably due to the fact that many scholars have traditionally defended that the quality in translation is a "subjective" concept, from Horguelin and Brunette (1998) to Stejskal (Stejskal, 2009). To solve this challenge, some researchers such as al-Qinai (2002) and Koby et al. (2014) recommend the use of the concept of "adequacy" instead of the degree of equivalence, as "quality is relative and absolutes of accuracy cease where the end-user (i.e. the client) imposes his own subjective preferences" (al-Qinai, 2002:498).

On the other hand, most of the literature from the industry (both from professional journals and standardisation bodies) defines the concept of "quality" as the ability of the translation to fulfil a client-defined set of parameters (Jiménez-Crespo, 2009). Furthermore, some industry experts such as Koby, Hague, and Lommel even describe five different approaches to quality applied to translation quality: transcendent, product, user, production, and value (Koby et al., 2014).

Given this lack of consensus, finding different definitions of "translation quality" is not surprising. Koby et al. (2014) propose to group these two types of definitions into "broad definitions" and "narrow definitions" as follows:

Broad definition of translation quality:

> A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs (Koby et al., 2014:416).

Narrow definition of translation quality:

> A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture (ibid).

We agree with Koby et al.'s affirmation that the broad definition seems to be better suited to define a translation quality management system rather than just translation quality, even though there is no consensus on which definition is better for the different types of stakeholders involved in the translation industry. However, there is a certain degree of consensus on some points, for example, that there are three key factors that should always be considered to assess the quality of a translation.

The first factor is that the concept of "quality" needs to be defined and agreed upon by the different parties involved (Fields et al., 2014; Koby et al., 2014; Mellinger, 2018). Here it is interesting to note that both academia and the industry agree that a quality translation is one that fits its purpose (Nord, 1997; O'Brien, 2012). The second step is to establish a

methodology with a quality assessment method that allows an objective measurement (Williams, 1989; Schäffner, 1997; Thelen, 2008). Finally, the assessment of the quality needs to be carried out in accordance with the definition of quality previously agreed upon and with the methodology chosen.

For the purpose of this doctoral dissertation, the narrow definition of translation quality mentioned above will be used, as it is also the one that reflects the most widespread trend observed in the industry and the standardisation bodies studied.

2.1.1 Lack of standardisation in translation quality definitions

Although translation quality management and assessment are considered essential topics in the areas of translation and localisation, industry representatives and academics tend to disagree on how to define and evaluate translation quality. On the one hand, "theorists and professionals overwhelmingly agree there is no single objective way to measure quality" (Drugan, 2013:35), and academic researchers typically focus more on theoretical and pedagogic approaches to translation quality (Castilho et al., 2018a). On the other, the translation and localisation industry is mainly focused on defining and applying different types of error typology models in order to be able to provide quantitative indicators of translation quality.

Another source of the lack of standardisation on this topic comes from the absence of alignment between the translation and the quality management disciplines. As some translation industry experts, such as Fields et al. point out, the translation industry and the discipline of generic quality management have not had much contact (Fields et al., 2014). Even though it is true that the translation industry, academia, and the standardisation bodies borrow some terms from quality management disciplines (for example, "quality assurance" and "quality control"), these terms are not used consistently. A more profound revision of the literature available on the current state of the translation quality management frameworks shows that there is also a lack of standardisation of this practice in the industry. Fields et al. consider that this disconnection between both dimensions might be caused by a limited relevance of the quality management concepts to the translation field.

In the third place, we can also observe a misalignment in the approaches taken to define and measure translation quality depending on whether this task is performed as part of the

production process or in a research study, also referred to as "micro-level" by Castilho et al. (2018a). If translation quality is considered within the scope of a production process (that is, in the industry environment), then its goal is "to ensure that a specified level of quality is identified, measured, and delivered to the client, buyer, end-user, etc., of translated content" (Castilho et al., 2018a:11). However, in a research environment, the goal is to analyse a change in quality from a previous translation or between different translation approaches.

Despite this absence of alignment and standardisation, there has been a joint effort coming from professional and academic bodies to apply and adapt the body of knowledge coming from the quality management discipline to the translation industry, and also to agree on specific methodologies and approaches, or at least to propose frameworks that would cater to the different goals expressed above. Part of this joint effort is probably due to the fact that this misalignment of terms and frameworks does not only affect the metrics or measures to be used in a translation management methodology but, more importantly, the design, workflow, and reporting of the tasks involved (Castilho et al., 2018a).

This negative impact is confirmed by the reports shared by professional organisations like Common Sense Advisory (CSA) (DePalma & Pielmeier, 2013), which mentions that the low level of agreement on the definition and measurements in translation quality management frameworks in the industry often tends to lead to mismatched expectations, lack of transparency and inconsistency in the translation quality management specifications. These risks are especially visible in human procedures, which have the potential to be more subjective and difficult to be measured in a consistent and repeatable way. However, we agree with researchers like Castilho et al. (2018b) or Jiménez-Crespo (2009), who consider that any evaluation method needs to constrain the inherent subjectivity of the human component as to minimise as much as possible any source of variance that might affect the measurement.

For the purpose of this doctoral dissertation, we will review the main approaches followed in three different dimensions: academic field, industry field, and regulatory field, to analyse what the common trends in these three dimensions are, and ultimately, how regulatory and academic practices can contribute to the advancement of the translation and localization industry.

2.1.2 Definitions of terms

As mentioned previously, one of the first challenges we encounter when reviewing the literature available on translation quality management is the terminological inconsistency coming from the professional world and academic research. For this reason, the definition of the different tasks associated with translation quality management varies depending on factors such as the approach (academic or professional), the goal (quality management and evaluation of the translation process as a whole, or just the translation product), or even the author of the resource (as different authors propose different terms and definitions of the same term).

After revising the most pertinent sources of information concerning translation quality management terminology, coming from academic, industry, and standardisation bodies of knowledge, we would like to propose the following compilation of the terminology that is currently more widely used in these three fields. This compilation of definitions includes sources such as:

Standards: ISO 9000 (a norm that defines the fundamentals and key terminology of quality management), ISO 17100:2015 (which establishes the requirements for translation services), ISO 20539:2019 (Translation, interpreting and related technology — Vocabulary), and ASTM F2575 (the standard guide for quality assurance in translation published by ASTM International). There are also references to WK46396, a forthcoming standard practice for analytic translation quality evaluation that is being developed by the ASTM Committee F43 on Language Services and Products. Most of the committee members are also part of the Multidimensional Quality Metrics (MQM) Community Group, and the definitions contained in the forthcoming standard practice are publicly available on the W3C consortium website.

Industry field: This dissertation makes use of the some of the definitions proposed by the MQM Community, a group of practitioners hosted by the W3C consortium, whose goal is to "foster the development of MQM for translation and localization quality assessment and its interoperability with W3C's Internationalization Tag Set (ITS) 2.0 recommendation" (W3C, 2021a). This group is led by Arle Lommel and Alan K. Melby, two of the most prominent scholars who specialised in translation quality research, and authors of several publications on this topic, some of which will also be referenced in the following section.

Research field: This dissertation also includes some of the definitions proposed by Lucia Specia (2010), a researcher specialised in machine translation quality evaluation and machine translation estimation systems.

We believe that the combination of these resources will allow the readers to have a clearer idea of the meaning and scope of certain terms that will be used throughout this doctoral dissertation and that sometimes are used inconsistently in the translation quality field.

To better understand the different translation quality tasks, the following sections of this chapter will group them in four categories (see Table 1): end to end quality processes, quality tasks that take place before the production phase, quality tasks that take place during the production phase, and quality tasks that take place after the production phase.

Translation quality management (TQM)		
Before production phase	**During production phase**	**After production phase**
Quality planning	Quality Assurance (QA)	Quality Assessment (TQA)/ Quality Evaluation (TQE)
MT Quality estimation (MTQE)	Quality Control* (QC)	

Table 1. Quality Management tasks, as outlined in the definitions of this article. *Note: Quality Control is considered in this article a linear step, as defined by the standard ASTM F2575.

2.1.2.1 End to end quality processes

Translation quality management (TQM): "The integration and coordination of management activities focused on ensuring the organization fulfils stakeholder requirements predictably, consistently, and reliably" (Lommel & Melby, 2018:4). The ASTM WK46369 proposal echoes the trend that is already followed by the principal actors in the translation industry and establishes that any translation quality management system should be compatible with the principles and key concepts contained in ISO 9000 (ASTM, 2021a). ISO 9000 also states that quality management "can include establishing quality policies (3.5.9)[1] and quality objectives (3.7.2), and processes (3.4.1) to achieve these quality objectives through quality planning (3.3.5), quality assurance (3.3.6), quality control (3.3.7), and quality improvement (3.3.8)." (ISO, 2015a).

---

[1] Note: The numbers between parenthesis refer to the section numbering of the standard cited.

2.1.2.2 Quality tasks that take place before the production phase

Quality planning: "Part of quality management (3.3.4) focused on setting quality objectives (3.7.2) and specifying necessary operational processes (3.4.1), and related resources to achieve the quality objectives" (ISO, 2015a). In the context of translation quality planning, these activities aim to design a system of policies, processes, and procedures that need to be followed to produce products (translations) that can meet stakeholder requirements (Lommel & Melby, 2018).

Machine translation quality estimation (MTQE): In the context of machine translation, quality estimation can be defined as a quality management task aimed at "estimating the quality of a system's output for a given input, without any information about the expected output" (Specia et al., 2010:40). In other words, quality estimation systems utilise automatic metrics to "predict whether a new source string will result in a good or bad translation" (Way, 2018:160) before the production phase, rather than assessing the MT segment after production by comparing how similar it is to different translation segments used as a reference (see definition of "Machine Translation Evaluation" later in this chapter).

2.1.2.3 Quality tasks that take place during the production phase

Quality assurance (QA): "Part of quality management (3.3.4) focused on providing confidence that quality requirements (3.6.5) will be fulfilled" (ISO, 2015a). In order to provide that confidence to the different stakeholders (management, customers, and even third parties), the assurance activities audit the quality processes and procedures put in place (Lommel & Melby, 2018). It is important to note that quality assurance is often used as a synonym for quality assessment in the industry and sometimes also in certain academic studies, and it is probably one of the terms that are used more inconsistently. Therefore, we have decided to follow the definition and scope proposed by regulatory bodies such as ISO and ASTM, whose published standards consider that the goal of the quality assurance workflows is to improve the product to the agreed quality. In contrast, the quality assessment activities (as we will see in the definition below) aim to evaluate the quality of the final product. For this reason, translation quality assurance activities usually take place during the production phase and can include the following sub-tasks:

- **Revision** (also referred to as the first step of the "editing" process in the standard ASTM F2575): "Bilingual examination of target language content (2.3.3) against

source language content (2.3.2) for its suitability for the agreed purpose" (ISO, 2015a). The standard ASTM F2575 also mentions that the main goal of the reviser is to check the accuracy of the translation and the correctness of the terminology (ASTM, 2014).

- **Review** (also referred to as the second step of the "editing" process in the standard ASTM F2575): "monolingual examination of target language content (…) for its suitability for the agreed purpose" (ISO, 2015a). According to ASTM F2575, the reviewer only focuses on the target text to check coherence and readability, although they can check the source text if necessary (ASTM, 2014).

- **Formatting and compilation**: This task might vary significantly depending on the characteristics of the project and the specifications, the applications used, and even the languages required (ASTM, 2014).

- **Proofreading and verification**: According to ASTM F2575, this can be a quality assurance step or part of the quality control step. It also can be performed after the editing phase or at the same time. In any case, the proofreader's mission is to focus on checking the target text for typographical errors, formatting issues, or incorrect spelling (ASTM, 2014).

Quality control (QC): "Part of quality management (3.3.4) focused on fulfilling quality requirements (3.6.5)" (ISO, 2015a). The standard ASTM F2575 considers that the translation QC step is linear and should consist "of random sampling or a full check of final deliverables or both as the last step in the process" (ASTM, 2014:10). However, more recent studies from authors such as Lommel and Melby (2018) consider that translation quality control activities should assess processes and performance in real-time, that is, during the whole production phase, to verify that the quality measures are being fulfilled.

2.1.2.4 Quality tasks that take place after the production phase

Quality assessment or evaluation (also referred to as "post-project review" or "post-mortem" in the standard ASTM F2575): "Performance evaluation procedure conducted at the end of a project to determine how well the project conformed to original specifications" (ASTM, 2014:3). Ideally, this step takes place before the delivery to the requester, although it can also be carried out by the requester when accepting the delivered translations to evaluate whether their quality requirements have indeed been fulfilled and compare the results against the Key Performance Metrics (KPIs) agreed. To avoid any confusion with the

abbreviation used for Quality Assurance (QA) and Translation Quality Assurance (TQA), some scholars (O'Brien, 2012; Doherty & Gaspari, 2013; Melby et al., 2014; Gladkoff et al., 2021) and forthcoming standards (ASTM, 2021a) lean towards the use of Translation Quality Evaluation (TQE) or Quality Evaluation (QE) instead of Translation Quality Assessment (TQA) or Quality Assessment (QA).

- **Analytic translation quality evaluation**: "Quality evaluation that identifies and tallies errors from an analytic metric and calculates quality measures and quality ratings using a suitable scoring model" (ASTM, 2021a; W3C, 2021b).
- **Holistic translation quality evaluation:** "Quality evaluation based on identifying overarching qualities such as readability and accuracy at the macro level" (ASTM, 2021a; W3C, 2021b).

Machine translation evaluation (MTE): Evaluation or assessment of Machine Translation systems via their output, either with human evaluations or with automatic metrics. "The main purpose of the state-of-the-art automatic evaluation metrics is to compare the output of an MT system, which are assumed to be good because they are human quality" (Castilho et al., 2018a:15). Some of the most popular automatic metrics used nowadays both in the industry and research projects are the Bilingual Evaluation Understudy, or BLEU (Papineni et al., 2001) and METEOR (Lavie & Agarwal, 2007). These automatic metrics can be used to evaluate how much effort would be required for post-editing, to assess the evolution and efficiency of different iterations of the engine used, to compare efficiency gains before and after an engine has been trained, and to assess how well different engines are suited to the type of text to be translated, the language pairs chosen, or the quality requirements of the translation project.

2.1.2.5 Other terms

Assessment: It refers to the process of evaluation. In translation, evaluators should adhere to four principles: using specific criteria; describing the purpose of the assessment; defining the levels of analysis to be carried out; and determining the indicators to be used in the process of assessment (Colina, 2008; Schäffner, 2011; Reiss, 2014).

Computer-aided translation (CAT): "Translation carried out using computer-aided translation tools" (ISO, 2019).

End-user (also referred to as "customer" in the standard ISO 9000:2015): ASTM F2575 defines customer as the "person who ultimately avails himself or herself of the translation, as opposed to various intermediate translation service providers who pass it on to the next client in a chain of suppliers" (ASTM, 2014:2).

Error: "Violation of a rule of good writing or good translation according to specifications" (ASTM, 2021a; W3C, 2021b).

- **Error annotation:** "Review by a human quality evaluator to find and flag errors manually in a writing product or translation product" (ASTM, 2021a; W3C, 2021b).

- **Error root cause:** "proximate cause at the end of a causal chain that leads to and is responsible for an error in the evaluation text" (ASTM, 2021a; W3C, 2021b).

- **Error severity level:** "one of a small set of error severity designations, ranging from neutral to critical, reflecting the effect of the error on the usability of the text" (ASTM, 2021a; W3C, 2021b).

- **Error type:** "Class of errors identified by error type IDs, error type names, definitions, and positions in a hierarchical organization" (ASTM, 2021a; W3C, 2021b).

- **Error typology:** "taxonomy of error types, including their error types, that can be assigned to errors in quality evaluations in order to characterize the nature of problems encountered" (ASTM, 2021a; W3C, 2021b).

Localization: ASTM F2575 defines localization as a "Cross-cultural communication process of preparing locale-specific versions of a product or service, consisting of translation of textual material into the language and textual conventions of the target locale" (ASTM, 2014:3). This standard also considers localization an "adaptation" not just of nontextual materials, but also of "input, output, and delivery mechanisms to meet the cultural, technical, and regulatory requirements of that locale" (ASTM, 2014:3). ASTM F2575 also mentions that localization includes the translation and adaption of computer software (although many practitioners include in this category other text types such as mobile applications, website pages, and videogames) and the preparation of a product or service for a particular locale and market.

Machine translation or automated translation (MT): "Mode of translation in which a computer program analyses a source text and produces a target text, typically without human

intervention at the actual time of translation" (ASTM, 2014:3). ASTM F2575 also explains that MT requires some type of human participation before the engine processes the source text (called "training" or "pre-editing") and after the translation is produced (called "post-editing"). See the definition of "post-editor" below.

Metric: "Procedure providing a means of measuring the quality of a product or service that results in a composite numeric value" (ASTM, 2014:3).

Post-editor: ASTM F2575 defines this role as a "translator who reviews a completed machine translation to validate the accuracy of the final target text with reference to the source text in order to ensure a defined degree of stylistic acceptability, and makes changes where necessary" (ASTM, 2014:3).

Quality score: "quality measure of a writing product or translation product, a multiple, usually 100, of the difference between 1 and the normed penalty total" (ASTM, 2021a; W3C, 2021b).

Requester: "Individual, department, company, or organization placing an order for a translation" (ASTM, 2014:4). This standard also mentions that the requesters can be external (representing outside clients) or internal (representing other departments inside the same company).

Specification: "Document that sets out detailed requirements to be satisfied by a translation product, and the procedures for checking conformity to these requirements (Based on: ISO 6707-2:2017(en), 3.2.22)" (ASTM, 2021a; W3C, 2021b).

Text type: According to ASTM F2575, a text type is considered a "c based on its function, format, or the specific intention of the author with respect to the target audience" (ASTM, 2014:4).

Translation modality: "Means by which a translation product is created, with respect to human and machine translation processes" (ASTM, 2021a; W3C, 2021b). The MQM group also mentions that translation modalities comprise three main groups: human translation, unedited machine translation (also called "raw" machine translation), and "post-edited" machine translation.

Translation product: According to the MQM group and the forthcoming standard ASTM F46396, translation as a product can be defined as the "translated content as formatted and

laid out in a document, web page, or application user interface, including text and complementary components, such as graphics, video, hyperlinks, and accessibility content" (ASTM, 2021a; W3C, 2021b).

Translation service provider (TSP) or Language service provider (LSP): ASTM F2575 defines a TSP as a "Company, department, or individual approached by the requester, providing professional translation services into one or multiple languages for the requester" (ASTM, 2014:4). ISO 20539:2019 also mentions that "the concepts of TSP and language service provider are connected by a generic relation, with the language service provider being the generic concept and TSP the specific concept" (ISO, 2019). According to this norm, translation service providers typically provide only translation services, although it recognises that they can also provide other language-related services

2.2 Academic approaches: Theoretical models for translation quality assessment

2.2.1 Academic approaches to quality from Translation studies

As mentioned earlier in the introduction of this thesis, translation quality has been an important topic in academic research since the beginning of the consideration of Translation Studies as a separate discipline. Moreover, almost from the very beginning, translation evaluation has been considered one of the most problematic areas as a field of studies, as "the activities of translation, interpretation and evaluation will always elude the grasp of objective analysis to some extend" (Holmes, 1988:190). However, this challenge has not discouraged translation theorists from trying to find approaches and methodologies to move away from "subjective, one-sided or dogmatic" quality assessments (House, 1997:84) and include not just product-based analysis (based on large multilingual corpora of translated texts), but also translation-process analysis and translation quality assessment models. As a consequence of this expanse in the scope, a great variety of approaches now coexist.

If we review the literature available and take a closer look at the first definitions of translation quality in the key theories that have guided the development of Translation Studies as an independent discipline, we can see that the focus was mostly towards translation criticism instead of empirical measurement (Castilho et al., 2018a). It would not be until the early years of the 21st Century that scholars such as House (2002) would try to swift this focus from subjective quality evaluation to an objective assessment of functional equivalence.

Drugan (2013) also states that translation studies theorists seem to disagree even on how many categories of translation quality assessment exists, a situation that, according to Castilho et al. (2018a) "suggests that research-oriented TQA models are heavily theoretically motivated, reflecting the assumptions, and to some extent the biases, of those who propose or adopt them".

2.2.1.1 Classification of approaches to translation quality assessment in academic research

Different scholars group translation quality assessment approaches in various ways, and there is widespread disagreement on how to classify these approaches (Drugan, 2013). We have summarised the main proposals to categorise in different translation quality approaches presented by different scholars from 1988 to 2007, following the analysis provided by Drugan (ibid).

	Category 1	Category 2	Category 3	Category 4
House (1977/1997)	Anecdotal and subjective	Response-oriented	Text-based	
Schäffner (1998)	Linguistic models	Supra linguistic models		
Lauscher (2000)	Equivalence-based	Functional		
William (2004)	Quantitative	Non-quantitative		
Chesterman (2007)	Retrospective	Prospective	Lateral	Introspective
Drugan (2013)	Theoretical	Applied		

Table 2. Comparison of classifications of translation quality assessment approaches based on Drugan's analysis (Drugan, 2013) [2].

The author also points out that none of the suggested classifications considers whether the researched approach is purely theoretical or has been extensively tested in the real world, and she proposes a new classification approach based on this criterion. One of the main reasons cited by Drugan is the fact that professional approaches tend to be excluded from translation theory or mentioned rather superficially (Drugan, 2013).

---

[2] The approaches marked in green cover some aspects that are present in the professional approaches to translation quality assessment.

2.2.2 Theoretical models of translation quality assessment

As discussed earlier in this chapter, researchers such as Drugan, Lauscher and Castilho et al. agree to point out the lack of theoretical translation quality assessment models that use a comprehensive and reproducible testing methodology. However, there are three main models that meet these criteria and that constitute a valuable starting point to illustrate the main theoretical approaches to translation quality assessment used and referenced in translation studies.

2.2.2.1 House´s model (1977 and 1997)

House's translation quality assessment model, published first in 1977 and revisited later in 1997, is undoubtedly one of the most influential and widespread models in the academic research field. House based her model on pragmatic theories of language use, Halliday's functional and systemic theories, and stylistics and discourse analysis (House, 1977/1981). The notion of "equivalence" is one of the core notions discussed in this model, as it is rooted in the understanding that translations are texts that are doubly constrained: to the source text and to the recipient's communicative environment. House's models also connect the notion of "equivalence" to the preservation of meaning in three different levels: semantic, pragmatic, and textual.

In terms of the practical application of the model, House outlines three different stages, as shown in Table 3 below:

House's Translation Quality Assessment Model (1977)		
Stage 1	Analysis of Source Text and Statement of Function	Key features across 5 main areas: - Medium - Participation - Social role relationship - Social attitude - Province
Stage 2	Source Text and Target Text Comparison	Types of errors: - Overly erroneous - Covertly erroneous
Stage 3	Statement of Quality	

Table 3. Stages of the translation quality assessment model proposed by House in 1977 (House, 1997/1981).

This first model was empirically tested with a corpus of eight pairs of English source texts and German target texts. The documents included "a scientific text, an economic text, a journalistic article and a tourist information brochure […] an excerpt of a sermon, a political speech" (House, 1997/1981).

The results of the test cases and the response from the academia led House to revise her model in 1997 to refine it by adapting the terminology used and incorporating new research that had appeared after the 1977 version was published.

House's Revised Translation Quality Assessment Model (1997)		
Stage 1	Analysis of Source Text and Statement of Function	Key features across 4 main areas: • Field • Tenor • Mode • Genre
Stage 2	Source Text and Target Text Comparison	Types of errors across 4 main areas: • Field • Tenor • Mode • Genre
Stage 3	Statement of Quality	

Table 4. Stages of the translation quality assessment model proposed by House in 1997 (House, 1997).

This revisited model was tested in four pairs of source texts and their target texts. The documents included a children's book, an excerpt from an autobiography, an essay on translation studies, and a passage from a history text (House, 1997). Three out of the four translations were from English into German, and the other was from German to English.

If we focus only on the methodology issued to test both versions of the model from a professional and scientific point of view, we can see several limitations and pitfalls. The first one is the restricted scope in terms of the number of language pairs, different types of texts, and size of the samples. Therefore, the test leaves out not only the majority of language pairs that are present in the translation industry nowadays, but it also omits types of texts and formats with specific quality challenges, such as software or multimedia texts. The second one is that there are no clear indications regarding the context in which the translation

process was carried out: was there a "translation brief" with instructions from the requestor of the translation? Was there a deadline? Which tools or resources were available for the translator? All of these factors influence the quality of a translation in the professional world, yet they are not considered in House's model. Finally, the main goal of this model is to find errors, "mismatches". However, as researchers like Drugan point out, the main focus of professional approaches to translation quality assessment is to validate whether a translation is an "adequate or acceptable product" (Drugan, 2013:54).

2.2.2.2 Larose´s model

Another model of translation quality worth mentioning is the one outlined by Robert Larose in 1998. Larose agrees with the *skopos* approach of certain translation theorists and considers that the most important aspect to be able to measure the quality of a translation is the purpose of that translation. However, unlike House or other theorists, Larose goes beyond textual comparison alone and considers the context of a translation and the professional constraints. Larose's model has two main domains: "textual elements" and "extra-textual elements" (Larose, 1989). This attempt to include extra-textual factors such as age or level of experience of the person in charge of the evaluation was somewhat pioneering at the time.

Another characteristic of Larose's model is that it only applies to the text itself and that he establishes three different levels with a hierarchical structure: microstructural (at a sentence and phrase level), macrostructural (discourse content above sentence level), and superstructural (the overall structure of discourse) (Larose, 1989). If we look at the practical application of this model, it is not as straightforward as House's proposal due mainly to the lack of detailed stages, sample texts, or translation-specific criteria (Drugan, 2013), which would allow to test or replicate this model in a professional environment. On the other hand, Larose shows a good understanding of the professional working conditions and recognises that the criteria for assessing the quality of a translation should consider the constraints of the professional practice to ensure that the approach is genuinely feasible (Larose, 1989). Larose also highlights that a translation quality assessment should not be confused with translation revisions and, unlike other theoretical approaches that are focused on finding errors (such as House's model mentioned earlier), he considers that the goal of a translation quality assessment should produce a statement of the quality of a translation.

However, researchers like Drugan (2013) and theorists such as House (2015) point out that some of the aspects of his model are not detailed and specific enough, which hampers his endeavour to produce a translation quality assessment model that can be relevant in the professional field. Other gaps in his model include the fact that it does not include examples of the conditions in which the translation was produced and that the model was "tested" using only one language pair and one text type (classical literary texts) (Drugan, 2013).

2.2.2.3 William´s model

A more recent approach to translation quality assessment was proposed by Malcolm Williams in 2004. Williams bases his proposal on argumentation theory (a branch of discourse analysis) and uses this theory to build a discourse-based framework that goes beyond microtextual approaches to error analysis (Williams, 2004). His model builds on the approaches proposed by theorists such as Larose and House "by integrating a macrotextual, discourse […] perspective, along with relevant aspects of pragmatics, into the assessment process" (Williams, 2004:17). In terms of the practical application of the model, Williams outlines four stages, as shown in Table 5 below:

Williams's Translation Quality Assessment Model (2004)		
Stage 1	Analysis of the original	Key features in 3 main areas: • Argument schema • Arrangement relations • Organisational relations
Stage 2	Analysis of the translated texts	Focused on overall coherence: • Overall arrangement • Readability issues • Acceptability issues
Stage 3	Comparative assessment	Types of argumentation parameters: • Relations (argument schema, arrangement, organisational) • Inference indicators (functions, conjunctives) • Types of arguments • Figures of speech • Narrative strategy
Stage 4	Overall quality statement	

Table 5. Stages of the translation quality assessment model proposed by Williams (Williams, 2004).

William proposes a preliminary translation quality assessment grid and tests his model on one language pair (French to English) and four pairs of source and target texts: two governmental texts on statistics and energy and two criminology and legal texts. William also tries to adopt a scenario that is closer to the professional environment than those taken by House or Larose. For this reason, he uses unrevised translations submitted by freelance translators to real clients (Drugan, 2013). Finally, William analyses the results of the test and proposes changes and enhancements to fine-tune his model. Some of the most relevant enhancements are a sample translation quality assessment with a detailed explanation on how to apply this approach in a practical scenario and a "rating scale" with four standards (Williams, 2004): publication standard, information standard, minimum standard and substandard.

As mentioned before, Williams covers the quality assessment methodology in professional environments in a more consistent and detailed manner than theorists such as House and Larose, and he claims that his model "covers all the significant elements in instrumental translation and places emphasis on quality according to translation function and end use" (Williams, 2004:17). However, and similarly to what has already been observed in House's and Larose's approaches, certain aspects of William's model fall short of real applicability to the professional practice of translation (Drugan, 2013). These shortcomings can be grouped into three main areas:

- **Focus on translation as a product**, with little or no consideration to the context, the processes, translators' competencies, or tools.
- **Limited scope of the test** in terms of text types, size, and language pairs.
- **Time-consuming application** as Williams considers that the translation quality assessment "would ideally entail detailed examination of all passages containing key elements of the argument schema"(Williams, 2004).

2.2.3 Other theoretical approaches from Translation Studies

Aside from the models presented before, there are several other approaches to translation quality assessment proposed by translation theorists to allow reproducible and intersubjective judgement (Lauscher, 2000). However, as Drugan and Lauscher point out, these approaches are purely theoretical (and therefore, they have not been tested in professional practices) or tend to be directed at training new translators (Drugan, 2013).

However, some of these theoretical approaches have had an undeniable influence on translation studies research, and therefore it makes sense to outline some of their main contributions to the field of translation quality assessment.

2.2.3.1 Equivalence-based approaches

The definition of the concept of "equivalence" has always been a recurrent theme in translation theories (Lauscher, 2000; Drugan, 2013; House, 2015). Equivalence-based theoretical approaches see translation as the attempt to reproduce a source text as closely as possible. However, due to its own nature, it is widely recognised that a translation can never be completely equivalent to a source text at all levels, which has led to several proposals of types of equivalence, for example, dynamic versus functional equivalence formulated by Eugene Nida (1964).

One of the first and most important systematic and equivalence-based approaches to translation quality assessment was presented in 1968 by Katharina Reiss. Reiss considers that in order to assess the quality of a translation, it is first necessary to determine its function and the text type of the source text, and she proposes four different text types, which can be determined by analysing the source text. According to Reiss, a translation can be considered acceptable if the target text and the target text units have the same "value" as the text unit in the source language, "considering the linguistic and situational context, the linguistic and stylistic level and the intention of the author" (Reiss, 1971:11. Translated by Lauscher, 2000:151).

Reiss also establishes a set of rules that apply to both the translation production and the evaluation processes and proposes an approach to translation quality assessment divided into two different steps (ibid).

- Analysis of the target text to assess "the appropriateness of target language use".
- Comparison of source and target texts to determine the degree of equivalence.

The approach suggested by Reiss has been very influential in academic translation research studies, mainly because it contains a comprehensive and systematic model to analyse texts for translation production and translation evaluation purposes (Nord, 1996). However, as pointed out by House, Nord and Drugan, Reiss´s ideas are mainly "programmatic" (House, 2015), as she does not explain how to establish the function of the text or the text type, nor she provides a precise definition of "optimum equivalence" (Lauscher, 2000).

Another drawback in terms of practical application is that, according to Reiss's approach, evaluating a translation involves reversing the translation process to deduce the translator's intent and reconstruct the translation strategy (Lauscher, 2000), rather than through having direct access to the real translation process or the translator (Drugan, 2013).

2.2.3.2 Functionalist approaches

In contrast with evidence-based approaches, functional approaches to translation assessment are based on the premise that the translation of a source text is a process of text production and that this process is mainly determined by factors related to the target culture (Reiss & Vermeer, 1984; Vermeer, 1996; Nord, 1997). The "function" of a translation is therefore determined by the requester and the translators themselves, who will need to take into consideration factors such as the use of the translation in the target culture and the situation of the target readers to determine the translation strategy that should be applied to each translation (Lauscher, 2000). Consequently, functionalist approaches change the focus from rating a translation as "good" or "bad" to considering whether it is "functionally appropriate" (Schäffner, 1997).

One of the most representative examples of this type of approach is the text analysis model proposed by Christiane Nord in 1991. Nord considers that her approach can be applied to professional practice and translator training, although she mainly focuses on the latter (Nord, 1991). Nord's proposal builds on Reiss's approach to propose a set of 76 questions to determine the function of the text and propose different translation grades that are determined by the purpose of both the translation "initiator" and the target text itself (Drugan, 2013). However, despite Nord's efforts to illustrate how her model could be applied in specific scenarios, her approach has been criticised by researchers such as House (2015), Pym (2009) or Williams (2004) for its lack of precision in its potential application to translation quality evaluation purposes.

2.2.4 Academic approaches to quality from Quality Studies

As we have seen in the previous section, there are many different academic approaches to define "quality" in the Translation Studies field, which hampers the development of further discussions on how to manage or evaluate quality. For this reason, experts such as Fields et al. (2014) have analysed existing proposals and frameworks coming from the quality management discipline to try to determine whether they could apply to the concept of quality

in the translation field. As a result of this analysis, Fields et al. (2014) and Kurz (2020) propose the application of Garvin's (1984) product quality framework into the translation field, as they consider that some of the approaches proposed in this quality framework might provide a valuable approach for evaluating the quality of the translation services.

Garvin's proposal synthesises the different definitions of product quality arising from fields such as philosophy, economics, marketing, and operations management; and describes five different yet complementary approaches to defining quality:

- **Transcendent Approach:** According to Garvin, this approach considers the quality of a product or service as "absolute and universally recognizable, a mark of uncompromising standards and high achievement" (Garvin, 1984:25). This approach focuses on a subjective relationship to some standard, and therefore, the capacity to determine this relationship can only be acquired through experience (Fields et al., 2014).

- **Product-Based Approach:** This approach focuses on the comparison between the inherent attributes of a service or product and its required or expected characteristics. Garvin also proposes eight ways to evaluate product quality: performance, features, reliability, conformance, durability, serviceability, aesthetics, and perceived quality. Even though some of these evaluation criteria might be more subjective than others, with this approach, "quality reflects the presence or absence of measurable product attributes" (Garvin, 1984:27) and, since it is not based only on preferences, as it happens with the transcendent approach, it allows to assess quality in a more objective way. Therefore, it is not surprising to see this approach used in industry standards such as the ISO 9000 series (Kurz, 2020).

- **User-Based Approach:** Galvin defines this approach as focused on the subjective view of the customer (or user), and it is based on the premise that quality "lies in the eyes of the beholder" (Garvin, 1984:27). According to this approach, quality is the degree to with the service or product is able to meet the needs or preferences of the user. Kurz (2020) considers it common to find examples of this approach in the translation industry, as it reflects the subjective and "unfounded" criticism that end clients or end-users of a translated product often express.

- **Manufacturing-Based Approach:** This approach defines quality as "conformance to requirements" (Garvin, 1984: 27) and aims to objectively measure the degree of

adherence to the pre-determined specifications and processes and minimise any deviations. Fields et al. (2014) propose to rename this approach as "Production-Based" as they consider that the original name might be misleading when applied to professional services, including translation, and it does not reflect the true scope and applications of this approach. Some examples of this approach in the translation industry can be found in certain steps of translation projects, in which the process to be followed is defined in great detail, and any deviation might be penalised by the end client or the translation agency. For example, the instructions to perform desktop publishing translation (DTP) or SEO translations or the settings that should be used in certain QA check tools.

- **Value-Based Approach**: This approach focuses on the relationship between cost and price, which results in a subjective view on quality since the quality of a product depends on whether the product "provides performance at an acceptable price or performance at an acceptable cost" (Garvin, 1984:27). For this reason, if a product or service does not provide the highest value, then it cannot be considered to be of the highest quality.

Despite the potential applications of the described approaches to the translation field, it is interesting to note that there does not seem to be a consensus as to the extent to which the approaches used in the quality-management discipline can be helpful in the translation field (Fields et al., 2014). For example, some of the authors of this article consider that there is no reason why quality-management approaches that are used in other similar industries (such as marketing) could not be applied to the translation field, while others believe that the translation services have certain unique characteristics and that certain aspects of quality management might not be applicable or relevant in the translation field.

2.3 Professional approaches: Translation quality methodologies in the translation industry

Translation quality is certainly evaluated and discussed in the professional world, especially since 1990, due to an increased awareness of the role of the translation industry as an enabler to other industries and the need to define and agree on what "translation quality" really means (Castilho et al., 2018b; Fields et al., 2014).

However, it is rather hard to find many in-depth publications that take into consideration both the reality of the industry and the academic research done in this field. One of the most remarkable efforts in this sense is the book *Quality in Professional Translation, assessment and improvement* published in 2013 by Joanna Drugan, from the University of East Anglia (UK). In this publication, the author aims to provide an analysis of the approaches to translation quality management observed in different research visits to several companies and to identify and explain the main academic and professional assumptions about translation quality. The extensive research carried out by the author of the book serves as a very sound academic base for the revision of the literature on professional approaches, and therefore it will be referenced several times throughout this section.

To understand the relevance of translation quality methodologies in the professional field and the increase of industry studies and publications on this topic in the past ten years, it is important to present three key trends that have driven (and continue driving) this focus: the proliferation of different translation standards, the changing context, requirements, and technologies in which professional translation is inserted, and, ultimately, the need of the companies or departments that require translations to be able to judge the quality of the service they are investing in (Drugan, 2013).

The following subsection of this chapter will analyse the similarities and differences between the academic and the professional approaches to quality, present a classification of different professional models, and discuss some of the latest trends in professional approaches. The last subsection will focus on the tools, technologies and metrics used nowadays in the industry.

2.3.1 Similarities and differences between the academic and the professional approaches to translation quality

As seen in chapter 2.2, academic research focuses on translation as a product rather than a process, and their primary interest is translation quality assessment. The industry approach, on the other hand, pursues a more comprehensive view of translation quality. The emphasis is not on the evaluation of a particular translation on its own, but to establish a comparative assessment, assess quality even after it has been "delivered" to the requestor, explore how quality can be measured most objectively and scientifically possible, and ensure efficiency

and consistency during the whole translation process, to mention just a few of the questions that tend to be out of the scope of an academic approach.

There is a particular point in which theorists' and practitioners' methodologies tend to agree, and that is the fact that a translation quality assessment should follow an error-based approach. However, in the professional world, errors have consequences that go far beyond "failing" an assessment. For freelance translators, a failed assessment can mean a withheld fee, reworking tasks free of charge, and even damage in the relationship with the requestor, who can decide to reduce the number of projects assigned to them. For clients and end-users, the consequences can also be very grave: damage to the client's reputation, loss of revenue for the client, or damage to the user's physical integrity due to a mistranslation or unclear translation.

According to Drugan (2013:70), the concern for translation quality in professional contexts is mostly "client-driven", as they are the ones that have a first-hand experience of the consequences and the impact (financial, legal, and in reputation) of the lack of quality of a translation. The author also mentions that this concern is caused by other factors, such as the lack of regulation in the profession, and the fact that, in her opinion, academic qualifications might not be enough to guarantee good results, as many of these qualifications are "assessed by essays on translation theory rather than hands-on practice" (ibid). Another concern coming from the industry, as reported by Drugan (ibid) and Roger Chriss in his book *Translation as a profession* (Chriss, 2006:140), is that even reputable, certified and well-established translation service suppliers might not always be able to provide a high level of quality consistently.

Another difference between the academic approaches to translation quality and the professional ones is that the first ones tend to consider translation quality assessment as a step that should take place after the translation has been delivered to the client. However, the professional approaches consider that clients want to know if the quality levels will meet their requirements before making a significant investment in a translation service or provider. As explained by Drugan (2013:70):

> Clients require assurances in advance that suppliers can produce the goods, the ongoing projects updates on how targets are being met. Agencies and translators have to bid against others to wind projects by demonstrating they can provide optimal quality levels at

competitive prices and be able to report to clients on key performance indicators as translation takes place.

However, from the point of view of the "client" or entity that requests a translation service, being able to assess the quality of this service presents at least two distinct considerable challenges. In the first place, they often cannot assess the quality of a translation by themselves, as the main reason for their need for this service is that they cannot translate the content. In the second place, ever-expanding translation needs make it very difficult for them to accurately assess quality according to each type of translation project's characteristics, priorities, limitations, and expectations. In addition to this, there are also very different types of content (marketing, legal, technical, user interface, to name just a few), language pairs and countries associated with each language pair, translation service providers (multi-language vendors, single language vendors, freelancers, hybrid-vendors), translation technologies (manual translation, computer-aided manual translation, raw machine translation, post-edited machine translation). Therefore, applying the same translation quality methodology across the board might seem to be an easy way to standardise and improve quality standards in the translation industry, but, as Drugan points out, it could also be counter-productive, as the translation quality management process could not be adapted to the needs, priorities and limitations of different clients and projects.

Another critical topic that is amply addressed in professional translation quality management approaches is the use of Machine Translation (MT) engines and the evaluation of its outcome, either by predicting the quality of an engine using automatic algorithms or by evaluating the output of an engine with human evaluators of automatic metrics (Vela-Valido, 2021). It is important to note that machine translation quality management workflows tend to differ from those used in human translations. Even though it is possible to use the same translation quality assurance tools, the quality assessment or evaluation methodologies present very significant differences compared to those used in human translation. Another challenge that concerns MT evaluation workflows is that the performance of the different MT engines continues increasing year after year, making the evaluation of the quality provided by these engines more complex. Although the industry considers human evaluation as the "gold standard" (Bojar et al., 2016:27), this type of evaluation has some limitations that have been highlighted as particularly problematic, such as the low agreement rates

among different evaluators in comparison with the automatic evaluation (Bojar et al., 2016), and the amount of time required to conduct evaluations of big sets of data.

Another challenge that is directly related to this topic is that, even though the decision on whether to use MT or not is usually driven by cost-benefit and speediness considerations, the time and effort needed during the post-editing phase to ensure that the final translation reaches the required levels of quality can vary enormously. Some of the factors that influence this variability are, for example, the language pair (uncommon languages and combinations that do not include English as the source of the target language tend to perform worse than widely spoken languages), the type of content (creative texts in which style and cultural nuances play a significant role perform worse than plain and standardised texts) and the resources and time available to perform the post-editing phase efficiently.

One final difference between the academic approaches to translation quality (mainly the most theoretical) and the professional ones is that the professional approaches tend to favour a functionalist methodology (Colina, 2008; Jiménez-Crespo, 2009; Calvo, 2018) to try to define different quality levels that would correspond to the degree of quality required by different clients for different types of content, depending on their function.

2.3.2 Classification of translation quality models

The complexity and constant evolution of the translation industry are tied to the need of the different translation service providers to adapt to their clients' needs. For this reason, there are multiple approaches to translation quality, almost as many as different types of clients, jobs, and providers. Drugan proposes two high-level groups of translation quality models that share some core features: top-down models (traditional hierarchical approaches) and down-top models (crowdsourcing approaches) (Drugan, 2013).

Top-down translation quality models (traditional hierarchical approaches)

According to Drugan (ibid), traditional hierarchical approaches are still predominant in the industry, especially in those sectors and organisations in which it is particularly important to be able to evaluate service quality in a structured and repeatable way. These approaches have these four elements in common:

- They put a particular emphasis on resources (both human and technological), as these two aspects have a significant impact on the quality delivered. In terms of human

resources, some of the strategies adopted include the use of entry qualifications, translation tests, probation or "review" periods, and ongoing feedback. For tools or automated quality tasks, Drugan (2013) mentions that TSPs tend to tailor their approach depending on the specific translation project, client, or other factors.

- They focus on optimising their structures and processes and building on their experience and knowledge to avoid future projects.

- They tend to address quality problems so that they can avoid waste of resources and provide just the right level of quality that is needed. Failing in doing so would mean an inefficient use of their resources and an impact on their desired value for money.

- Consequently, all the top-down models studied by Drugan consider different quality levels so they can find a balance between quality and adherence to other requirements (for example, time and money).

Drugan also highlights that the case studies from "real-world providers" show very distinctive differences in relation to the academic research and describes some of these differences (see table below).

Academic research	Professional case studies
Tests carried on students or untested volunteers	Providers aim to recruit the most suitable resources
Subjects are observed in artificial conditions	Resources build long-term relationships and expertise
Focus on the quality of the text, ignoring the impact of production conditions	Both quality of the text and production conditions are addressed
Focus on errors and tools used	Focus on efficient allocation of resources and compliance with client preferences

Table 6. Differences between academic research and professional case studies from top-down models, as outlined by Drugan (2013).

Down-top translation quality models (crowdsourcing approaches)

In contrast to the traditional top-down translation quality models, the down-top models aim to adapt to some of the most recent constraints and also opportunities of the Age of Digital Transformation, such as the downward pressure on costs, the different technological advances, new types of user-generated content, new types of "clients" (organisations that could not afford to pay for these services in the past), or increasing demand in translation coming from the end-users themselves (communities of speakers of certain languages considered not big enough to justify the commercial translation of products or services to their native languages).

As Drugan highlights, "these new approaches overturn core industry tenets (e.g. that translation should be into the mother tongue, domain experts are needed in technical translation, quality should be measured and controlled)" (Drugan, 2013:159).

Similarly to what happens with top-down models, Drugan also highlights certain core features that down-up models share:

- They emphasise resources, although with an approach that differs from the one adopted by top-down models: here, companies focus primarily on technology and not so much on the qualifications or skill of the human resources.
- They focus on optimising their structures and processes, encourage community support to address any problems that might arise, and trusts users' ability to judge the quality for themselves.
- They feel comfortable addressing any quality problems after the translation has been delivered, making use of user input compiled post-delivery
- They adapt to their human resources (frequently referred to as "contributors"), who can influence the environment of each language effort.
- Contributors have different strengths that compensate for any lack of linguistic competencies, such as low or no cost, commitment to project values or goals, or willingness to learn.
- They emphasise ongoing feedback loops to keep contributors engaged and allow for continuous improvement of the quality of the translations.
- They feel comfortable defining their limitations, as contributors are encouraged to report challenges in forums or discussion boards and work with other members of the same "community" to overcome them together.

The classification of translation quality models provided by Drugan is particularly relevant to establish a general framework on how translation quality is managed (top-down or down-up). However, it does not provide specific information on the different types of translation evaluation models used in the industry nowadays, as these models follow specific approaches that need to be considered separately. Therefore, the following section will be providing a classification of translation evaluation models based on the most common practices of the industry.

2.3.3 Classification of translation evaluation models

If we analyse the different evaluation approaches followed by the translation industry and referenced in different professional publications, it is possible to identify two high-level models: the analytical approach and the holistic approach. Both have their advantages and disadvantages, although it seems that big translation service providers predominantly use the analytical model, given that it allows an objective and robust statement of the quality of translation as a product according to the definitions of standards such as ISO 9001. One of the main differences between these models is that, while the analytical evaluation focuses on the detection and evaluation of errors to determine the level of quality of a translation, the holistic evaluation aims to get an overall impression of the quality of a text based on the experience and impressions of the evaluator.

The analytical evaluation model

This model entails a detailed comparison between the translation and the requirements, and it counts every non-compliance as an "error". Each error is assigned an error category, and each category has a pre-defined set of error points or penalties. This method helps the evaluator to understand which areas have a higher number of errors and, therefore, should be addressed and improved. Another advantage of the analytical method is that the evaluation methodology is highly standardised, which provides a higher level of objectivity in comparison with the holistic approach. However, this model requires in-depth training of the evaluators to ensure that they have a deep understanding of not just the logic of the model but also the entire quality management and evaluation process. For this reason, the training phase is considered costly and time-consuming (Kurz, 2020).

The holistic evaluation model

In contrast with the analytical evaluation model, the holistic evaluation approach relies on the experience and personal impression of the evaluator. The overall impression of the quality can be expressed in different ways: with ratings (three stars out of five), binary results (thumbs up/thumbs down, pass/fail), or with a score (X out of 5 or X out of 10). One of the main advantages of this evaluation method is that it is faster than the analytic one, and it allows the evaluator to present a first estimation of the quality of big translation projects that spans several weeks and months and confirms whether there are major quality errors that might grant further measures to correct any quality concerns before the project is delivered

to the end-client. However, the results of this evaluation model depend entirely on the impression of a specific person, which can distort the actual quality of the translation provided and lead to contentious discussions on whether the evaluator's judgment was justified.

Although analytic and holistic evaluation methods are used to evaluate both human and machine translation quality, the application of these methods in Machine Translation studies focused on translation quality evaluation presents some specific characteristics that will be presented in the next section.

2.3.4 Approaches for translation quality evaluation of machine translation quality (MTE)

Another significant branch on translation quality can be found in Machine Translation studies. The majority of the approaches to translation quality assessment in this field are specifically designed to assess, compare, and improve the quality of machine translation engines only. Consequently, they present very significant differences compared to the evaluation approaches used for human translation (Vela-Valido, 2021). One of the main differences is the widespread use of automatic evaluation metrics, such as the Bilingual Evaluation Understudy, or BLEU (Papineni et al., 2001) or METEOR (Lavie & Agarwal, 2007), instead of or in combination with human evaluation. It is interesting to note that, even though academics and professionals agree that human evaluation is the "gold standard" for machine translation evaluation (Birch et al., 2016), it is also evident that this type of evaluation has some important limitations, such as the relatively low agreement rates among different evaluators in comparison with the automatic evaluation, the amount of time and effort required to evaluate big sets of data, and the lack of scientific studies to assess human-machine parity in language translation (Läubli et al., 2020). Given that automatic evaluation methods and metrics are out of the scope of this dissertation, the following section will outline the main approaches and trends in human assessment of machine translation quality from the 1960s to the present day.

2.3.4.1 Human assessment of machine translation quality

First approaches to human assessment of machine translation (1960 – 2000)

One of the first approaches to human assessment of machine translation outputs was proposed in 1966 by John B. Carroll at the Automatic Language Processing Advisory Committee (ALPAC) (Carroll, 1966). This approach was focused on two aspects: intelligibility (the output should be understandable and read as if it was originally produced in the target language) and fidelity (the output should retain the same meaning intended by the original).

Almost three decades later, the Advanced Research Projects Agency (ARPA) published a new methodology to evaluate machine translation outputs using three different criteria: adequacy, fluency and comprehension (Church & Hovy, 1993), which was later adapted by White et al. (1994). This methodology also specifies the grading method that should be used in each of the three criteria:

- **Adequacy**: Evaluators are asked to look at each sentence of the machine translation output and grade the adequacy using a scale from 1 to 5.
- **Fluency**: Evaluators are asked to look at each sentence of the machine translation output and judge whether the segment is well-formed and fluent in the target language.
- **Comprehension**: The evaluators are presented with different multiple-choice questions to determine whether the overall output retains sufficient information compared to the original text.

Further developments in human assessment of machine translation (2000 – Nowadays)

In 2000 Bangalore et al. (2000) introduced the concept of "accuracy" divided into different subcategories and King et al. (2003) proposed FEMTI, a new machine translation framework born from the collaborative work carried out within ISLE, a project funded by the European Union, the National Science Foundation (USA) and the Federal Office for Education and Science (Switzerland). The main goal of this proposal was to gather into one place all the accumulated experience of machine translation evaluation and propose additional machine translation evaluation methods that would cover the actual practical use of the machine translation systems in a professional environment. Some of these new criteria include concepts such as *suitability* of the results in the context where the engine will be used,

interoperability with other software or hardware platforms, *reliability* of the engine, *usability* of the interface, *maintainability* of the engine in order to adapt it to different users, or *portability* from one version of the engine to a new one (Han et al., 2021).

In 2006, the first machine translation evaluation campaign organised by the Conference on Machine Translation (or WMT) decided to use the adequacy and fluency criteria on a 5 point scale as their main metric (Koehn & Monz, 2006). One year later, Vilar et al. proposed a ranking-based evaluation approach (Vilar et al., 2007) in the WMT metrics task using human evaluators to assess the output of different machine translation engines at the same time (Bojar et al., 2013). In this type of approach, the outputs of the different engines are randomised and the human evaluators are presented with the source segment and the candidate translations of each of the engines. The evaluators are asked to rank each translation from 1 to 5, and the compiled results are used to assign a final score to each engine. This raking-based evaluation approach became the official metric from 2008 to 2016 (Bojar et al., 2016).

Around 2013, after assessing the results obtained in the WMT segment metrics tasks using the ranking method described above, some researchers noticed that there were very low human inter-agreement scores (Han et al., 2021) and started exploring different human evaluation methods that could give more consistent results in segment level rankings. One of the resulting proposals to solve this issue was suggested by Graham et al. in 2013, and it involved continuous rating scales for human evaluation using the fluency criteria and the use of crowd source intelligence evaluation (Graham et al., 2013). According to the authors of the proposed evaluation method, one of the main advantages of this direct estimation method in comparison with the segment ranking method used in previous WMT tasks is that it includes not only information regarding which outputs are better than others, but also the degree to which a specific output is better or worse than others. In order to achieve the desired volume of assessments from different evaluators, measure the level of agreements and, therefore, confirm the reliability of the evaluation method, Graham et al. (2013) proposed the use of crowdsourcing services such as Amazon's Mechanical Turk (AMT). In their research, the authors recognise some of the risks associated to the use of crowdsourcing solutions, such as a potential increase in the inconsistency of the results, and the lack of visibility over the experience, skills, or level of knowledge of the crowd performing the evaluation. To address these risks, Graham et al. designed and tested a "judge-intrinsic"

quality control method, which, according to the authors of this research, increased the intra-annotator consistency. As a consequence of these results, the WMT decided to adopt the continuous direct assessment approach, which, due to budget constraints, is carried out by researchers or crowdsourcing solutions (Freitag et al., 2021).

Interestingly enough, the use of researchers, students or crowdsourcing solutions has been questioned in the last few years by different researchers (Toral et al., 2018; Läubli et al., 2020) who have reassessed some studies that claimed that machine translation had reached (or even surpassed) human parity in certain text types and language pairs and compared the evaluation results of these studies against assessments done by professional translators. As a result of this comparison between assessments done by professional and non-professional translators, these researchers showed that professional translators are able to distinguish between human and machine translations where non-professional translators could not do so (Freitag et al., 2021), casting some reasonable doubts over the reliability of the current human evaluation methodologies used to evaluate the performance of machine translation engines, especially in human-machine parity tests.

New paradigms in human assessment of machine translation

As we have seen throughout this section, all the approaches proposed to evaluate machine-translated texts are adapted to the specific needs and limitations of the machine evaluation workflows, and therefore tend to use holistic and qualitative methods that do not seem to correspond with the approaches to human evaluation for human translations proposed by the academia or the industry. Furthermore, even though some of the latest models proposed by the industry (such as the MQM framework and the harmonised MQM-DQF model) do mention that they can be used to evaluate machine translation outputs (DFKI, 2015), there has been little to no research on how these explicit error analysis models would work on a professional machine translation evaluation setting.

However, one very recent and promising effort in this direction is the large-scale study on human evaluation published by several members of Google Research in 2021 with the goal to "contribute to the evolution of standard practices for human evaluation of high quality MT" (Freitag et al., 2021:1). The authors of this research consider that "any scoring or ranking of translations is implicitly based on an identification of errors" and also that "making such an identification explicit by enumerating errors provides a *platinum standard*

from which various gold-standard scorings can be derived" (Freitag et al., 2021:1). In their research, Freitag et al. adopted a specific hierarchy from MQM that they considered appropriate for machine translation evaluation purposes, employed professional evaluators with access to the context of the documents they would be assessing and asked them to evaluate the outputs of 10 machine translation systems in two language pairs: English to German and Chinese to English. The team also collected different rating methods, including scalar ratings from professional translators, crowdsourcing scores obtained in WMT tasks and automatic metrics, and compared them against the results obtained through the MQM scoring system. This large-scale study showed several notable results that can be summarised in three main areas:

- The evaluations done through crowdsourcing solutions showed a low correlation with the MQM-based evaluation, which, according to the authors, puts into question some of the conclusions drawn in other research papers and that were based on previous crowdsourced assessments.
- The MQM ratings challenge the original WMT ranking of translations, as they show a clear preference for human translation over machine translation and put some low-ranked machine translation engines in higher positions.
- Most automated metrics correlate better with MQM ratings than with the WMT human scores.

Overall, these results seem to add further evidence and weight to the concerns expressed by researchers such as Toral and Läubli regarding the adequacy and accuracy of the current human evaluation methodologies to assess machine translation outputs.

2.3.5 Tools and technologies

The translation industry and its main actors have traditionally been very active in their use of new technologies to increase the efficiency of the translation process (Doherty, 2016), and this increase in the adoption of Computer Assisted Translation (CAT) tools and machine translation engines has also influenced the evolution and adoption of different tools to support translation quality assurance and translation quality assessment tasks (Doherty et al., 2018).

However, the tools available today on this front are still not at the same level in terms of development and functionalities as the tools used for other translation-related tasks (such as

translation management tools, memory tools, terminology tools or subtitling software, to name just a few). Indeed, a study conducted in 2007 by Makoushina (2007) concluded that the quality assurance tools used in the industry were at the time at least ten years behind other translation tools and, although improvements have taken place in the interim, a distinction in functionality can still be detected.

It was precisely around that time when the dominant translation tool in the industry (Trados, nowadays renamed as SDL) introduced quality assurance features in 2006, and the first specialised QA tools appeared in the market. These tools were conceived as stand-alone products that could perform QA-related tasks suitable for automation (such as spell-checkers, terminology blacklists, or whitelists). These first tools and add-ins bundled with CAT software were rapidly adopted by the industry, in particular the TSPs (Makoushina, 2007), and, some years later, by freelance translators too (Drugan, 2013).

Nowadays, there is a wider variety of add-ins bundled with CAT software and standalone translation quality assurance tools (such as Xbench, QA Distiller, or Verifika), which are used as support tools by translators, reviewers, and linguists in charge of the translation quality assurance step. This is due to several reasons: first, there are still specific categories of mistakes that these tools are not able to detect automatically (for example, meaning, tone of voice[3] or style); secondly, these tools tend to produce what is known as "false-positive" errors; and thirdly, the support for certain languages (especially if they are non-European languages or minority languages or locales) is not at the same level as the one available for bigger languages such as French, German, Spanish or Italian.

Due to these limitations, quality assurance specialists or translators using these tools need to make sure to "select the most appropriate [settings] for each type of language pair and translation requirements, fine-tune [the tools] to try to reduce these "false positives" and analyse the reports generated by these tools to detect which errors are "false positives" [and] mark them as such" (Vela-Valido, 2021:4).

As we have seen, the translation quality assurance tools and add-ins mentioned above are focused on the detection and correction of errors. However, there is also a group of tools that

---

[3] "False-positive" errors are incorrectly identified errors, such as a difference in length from source to target (even though different languages have different semantic and morphological structures) or a difference in spacing rules (for example, a number and its unit measures are written without a space in-between in English while in Spanish this space is mandatory).

allow translation quality evaluation (either as a standalone step or in combination with the quality assurance step). These tools (such as ContenQuo) or add-ins (such as the ones provided by translation management systems like XTM, Wordbee's TQI, or Smartling's QCS) are designed to help human evaluators to annotate, classify and compile errors and provide a quality score based on a specific metric (see section 2.4.4). These tools (as the purely linguistic quality assurance tools mentioned above) can spot some errors automatically and suggest certain error categories. However, the evaluator still needs to check the source and the target texts to identify those errors that the tool cannot detect automatically and assign them a category and severity. Once done, the tool generates an automatic quality score based on the total score of all the errors, the number of words, and the weight assigned to each type of error. All these parameters can be modified for different translation projects, clients, or quality expectations.

Finally, there are also proprietary translation quality assurance and quality evaluation tools developed over the years by large TSPs and companies. These tools can use completely bespoke quality metrics or utilise some of the quality metrics used in the industry. Due to confidentiality and intellectual property regulations, there is not much publicly available information as to what is design or capabilities of these tools, and the different scholars and practitioners that mention them (O'Brien, 2012; Drugan, 2013; Esselink, 2019) do not provide any specific information about them.

2.3.6 Metrics

As discussed in previous sections, the industry seems to have a clear preference for quantitative systems to perform translation quality evaluations, commonly referred to as "metrics". These systems are based on a typology of errors and apply a point-subtraction scheme depending on the type and severity of each error, which gives a final score correlated with a defined quality scale. In the following sections, we will review the two groups of metrics currently used in the industry: manual metrics (quantitative systems based on human evaluation) and automatic metrics (quantitative systems based on automatic evaluation).

2.3.6.1 Manual metrics

SICAL

One of the first efforts recorded to create a systematic and quantitative model to measure translation quality was SICAL (Système Canadien d'appréciation de la Qualité Linguistique), a set of metrics created in the 1970s by the Canadian government's Translation Bureau. This system was developed as an examination tool and as an evaluation framework to assess the quality of the translations the organisation delivered (Williams, 2001). Regarding its application to the evaluation of translation quality, the system established a revision process through a contrastive linguistic analysis of the source text and the target text. It also contained a defined error typology and a set of reference criteria with which to compare the characteristics of the final translation, which allowed it to establish the concept of "acceptability", that is, the level of linguistic quality required to accept a translation. In 1986 SICAL reduced the number of error categories from around 100 to two: *transfer* and *language*, which could have two different levels of severity (*minor* and *major*) (Williams, 1989). This system also introduced four quality "grades" to establish different quality thresholds depending on the number and severity of errors found in the final translation: *superior*, *fully acceptable*, *revisable*, and *unacceptable* (Williams, 2001). It is interesting to note that the SICAL model was sample-based, which means that all the translations were evaluated using a randomly chosen excerpt of around 400 words (Secară, 2005), without any further scientific or statistic research on whether the quality of the sample would be a genuinely representative result of the quality of the whole translation (Mateo, 2014). In any case, SICAL is regarded as one of the first milestones in translation quality assessment metrics, and it had a significant impact on the systems that were developed afterwards.

LISA QA Model

In 1995, the Localization Industry Standards Association (LISA) developed and published the LISA Quality Assurance (QA) Model. This model was released as a spreadsheet and later as a stand-alone tool to evaluate the quality of the localisation of text types such as product documentation, help, or user interface files (Parra-Galiano, 2005). This tool had an interface with templates, forms, and reports built into an Access database and contained a predefined list of error categories, severity levels, weights, and criteria to define a Pass/Fail

grade (Stejskal, 2009). One of the advantages of this model is that it was contained in a tool that could be customised according to different criteria. LISA continued improving and expanding the capabilities of this tool and launched a second version (LISA QA Model 2.0) in 1999 with more error types, and a third version in 2004 (LISA QA Model 3.0). This version was aimed to allow customers and vendors in different vertical industries to "define and experiment with their own quality metrics" (Cadieux, 2004), which would encourage the development of different industry-specific quality metrics. The LISA QA model also defined several steps to implement the evaluation process, including sampling, quality assessment using a template, error annotation, and full revision (also called quality control) (Parra-Galiano, 2005). Given that this model was intended to evaluate localisation projects as a whole, it had eight dimensions, although only one of them covered linguistic matters (Jiménez-Crespo, 2009). This dimension defined seven error types (*Mistranslation, Accuracy, Terminology, Language, Style, Country, and Consistency*) and three severity levels (*minor*, *major,* and *critical*) (Parra-Galiano, 2005). Although this model was never released as a standard, it was considered a *de facto* industry standard (Doherty & Gaspari, 2013) by the time LISA was dissolved in February 2011. In 2015, a study conducted by Snow revealed that it remained the most common model used in the industry, although modified to some extent to meet the needs of its users (Snow, 2015).

Figure 1. LISA QA 3.0 model (©LISA, 2004).

SAE J2450

A couple of years after the LISA QA Model 2.0 was released, a working group made up of SAE (Society of Automotive Engineering) and General Motors representatives introduced SAE (Surface Vehicle Recommended Practice), a practice that was turned into a standard in 2005. The initial goal of this model was to assist in the revision of automobile documentation and provide "a consistent standard against which the quality of the automotive service information can be objectively measured" (SAE, 2001:1). Given the very defined scope of this model (the automotive industry), SAE J2450 does not cover errors related to style, register, or tone voice, which limited its application to sectors in which these aspects were not particularly important (such as medical or pharmaceutical) (Mateo, 2014). Other aspects that are also missing from this model are sampling criteria, quality thresholds, quality improvement, or root causes analysis, as it mainly focuses on error detection and categorisation.

SAE J2450 contemplates seven error types (*Wrong Term, Syntactic Error, Omission, Word Structure or Agreement Error, Misspelling, Punctuation Error, Miscellaneous* and *Error*) and two severity levels (*minor* and *major*) (SAE, 2001).

One of the most remarkable aspects of this model is that it aims at limiting the potential subjectivity of the evaluators and provides two specific rules to be applied whenever the evaluators are not sure of which category or severity to apply (SAE, 2001:4):

1. When an error is ambiguous, always choose the earliest primary category.

2. When in doubt, always choose 'serious' over 'minor.'

Finally, SAE J2450 establishes a review process consisting of five steps that need to be followed in chronological order and includes tasks such as error annotation, error evaluation, and score calculation.

QAT

Another noteworthy quantitative assessment model that was used in conjunction with a tool was the Quality Assessment Tool (QAT) (Mateo, 2014), developed by the European Commission's Directorate-General for Translation (DGT) to help their revisers to conduct translation assessments. The DGT is not only the world's largest translation service, but its quality assessment and evaluation processes are also considered as the gold standard

(Drugan et al., 2018). This tool was based on the LISA QA Model, which means that it followed the analytical assessment approach.

The prototype of the tool was presented to all the Language Departments within the DGT in 2009. However, after a trial period, only five departments found the tool useful (Strandvik, 2017). One of the main objections reported by Strandvik, who works as a Quality Manager in DGT, is that the objectivity of the assessment produced with the tool is limited to "an objective calculation of error points resulting from a subjective identification of errors" (Strandvik, 2017:129). Therefore, as Strandvik points out, there needs to be a common understanding of the three key components of the analytical evaluation: the guiding principles for the evaluation itself, the error categorisations, and the severity levels. Only then the quality assessment can be really consistent.

The error typology used in the tool was based on the one that was already used by the Translation Centre for the bodies of the European Union, and it contemplates eight error types (*Sense, Omission, Terminology, Reference Documents, Grammar, Spelling, Punctuation,* and *Clarity*) and two severity levels (*minor* and *major*).

This model also proposed a sampling methodology (based again on the Translation Centre) of around 10% of the translated text and established a minimum size (2 pages) and a maximum size (10 pages) (Mateo, 2014). Like the LISA QA tool, the TQA tool can be customised according to different criteria, and it even allows multi-users. However, the TQA tool is more advanced in terms of reporting features and can generate reports with information such as the type and number of errors detected and the final score.

MQM/DQF Model

Around 2012, two independent projects were started to address the need for standardised methods to evaluate translation quality: the Multidimensional Quality Metrics (MQM), published by the QTLaunchPad, an EU-funded collaborative research project; and the Dynamic Quality Framework (DQF), developed by the Translation Automation User Society, a translation industry organisation. After a couple of years, both approaches were harmonised and integrated into one shared model called the "DQF/MQM" (Lommel, 2018; TAUS, 2015) or "MQM-DQF" (Lommel & Melby, 2018), a collaborative effort taken by the German Research Center for Artificial Intelligence (DFKI) and TAUS within the project QT21 (QT21 Consortium, 2015), funded by the European Union.

During this harmonisation process, both models were reviewed to incorporate the strengths of each in the new MQM/DQF Error Typology. This harmonised model focuses on two main aspects: identifying the cause of the problems to prevent or correct them and relating problems to a predefined list of issue types (Lommel & Melby, 2018).

In terms of scope, both the MQM model and its subset DQF cover the quality assessment of translated content as a "product", and therefore they do not apply to the assessment of translation processes or projects. Furthermore, both consider "translated content" as any text or graphic that might be translated or adapted for multiple locales. Due to this broad interpretation, these models are applicable not only to traditional translation tasks but also to the localisation of software, videogames and websites, and creative adaptation of content for different audiences and purposes, such as adaptation of marketing materials and multimedia content. They are also designed to be applied to monolingual source texts and translated target texts and to human translation evaluation and machine translation analysis (Freitag et al., 2021) through a specific subset focused on the *Accuracy* and *Fluency* categories.

Some of the main changes and improvements highlighted by Arle Lommel, one of the participants in this project, were (Lommel, 2018):

- The dimensions of MQM were modified so that they would match the top-level categories of DQF.
- MQM's "null" severity level was incorporated to allow for issues to be flagged without any penalty assigned.
- DQF increased its catalogue of issues so that they could be tied to the existing MQM issue types and became a subset of MQM.
- The issue types *Internationalisation* and *Verity* were added to the DQF subset.

The resulting hierarchy was much more condensed than the original proposed by MQM and is also more focused on the issues considered more important for buyers of translation and localisation services (Lommel, 2018).

To simplify the application of this model, MQM defines a "core" of 20 issue types that cover the most common issues arising in quality assessment and that presents a high level of granularity so that it can be used for different tasks. Users of this model can define subsets

of the core to suit their individual needs, although the minimum recommended is *Accuracy* and *Fluency* (DFKI, 2015).
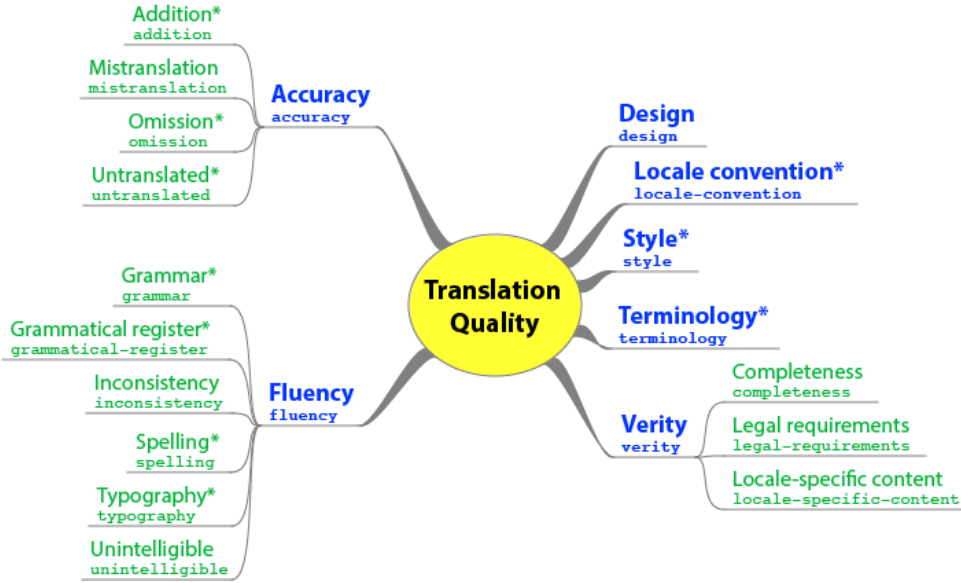
Figure 2. MQM Core (©DFKI, 2014 under CC BY-ND 4.0). Source: https://www.qt21.eu/mqm-definition/definition-2015-06-03.html.

As mentioned earlier, the DQF Error Typology was reviewed and incorporated into MQM as a recognised subset, and it supports additional levels of issues and additional features, as can be seen in the graphic below:
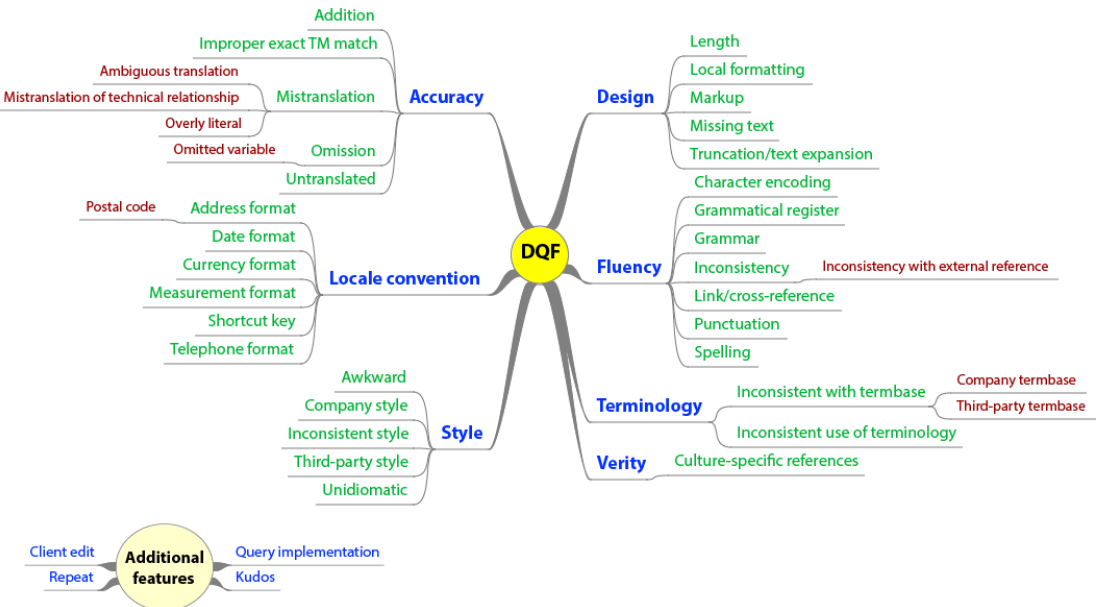
Figure 3. DQF subset (©DFKI, 2014 under CC BY-ND 4.0). Source: https://www.qt21.eu/mqm-definition/definition-2015-06-03.html.

The "additional features" include four different types of issues that can be marked with a severity level "none" to flag elements that require special attention that should not count negatively against the quality of the translation (DFKI, 2015). These four categories are:

- Client edit (changes requested by the client)
- Query implementation (changes made in response to a query)
- Repeat (systematic repetition of errors that should not be counted)
- Kudos (positive remark regarding the quality of the translation)

Another remarkable aspect of this model is that, even though it proposes default severity levels (*none*, *minor*, *major,* and *critical*) and an algorithm for error-count metrics, it also proposes exhaustive guidance for creating other MQM-compliant metrics. This section helps the users of this model to determine what sort of metric will be better suited to their needs based on five criteria:

- What is being assessed (product or system)
- Who does the assessment (expert evaluator, end-user, native speaker, automatic)
- Where the assessment takes place (industry, academic, research)
- When the assessment takes place (draft version, after revision step, after delivery)
- Why the assessment is taking place (acceptance testing, avoid problems, improve processes)

Based on this criterion, the users can select one or more methods to assess the quality of the translation, and the model mentions four main options: analytic methods, holistic methods (mono-dimensional or multi-dimensional), task-based testing and functional testing.

Finally, the model also provides guidance on the process to follow to create analytic metrics, and it even includes some guidelines for holistic assessment methods, as it considers that most of the MQM issue types can be used for holistic evaluation.

As of 2018, DQF/MQM is considered one of the most widespread metrics in the industry. SDL incorporated it in its Translation Quality Assessment (TQA) functionality within SDL Trados Studio in 2016 (Heyn, 2016) and was followed by other CAT tools and Translation Management Systems (TMSs) such as XTM International (XTM International, 2021) or Memsource (Memsource, 2021). Other enterprises and translation service providers have

also announced the adoption of the MQM/DQF metric, such as Microsoft, Mozilla, Lionbridge and Welocalize (Lommel & Melby, 2018).

MQM 2.0 (forthcoming)

The effort to continue developing and enhancing the harmonised DQF/MQM model has been taken within ASTM Committee F43. The new Multidimensional Quality Metrics (MQM) 2.0 is currently being developed as an ASTM International standard, the ASTM WK46396 – Standard Practice for Analytic Evaluation of Translation Quality (ASTM, 2021b). This committee is composed of different representatives from the industry, such as end-clients, translation service providers and government agents to ensure the applicability of this updated model (Lommel, 2018), and has agreed to keep the specific error typology free and open to the public, with several supporting materials being shared in a Community Group of the World Wide Web Consortium (W3C) (W3C, 2021a) and at Tranquality.info, a translation-related public website (Tranquality, 2021).

JTF's Translation Quality Assessment Model

In 2018, the Japan Translation Federation (JTF), the largest industry federation of Japan, published the Translation Quality Evaluation Guidelines, which adapted some concepts of the MQM metrics to the specific context of the Japanese translation industry (Nishino, 2020). One aspect of this model is that, even though it follows an error-based analytical approach, it recognises that this method can assess only certain aspects of quality. Therefore, evaluators should complement this model with other approaches depending on different factors. The guidelines also stated that the proposed metrics should be considered static and immutable, but rather an ever-evolving framework that should be "continuously improved through communication between the parties" (Nishino, 2020:160).

As mentioned before, the error typology follows the top-level categories of the MQM model (*Accuracy, Fluency, Terminology, Style, Locale convention, Design, Verity* and *Other*), with some new subcategories, for example, "homophone error" as these types of errors are particularly present in the Japanese language. The model also contemplates the same four severity levels proposed by the MQM approach: *critical, major, minor* and *none.*

Finally, the guidelines provide specific examples and information to help evaluators understand and implement the evaluation model in different environments. For example, it contains examples of sample weights for each error category and document type, which can

be modified as needed. The document types are grouped into nine categories: Patent specifications, Manuals and technical documents, Academic papers, Financial statements, Medical documents, Contract and legal documents, Reports and white papers, User interface texts, and marketing materials such as ads, PR and websites. The guidelines also offer a sample evaluation sheet to show how to use the model in a translation evaluation project.

2.3.6.2 Automatic metrics

Although the industry and academia seem to agree that human evaluation is the gold standard for both human and machine translation (Birch et al., 2016), several professionals such as Bojar et al. and Han et al. (2016; 2021) have also pointed out that human evaluation is also time-consuming, labour-intensive expensive and not reproducible. This means that this type of evaluation is not especially suitable for massive-scale experiments designed to evaluate the performance of different MT engines across different languages and scopes (Vela-Valido, 2021). This lack of scalability has been addressed by the use of automatic approximations to human judgment, called "automatic metrics", which are widely used for machine translation evaluation (Papineni et al., 2001; Görög, 2017; Way, 2018; Läubli et al., 2020; Han et al., 2021). For this reason, and as pointed out by Specia and Shah (2018:25), the purpose of these automatic evaluation metrics is to automatically compare the output of an MT system with the human translation of the same text. The level of quality of the MT output is then stablished according to how similar both translations are. It is also important to note that these automatic evaluation metrics do not provide any information about the severity of the errors produced by the system, and their correlation with human judgment varies significantly depending on factors such as the language pair, the domain, or even the skillsets of the evaluators used as a reference (Doherty & Gaspari, 2013).

As explained in the introduction, the scope of this research is human translation quality management and evaluation of both human and machine-generated translations. However, even though this dissertation does not aim to analyse automatic quality management and evaluation systems, the fact is that most common metrics to evaluate machine translation outputs are, as we have seen, generated automatically. For this reason, we believe it is pertinent to briefly present BLEU and METEOR, the most referenced automatic metrics used by the industry nowadays (Specia et al., 2010; Bojar et al., 2016; Han et al., 2021), as this will allow us to have a complete overview of the different types of metrics used in translation evaluation.

BLEU (Bilingual Evaluation Understudy): This evaluation algorithm was presented by Papieni et al. at the ACL Journal in 2001 as a "method of automatic machine translation evaluation that is quick, inexpensive, and language-independent" (Papineni et al., 2001:1). As mentioned before, quality is considered here as the correspondence between a machine's output and that of a human so "the closer a machine translation is to a professional human translation, the better it is" (Papineni et al., 2001:1). Due to the fact that BLEU is based on the degree of overlap between the strings of words produced by the machine translation engine and the human translation references, it needs a corpus of human translation for the same source text that will be translated by the engine. In 2002, it became the official metric of the Machine Translation Evaluation campaigns of the American National Institute of Standards and Technology, and it is still considered the *de facto* standard for research initiatives (Specia et al., 2010; Esselink, 2019).

METEOR: This algorithm was designed by Agarwal and Lavie (2007) as a more comprehensive approach to automatic evaluation by comparison than BLEU. It recognises non-exact matches such as synonyms and paraphrases and additional linguistic knowledge and, according to Agarwal and Lavie, not only shows a high correlation with human evaluations, but it can also outperform BLEU. However, it also relies on sophisticated resources and language processing tools that are only available for a limited amount of languages (Doherty & Gaspari, 2013).

Another way of assessing machine translation outputs through automatic metrics is the machine translation quality estimation (MTQE). The main difference between machine translation evaluation and machine translation estimation is that the second aims to predict machine translation quality in a completely automated way (that is, without any human reference) and that MTQE can be applied at different levels, from word-level to document-level (Vela-Valido, 2021). There are a couple of commercial solutions to perform machine translation quality estimations and that have been presented in the past few years, such as COMET (developed by Unbabel in 2016) and an AI-powered quality estimation module integrated into Memsource, a translation management system, in 2018. However, as shown in section 2.3.4.1 there is a very heated debate on what is the real level of accuracy of these estimations and metrics that are provided without any type of human reference (Sun et al., 2020), which has led to several recent studies aiming to compare automatic assessments with human assessments (Bojar et al., 2016; Freitag et al., 2021; Ziganshina et al., 2021).

2.3.7 Latest trends in professional approaches to translation quality

Some of the most recent publications on translation quality from the point of view of the industry take a step forward and claim that the "normative" or traditional models used in the industry are too rigid and do not take into account the reality of the businesses that use translation services (Beste, 2020; Johnson, 2021).

One of the limitations of the normative models mentioned by practitioners is that they look at content in a linear and unbiased way and that they are based on a system that assigns error points depending on the error type and severity. Therefore, the quality of a translation is directly connected to the number of errors it contains, without any consideration to how texts will be used and why. For this reason, there is a push coming from different actors from the industry to develop user-centric quality evaluation models that take into account how the translation is used (reader or customer reception) and why the source text was created in the first place (business goals).

In this regard, Beste (2020) proposes five factors that should be considered when evaluating the quality of a source or translated text: the user, the content, the consistency, the organisation, and the intent. Beste also proposes some metrics to measure content quality, although they are limited to web texts, limiting his proposal to other types of texts. Among the metrics proposed by Beste, we can find user-behaviour data (number of visitors, number of visits, number of page views), audience engagement data (time on page, scroll depth, bounce rate, social media analytics), and commercial data (return of investment and conversion rate).

Johnson (2021) takes a step further and proposes a new type of user-centric quality assessment for the localisation industry called "Experiential Localization Quality Assessment" or xLQA. In his online article, the author describes several trends that influence how global companies manage and measure the effectiveness of their content in different markets and the effect of these trends in the language and localization industry.

Trend 1: Quality as a design principle

A growing number of companies look at quality not just as a step but as a whole mindset. This user-centric design system brings with it a cross-functional approach that takes into consideration the quality of the localised end-product experience.

Trend 2: Quality is "experienced"

Nowadays, companies have access to a significant amount of information about the users' experience and feedback in different countries and languages, which can complement the translation quality assurance and quality evaluation processes to have a more holistic view of the quality of the translated content from different angles.

Trend 3: Paraphrasing approach

According to Johnson (2021), there is a general shift away from formality favouring a more accessible, inclusive, and personal "voice". For this reason, there is an increasing preference for "true" localization services (such as transcreation, creative translation, copy-writing, or cultural review) over strict adherence to source texts.

The author considers that these trends question whether an objective measurement of the quality of a translation as judged by a linguist is still relevant. Instead, he proposes to put the focus on the customers and understand what quality means for them. In his opinion (Johnson, 2021):

> Today, mistakes are not important. Customers are important. Globalization teams at major brands are trading in their LQA spreadsheets for user surveys. Localizing surveys into customer languages and asking them about their experience is a great place to start.

2.4 Regulatory approaches: Norms, standards, and metrics

Especially in the context of the translation industry, standardisation and regulation seem to have played a key role, as they help experts and stakeholders agree on how the activity of translation would be best carried out and homogenise key terms, concepts, and working methods. Standards are also a valuable source of knowledge, as they intend to set up a minimum necessary baseline. These two factors allow the different actors to "talk the same language" and have a clear understanding of what to expect from each other. Moreover, standards can provide specifications to help companies requiring translation services in their evaluation and contracting of providers of such services (be it freelancers, translation service providers, or translation service providers). Finally, some standards, such as F2575 (ASTM, 2014), are used by translation service providers to structure service level agreements (or SLAs) with their clients.

However, the issue of certification according to the current standards presents some particular challenges. For example, as we will see a little bit later in this chapter, ISO 11669 and ISO 5060 are guidance standards, not requirement standards, which means that it is not possible for any company to get certified for these standards. Another challenge observed is that some standards, such as ISO 17100, can be pretty broad in scope, which makes them not very suitable nor effective as a basis for building a certification. On the other hand, over-specification can also have a negative impact as other standardisation bodies can then claim that those specifications violate the "principle of global relevance" (ISO & TMB, 2004:2).

It is also worth noting that the majority of the certifications available in the field of translation and interpreting are only valid at a national level and that they are typically granted by an avowed major professional organisation, for example, the American Translators Association (which offers the ATA certification for professional translators) (ATA, 2021) or the Certification Commission for Healthcare Interpreters (CCHI, 2021). One relevant exception to this trend is the standard ISO 17100, certified through the International Network for Terminology (TermNet).

In the following sections of this chapter, we will present a summary of some of the most relevant international standards and norms that cover the different aspects of translation quality: from the management of the translation process to final product quality to quality control procedures, among others.

After that section, we will also give an overview of the most pertinent metrics used nowadays to measure the quality of both human and machine translation.

2.4.1 International standards and norms (in order of publication)

The majority of the standards that cover the topics of translation and quality management and that are currently used in the industry have been published by ISO between 2021 and 2019.

- **EN 15038** (deprecated in 2015): EN 15038 was a specific European standard for translation services approved by the European Committee for Standardization (CEN) (CEN, 2006). This standard defined the translation process and stated that the quality was guaranteed in the translation step itself, but also by the fact that the translated

text is reviewed by a different person. This standard also specified the competencies of each participant in the translation process.

- **ISO/TS 11669 Translation Projects – General Guidance:** This technical specification aims to provide guidance or all phases of a translation project and provides a framework for developing structured specifications for translation project (ISO, 2012)

- **ASTM F2575 - 14 Standard Guide for Quality Assurance in Translation:** This standard focuses on using standardized translation parameters to develop specifications (ASTM, 2014). It is currently under revision and will be replaced by the new standard developed by the ASTM workgroup WK47362.

- **ISO 17100:2015 - Translation services — Requirements for translation services:** This international standard was published in 2015 by the International Organization for Standardization (ISO) (ISO, 2015b) and replaced the EN 15038. This standard also provides requirements for the overall translation processes and resources and other aspects that are considered necessary for the delivery of a quality translation service. It also provides the means by which a translation service provider can demonstrate its capability to deliver a translation service that meets the client's requirements.

- **ISO 9000:2017 - Quality management systems - Fundamentals and vocabulary:** This norm defines the Quality Management Systems (QMS) and the necessary procedures and practices for organizations to be more efficient and improve customer satisfaction (ISO, 2015b).

- **ISO 18587:2017 (en), Translation services — Post-editing of machine translation output — Requirements:** This standard defines the requirements for full post-editing published (primarily based on the TAUS guidelines) and provides some guidelines for light post-editing (ISO, 2017).

- **ISO 20539:2019 (en), Translation, interpreting and related technology — Vocabulary:** This document outlines the key vocabulary for translation, interpreting and related technology standards (ISO, 2019). It will be replaced by ISO/AWI 20539, which is currently under revision.

2.4.2 Forthcoming standards

In addition to the standards mentioned above, there are five documents that are currently being developed or updated by ISO and ASTM:

- **ASTM WK46396 — New Practice for Analytic Translation Quality Evaluation:** This standard is currently under development (ASTM, 2021b) and has three main goals: propose a taxonomy of translation errors, establish a process to move from translation specifications to task-specific metrics, and define a scoring method to produce relevant numeric indications of translation quality (Dzeguze et al., 2018).

- **ASTM WK54884 ("HQuest") — New Practice for Holistic Quality Evaluation System for Translation:** This standard is currently under development and aims to "document the core principles necessary to objectively measure translation quality using a holistic approach" (ASTM, 2021c).

- **ISO/CD 5060 — Translation Services – Assessment of Translation Output – General guidance:** The goal of this standard, which is being developed by the Technical Committee ISO/TC 37/SC 5 (Translation, interpreting and related technology), is to provide guidelines and recommendations for the assessment of human translation output and post-edited machine translation output. Its focus lies on an analytical translation assessment method using error categories and error points producing an error score (ISO, 2021).

- **ISO/AWI 11669— Translation projects – General guidance:** This standard will replace the ISO/TS 11669:2012 (ISO, 2012). The original technical specification provided the basis for qualitative assessment, but it did not cover any procedures to carry out quantitative measures.

2.4.3 Other documents (online repositories)

Along with the standards and guidelines published by organisations such as ISO and ASTM, there are some worth noting efforts in the industry to share information on translation quality methodologies through online repositories with the aim to facilitate the access to several initiatives and research projects to wider audiences. These are the most relevant repositories that have are currently available online:

- Multidimensional Quality Metrics (MQM) Error Typology: https://www.w3.org/community/mqmcg/mqm-top- level-2019-04-11/

- Tranquality: Translation Quality: https://www.tranquality.info/

- European project QT21: https://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf

- Harmonized DQF-MQM Error Typology: https://www.taus.net/qt21-project#harmonized-error-typology

- MQM (Multidimensional Quality Metrics): https://www.theMQM.org

3 Methodological approach

The review of the academic, professional, and regulatory approaches to translation quality offers us three different angles on how translation quality should be managed and evaluated in professional contexts. However, each presents different limitations in scope and applicability, and it is difficult to find examples of how these approaches are applied in the translation industry. For this reason, this dissertation aims to propose an interdisciplinary and empirical research aimed to describe the current quality management and evaluation practices in the translation industry and analyse the adoption, impact and suitability of the approaches proposed by academic, professional, and regulatory bodies.

3.1 Introduction

Before discussing the detailed methodological approach to collecting and analysing the data for empirical research, this section will outline the suitability of the four different groups of participants selected and provide the rationale for their inclusion in the study.

As shown in the literature review conducted in the first part of this dissertation, the translation industry operates at three fundamental dimensions: the "buyers" or requestors (that can also be the final clients or an intermediary such as an TSP), the "providers" of the translation service (which can be TSPs or freelancers) and the tools and technologies that allow the translation to take place. The different descriptive studies on the ownership and configuration of the translation quality workflows in professional environments also show that these three dimensions are also present here: translation quality can be managed or evaluated by those final clients that have an internal translation quality management program or by the intermediary providers (multi-language translation service providers that employ regional translation service providers, or generic translation service providers that work with freelance translators). In all the cases, the translation quality management and evaluation tasks are enabled by the use of specific tools and technologies. Therefore, the purpose of this research is to gain insights into the current translation quality practices from the point of view of the representatives of final clients, translation service providers and translation technology companies. These three groups cover the main actors involved in translation quality management and evaluation in the industry and can provide accurate information on the current practices (real scenarios). To complement this panoramic view, this research includes a fourth group composed of experts and researchers specialised in translation

quality methodologies in professional environments that can provide experience-based information on what the translation quality practices should be (ideal scenarios).

The following sections offer an overview of the overall research design and detail the rationale and justification of the methods selected, together with other methodological considerations. They will also detail the components of the research and the methods employed and will conclude with a description of data collection and analysis procedures followed.

3.2 Overview and overall research design

After reviewing the available literature on translation quality management and evaluation practices, no clear framework or methodology that could serve as a guide for this empirical research could be identified. However, different specific methods were identified as potentially suitable for this research, although none of them was deemed to be enough on its own to approach the variety of the questions of this research. Consequently, a combination of qualitative and quantitative methods was used, or the "mixed methods" approach (Creswell et al., 2003).

For the purpose of this study, empirical data was collected utilising survey and interview methods. Both methods include four separated but closely connected questionnaires (one per group of respondents) with close and open-ended question formats. The content analysis and data generation phases are approached from a quantitative and qualitative angle.

Some of the main sources of information used as a reference for the design of this survey were translation and language industry surveys conducted by industry research organisations, such as Nimdzi (Nimdzi, 2021) and the ELIS European Language Industry Survey (ELIA et al., 2021), co-organised by EUATC, ELIA, FIT Europe, GALA, and the EMT university network. These two surveys were particularly relevant in the design of the demographics of the study and provided some guidance for certain questions related to norms and certifications, staff composition, major developments, operational practices, technology and trends.

Another reference used for the design of the questions aimed to help the participants to define the maturity levels of their companies in terms of translation processes, translation analytics

and translation quality management was the Localization Maturity Model developed by CSA Research (CSA Research, 2021a).

Finally, this research counted with different methods of data collection and analysis, which were performed separately. The results of each of the data sets collected were subsequently merged in the analysis and discussion phase. The approach offered the advantage of yielding different perspectives on the topic of translation quality in professional environments and therefore provided further insights and details that were relevant at different angles.

3.3 Outline of the design of the research

The following section will provide a summary of the overall design of the research and will present the foundations for the discussion of the rationale and methodological considerations that will be introduced later.

3.3.1 Surveys

In survey-based research, the information is collected from different cases on the same variables, and it usually involves the use of questionnaires and other data collection methods (de Vaus, 2013). In this research, the two main data collection methods were questionnaires and interviews. One of the main benefits of the survey method is that the information is structured in a variable by case grid, which allows for the comparison of different cases and the identification of casual relationships (ibid). This method is particularly beneficial when the goal of the research is to observe naturally occurring variations in different variables without any intervention from the researcher (ibid).

Finally, this research method is especially suitable for quantitative analysis and systematic data collection from a large number of participants, which could allow for the extrapolation of the results to a larger group than the originally targeted (Oates, 2006).

3.3.2 Interviews

Interview-based research is typically used for qualitative analysis, as this technique allows the researcher to compile information, insights and experiences that might easily be observed if the participants are questioned or somehow guided (Salmons, 2010). Interview-based studies can serve different types of goals, from simply getting information from the respondents to collecting detailed information about processes that cannot be observed.

In structured interviews, the conversation is guided by a set of questions that are prepared before the interview. These questions should be asked in the same order so that responses can be placed in similar categories (Oates, 2006).

For this research, the goal of the structured interviews was to supplement and expand the quantitative data compiled through questionnaires and gain deeper insights into the replies to the previously defined questions. The interviews were conducted online through a videoconference system, and the data collected was transcribed in real-time and integrated into the same data collection system as the one used for the questionnaire. The data collected following this specific method provided further context and extended details and helped to the interpretation and analysis of the other elements of this study.

3.3.3 Questionnaires

Questionnaire-based research can be used for both qualitative and quantitative analysis. This technique is typically associated with survey research (de Vaus, 2013). In the first phase of a questionnaire-based study, the researcher defines the concepts that will be measured and that will be used to perform those measurements (ibid). In the second phase of this type of research, the participants are asked to respond to a series of open-ended or closed-ended questions.

Close-ended questions are particularly useful to compile specific information and allow the researcher to control different variables to a large extend, while open-ended questions are often used to expand and contextualise the quantitative information compiled through the close-ended questions.

One of the main advantages of the use of questionnaires in empirical research is that they are easy to distribute and, in the case of web-based questionnaires, there are several online tools available that enable the collection and automatised analysis of large volumes of data. On the other hand, the design phase of a questionnaire-based study must take into consideration and address the main challenges of this technique, which are mainly related to the potential of misunderstanding or bias from the participants. To prevent these difficulties, this study included a testing phase prior to the publication of the online questionnaire.

In order to conduct this research, a series of self-administered online questionnaires were designed and participation was requested from four different groups of informants (Table 7).

The goal of these questionnaires was to gather insights and perspectives on the current and future practices of translation quality management and evaluation in professional environments from a large number of informants that are actively involved in these tasks and to address the six research questions presented in section 1.3:

1. What is the demographic and professional profile of the companies and participants that carry out translation quality management and evaluation tasks?
2. What is the level of maturity in terms of translation processes and analytics?
3. What is the level of maturity in terms of translation quality management processes?
    3.1. What is the overall level of maturity in translation quality management?
    3.2. What is the overall level of knowledge and use of translation quality management norms and standards?
    3.3. What are the documented processes and human resources used in translation quality management programs?
        3.3.1. How does the translation quality management program work?
        3.3.2. What are the translation quality evaluation methodologies used?
        3.3.3. How does the translation quality evaluation process work?
        3.3.4. What kind of human resources are involved in translation quality management and evaluation processes?
    3.4. What are the main tools and technologies used in translation quality management and evaluation processes?
4. What are the main pain points and limitations in translation quality management and evaluation processes?
5. What solutions and workarounds are applied in the industry to overcome the current pain points and limitations?
6. What are the future challenges and trends in translation quality management, and what kind of initiatives would help the advancement of the translation quality management practices in professional environments?

A master questionnaire covering the six research questions mentioned earlier was explicitly developed for this research, and four variations were derived to suit the different groups of informants. Questions were designed based on sources such as the academic literature reviewed in section 2.2, the professional methodologies presented in section 2.3 and the regulatory approaches described in section 2.4.

This procedure resulted in four questionnaires that presented slight variations and were used to compile information from a comprehensive set of target audience groups.

Nr.	Target audience group	Target audience sub-group (profile of the company)	Target audience sub-group (profile of the respondent)
1	Corporations and organisations (buyers)	Group 1: Translate 5 languages or less Group 2: Translate between 5 and 10 languages Group 2: Translate more than 10 languages	Group 1: CEO Group 2: Translation/Localisation Manager or Director Group 3: Translation/Localisation Program Manager Group 4: Translation/Localisation Quality Manager or Director Group 5: Language services Manager or Director Group 6: Researcher (academic) Group 7: Researcher (non-academic) Group 8: Independent consultant Group 9: Other
2	Translation service providers (providers)	Group 1: Translate between 1 and 3 languages Group 2: Translate between 4 and 10 languages Group 2: Translate more than 10 languages	
3	Translation technology providers	Group 1: Translation management tools Group 2: Translation quality management tools	
4	Subject matter experts and researchers	Group 1: Professional services in translation quality management Group 2: Academic researcher or teacher	

Table 7. Overview of target audiences included in the survey part of this research.

All the questionnaires were published together in one online survey and started with a set of shared demographic questions.

After this first set of questions, each questionnaire branched off depending on the professional background of the responses. The goal of this first shared demographic section was to relate the demographic of the profile of the respondents and the entities with the answers provided to the research questions.

3.4 Rationale for using the mixed methods approach

As presented in the introduction of the methodological approach of this research, this study aims to study the current practices in translation quality management and evaluation from different angles, which presents different levels of complexities. One of the complexities observed is that translation quality management and evaluation tasks are performed or facilitated by a wide range of different target groups, both in terms of the type of entities involved and the professional background and experience of the professionals implicated in these processes (Table 7). This level of complexity is not unusual in social research studies, which tend to use mixed methods to be able to fully address the challenges of this complexity (Creswell et al., 2003). One of the main advantages of the approach chosen to conduct this research is that it allows the integration of different data sources and perspectives. Structured interviews provide us with qualitative data that include personal experiences and opinions, while questionnaires with open-ended and closed-ended questions offer both quantitative and qualitative data covering both experiences and processes. Thanks to the combination of the strengths of both methods, we can isolate and analyse generalisable trends and gain a better understanding of the topic of this research than what we could have achieved through only one approach.

3.5 Methodological considerations

There are multiple considerations that justify the use of an online approach to conducting this research, the main ones being that this type of approach allows the research to reach different groups of target audiences that would be very difficult (or almost impossible) to reach due to geographic boundaries, and that online interviews and questionnaires are quicker and cheaper to design, distribute and conduct than offline or in-person interviews and questionnaires. Given the fact that the translation and localisation industry is dispersed

globally and that the current research does not take into consideration geographical differences as a variable to be studied (although this could be an aspect that could be developed in further studies), the two advantages mentioned seem to be particularly relevant for the purpose of this research.

However, due to the reliance on online research methods of this research, special consideration must be given to certain issues that are specific to these types of methods. The following sections will provide more detailed information on the challenges and ethical considerations applied to this study.

3.5.1 Online interviews

Online interviews can have multiple advantages, but they also present specific challenges and disadvantages that should be considered. For this research, the interview technique chosen was the synchronous communication method via Voice over IP (VoIP) technologies, as it facilitates the exchange of information in real-time, and it has some of the advantages of in-person interviews (Salmons, 2010). All the interviews were performed synchronously following the same structure and questions that were designed for the questionnaire method.

One of the main benefits of conducting online interviews instead of traditional face-to-face conversations is that it gave the researcher access to a wide range of professionals and scholars from different countries and cultures and that this method allowed the participants to choose the most suitable time and date to accommodate to their time schedules. This advantage proved to be particularly important during the course of this research, given the fact that the interviews were conducted between 2020 and 2021, when travel restrictions and health-related measures, such as lockdowns and gathering restrictions, were in full effect in many countries due to the COVID-19 pandemic.

The access to a variety of respondents from different geographical locations was particularly valuable for a global research in translation and localisation such as this one, as it helped lower the potential impact of culturally or geographically marked variables.

A final consideration regarding the use of online interviews is that all the different groups interviewed were quite familiar with this form of synchronous communication from their daily work and personal life. Moreover, none of the participants expressed having any difficulties or feeling uncomfortable about speaking with the researcher using this method.

3.5.2 Online questionnaires

The questionnaires were designed and published through an online survey tool (Google Forms[4]), and the respondents were recruited through different online channels, such as email, LinkedIn, Twitter, and distribution lists and social media channels from academic organisations (CenTraS, TRANSLATIO, AESLA), and professional entities (GALA, Nimdzi, ATRAE). The variety of online promotion channels allowed the questionnaire to be easily distributed among a wide group of target groups. In addition to this, participants had the opportunity to fill in the online survey at their convenience without any intervention or help from the researcher.

Despite the increasing use of online questionnaires in research studies due to the low cost and high efficiency of this method, some specific challenges should be taken into account and addressed in the design phase of the study.

The first challenge is that online surveys tend to get lower response rates than other traditional methods (Manfreda et al., 2008). To overcome this challenge and increase the number of responses, different scheduled reminders were sent after the initial launch of the survey. In addition, the amount and frequency of the reminders were adapted to the type of channel used (emails, social media channels or distribution lists) and the approach (direct communication with potential participants versus generic messages in groups or communities).

The second challenge that needs to be considered is the possibility of having one participant respond to the same questionnaire several times, which could skew the data. Since the questionnaire did not gather personal data such as names, titles, or location to ensure the confidentiality of the data gathered and the anonymity of the participants, one partial solution was to configure the settings of the questionnaire platform (Google Forms) to restrict the number of valid responses to only one per IP address. These IP addresses were detected directly by the platform, but they were not visible or accessible in any shape or form to the researcher.

---

[4] https://www.google.com/intl/en-GB/forms/about/

3.6 Ethical considerations

When conducting an empirical study involving human subjects, it is imperative to consider and apply certain best practices and guidelines that guarantee the ethics of the research in general and the use of the most appropriate methodologies for data compilation and analysis. These guidelines state that the individuals that are asked to participate should have the right to decline the invitation and to withdraw at any point if they wish so. The participants should also have the right to give their informed consent, the right to anonymity and the right to confidentially (Oates, 2006). For this reason, researchers should ensure that they provide to the participants an explanation of the main characteristics and goals of the research, and they should design their research in a way that allows the anonymisation of names of individuals or organisations while avoiding the compilation of confidential or irrelevant information (ibid).

3.6.1 Interview ethics

Online interviews tend to follow the same ethical consideration as face-to-face interviews, although a couple of additional considerations were taken into account before, during, and after the interview process.

Before the interview started, participants were provided with a summary of the characteristics and goals of the research and were asked to confirm their consent, which was recorded through the survey platform (see Appendix B).

Once the session started and before proceeding with the first questions, the participants were reminded that they could withdraw at any point and that the statements they provided could be used in subsequent publications but that their names, the names of their companies or any other identifying information would be dutifully omitted.

During the interview, special care was taken as not to ask irrelevant questions to the participants, and the questions on demographics were limited to professional experience and general size and characteristics of the companies the participants worked for.

Another guideline for interview-based research proposed by Oates (2006) is that participants should be allowed to confirm that the facts transcribed by the interviewer correspond to what they intended to express. This goal was achieved by using the screen functionality of the videoconference tool to show the real-time transcription of the replies provided by the

participants. This allowed them to correct or expand the recorded information and confirm that they were satisfied with the accuracy of the transcription of each question before proceeding to the next one.

3.6.2 Questionnaire ethics

The questionnaire-based research followed the same principles and best practices outlined in the previous sections. Consequently, the first page of the online questionnaire contained an introduction with the main characteristics and goals of the research and a specific consent section that reminded the participants of their right to withdraw at any point. The last section of the consent section prompted the participants to confirm they had understood the information provided and asked them to confirm their consent by clicking "Yes" to be able to proceed with the questionnaire itself (see Appendix B). As outlined in section 3.5.2, the survey platform used only allowed one valid response per IP address. These IP addresses were not visible to the researcher, and therefore they were not part of the data compiled for this research. No other personal information was collected, except email addresses, which were facilitated voluntarily by respondents who declared their interest in receiving a summary of the results of this study. This piece of data was therefore excluded from the data analysis phase. The survey platform only records the questionnaires that have been fully completed by the participants. Therefore it was not possible to determine the withdrawal rate of this research in relation to the number of entries received.

3.7 Data collection

In total, more than 4,000 data points were collected from the 68 participants that took part in the study.

The following sections will present the approach taken to data collection and data analysis and will also dive into the specific details pertaining to sampling techniques, categorisation of participants, and design of the structure of questionnaires and interviews.

3.7.1 Sampling techniques

In order to ensure a well-rounded set of perspectives on the topic of translation quality management and evaluation from different angles, this research used three different but complementary sampling approaches: nomination, convenience, and snowball sampling. We will briefly describe the rationale behind the use of each of these approaches before we

outline the main characteristics of the categorisation of the participants, the collection of data, the structure of the interviews and questionnaires and the analysis of the results.

3.7.1.1 Nomination sampling approach

Part of the participants was selected through existing records or using the nomination sampling approach in which other participants or experts recommend potential candidates, which are then identified based on common patterns (Salmons, 2010). As a result of these two sources of information, 76 candidates were invited to take part in this study, of whom ten chose to participate via an online interview, and 37 preferred to answer the online questionnaire. The remaining 29 did not respond to the researcher's request. It is important to note that both the interview and the questionnaire designs were tested in a small pilot conducted with the help of two volunteers to refine the question phrasing, confirm the relevance of the questions selected in the design phase and adjust the time required to complete them. The results of the pilot were not included in the analysis of the data compiled.

3.7.1.2 Convenience and snowball sampling approaches

With the aim of complementing the results obtained by the nomination sampling approach described in the section above and being able to reach a more extensive set of individuals, communities, companies, and organisations involved in translation quality management and evaluation processes, this research makes use of two additional techniques: convenience and snowball sampling. Consequently, the online form used for the questionnaires and interviews was open to the public (Ackland, 2013) and the link to the survey was sent to several social media platforms (LinkedIn, Twitter and Facebook), distribution lists from academic and professional organisations (CenTraS, TRANSLATIO, AESLA, ATRAE), and social media channels of industry-related entities (GALA and Nimdzi).

To avoid any skewed results due to the participation of individuals that did not belong to any of the main four target groups identified in the design of the research, the demographic section of the questionnaire and the interview had a specific question to enquire about the participant's connection to translation quality management or evaluation processes. This question had a special logic associated so that any participant who indicated that they had no connections to the topic of the research could be automatically excluded from the survey.

One of the main disadvantages of the convenience and snowball techniques is that they are non-probability approaches (Taherdoost, 2016), and therefore they are not suitable to make

statistical inferences in relation to the wider population or to get information about response rates. However, they are very effective in research studies that want to focus on small samples of the population (such as, in this case, the translation quality management and evaluation practitioners) and that are intended to analyse real-life experiences and perspectives (ibid).

These two sampling approaches provided 21 additional participants, of whom 2 scheduled an online interview with the researcher, and 19 preferred to answer the online questionnaire.

3.7.2 Categorisation of participants

In order to be able to conduct a study that will make use of interviews and questionnaires as the primary sources of information, the researcher needs to make sure that the participants can offer meaningful, relevant, and accurate data about the topic subject of the research and are willing to participate not just in the interview or questionnaire itself, but also in any other related communications (Salmons, 2010). For this reason, the participants were selected using purposive sampling techniques (ibid) and divided into four different groups:

- Representatives of internal corporate translation or quality management teams (buyers)
- Representatives of translation service providers (translation providers)
- Representatives of language technology companies (technology providers)
- Translation quality management and evaluation experts (subject matter experts)

The selection of these four categories aimed to present a multifaceted set of perspectives on the topic of translation quality management and evaluation from both the buyer and the provider side and academic and professional angles.

3.7.3 Data gathering

Online questionnaires and interviews were designed with a common section (demographics) and branched off into four separated sets of questions depending on the primary role of the respondent (buyer, provider, technology provider or subject matter expert). However, most of the questions were generally the same, with some questions being removed or added according to their relevance to the different target groups (see Appendix C).

All practitioners, researchers and subject matter experts involved in translation quality management and evaluation process were welcomed to participate in the study, which was disseminated using the sampling techniques described in section 3.7.1.

The following sections will outline some of the specific characteristics of the data gathering process in the interviews conducted for this study.

3.7.3.1 Online interviews

As described in section 3.5.1, this research included several structured interviews that were conducted synchronously via VoIP. The main tools used for this phase of the research were a videoconference plug-in (Google Meet[5]) and an online scheduling tool (Calendly[6]) that allowed participants to select the most convenient meeting slots according to their availability and that of the researcher's. This online scheduling tool was configured to send automatic meeting invitations to both the participant and the researcher with a fixed duration of one hour. The invitations included a summary of the structure of the interview and the main goals of the research, as well as an additional note indicating that the interviews would take around 45 minutes. However, during the interview with the researcher, conversations were allowed to run over the estimated 45 minutes in those cases in which the participant indicated they were interested in extending the interview duration.

To allow for a faster compilation and systematic analysis of the information obtained through this method, the interviews had the same structure and questions as the questionnaires, and the researcher filled the web form in real-time with the information provided by the participants in each session.

3.7.4 Questionnaire design

The questionnaires designed for this research consisted of five different sections (see Appendix C): one section shared by all the target groups (demographics) and four separate sets with variations of the questions depending on the primary role of the respondent (translation processes and analytics, translation quality management and evaluation, current challenges, solutions, and the future of translation quality management) and one final shared

---

[5] https://meet.google.com/
[6] https://calendly.com/

section (general comments and feedback). The master questionnaire contained 250 questions that were distributed per target group as follows:

Sections	Translation buyers	Language Service providers	Technology providers	Subject matter experts	Total
Section 1: Demographics	7	7	7	7	28
Section 2: Translation processes and analytics	2	1	0	0	3
Section 3: Translation quality management and evaluation	51	53	21	42	167
Section 4: Current challenges	8	8	8	8	32
Section 5: Solutions	1	1	1	1	4
Section 6: The future of translation quality management	3	3	3	3	12
Section 7: General comments and feedback	1	1	1	1	4
Total	73	74	41	62	250

Table 8. Distribution of the questions per section and target group.

Shared section: Demographics

The demographic section included questions to assess the profile of the respondents' company (size, sector, annual translation volumes) and define the profile of the respondents themselves (job title and description of their role in the translation industry). Possible answers to this last question, which was used to determine the research category to which the respondents would be assigned and the subsequent sets of questions, were:

- I am part of an Internal Corporate Translation/Localisation Team (Group 1).
- I work for a Translation or Language Service provider (Group 2)
- I work for a Language Technology Company (Group 3)
- I offer consulting or research services in translation quality management (Group 4)
- I work as a teacher or researcher (Group 4)

Branching sections

After branching, all participants were presented with four separate sets with variations of the questions depending on the primary role of the respondent. Each section covered the following topics:

- Translation processes and analytics (Groups 1 and 2)
- Translation quality management and evaluation (All groups)
- Current challenges (All groups)
- Solutions (All groups)
- The future of translation quality management (All groups)

Question design

The goal of some of the questions included in the questionnaires was to explore evaluative attitudes, while other questions covered cognitive attitudes. To guide the respondents through these differences, the questions related to evaluative attitudes were signalled with phrases such as "In your opinion" or "What/Who/When/How/Where do you think". At the same time, the questions related to cognitive attitudes typically would include a definition of the concepts mentioned. The questionnaire also included several types of close-ended questions (multiple-choice, checkboxes, and ranking questions) and contained several open-ended questions, which allowed the participants to expand their answers without any influence from the researcher (see Appendix D).

In some cases, participants were asked to share their opinions or experiences in an open-ended format before being asked a similar question through some questions with a closed-ended format, or vice versa. This approach was used first to incite a spontaneous response from the participants without any guidance or influence and then use the set of closed-ended questions on the same topic to foster a deeper reflection on the research topic. This allowed the researcher to compare whether both types of responses were consistent and the differences between them.

During the design phase of the questionnaire, and due partly to the sampling approach taken (see section 3.7.1), it was decided to prioritise high-quality responses and rich information from fewer highly engaged professionals over a larger number of low-quality, more generic responses. For this reason, special care and consideration were taken to make it easier for the respondents to follow and complete all the questionnaire sections. Some of the techniques used to reduce the level of fatigue included choosing different question types to keep respondents' attention, automatic branching, and filtering to avoid redundant questions and indications of the progress of the questionnaire.

Once the design of the questionnaire was completed, the overall structure, logic and duration of the process were tested in a pilot. During the pilot, a first version of the questionnaire was uploaded to an online platform (Google Forms) and tested by the researcher and two additional scholars. Based on the observations compiled, some changes were introduced to ensure that all the questions and definitions were worded clearly and did not lead to confusion, to make sure all the questions were relevant.

3.7.5 Structure of the interviews

Interviews were conducted individually between the respondent and the researcher, who also acted as the interviewer. They were always held in English and, even though several participants were not native English speakers, the use of English as a vehicular language did not cause any communication issues. As stated by Oates (2006), one of the most important tasks before the start of the research interview is to make sure that the respondent knows the purpose of the interview and to provide assurances about the confidentiality and anonymity of their responses. For this reason, all the interviews had an introductory section with these three steps:

1. First, the researcher provided a general overview of the topic and the characteristics of the research.
2. Secondly, the researcher shared her screen to show the contents of the consent section and asked the respondents to confirm their willingness to participate in the research.
3. After that, participants were asked to introduce themselves, including their experience in translation, their connection to translation quality management and evaluation processes and their actual role.

After this introduction, the researcher showed the online form with the different sets of questions previously defined and proceeded to formulate them in order. In the case of the open-ended questions, the researcher transcribed the replies provided by the participants in real-time and asked follow-up questions to confirm she had captured the participant's feedback accurately and allow for richer data to be captured.

Finally, once the interview had covered all the predefined questions, participants had the opportunity to add any further insights or comments about the research topic. Some participants did choose to expand on specific points previously raised, while others declared they were satisfied with the transcription of their replies shared by the researcher.

4 Data analysis and research findings

This chapter describes the findings of the questionnaires and interviews that were part of this research on the current practices on translation quality management and evaluation in the industry. Section 4.1 will present the demographics of all the participants in the study, and sections 4.2, 4.3, 4.4 and 4.5 will focus on the results of each of the categories: buyers, TSPs, technology providers, and experts in translation quality management. These sections are divided into six topics that correspond to the main research questions presented in section 1.3:

- Translation processes and analytics
- Translation quality management
- Current challenges in translation quality management and evaluation
- Solutions
- The future of translation quality management

Questionnaires were analysed in Google Forms and Microsoft Excel[7] using two main techniques: frequency analysis and cross-tabulation with pivot tables.

4.1 Demographics of the study

As described in Chapter 3, one of the main goals of this research is to get a wide variety of perspectives on the current practices on translation quality management and evaluation in the industry from both the buyer and the provider side and academic and professional angles. Therefore, in this section, we will present the general demographic distribution of the participants and the companies they represented.

---

[7] https://www.microsoft.com/en-us/microsoft-365/excel

4.1.1 Profile of the participants

What is your job title?
68 responses

- Translation/Localisation Quality Mana…
- Translation/Localisation Manager or D…
- Language services Manager or Director
- Translation/Localisation Program Man…
- CEO
- Researcher (academic)
- Researcher (non-academic)
- Independent consultant

1/3
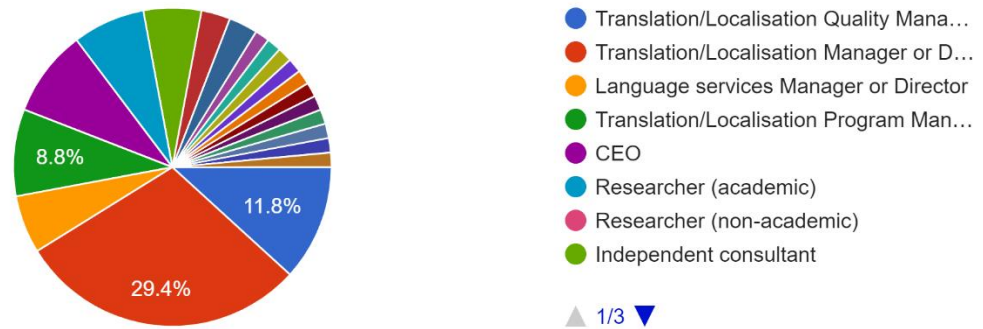
Figure 4. Profile of the participants.

This question provided eight different options for the respondents to choose from, but it also allowed for the inclusion of additional job titles. The eight predefined answers had a direct correlation with the categorisation created during the design phase of the research, which was presented in Table 7.

- Option 1: CEO
- Option 2: Translation/Localisation Manager or Director
- Option 3: Translation/Localisation Program Manager
- Option 4: Translation/Localisation Quality Manager or Director
- Option 5: Language services Manager or Director
- Option 6: Researcher (academic)
- Option 7: Researcher (non-academic)
- Option 8: Independent consultant
- Option 9: Other

Of the 68 respondents, approximately half of them declared that their job title was either Translation/Localisation Manager or Director (29.4%), Translation/Localisation Quality Manager (11.8%) or Translation/Localisation Program Manager (8.8%), which are indeed the types of roles that tend to have a closer contact with translation quality management and

evaluation processes. Fifteen participants chose the category "Other" and specified different job titles, which are shown in Table 9.

Job Title	Numbers	Percent
Translation/Localisation Manager or Director	20	29.41%
Translation/Localisation Quality Manager	8	11.76%
CEO	6	8.82%
Translation/Localisation Program Manager	6	8.82%
Researcher (academic)	5	7.35%
Independent consultant	4	5.88%
Language services Manager or Director	4	5.88%
Language Lead	2	2.94%
Senior localization specialist	2	2.94%
Language Specialist Team Lead	1	1.47%
Localization coordinator	1	1.47%
Localization Quality and Vendor Manager	1	1.47%
Localization Quality Control	1	1.47%
Product Owner (software industry)	1	1.47%
Quality Strategist	1	1.47%
Senior Localisation Quality Engineer	1	1.47%
Sr. Localization Quality Strategist	1	1.47%
Translation Coordinator	1	1.47%
Translation Quality Intern	1	1.47%
Translation/Localisation Project Manager	1	1.47%
Grand Total	**68**	**100.00%**

Table 9. Extended view of job titles of the participants.

As we can see, there seems to be a wide variety of job titles that present variations from the main job roles presented in these questions. For example, some participants have "Language" instead of "Translation" or "Localisation"; others have specific functions specified in their roles, such us "Lead" or "Team Lead", "Coordinator", or "Strategist". Finally, some candidates had certain job titles that the researcher did not consider initially, such as "Translation/Localisation Project Manager" or "Localisation Quality Engineer".

4.1.2 Categorisation of the participants

Which description best fits you?
68 responses

- I am part of an Internal Corporate Translation/Localisation Team (client or buyer)
- I work for a Translation or Language Services provider (services provider)
- I work for a Language Technology Company (technology provider)
- I offer consulting or services in translation quality management
- I work as a teacher or researcher
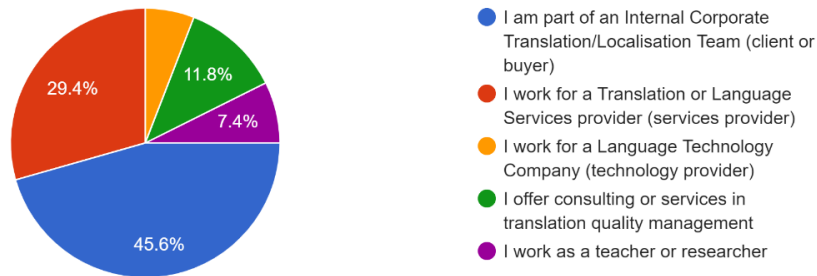
Figure 5. Categorisation of the participants.

As shown in Figure 5, there was a high percentage of participation on the buyer's side (45.6% of participants), followed by the provider's side (29.4%). Experts in translation quality management accounted for 19.2% of participants, of which 11.8% were non-academic, and 7.4% were teachers or researchers. Translation technology providers were the smallest group in terms of representation and accounted for 5.9% of the overall responses.

4.1.3 Categorisation of the organisations

In order to have a better overview of the characteristics of the organisations in which the respondents worked, the research included three different questions aimed to discover the organisation's size (in terms of overall staffing resources and annual translation volumes) and the sectors in which the organisations operate.

What is the size of the staff of your organisation?
68 responses

- 1-25
- 26-49
- 50-99
- 100-199
- 200+
- Not sure/Not applicable
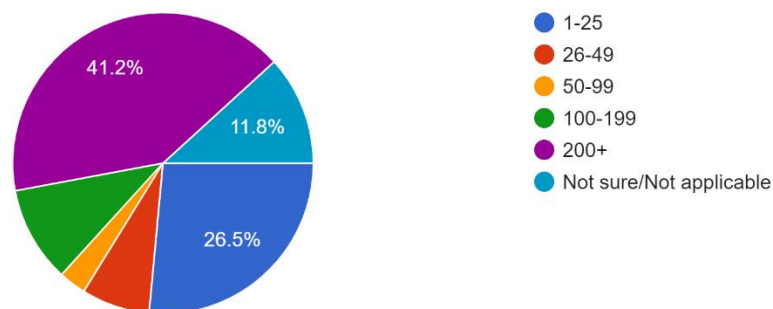
Figure 6. Size of the staff of the organisations represented by the respondents.

41.2% of the respondents indicated that they worked in companies with 200+ human resources, while 26.5% declared that their companies had 25 employees or less. This disparity in the size of the organisations might be partially caused by the fragmentation of the translation industry, which is composed of very large multi-language vendors (MLVs) but also small single-language vendors (SLVs). Proof of this divide between these two types of language services companies is that the top 100 large translation service providers accounted for only 15% of the overall revenue of the language industry in 2020, according to the latest market research report published by Nimdzi (2021).

What are the annual translation volumes of your organisation?
68 responses
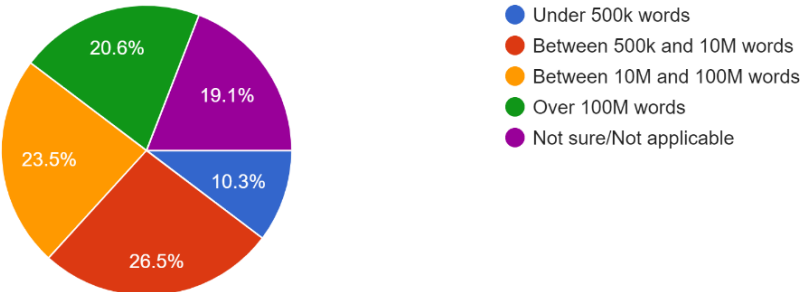
Figure 7. Annual translation volumes.

If we take a look at the annual translation volumes of the organisations, we will find a balanced distribution of companies with more than 100M words (20.6%), organisations that have between 10M and 100M words (23.5%), companies that report between 500k and 10M words (26.5%) and entities with less than 500k words (10.3%), which will be very useful to examine the different realities that take place in buyers and providers with in small, medium and big annual translation volumes.

In which sector(s) does your organisation operate?

68 responses
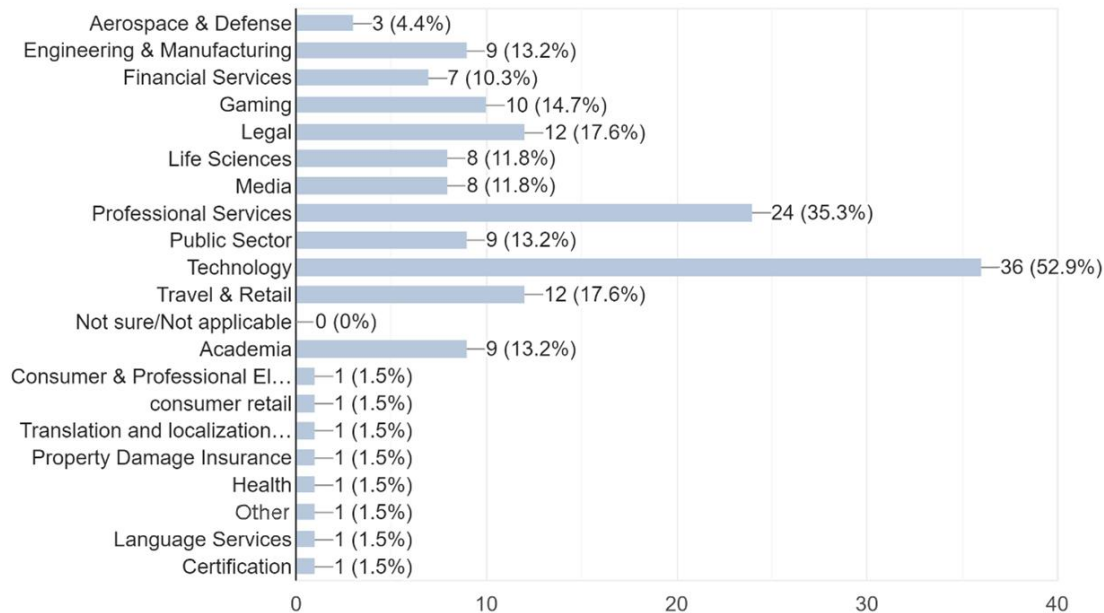
Figure 8. Sectors of the organisations.

The variety for the participants of this research is also reflected in the number of different sectors selected in this question, which allowed multiple-choice selections and the addition of new categories. If we temporarily set aside the categories "Technology", which many companies use as an umbrella term that complements the other sectors, and the category "Professional services", which is also used as an umbrella term by many TSPs, it is possible to identify eight main sectors that account for roughly 80% of the results:

- Engineering & Manufacturing (13.2%)
- Financial services (10.3%)
- Gaming (14.7%)
- Legal (17.6%)
- Life Sciences (11.8%)
- Media (11.8%)
- Public sector (13.2%)
- Travel and retail (17.6%)
- Academia (13.2%)

82

4.2 Buyers of translation services

Out of the 68 participants in this research, 31 (45.6% of the total) worked in internal translation or localisation teams on the buyer's side. In the following sections, we will present and analyse the data collected from this target group in each of the seven parts of the questionnaires.

4.2.1 Demographics
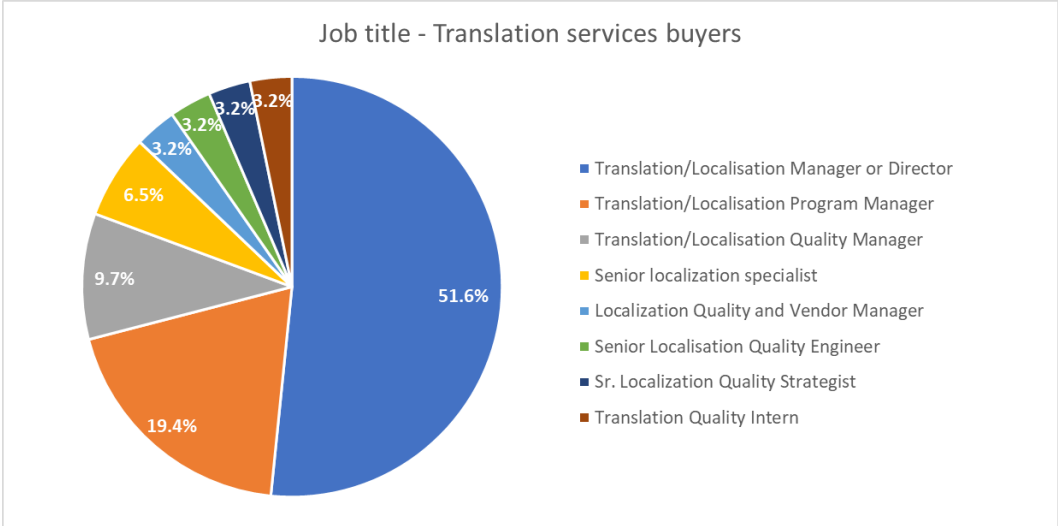
Figure 9. Job titles – Translation service buyers.

Similarly to the results observed on the overall demographics section, the majority of the participants from the buyer's side were Translation/Localisation Managers or Director (51.6%), followed by Translation/Localisation Program Managers (19.4%) and Translation/Localisation Quality Managers (9.7%).
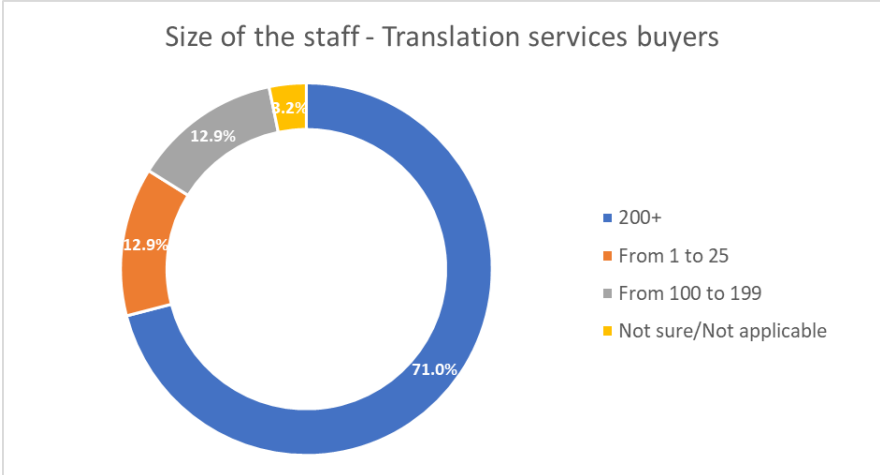
Figure 10. Size of the staff – Translation service buyers.

More than 70% of the respondents on the buyer's side indicated that they worked in companies with 200+ human resources, while 19.5% declared that their companies had between 100 to 199, and another 19.5%, that their companies counted with 25 employees or less.
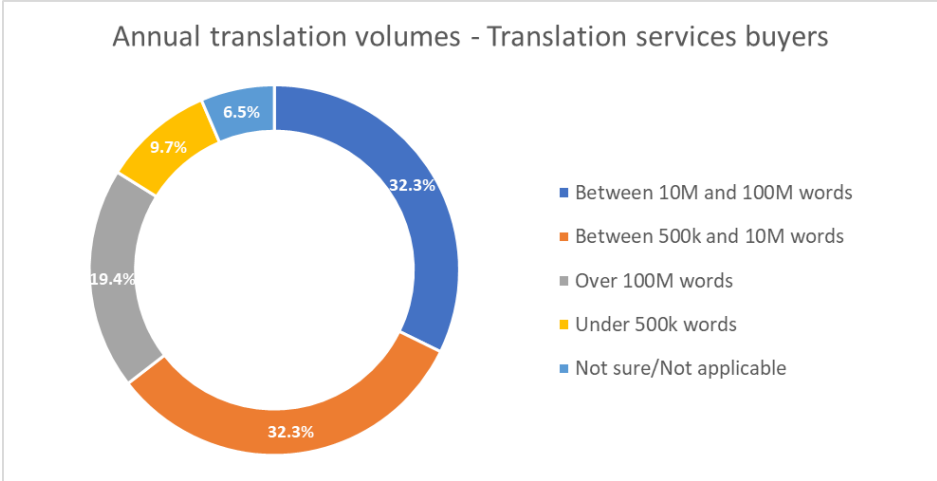
Figure 11. Annual translation volumes – Translation service buyers.

Despite having a majority of companies with 200+ employees, if we look at the annual translation volumes reported by the respondents, we can notice that a bigger size in terms of staffing does not necessarily mean bigger translation volumes. Instead, there is a distribution of companies with "large" translation volumes (19.4% reported translating more than 100M words per year), "medium" translation volumes (32.3% reported between 10M and 100M words, and another 32.3%, between 500k and 10M), and "small" translation volumes (9.7% reported less than 500k words per year).
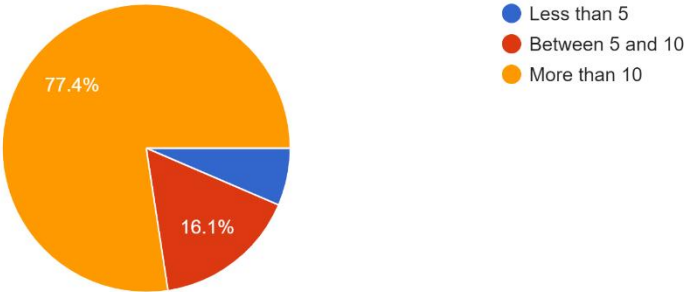
Figure 12. Languages translated – Translation service buyers.

Interestingly enough, despite the different realities in terms of translated volumes observed in the previous figure, 77.4% of the respondents reported that their companies worked with or outsourced more than ten languages and only 6.5% that their companies worked with or outsourced less than five languages.

4.2.2 Translation processes and analytics

How would you score the actual level of maturity of your organisation in terms of translation and localisation processes?
31 responses

Figure 13. Maturity level in terms of translation and localisation processes – Translation service buyers.

None of the participants in this research considered that the companies that they worked for were in the lowest maturity tiers (levels 1 and 2), while more than 70% of the respondents thought that their companies were either on level 3 (19.4%) or level 4 (54.8%). Finally, 25.8% of the participants reported that the level of maturity of their translation and localisation processes was the highest of the scale.

How would you score the actual level of maturity of your organisation in terms of translation analytics?
31 responses

Figure 14. Maturity level in terms of translation analytics – Translation service buyers.

However, if we analyse the results of the following question, which asked the participants to rate the maturity level of their organisations in terms of translation analytics, we find that the picture is slightly different as the majority of the respondents considered that their companies were either on levels 2 (35.5%) or 3 (38.7%). 12.9% of the participants reported that their company was on level 1, and the same percentage of participants considered that their company was on level 4. None of the respondents chose level 0 or 5.

4.2.3 Translation quality management

The third part of the questionnaire treated different elements of the translation quality management process followed by the subjects of the study and was divided into four main topics: overall level of maturity in translation quality management processes, knowledge and use of norms and standards, documented translation quality management and evaluation processes, and tools and technologies. We will see each of these topics in detail in the following sections.

4.2.3.1 Overall level of maturity in translation quality management processes (self-assessment)

How would you score the actual level of maturity of your organisation in terms of translation and localisation quality management processes?
31 responses

- 1 - Manual: My organisation uses manual and/or legacy processes
- 2 - Automated: My organisation has started to integrate some automation i…
- 3 - Agile: My organisation has automated and scalable quality mana…
- 4 - Centralised: My organisation practices strategic and centralised tra…
- 5 - Expert: My organisation benefits from translation as a revenue center and is…
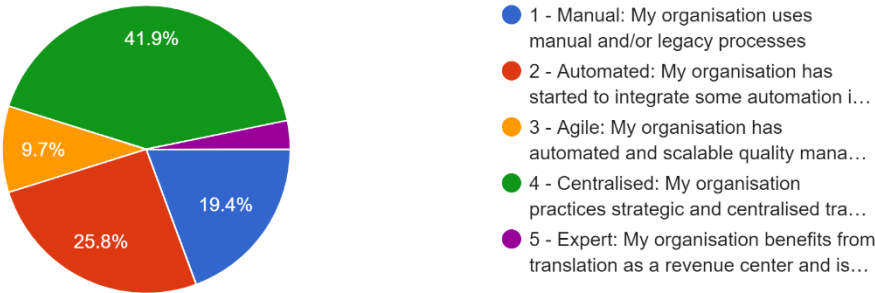
Figure 15. Maturity level in terms of translation quality management processes – Translation service buyers.

In this question, the participants were asked to self-assess the level of maturity of the quality management process of their companies according to a similar scale to the one used in the previous two questions. As we can see in Figure 15, the results revealed two polarised tendencies: 45.1% of the respondents considered that their companies were in the highest

86

levels (41.9% in level 4, and 3.2% in level 5), while the other half (45.2%) reported that their companies were on the lowest tiers (25.8% in level 2, and 19.4% in level 1).

4.2.3.2 Knowledge and use of norms and standards
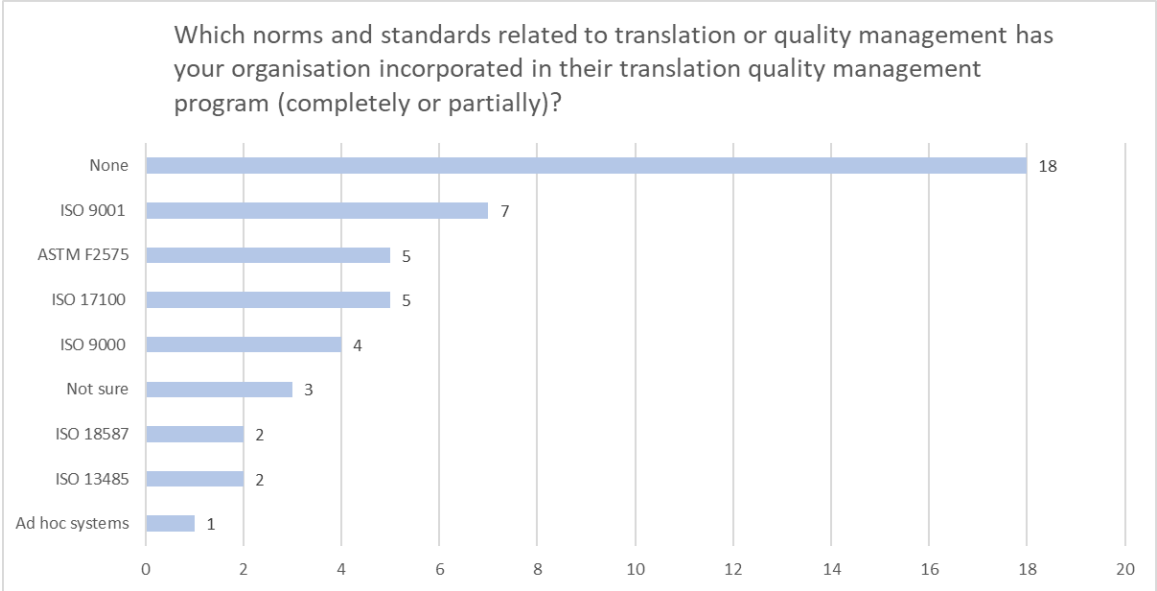
Figure 16. Knowledge and use of norms and standards – Translation service buyers.

Surprisingly (or maybe not so much so, if we take into consideration that translation service buyers do not need to have a certification in translation services or translation quality management), more than half of the participants on the buyer's side (58.1%) declared that their translation quality management program did not follow any norm or standard. Those that took into account one or more of the current quality management or translation services standards mentioned the ISO 9001 and 17100 norms and ASTM F2575.
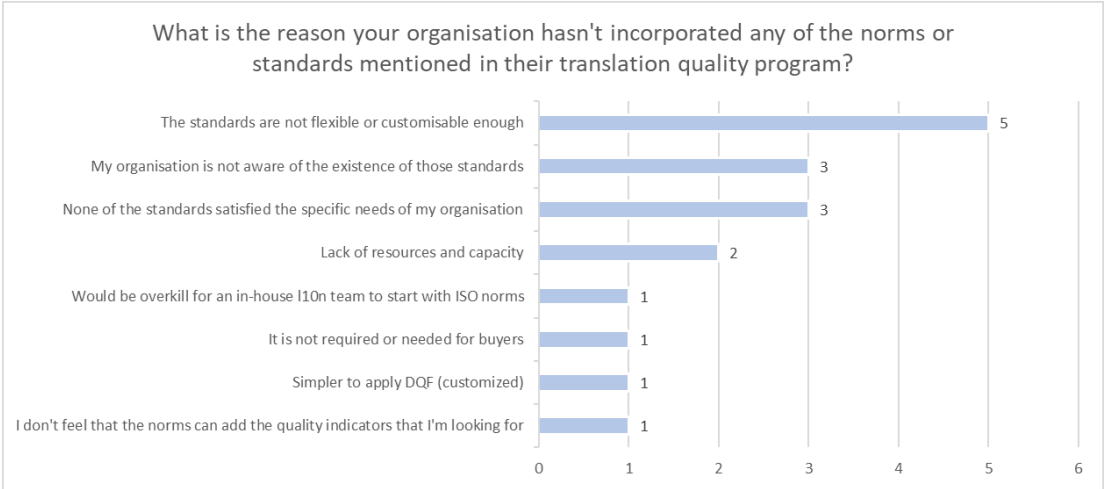
Figure 17. Reasons to not use any norms or standards – Translation service buyers.

Respondents that had selected "None" in the previous question were prompted to a follow-up question that allowed them to specify one or more reasons why the translation programs of their companies did not make use of any translation quality management or translation services norms or standards. Some of the reasons more commonly cited were that the standards were not flexible or customisable enough, that their companies were not aware of the standards, or that they considered that the standards did not meet the specific needs of their organisations.
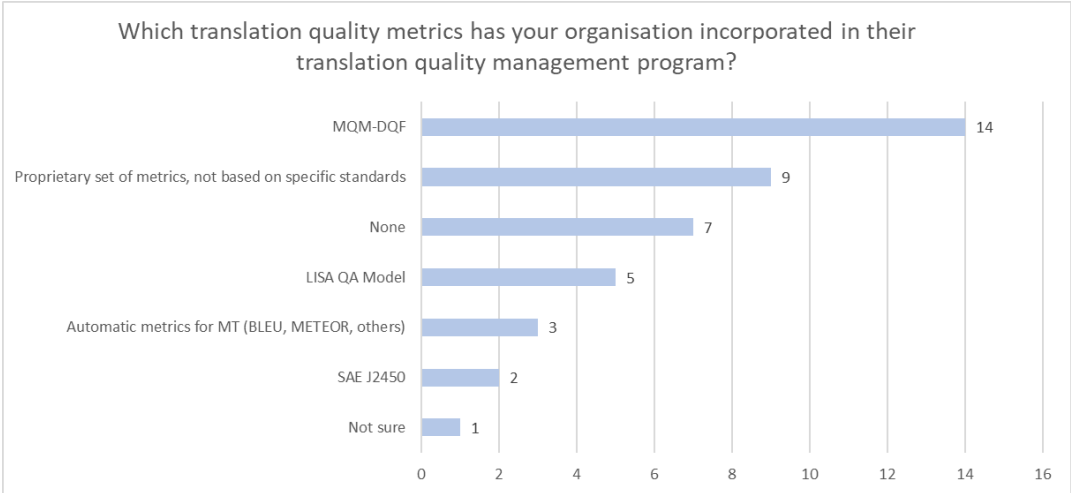
Figure 18. Use of translation metrics – Translation service buyers.

45% of the participants on the buyer's side mentioned MQM-DQF as one of the translation metrics used in their translation, and a high number of respondents reported that they used proprietary metrics that were not based on any specific standards (29%) or any metrics at all (23%). Interestingly enough, around 22% of the respondents mentioned they were still using older metrics such as LISA or SAE J2450, and only 10% reported using specific metrics for machine translation.
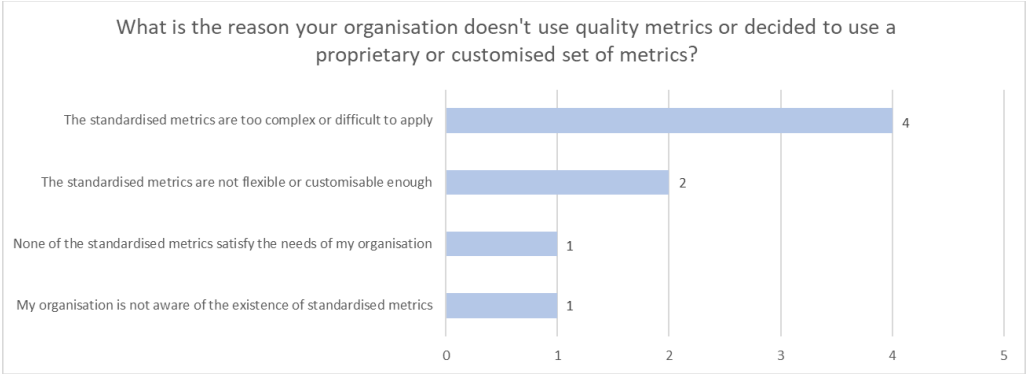
Figure 19. Reasons to not use standardised translation metrics – Translation service buyers.

Respondents that had selected "None" in the previous question were shown a follow-up question that allowed them to specify the reasons why the translation programs of their companies did not make use of any standardised translation metrics. The two reasons more selected were that the standardised metrics were too complex or difficult to apply and that they considered that standardised metrics were not flexible or customisable enough.

4.2.3.3 Documented translation quality management and evaluation processes

The aim of this part of the survey was to get deeper insights into the processes that the translation service buyers follow, and it was divided into three different topics: overall translation quality management processes, specific translation quality evaluation processes and processes related to human resources, training, and qualifications.

Overall translation quality management processes

What types of content are included in the translation quality management process?
31 responses
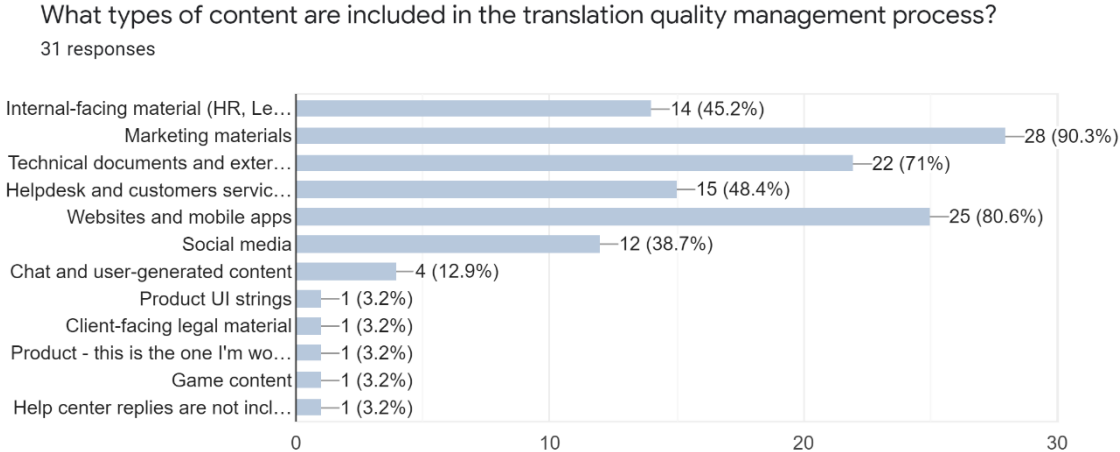
Figure 20. Types of content included in the translation quality management process – Translation service buyers.

According to the participants of the research, the types of content that were usually included in the translation quality management processes of their companies included marketing (90.3%), websites and mobile apps (80.6%), technical documents and external documentation (71%), helpdesk and customer service content (48.4%) and internal-facing material (45.2%).
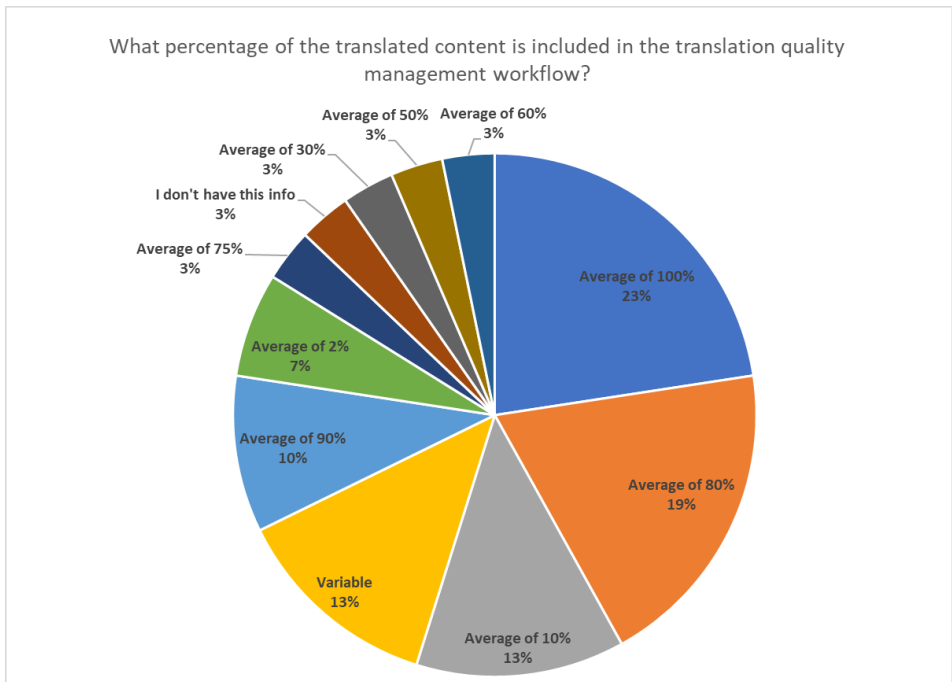
Figure 21. Percentage of content included in the translation quality management process – Translation service buyers.

While more than 50% of the respondents revealed that their companies included between 100% and 80% of the translated content in their translation quality management processes, 26% reported that their companies included between 50% and 2% of their translated content. An additional 13% mentioned that this percentage varied significantly depending on different circumstances.
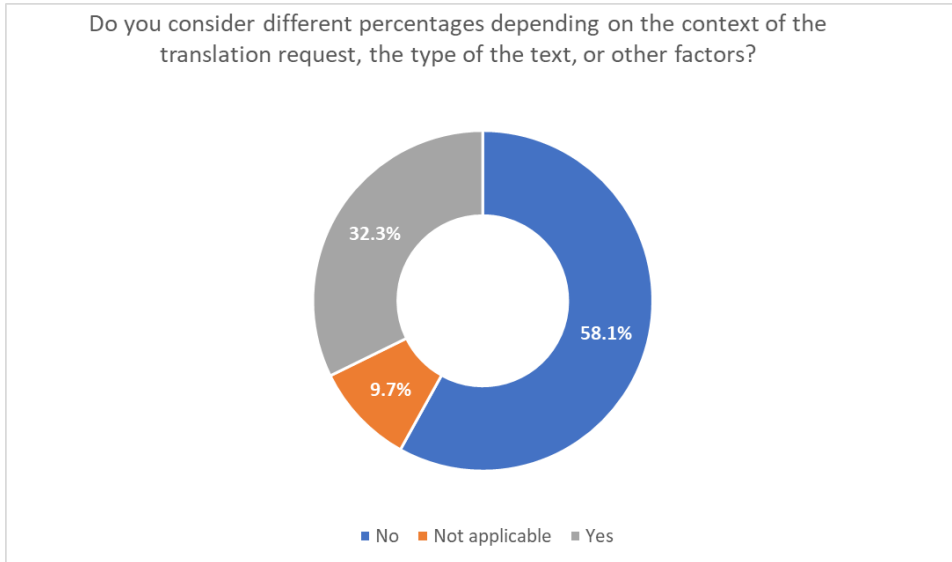
Figure 22. Percentage of companies that contemplate different percentages of translated content in their translation quality management programs – Translation service buyers.

58.1% of the participants in the research said that their companies did not contemplate different percentages of translated content in their translation quality management programs, while 32.3% did mention that their companies applied different percentages depending on different factors, such as:

- Content type (business value, priority or impact associated with it)
- Audience
- Level of trust in the quality of human resources used
- Business Key Performance Indicators (KPIs)

Translation quality evaluation processes
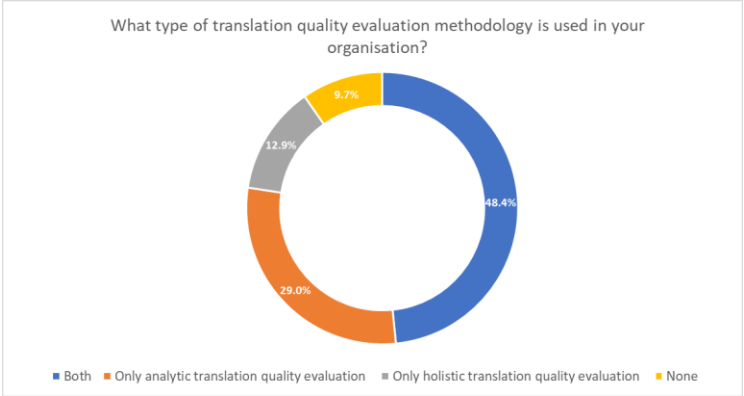
Figure 23. Type of translation quality evaluation methodology used – Translation service buyers.

Almost half of the respondents (48.4%) declared that they used both analytic and holistic translation quality evaluation methods, and for those that used only one, the analytic method was the preferred one (29.0%) over the holistic (12.9%).
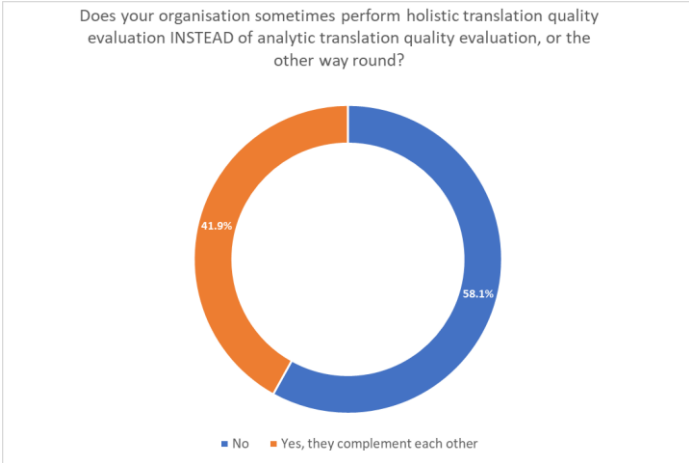
Figure 24. Use of holistic versus analytic translation quality evaluation – Translation service buyers.

However, when asked whether there were cases in which one type of evaluation was performed instead of the other, there was not a clear trend. 41.9% of the respondents said that they would only use one of the methods, while 58.1% reported that both methods complemented each other and explained that the analytic evaluation was typically used as a baseline, and that holistic was added to collect additional information such as user sentiment on quality, the influence of the quality of the translation in the user experience, and the level of compliance of the translation with local markets.

In those cases where the respondents declared that they only used one of the methods, they provided examples of situations or requirements in which they considered one of the methods worked better than the other. For instance, holistic evaluation was used by some of the participants to evaluate translations that required creative approaches and a certain style or tone of voice (such as marketing texts) as a cheaper and faster alternative to evaluate translations for markets and languages considered with a lower return of investment for the business, or to provide a high-level quality evaluation of translations with low-risk levels. On the other hand, analytic evaluation was considered by the majority of the respondents as the baseline of their translation quality management evaluation methodology.
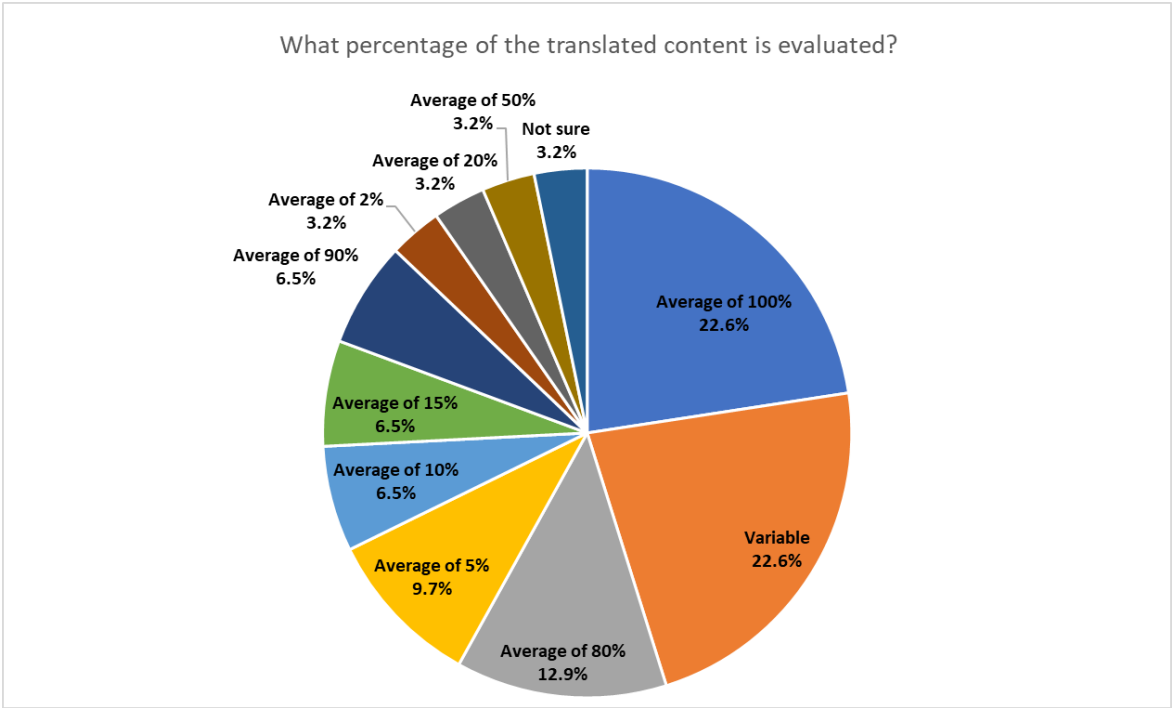

Figure 25. Percentage of content included in the translation quality evaluation process – Translation service buyers.

While 42% of the respondents revealed that their companies included between 100% and 80% of the translated content in their translation quality management processes, 45% of them reported that their companies included between 50% and 2% of their translated content. An additional 22.6% mentioned that this percentage varied significantly depending on different circumstances, such as the type of text or type of evaluation used (holistic versus analytic).
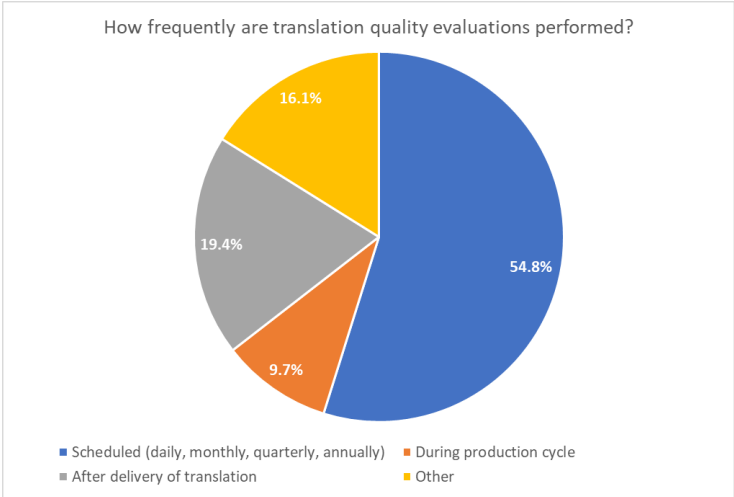
Figure 26. Frequency of translation quality evaluations – Translation service buyers.

More than half of the participants in the research (54.8%) reported that their companies performed translation quality evaluations with a scheduled frequency. In comparison, 19.4% mentioned that the evaluation would usually take place right after the translation was completed, and 9.7% said the evaluation was done while still in the production cycle. Around 16.1% of the respondents mentioned different or more complex approaches to the frequency of the evaluations
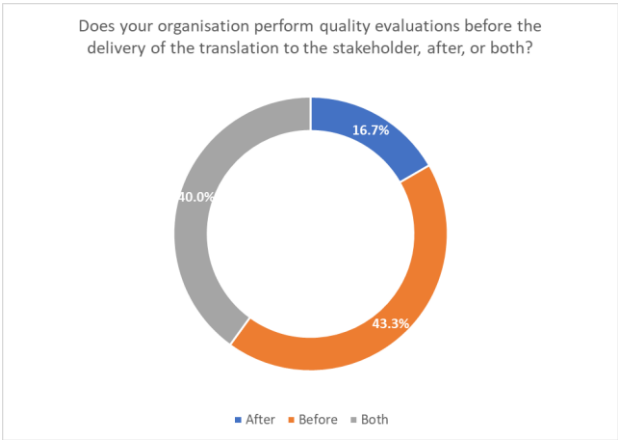
Figure 27. Timing of translation quality evaluations – Translation service buyers.

The vast majority of the participants in this research reported that their companies perform translation quality evaluations before it is delivered to the internal teams that commissioned it (43.3%) or that they follow a mixed approach in which some translations are evaluated before the delivery, and others, after (40%). When asked about the percentage of the translated content that is evaluated before the delivery versus the one that is evaluated after the delivery to the internal requestors, the majority of them explained that most of the content was evaluated before the delivery and provided an estimate of between 60% and 90% of content that was evaluated following this criterion.

Does your organisation evaluate whole texts or select a sample for evaluation?
31 responses
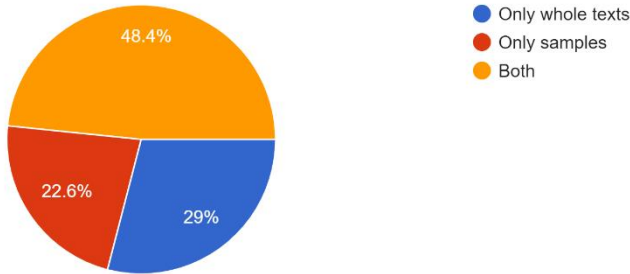
Figure 28. Sampling in translation quality evaluation – Translation service buyers.

When asked whether their organisation used sampling techniques for evaluation purposes or only whole texts, 48.4% of the respondents mentioned that they used both methods, while 29% of the participants reported that their organisations only evaluate whole texts.

How are samples selected?
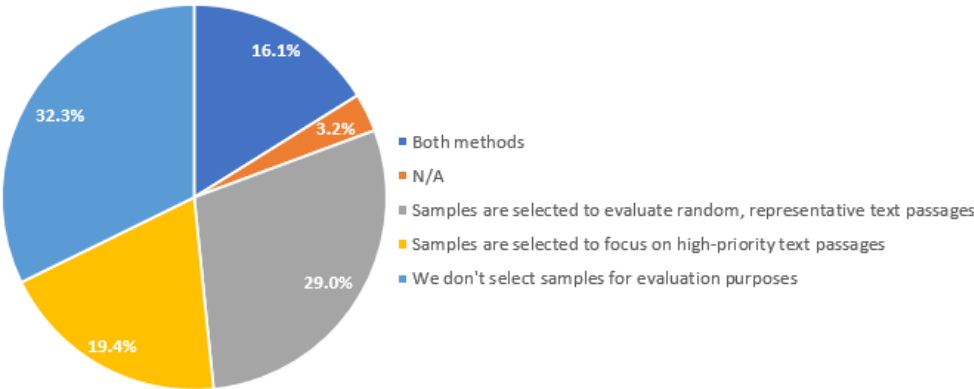
Figure 29. Sampling methods in translation quality evaluation – Translation service buyers.

94

In terms of sampling techniques, 29% of the participants reported that the samples were selected to evaluate random texts passages, while 19.4% preferred to focus on high-priority texts passages, and 19.4% said that their companies used both methods.
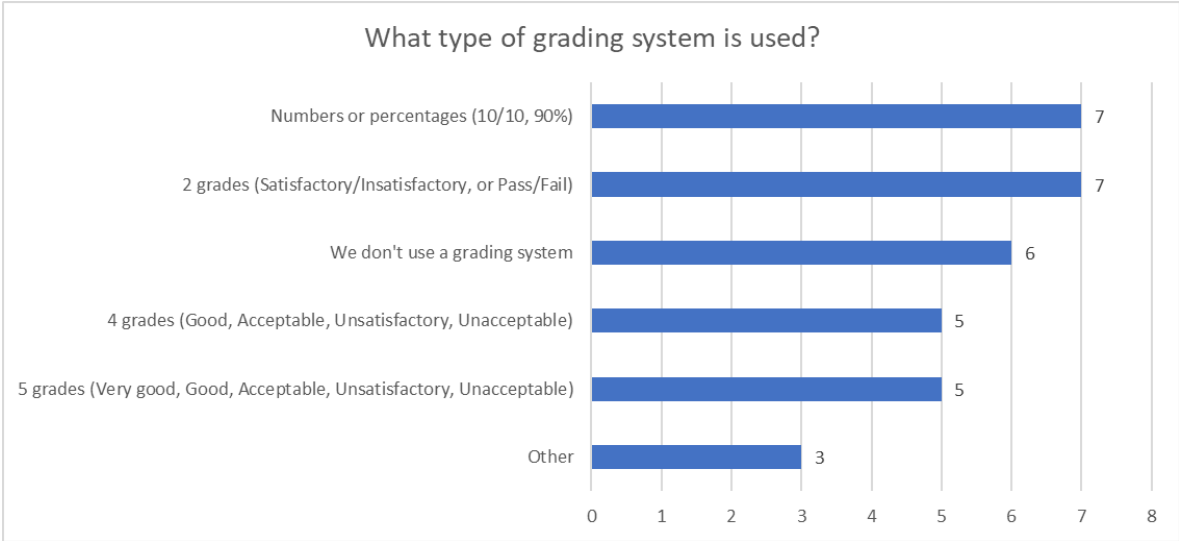
Figure 30. Grading systems used in translation quality evaluation – Translation service buyers.

The majority of the participants in the research reported that they used a dual grading system that had a number or percentage and a certain number of "grades" associated with the score obtained in the evaluation. Around 18% of the respondents said that they did not use a specific grading system.
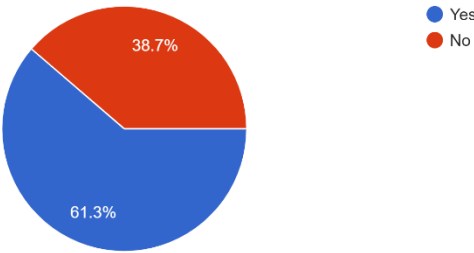
Figure 31. Uniformity of evaluation metrics in translation quality evaluation – Translation service buyers.

When asked whether the evaluation metrics used in their companies were uniformly applied to different text types, the majority of the participants in the research (61.3%) said that there were different evaluation metrics (such as error types, severity levels and severity

multipliers) applied attending to criteria such as the text type, or the specific quality requirements.

Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)?
31 responses

Figure 32. Uniformity of severity multipliers in translation quality evaluation – Translation service buyers.

However, when asked whether the severity multipliers were always the same for all error types, 54.8% of the respondents did mention that their companies used uniform severity multipliers distributions.

What error types does it include? Please, select all the applicable options.
31 responses
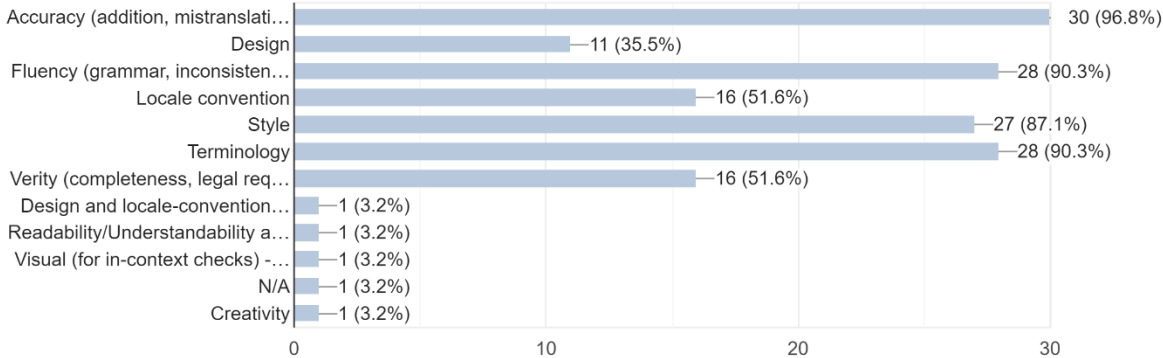
Figure 33. Error typology in translation quality evaluation – Translation service buyers.

Unsurprisingly, three main error categories listed by the participants coincided with the ones proposed by the standardised metrics presented in section 2.3.7: Accuracy (96.8%), Fluency (90.3%) and Terminology (90.3%), followed by Style (87.1%) and Verity (51.6%).

What severity levels does it include? Please, select all the applicable options:
31 responses
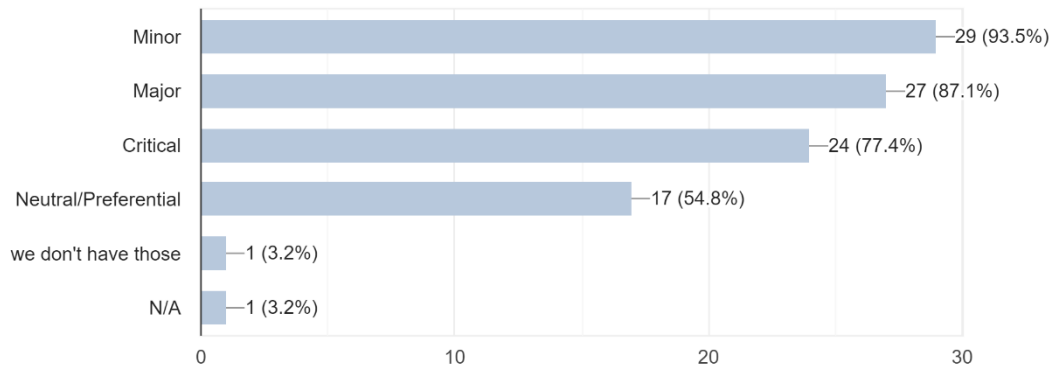
Figure 34. Error severity in translation quality evaluation – Translation service buyers.

Something similar happens with the error severity types reported, as all the participants that used some form of error severity typology confirmed that they followed the same nomenclature proposed by LISA QA Model or MQM-DQF. From the four categories proposed by these models, the two more used are "Minor" (93.5%) and "Major" (87.1%), while "Critical" and "Neutral/Preferential" seem to have a little less widespread use.

Does the evaluation include a pass/fail threshold?
31 responses
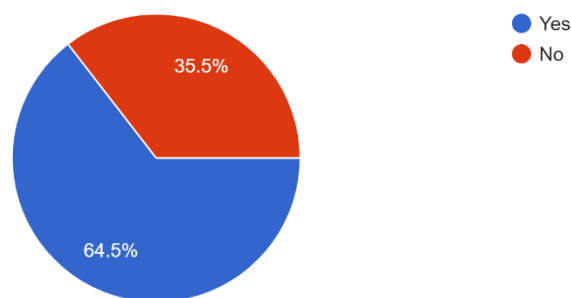
Figure 35. Use of Pass/Fail thresholds in translation quality evaluation – Translation service buyers.

Although 64.5% of the respondents confirmed that their companies used a pass/fail threshold in their quality evaluation programs, 35.5% indicated that they did not have a specific threshold that defined what could be considered a "pass" or "fail". Some of the reasons for this lack of standardised threshold were that the assessments were done in a flexible or

holistic way, that the "pass" was up to the reviewer's judgement, or that their quality evaluation system was more focused on reducing the number of critical or major errors.

Is the quality threshold uniform, or there are different thresholds for different content types or quality criteria?

31 responses
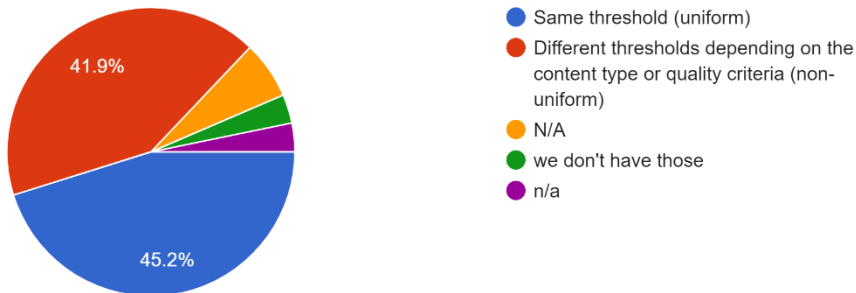
Figure 36. Types of Pass/Fail thresholds in translation quality evaluation – Translation service buyers.

There does not seem to be a clear trend in terms of uniformity of the quality threshold used. While 45.2% of the participants recognised that the same quality threshold was used for different content types or quality criteria, another 41.9% reported having different thresholds to evaluate the quality of the translated content.

How would you rate the current level of confidence in the adequacy of the quality scores of the quality evaluation methodology used?

Figure 37. Level of confidence in the quality scores used – Translation service buyers.

Almost 50% of the participants in this research reported feeling only moderately confident on the adequacy of the quality scores used, which could mean that they feel there are certain gaps or opportunities for improvement in this aspect.

Human resources and qualifications

What kind of resources are involved in the translation quality management process?
31 responses
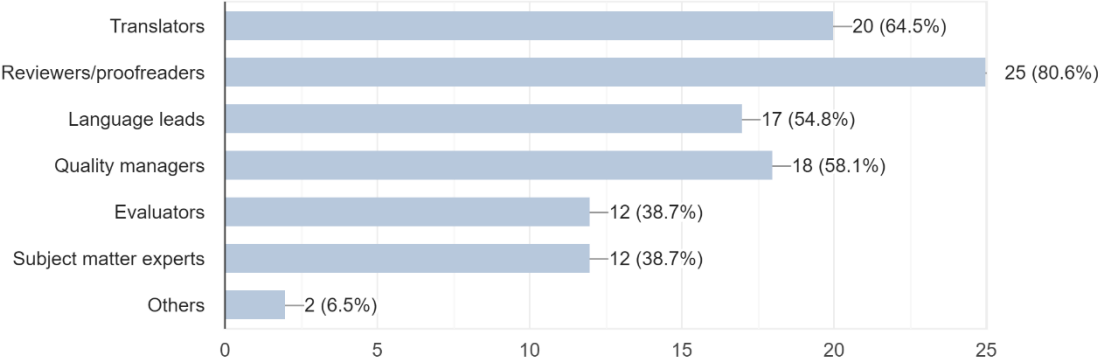
Figure 38. Human resources involved in translation quality management – Translation service buyers.

The type of human resources most commonly involved in translation quality management processes on the buyer's side are reviewers/proofreaders (according to 80.6% of the participants), followed by translators (64.5%), quality managers (58.1%) and language leads (54.8%). Only 38.7% of the respondents mentioned other types of profiles, such as evaluators or subject matter experts.
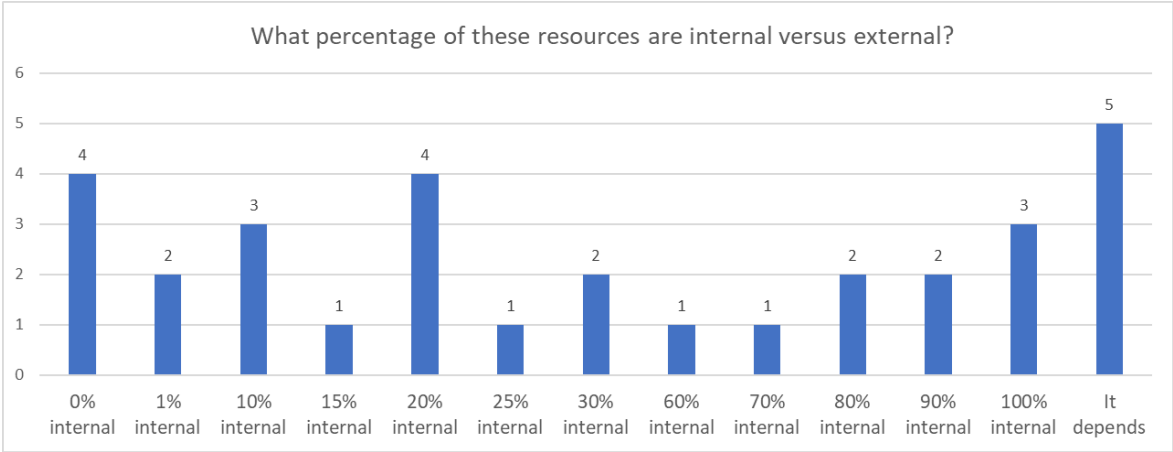
Figure 39. Percentage of internal human resources versus external resources – Translation service buyers.

When asked about the percentage of the human resources involved in translation quality tasks employed directly by the company, there was no clear trend. While 16% of the participants reported that the percentage of in-house resources involved in translation quality varied depending on factors such as the languages, the type of translated texts or whether the tasks were performed before the delivery to the internal stakeholders or after, the rest of the respondents mentioned different percentages that did not seem to follow a straightforward pattern. However, some commonalities were found as to the main reasons to outsource translation quality tasks. The main ones cited by the participants were:

- All language-related work (including quality management) is performed by the TSPs employed by the translation service buyers.
- Lower costs and better scalability.
- Lack of internal resources dedicated to translation quality management.
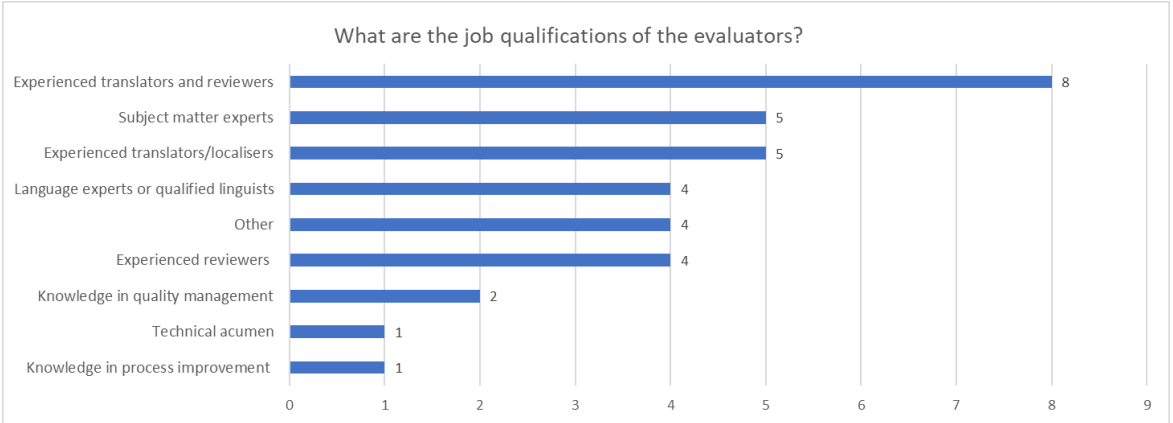
Figure 40. Skills and qualifications of evaluators – Translation service buyers.

24% of the participants in the study reported that evaluators were required to have experience both in translation and review, while 15% only required experience in translation, and 12% mentioned that evaluators did not need to have experience in translation as long as they were experienced reviewers. Another qualification mentioned by 15% of the respondents was experience as subject matter experts of the industry or the type of content that would be evaluated.
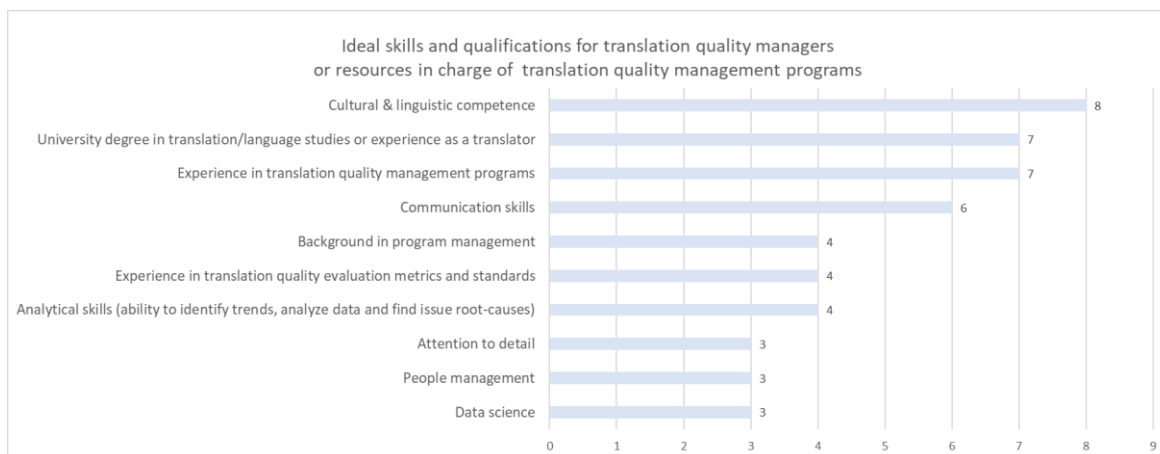
Figure 41. Skills and qualifications of translation quality managers – Translation service buyers.

The last question of this part of the study was what the ideal qualifications and skills of translation quality managers or other human resources in charge of translation quality programs would be. This question had an open-ended format, which allowed the respondents to share their thoughts freely. The replies were coded and grouped into main themes, resulting in the 10 top qualifications and skills shown in Figure 41.

4.2.3.4 Tools and technologies
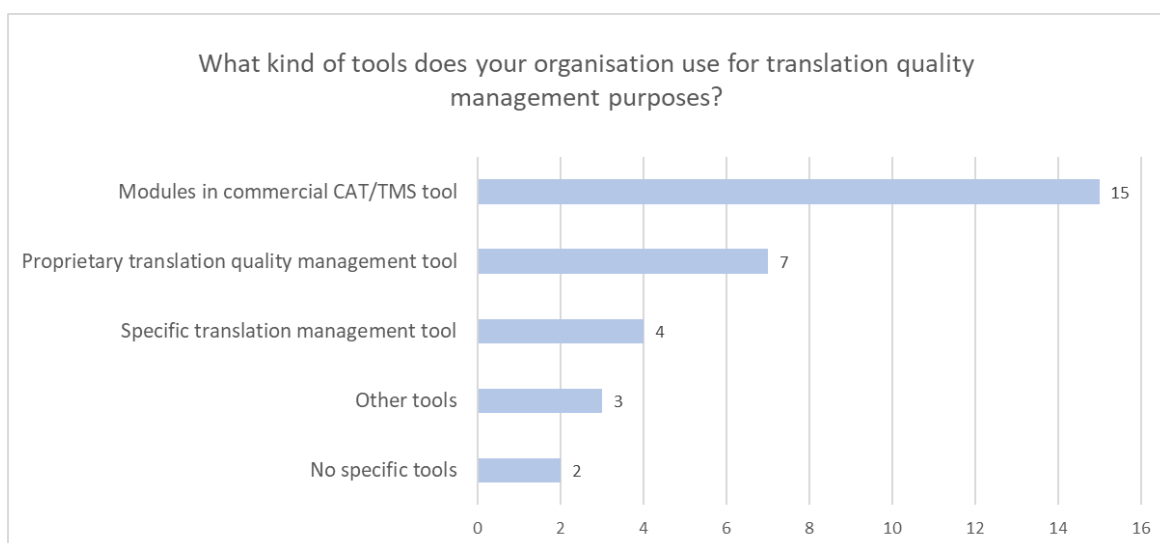
Figure 42. Tools used for translation quality management purposes – Translation service buyers.

48% of the respondents reported that their companies used the modules and features available in the commercial CAT/TMS tools for translation quality management purposes. The most popular ones were Memsource, Smartling and XTM. 23% of the participants declared that their companies had created proprietary translation quality management tools to suit their needs, and 13% mentioned that their companies were using a standalone

translation quality management tool (ContentQuo). The rest of the participants mentioned that they were using only Excel for translation quality management purposes (10%) or that they did not have any specific translation quality management tool.
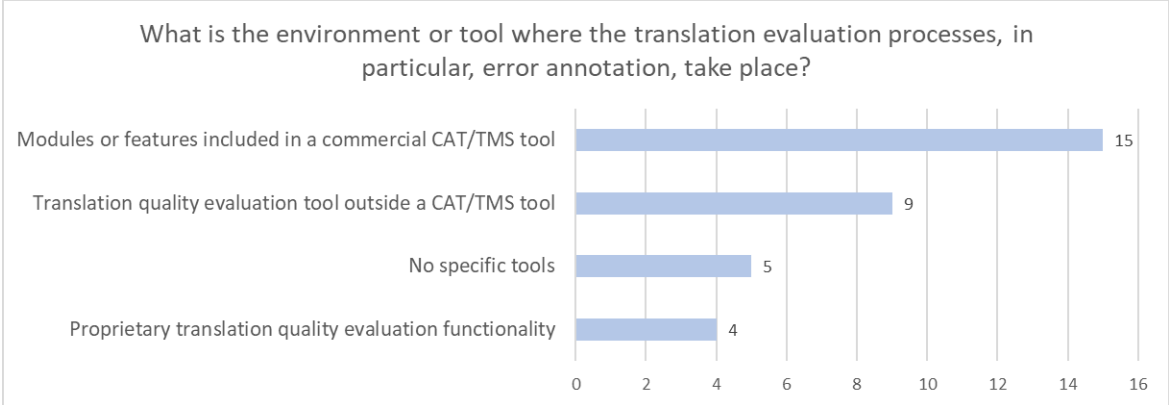
Figure 43. Tools used for translation quality evaluation (including error annotation) purposes – Translation service buyers.

Similarly to what was reported in the previous question, 45% of the respondents said that their companies used the modules and features available in the above-mentioned commercial CAT/TMS tools for error annotation. The second most popular choice was the use of tools outside the CAT/TMS environments, as 27% of the participants declared that their companies used tools such as ContentQuo, Excel or Smartsheet for error annotation. The rest of the participants reported that they did not have any specific tool for annotation purposes (15%) or that their companies had developed proprietary solutions for quality evaluation and annotation (12%).

4.2.4 Current challenges in translation quality management and evaluation

In this part of the survey, the participants were asked to rate, on a scale of 1 to 5, their level of satisfaction with the current translation management program in their organisations according to seven parameters discussed in the previous parts of the survey.
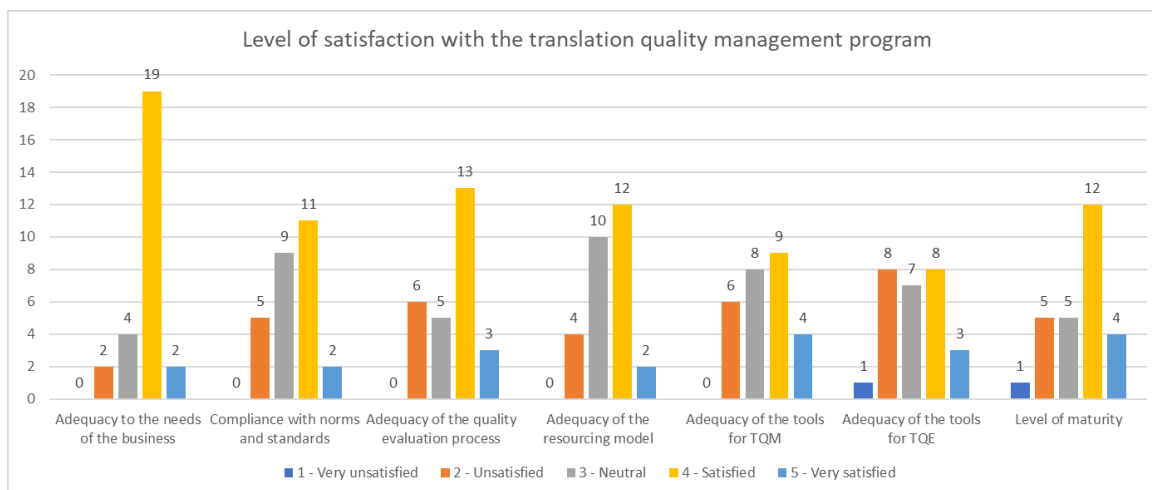
Figure 44. Level of satisfaction with the translation quality management program in place – Translation service buyers.

The three aspects with a higher level of satisfaction reported by the respondents were the adequacy to the needs of the business and stakeholders, the level of maturity of the translation quality management program, and the adequacy of the quality evaluation process. On the other hand, the two categories with a lower level of satisfaction were the adequacy of the tools used for translation quality evaluation and the adequacy of the tools used for translation quality management.

Main pain points in translation quality management and evaluation	Nr of replies
Insourced TQM: Lack of centralised data or dashboards with visibility on trends	4
Insourced TQM: Lack of clearly defined TQM process	1
Insourced TQM: Lack of dedicated human resources	5
Insourced TQM: Lack of support or interest from internal stakeholders	1
Insourced TQM: Lack of technical resources to implement advanced tools	1
Insourced TQM: Lack of training and competences	1
Insourced TQM: Lack ot automatisation	7
Insourced TQM: Level of maturity of the translation/localisation team	2
Insourced TQM: Missing requirements or guidelines from internal stakeholder	1
Outsourced TQM: Lack of visibility on how the QM process is performed	2
Outsourced TQM: Maintaining a stable quality (especially freelancers)	1
Quality improvement: Communication loops and feedback channels	3
Quality improvement: Time constrains	2
Tools: Ability to review content easily in context (app and web content)	2
Tools: More integration required into TMS (TQM and TQE)	2
Tools: No tools allow a macro LQA view	2
Tools: Use of proprietary tools that cannot easily be updated or improved	1
TQE: Alignment and calibration among different reviewers	1
TQE: High costs, time-consuming processes and rigidity of the evaluation model	1
TQE: Lack of clear, scalable and customisable evaluation metrics and thresholds	4
TQE: Use of very unstructured holistic models without any grading or sampling	3

Figure 45. Main pain points in translation quality management and evaluation – Translation service buyers.

In the following question, the participants were asked to list the main translation quality management and evaluation pain points their organisation faced. This question has an open-ended format, which allowed the respondents to share their insights. The replies were coded and grouped into five main topics: Challenges in insourced translation quality management programs, challenges in outsourced translation quality management programs, challenges related to quality improvement, challenges related to the tools used and specific challenges in translation quality evaluation. The main findings are summarised in Figure 45 above.

4.2.5 Solutions

The candidates were asked to share what were the solutions or workarounds that their organisation had applied to try to solve the pain points mentioned in the previous question of the survey. This question had an open-ended format too, and the replies were coded and added to a table to show any correlations found between the answers to both questions.

Main pain points in translation quality management and evaluation	Applied solutions and workarounds
Insourced TQM: Lack of centralised data or dashboards with visibility on trends	Automatise as much as possible (macros, proprietary tools) Weekly meetings to review data Moving to a TMS that has dashboard
Insourced TQM: Lack of clearly defined TQM process	Team is working on finding solutions to these gaps
Insourced TQM: Lack of dedicated human resources	Translators help with evaluations Content audits with different type of contents and quality levels per content
Insourced TQM: Lack of support or interest from internal stakeholders	
Insourced TQM: Lack of technical resources to implement advanced tools	Internal automatic checks with TMS
Insourced TQM: Lack of training and competences	Learning from our translation vendors
Insourced TQM: Lack ot automatisation	Migrating to a new TMS to gain flexibility Regular updates of excel dashboards to make sure the data is available at any given time
Insourced TQM: Level of maturity of the translation/localisation team	Translators help with evaluations Introducing DQF model
Insourced TQM: Missing requirements or guidelines from internal stakeholder	
Outsourced TQM: Lack of visibility on how the QM process is performed	Quality management policy with KPIs to assess service received
Outsourced TQM: Maintaining a stable quality (especially freelancers)	Metrics, monitoring, autoQA
Quality improvement: Communication loops and feedback channels	Dedicated email and ticketing system Slack discussions, informally, between linguists peer to peer
Quality improvement: Time constrains	Automatise as much as possible (macros, proprietary tools)
Tools: Ability to review content easily in context (app and web content)	Creating manual workaround with the third-party reviewer vendor's help Working with TQE tool that can display same content that is sent to translators
Tools: More integration required into TMS (TQM and TQE)	Localisation engineering providing interim solutions with CAT tool exports Planning to integrate quality reporting into TMS
Tools: No tools allow a macro LQA view	
Tools: Use of proprietary tools that cannot easily be updated or improved	Using scorecards to issue quality reports and assess quality in a more granular way
TQE: Alignment and calibration among different reviewers	SLAs with reviewers, survey based model rather than out of the box Calibration sessions with all the senior localisation specialists
TQE: High costs, time-consuming processes and rigidity of the evaluation model	Crowd-based testing by non-linguists, linguistic testing on a bigger scale only with indicator
TQE: Lack of clear, scalable and customisable evaluation metrics and thresholds	Introducing DQF model
TQE: Use of very unstructured holistic models without any grading or sampling	Currently working on a process improvement plan

Figure 46. Pain points and solutions applied in translation quality management and evaluation – Translation service buyers (see Appendix E).

4.2.6 The future of translation quality management

In this section of the study, candidates had the opportunity to mention what were the main challenges that they believed would impact their current translation quality management

processes. This question had an open-ended format, and the replies were coded to show the most recurrent topics mentioned by the respondents.

Future challenges in translation quality management
Increase capacity and scalability to cover bigger volumes and more languages
New types of TQE metrics: user centric, holistic, post-editing effort, MT evaluation
Automation and integration of tools for TQM and TQE purposes
Setting or reshaping the whole quality management strategy
New text types in the TQM workflows: Transcreation, creative translations, MT evaluation, post-editing
Translation quality management business analytics

Figure 47. Future challenges in translation quality management and evaluation – Translation service buyers.
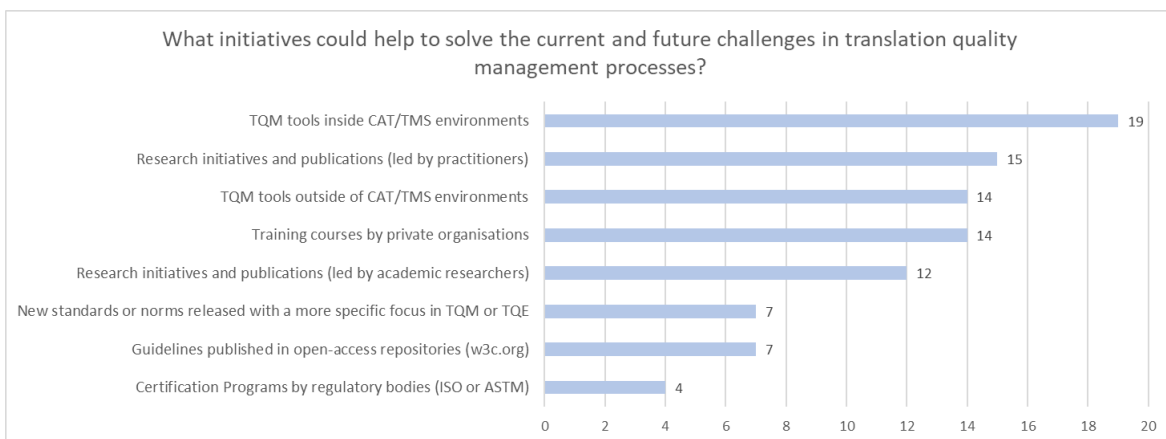
Figure 48. Useful initiatives in translation quality management and evaluation – Translation service buyers.

The last closed-ended question of the survey was aimed at discovering what kind of initiatives could help the advancement of the translation quality management programs in professional environments.

- In terms of technological solutions, most of the respondents preferred translation quality management solutions that would work inside their existing CAT/TMS environments instead of standalone ones.

- Participants also favoured research initiatives led by practitioners and industry-related organisations over research initiatives led by universities.

- The fourth most voted initiative was related to training courses on translation quality management or translation quality evaluation.

- Unsurprisingly, especially if we take into consideration some of the results already analysed in previous parts of this survey, respondents from translation service buyers

did not consider that further certifications from regulatory bodies such as ISO or ASTM were particularly interesting or valuable for them.

4.2.7 General comments and feedback

Out of the 31 participants on the buyer's side, only five chose to provide some additional feedback. From these five comments, 3 were positive remarks about the design of the research in general, and 2 contained constructive criticism about the research, as well as some final thoughts the participants wished to share:

> "The survey was quite centered on MQM and left little space or different scenarios e.g company-own evaluation scenarios or company-own score cards. The range of answers was therefore too limited and reflected to a great extend the scenarios at TSPs. Therefore it looked quite academic in the first part. However, having obtained my PhD […] on translation quality management, I know that academic research has often very little to do with the real life. So I was surprised to find the questions in the latter part of the survey pretty close to everyday translation business's reality. That was really a very pleasant surprise." [Questionnaire respondent]

> "I believe it is important to think about the future, the long term, and try to see and fix the errors and flaws of any translation quality program. I also think it is important to reinforce the human side of the quality management program, translations are done by humans, and the industry should always follow ethical practices." [Questionnaire respondent]

4.3 Translation service providers

Out of the 68 participants in this research, 20 (29.4% of the total) worked for translation service providers of different sizes. In the following sections, we will show and discuss the data compiled from this target group in each of the seven parts of the surveys.

4.3.1 Demographics
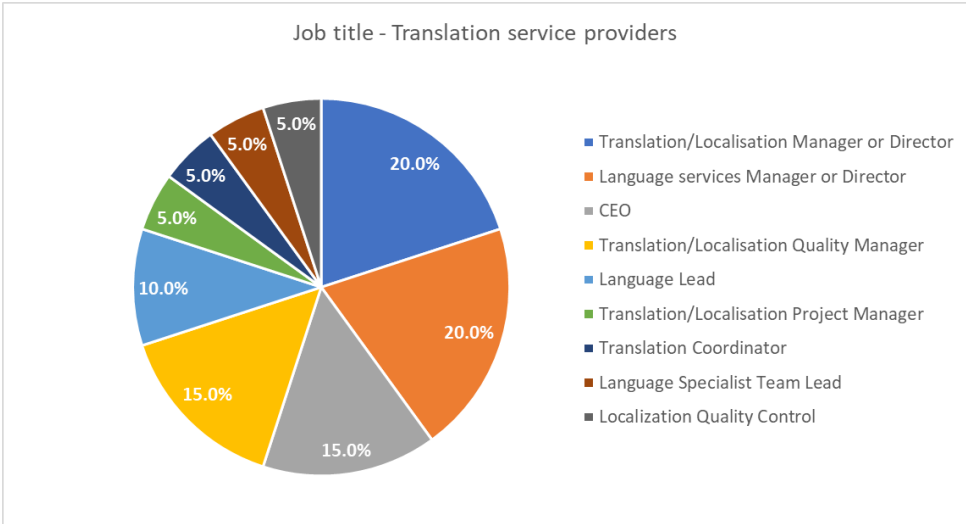
Figure 49. Job titles – Translation service providers.

The demographics of the translation service providers representatives were similar to the demographics of the overall list of participants and the demographics of the respondents of the translation service buyers. However, the main difference was that the representation was much more balanced and that around 20% of the respondents were also the CEOs of the translation service provider.
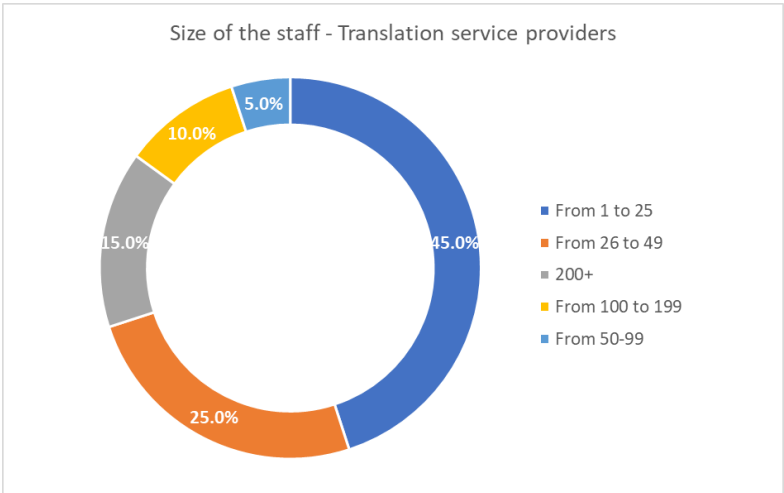
Figure 50. Size of staff – Translation service providers.

107

Unlike the demographics of the buyers' respondents, the size of the staff of the translation service providers was much more varied and predominantly smaller, as 45% of the participants said that their companies had 25 employees or less. 30% of the companies had a medium size (from 26 to 100 employees) and 25% a big size (from 100 to more than 200 employees).
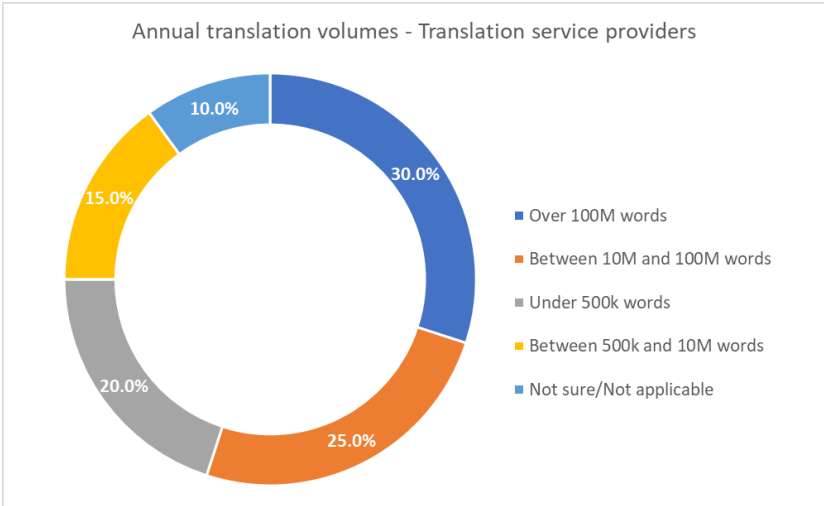
Figure 51. Annual translation volumes – Translation service providers.

This distribution of small, medium, and big translation providers can also be seen in annual translation volumes shared by the participants, as 30% of them reported an annual translation volume of over 100M words (which would correspond to the highest tier), while 40% said their companies translated between 500k and 100M words per year (medium size tier) and 20% that their annual volumes were below 500k words translated per year (small size tier).
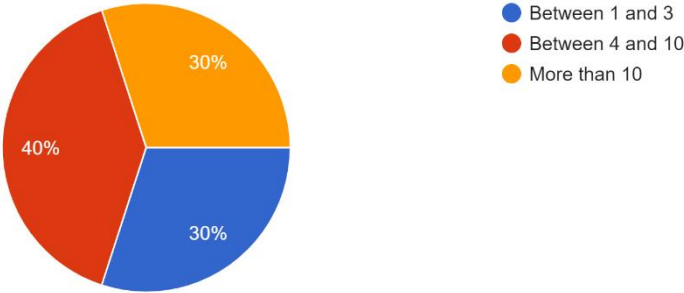
Figure 52. Languages translated – Translation service providers.

This balance among small, medium, and big companies can also be seen in the number of languages translated, as 30% declared their companies provided translation services for three languages or less, 30% that they worked with more than ten languages, and 40% that offered translation services in 4 to 10 languages.

4.3.2 Translation processes and analytics

How would you score the actual level of maturity of your organisation in terms of translation analytics?
20 responses
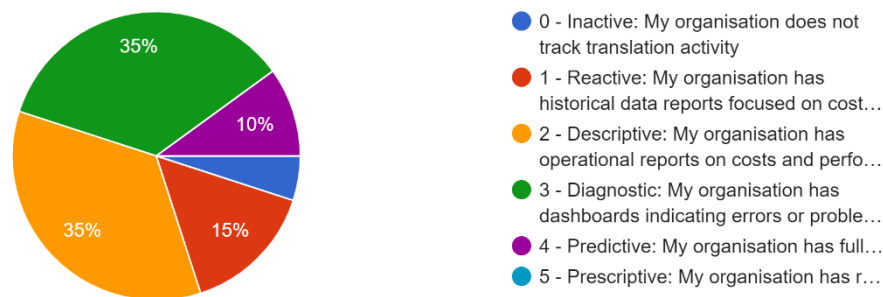
Figure 53. Maturity level in terms of translation analytics – Translation service providers.

When asked to choose the level of maturity of their organisation in terms of translation analytics, none of the participants considered that their companies were at the highest tier (Prescriptive), and 70% of them reported their companies were either on level 2 (35%) or 3 (35%). Finally, 10% of the respondents considered that their level of maturity in translation analytics was the highest of the scale.

4.3.3 Translation quality management

The third part of the survey focused on the different elements of the translation quality management processes implemented by the translation service providers, and it covered the same four topics as the version that was prepared for the translation service buyers: overall level of maturity in translation quality management processes, knowledge and use of norms and standards, documented translation quality management and evaluation processes, and tools and technologies.

4.3.3.1 Overall level of maturity in translation quality management processes (self-assessment)

How would you score the actual level of maturity of your organisation in terms of translation and localisation quality management processes?
20 responses

- 1 - Manual: My organisation uses manual and/or legacy processes
- 2 - Automated: My organisation has started to integrate some automation i…
- 3 - Agile: My organisation has automated and scalable quality mana…
- 4 - Centralised: My organisation practices strategic and centralised tra…
- 5 - Expert: My organisation benefits from translation as a revenue center and is…
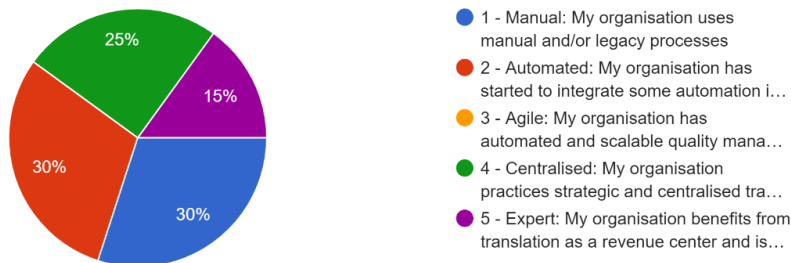
Figure 54. Maturity level in terms of translation quality management processes – Translation service providers.

In this question, the participants were asked to reflect on the level of maturity of the quality management processes implemented in the companies using a similar scale to the one used in the previous question. Interestingly enough, even though 70% of the respondents had reported that their companies were in the intermediate level in terms of translation analytics, in this question, the replies gravitated more towards the lower levels, as 30% of the participants considered that their companies were still using manual or legacy processes (level 1). Another 30% declared that their companies had certain automated processes for translation quality management (level 2). On the other hand, 15% of the respondents claimed their companies were already at the highest maturity level (level 5).

4.3.3.2 Knowledge and use of norms and standards

Which ones of the following norms and standards related to translation or quality management has your organisation incorporated in their translation quality management program?
20 responses
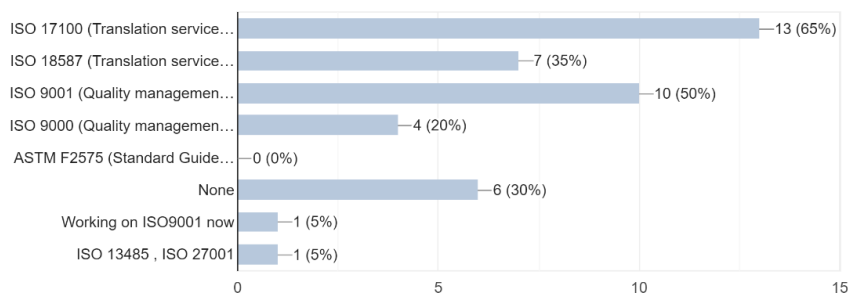
Figure 55. Knowledge and use of norms and standards – Translation service providers.

It should be noted that the replies to this question from the providers' side were considerably different from the ones provided by the buyer's side, as 70% of the respondents reported using or following one or more translation or quality-related standards. The standards more frequently mentioned were ISO 17100 (65%), ISO 9001 (50%) and ISO 18587 (35%).

Those respondents that replied "None" were shown a follow-up question to allow them to select one or more reasons for this decision. The reason more commonly chosen was that their companies were not aware of the existence of those standards.
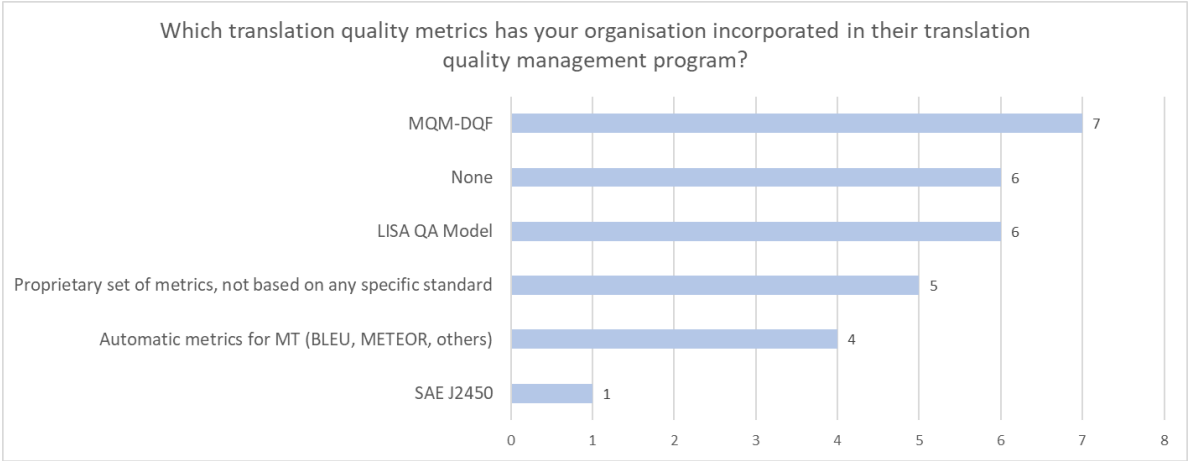
Figure 56. Use of translation metrics – Translation service providers.

24% of the participants mentioned MQM-DQF as one of the main translation metrics used, and another 21% said they were using the LISA QA Model. Conversely, 21% of the participants reported that they were not using any metrics at all, and an additional 17% were using a proprietary set of metrics that were not based on any specific standards.

Respondents that selected "None" were directed to a follow-up question asking them to specify the reasons. Some of the reasons more commonly cited were that the metrics were not too complex or difficult to apply, that their companies were not aware of the metrics, or that the standardised metrics did not meet the specific needs of their organisations.

4.3.3.3 Documented translation quality management and evaluation processes

This subsection of the survey delved into the specific translation quality management and evaluation processes followed by the translation service providers and, like the version prepared for the translation service buyers, it covered three different topics: overall translation quality management processes, specific translation quality evaluation processes and processes related to human resources, training, and qualifications.

Overall translation quality management processes
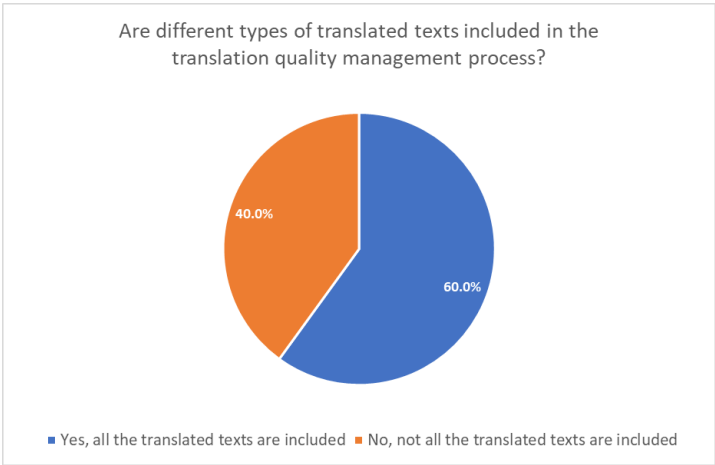
Figure 57. Content included in the translation quality management process– Translation service providers.

60% of the participants said that all the translated texts produced by their companies were included in their translation quality management workflows, while the other 40% revealed that there were exceptions. Two of the main reasons mentioned by this 40% were the fact that some of their clients do not require (and therefore, they would not pay for) translation quality assurance services and that raw machine translation was generally excluded too.
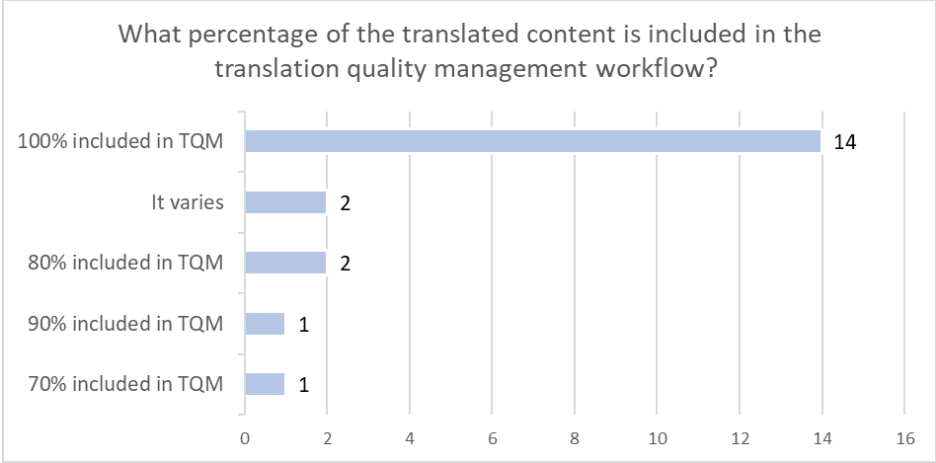
Figure 58. Percentage of content included in the translation quality management process – Translation service providers.

For those projects that were included in the translation quality management processes of the translation service providers, 70% of the respondents confirmed that all the content translated was included in the workflow, and 15% reported having lower percentages (between 90% and 70% of the translated content).
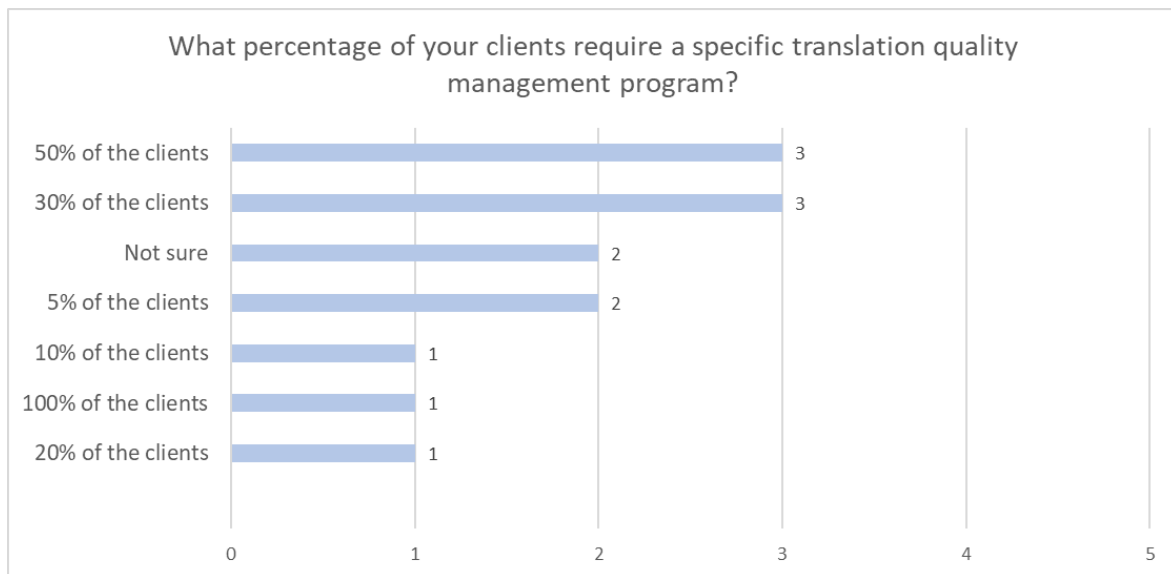
Figure 59. Client's requirements of specific quality management programs – Translation service providers.

The results varied greatly when asked what percentage of their clients required a specific translation quality management program. This could be partly explained by the fact that, as several respondents revealed, bigger translation service providers that outsource some of the translation services to other providers tend to have specific translation quality programs in place that need to be implemented by their vendors. In contrast, direct clients tend to rely on the translation quality management programs of their vendors.

Figure 60. Client's requirements and expectations – Translation service providers.

60% of the participants on the providers' side considered that the quality requirements and expectations from their clients helped their companies build their quality management programs, while only 10% thought that these requirements hindered or limited them.

Translation quality evaluation processes
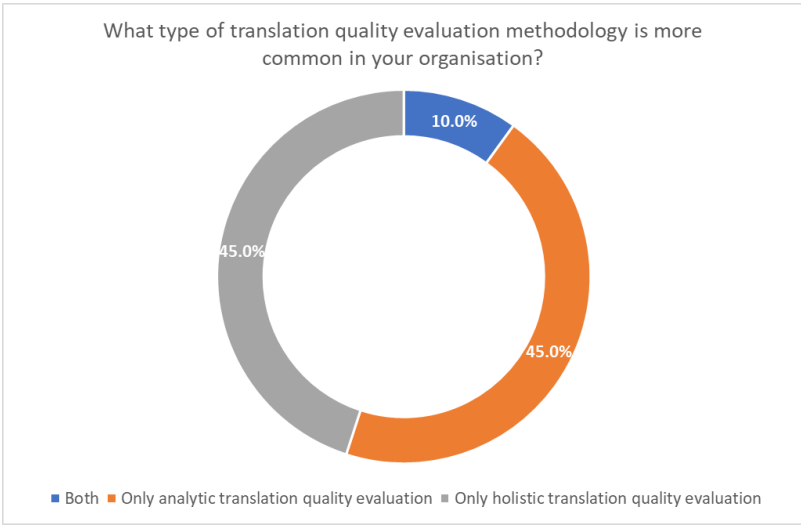
Figure 61. Type of translation quality evaluation methodology used – Translation service providers.

In contrast to the data compiled from the translation service buyers, only 10% of the respondents from the provider's side reported using both holistic and analytic evaluation methods. Additionally, each method was chosen by half of the remaining participants.

It is interesting to note that, when asked whether there were cases in which one type of evaluation was performed instead of the other, most of the participants revealed that these cases were usually triggered by specific asks from their clients, for example, target readability proofreading services or overall estimations of the quality of a translation that is already inserted in a product (also called in-context reviews).
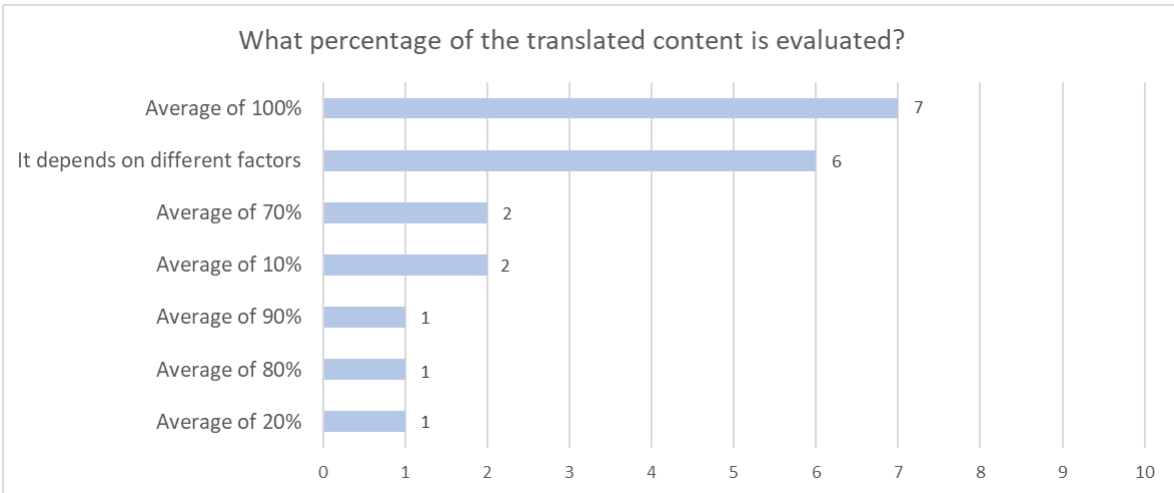
Figure 62. Percentage of content included in the translation quality evaluation process – Translation service providers.

While 35% of the respondents reported that their companies evaluated 100% of the translates content, 30% of the participants said that this percentage could vary significantly depending on different factors, for example, the reliability of the resources that performed the translation, the quality risks identified during the quality planning phase, or whether the translation had been done with machine translation engines or not. Some of the respondents also explained that in their translation quality management programs, the evaluation step was performed separately from the review step and that, even though all the content was reviewed in compliance with the ISO norms, not all the content was evaluated.
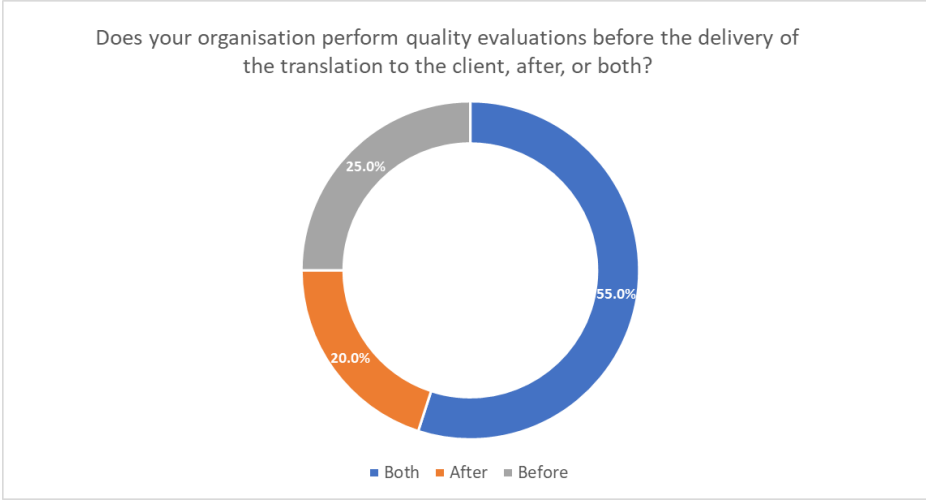
Figure 63. Timing of translation quality evaluations – Translation service providers.

In contrast with the information obtained from the participants representing the buyer's side, the majority of the respondents on the provider side (55.0%) revealed that they performed evaluations before and after the delivery of the translation to their clients. Some of the reasons for this flexibility in their approach were that the post-delivery evaluations were part of their internal auditing system, time constraints and, yet again, compliance with the requirements of their clients.

Does your organisation evaluate whole texts or select a sample for evaluation?
20 responses

- Only whole texts
- Only samples
- Both
- In the case of reviews after delivery, usually a sample is selected.

65%
30%
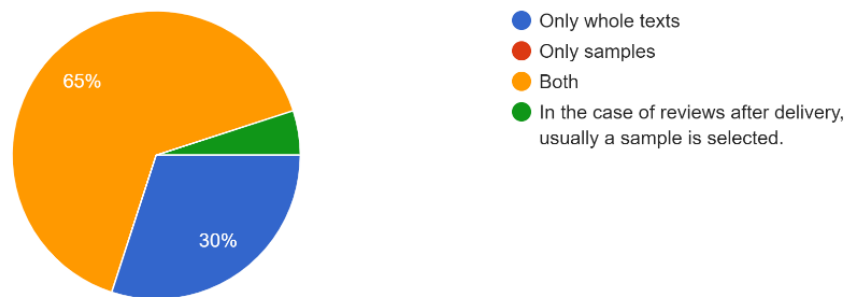
Figure 64. Sampling in translation quality evaluation – Translation service providers.

When asked whether their companies used sampling techniques to perform translation quality evaluations, 65% of the participants said that their programs included both sampling and whole texts, and 30% said that evaluation programs only evaluate whole texts.

How are the samples selected?

- Both methods
- Not sure
- Samples are selected to evaluate random, representative text passages
- Samples are selected to focus on high-priority text passages
- We don't select samples for evaluation purposes

30.0%
25.0%
5.0%
5.0%
35.0%
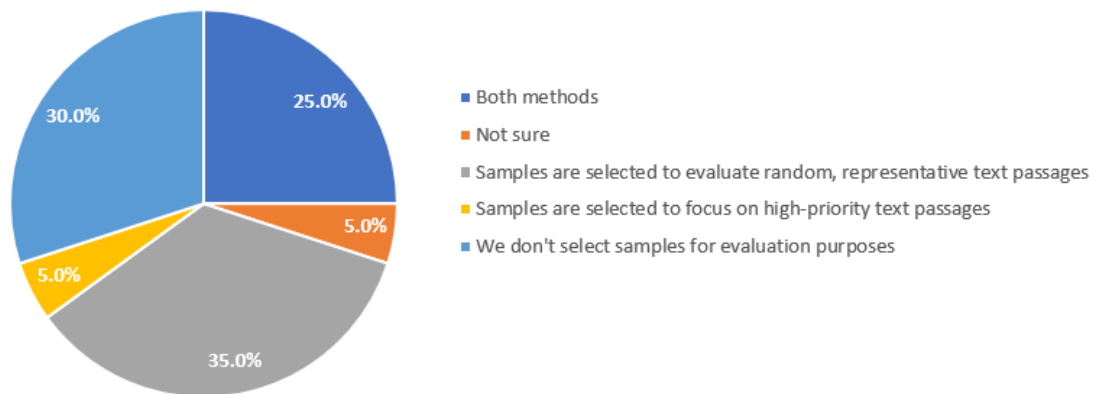
Figure 65. Sampling methods in translation quality evaluation – Translation service providers.

According to the respondents on the buyer's side, the sampling technique most used by their organisations was random sampling (35.0%), although another 25.0% mentioned their companies preferred using a mixed sampling approach.
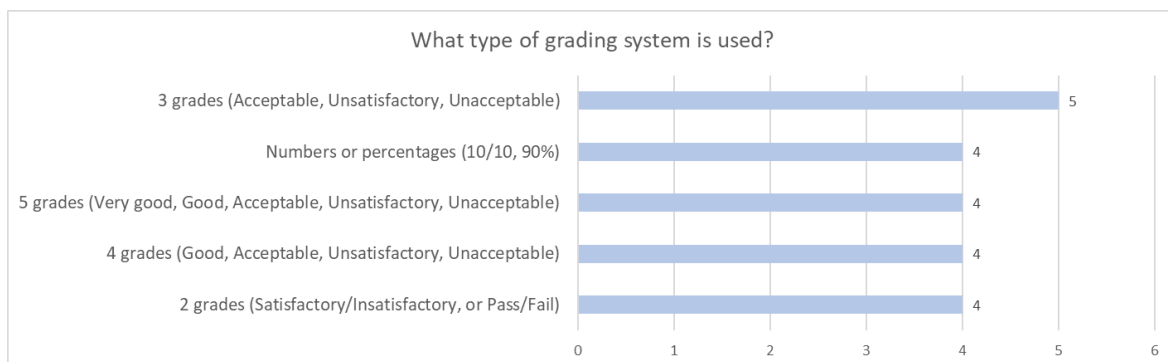
Figure 66. Grading systems used in translation quality evaluation – Translation service providers.

All the participants reported using a grading system in their evaluation programs, and several of them mentioned using more than one, depending on their client's requirements. Even though it seems the 3-grades system was slightly more popular, there is not enough data to conclude whether there is a clear trend among translation service providers in this regard. Some respondents also provided additional information about the naming used in their grading systems, for example:

- Does not meet expectations, Meets expectations, Exceeds expectations (3 grades)
- Good, Needs Improvement, Poor (3 grades)
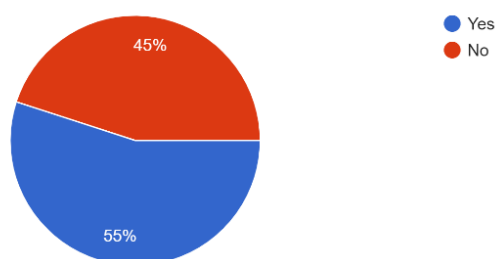
Figure 67. Uniformity of evaluation metrics in translation quality evaluation – Translation service providers.

Similarly to the information provided by the representatives of the client's side, most of the service translation provider's (55%) confirmed that they applied different evaluation metrics depending on the type of text (non-uniform evaluation metrics).

117

Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)?
20 responses

- ● The severity multipliers are the same for all error types
- ● The severity multipliers are NOT the same for all error types
- ● NA
- ● Not applicable
- ● It depends on client needs
- ● We have 4 categories measured with different weight -> composite score

Figure 68. Uniformity of severity multipliers in translation quality evaluation – Translation service providers.

However, the main trend was not clear to identify when it comes to the uniformity of the severity multipliers depending on the text type, as 40% of the participants reported using uniform severity multiplier distributions, and the other 40% said their companies applied non-uniform severity multiplier distributions.

What error types does it include? Please, select all the applicable options.
20 responses

Error type	Count (%)
Accuracy (addition, mistranslati…)	19 (95%)
Design	8 (40%)
Fluency (grammar, inconsisten…)	19 (95%)
Locale convention	12 (60%)
Style	18 (90%)
Terminology	19 (95%)
Verity (completeness, legal req…)	10 (50%)
Type of mistake is not specified	1 (5%)
Consistency	1 (5%)
Tag management	1 (5%)
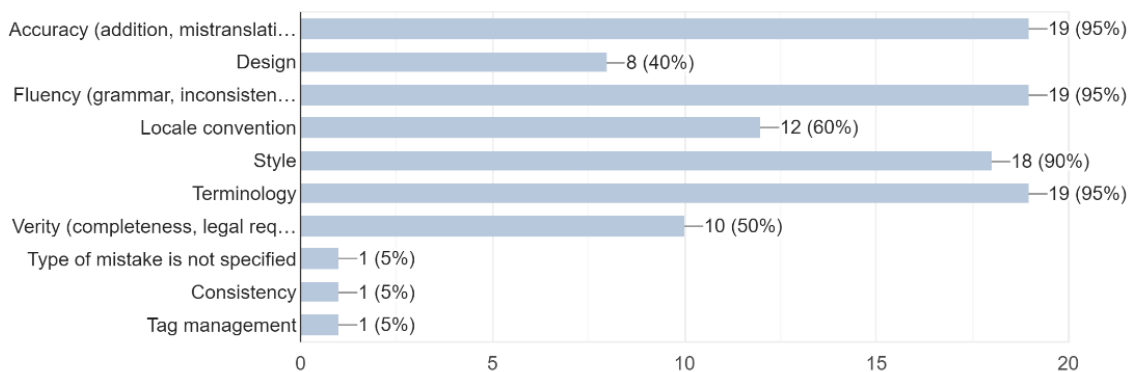
Figure 69. Error typology in translation quality evaluation – Translation service providers.

The influence of standardised quality metrics summarised in section 2.3.7 can be seen in the results compiled from the representatives of the buyer's side, given the fact that the three main error categories listed by the participants are indeed, Accuracy (95%), Fluency (95%), and Terminology (95%), followed by Style (90%), and, to a lesser extent, Locale Convention (60%) and Verity (50%).

What severity levels does it include? Please, select all the applicable options:
20 responses
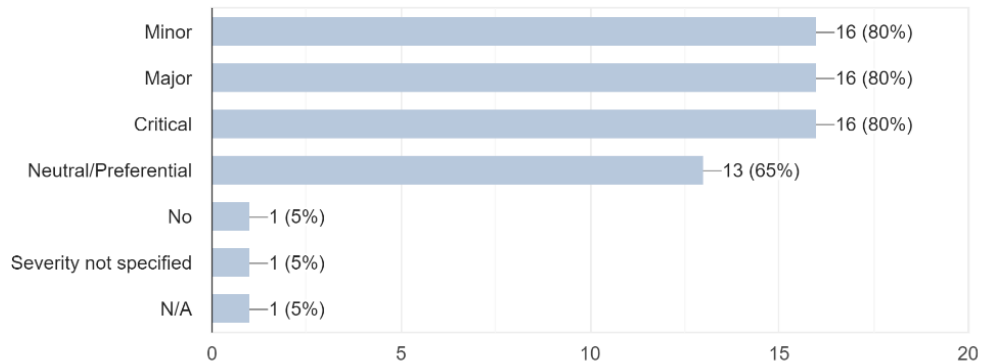
Figure 70. Error severity in translation quality evaluation – Translation service providers.

A similar phenomenon can be observed in the error severity types selected by the participants, as they follow the same nomenclature and the three categories proposed by LISA QA Model or MQM-DQF (Minor, Major and Critical).

Does the evaluation include a pass/fail threshold?
20 responses
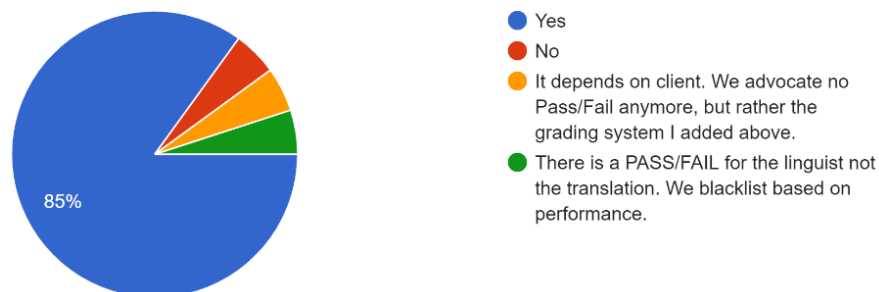
Figure 71. Use of Pass/Fail thresholds in translation quality evaluation – Translation service providers.

The majority of the respondents from the translation service providers (85%) confirmed their translation evaluation program did include a Pass/Fail threshold, and a small percentage (2%) reported that this threshold depended on the client or that it was not applied to the translation itself, but to the linguist that had procured the translation in order to measure their performance.

119
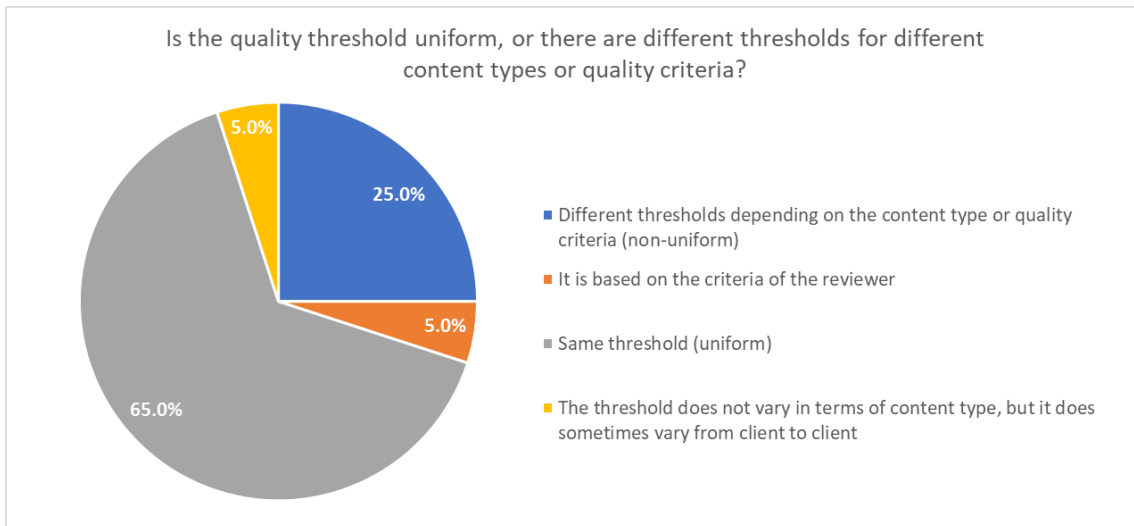
Figure 72. Types of Pass/Fail thresholds in translation quality evaluation – Translation service providers.

The majority of the respondents (65%) also confirmed their companies applied a uniform quality threshold used for all content types and clients, while 37% of the respondents reported using different thresholds depending on the content type and the reviewer's criteria or the client's requirements.

Figure 73. Level of confidence in the quality scores used – Translation service providers.

60% of the representatives of the translation service providers declared feeling very confident on the adequacy of the quality evaluation methodology used, and an additional 25% felt moderately confident.

Human resources and qualifications

What kind of resources are involved in the translation quality management process?
20 responses
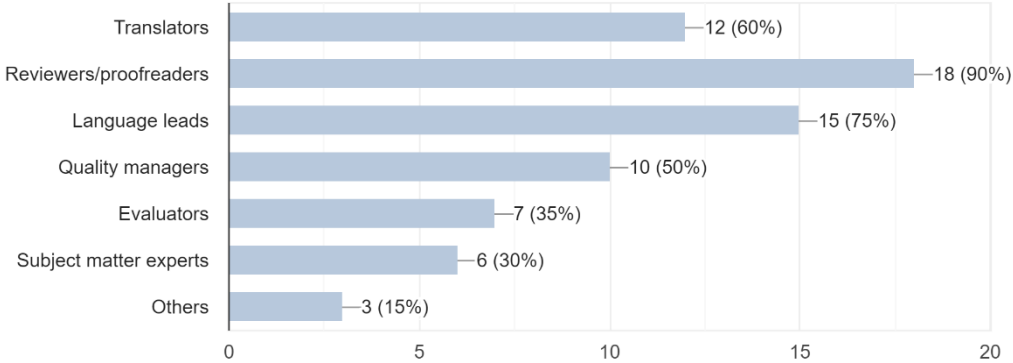
Figure 74. Human resources involved in translation quality management – Translation service providers.

Similarly to the results compiled from the survey filled by the representatives of the translation service buyers, the type of human resources most commonly involved in translation quality management processes on the provider's side are reviewers/proofreaders (90.0%), followed by language leads (75%), translators (60%), and quality managers (50%) and language leads (54.8%). Other profiles such us evaluators or subject matter experts were mentioned only by 35% and 30% of the participants.
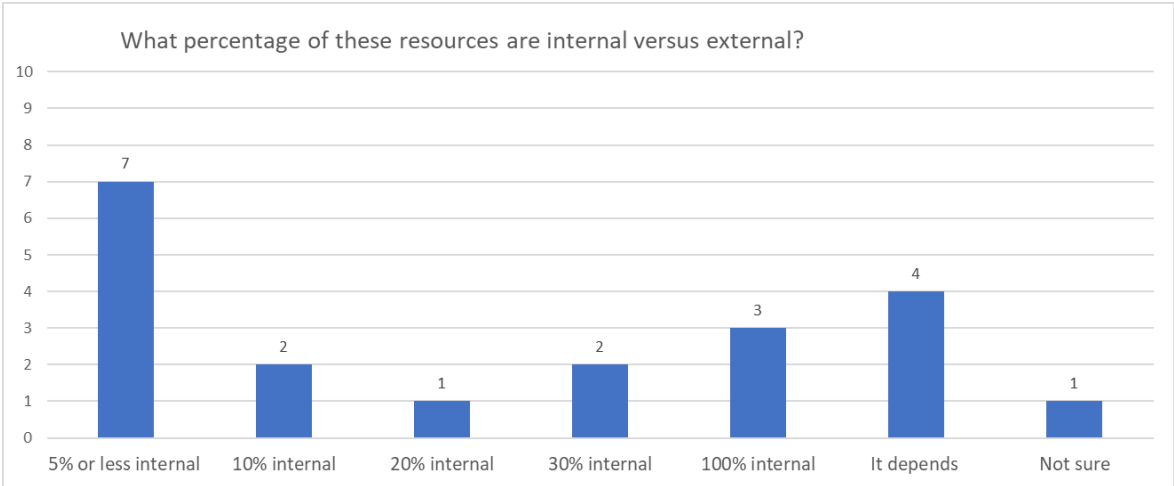
Figure 75. Percentage of internal human resources versus external resources – Translation service providers.

When asked about the percentage of the internal human resources that were involved in translation quality tasks, two distinctive trends were identified. Either the resourcing model of the translation service provider favoured the externalisation of these tasks (45% of the

candidates revealed that 10% or less of these resources were internal), or it adopted a flexible approach in which translators and reviewers were usually external, but language leads and evaluators were in-house workers (16% of the candidates mentioned this approach). In addition, 13% of the respondents mentioned a third approach in which all the resources involved in translation quality tasks tend to be employed directly by the company.

The study also found some commonalities as to which were the main reasons for the different translation providers to choose one approach over the other.

Advantages of having most of the translation quality management resources in-house mentioned by the participants:

> "It is better to handle the time and quality with internal sources." [Questionnaire respondent]

> "We perfectly know our in-house linguists and, as employees, they do not have time/money pressure." [Questionnaire respondent]

> "We don't have any trained external resources to handle translation quality management and we feel more confident keeping this internally." [Questionnaire respondent]

Advantages of having most of the translation quality management resources outsourced mentioned by the participants:

> "Scalability and talent within native country - Business Model." [Questionnaire respondent]
> "It is more efficient for an MLV" [Questionnaire respondent]

Advantages of having a mixed approach mentioned by the participants:

> "For us it is easier to have a majority of translators and reviewers that are external, while the majority of language leads/evaluators are internal" [Questionnaire respondent]

> "For the language combinations that are not our main ones, we create reliable teams and externalize everything." [Questionnaire respondent]
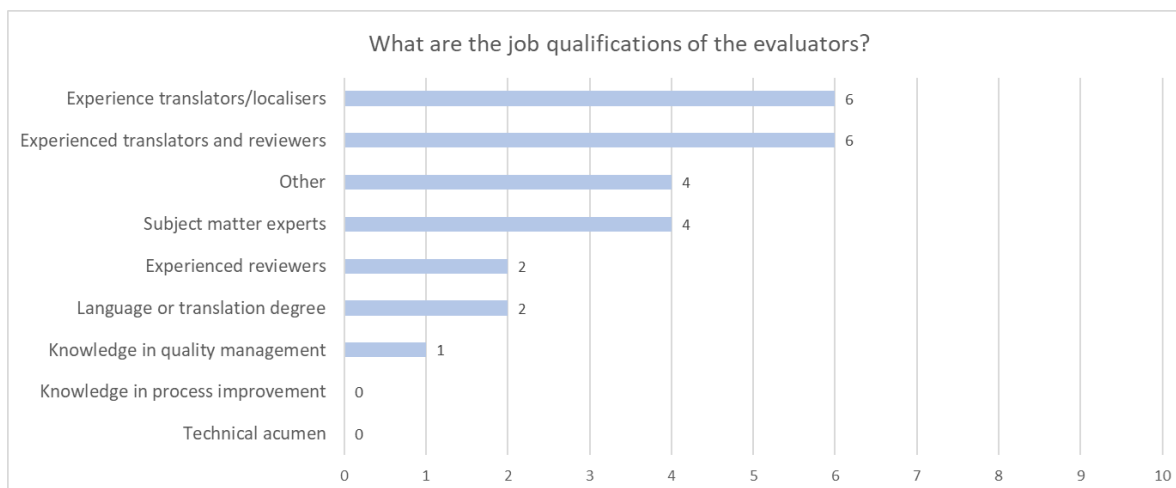
Figure 76. Skills and qualifications of evaluators – Translation service providers.

24% of the participants on the provider's side declared that evaluators were mostly required to have experience in translation, while another 24% reported that evaluators had to have proven experience both as translators and reviewers. Another qualification mentioned by the respondents was experience as subject matter experts in the area of specialisation or the client's account. Language or translation degrees were cited as a qualification in 8% of the replied. Another finding of this question was that 16% of the replies mentioned a different criterion followed by the providers. Many of the respondents that chose "Other" specified that they had a specific training path for reviewers and evaluators, which were selected among the most experienced and high-quality translators of their databases.
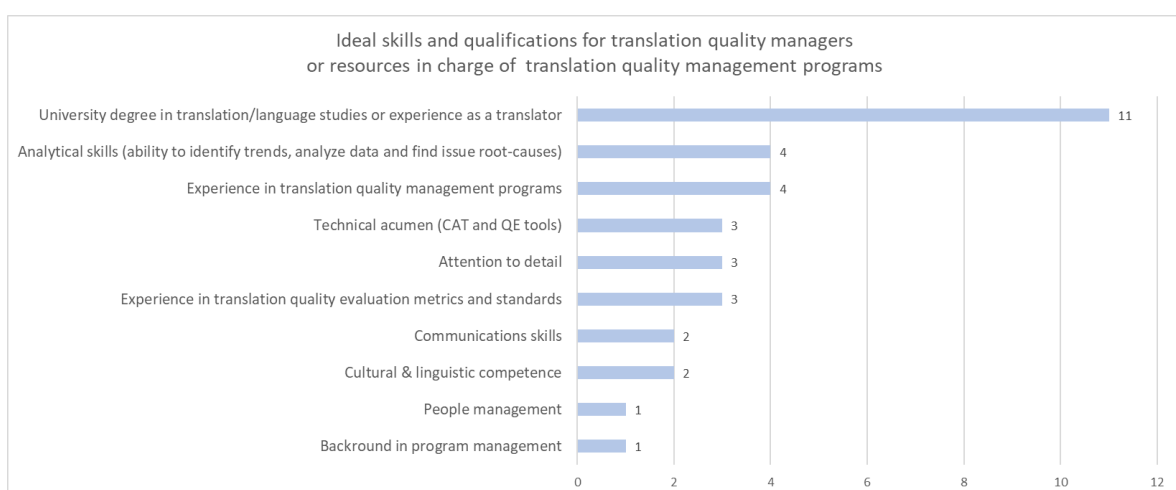
Figure 77. Skills and qualifications of translation quality managers – Translation service providers.

The last question of this part of the study aimed to get some insights regarding the most valued skills and qualifications for translation quality managers or resources in charge of

123

translation quality management programs. This question had an open-ended format, and the same coding parameters used in the version provided to the translation service buyers were applied here too. As a result, there were 10 top qualifications detected (Figure 77). The most valued qualification was having a university degree in translation or experience as a translator, followed by analytical skills or experience in translation quality management programs.

4.3.3.4 Tools and technologies
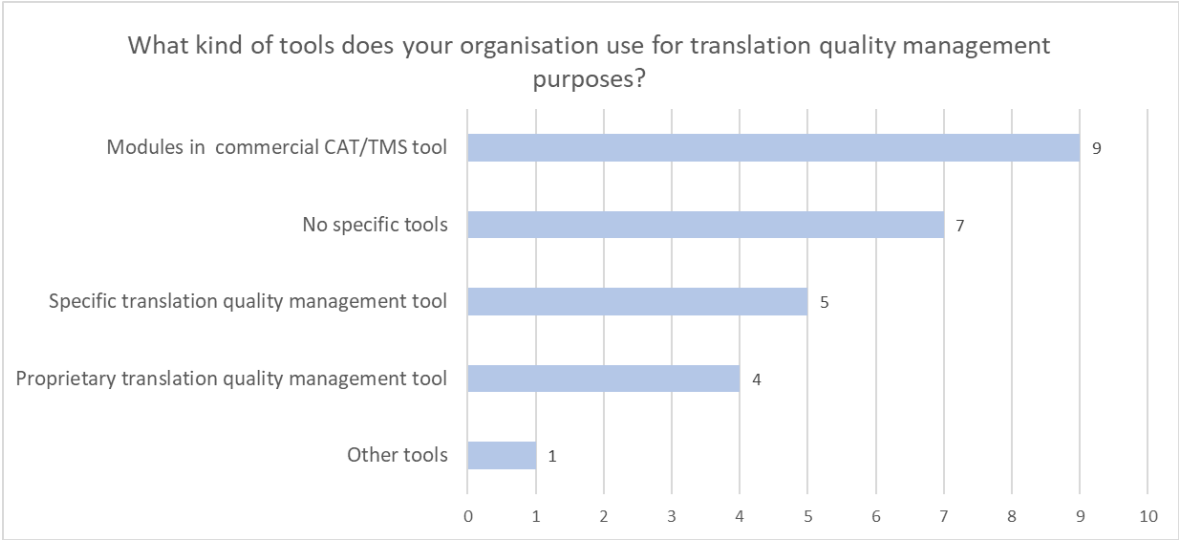
Figure 78. Tools used for translation quality management purposes – Translation service providers.

35% of the participants on the provider's side reported that their companies used modules and features available in the commercial CAT/TMS tools for quality management purposes (Memsource, XTM, Trados Studio, Matecat, MemoQ, Wordfast and XTRF). Contrary to the information provided by the translation service buyers, of which only 6% stated that they did not use a specific translation quality management tool, 27% of respondents from the provider's side declared that their companies did not have any specific tools for this purpose. In addition, 19% of the participants reported that their companies used standalone translation quality tools such as ContentQuo, Xbench and Verifika. Finally, 15% of the respondents revealed that their companies had developed proprietary tools for translation quality management purposes.
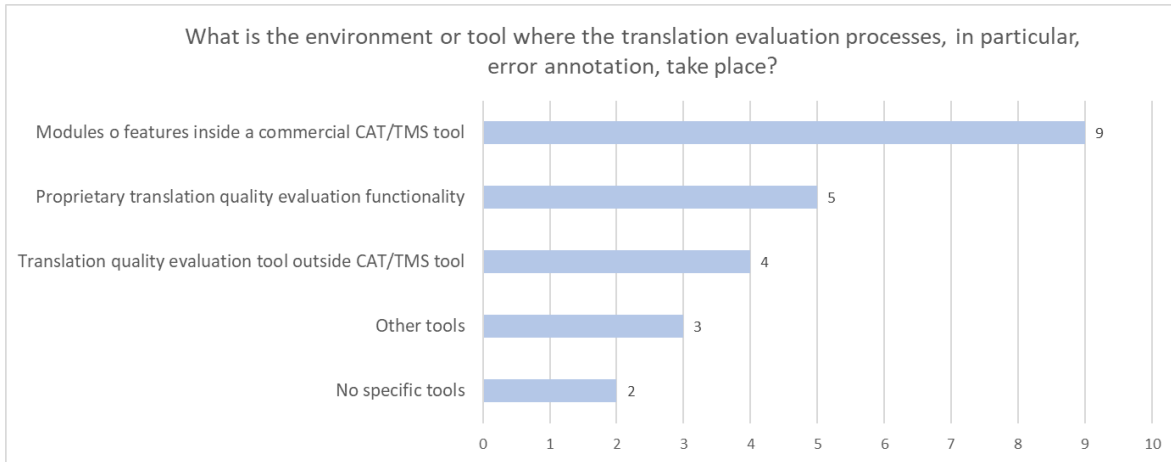
Figure 79. Tools used for translation quality evaluation (including error annotation) purposes – Translation service providers.

In line with the information provided by the participants in the previous question, 39% of the respondents stated that their companies made use of the features available in the above-mentioned commercial CAT/TM tools for error annotation. The second most popular option was the use of proprietary translation quality evaluation tools (22%), and the third option (17%) was the use of stand-alone evaluation tools outside CAT/TMS environments. 13% of the respondents declared that their companies used other tools for evaluation purposes (primarily spreadsheets), and 9% said their companies did not use any specific tools for error annotation.

4.3.4 Current challenges in translation quality management and evaluation

The aim of this part of the survey was to get information about the current level of satisfaction with the translation management program in place and compile additional insights regarding the current challenges faced by the translation service providers.
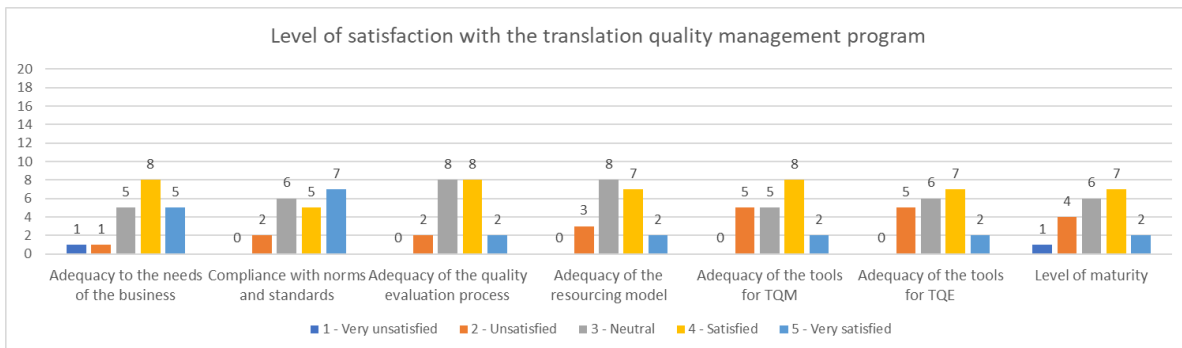
Figure 80. Level of satisfaction with the translation quality management program in place – Translation service providers.

The two parameters with a higher level of satisfaction stated by the participants were the adequacy to the needs of the business and the compliance with norms and standards. On the other hand, the two categories with the lowest level of satisfaction were the adequacy of the tools used for translation quality evaluation and the adequacy of the tools used for translation quality management.

Main pain points in translation quality management and evaluation	Nr of replies
TQM: Lack of centralised data or dashboards with visibility on trends	1
TQM: Financial limitations to invest in tools or dedicated resources	1
TQM: Lack of cost-effective, scalable and flexible TQM methodologies	7
TQM: Lack of training and knowledge	2
TQM: Lack of automatisation	5
TQM: Lack of access to practical information on TQM methodologies	2
Clients: Lack of visibility of upcoming work	1
Clients: Unrealistic, inconsistent or unclear quality expectations	2
Clients: Aligning and educating clients on quality management best practices	1
Clients: Lack of content strategy	1
Quality improvement: Time constrains	3
Tools: Ability to review content easily in context (app and web content)	1
Tools: More integration required into TMS (TQM and TQE)	3
Tools: Excel still used as main tool for TQM and TQE	1
Tools: Lack of advanced tools for TQM and TQE	1
TQE: Alignment and calibration among different reviewers	1
TQE: High costs, time-consuming processes and rigidity of the evaluation mod	1
TQE: MT evaluation (raw and post-edited)	1
TQE: Use of unstructured holistic models without objective grading	2

Figure 81. Main pain points in translation quality management and evaluation – Translation service providers.

Next, participants were asked to mention the main three translation quality management and evaluation pain points currently faced by their organisations. This question had an open-ended format, and the replies were coded and grouped into five main topics: Challenges in translation quality management, challenges related to the collaboration with their clients, challenges related to quality improvement, challenges related to the tools used, and specific challenges in translation quality evaluation. The main findings are summarised in Figure 81 above.

4.3.5 Solutions

In this part of the research, candidates shared the solutions and workarounds applied by their organisations to solve the pain points mentioned in the previous question. The replies to this open-ended question were coded and added to the table with the compilation of pain points to show the correlation between the answers to both questions. As can be seen in the figure

below, two of the most common solutions adopted by the participants are the research and adoption of new translation quality management or evaluation tools, and the transfer of knowledge using different strategies, such as hiring consultants or quality managers, learning from colleagues or in conferences, or helping less experienced clients.

Main pain points in translation quality management and evaluation	Applied solutions and workarounds
TQM: Lack of centralised data or dashboards with visibility on trends	Spreadsheet to register the results of each evaluation
TQM: Financial limitations to invest in tools or dedicated resources	Researching affordable tools that would mean an improvement to the TQM processes
TQM: Lack of cost-effective, scalable and flexible TQM methodologies	Hiring a Quality Manager
TQM: Lack of training and knowledge	Hiring consultants to help us optimise our TQM processes
TQM: Lack of automatisation	Investing in new automatisation tools
TQM: Lack of access to practical information on TQM methodologies	Learning from colleagues and similar companies
Clients: Lack of visibility of upcoming work	Forecast requests, scalable teams
Clients: Unrealistic, inconsistent or unclear quality expectations	Setting up specific processes per client to have clear instructions
Clients: Aligning and educating clients on quality management best practices	Proactive communication towards clients concerning expectation management
	Helping less experienced clients understand the benefits of analytical quality monitoring and to correctly interpret data
Clients: Lack of content strategy	Helping clients create a Content Matrix, define quality expectations with their stakeholders
Quality improvement: Time constrains	
Tools: Ability to review content easily in context (app and web content)	
Tools: More integration required into TMS (TQM and TQE)	Research on available TQM and TQE tools that can be plugged into our TMS
Tools: Excel still used as main tool for TQM and TQE	Research on specific translation quality management tool to replace Excel
Tools: Lack of advanced tools for TQM and TQE	Research on available TQM and TQE tools
TQE: Alignment and calibration among different reviewers	Hiring a Quality Manager
TQE: High costs, time-consuming processes, rigidity of the evaluation model	Using alternative approaches (such as holistic approach or, fit for purpose evaluation)
TQE: MT evaluation (raw and post-edited)	Trying to harmonize MTPE evaluation instructions
	Going to conferences to try to learn better ways to do MT evaluation
TQE: Use of unstructured holistic models without objective grading	Hiring a Quality Manager

Figure 82. Pain points and solutions applied in translation quality management and evaluation – Translation service providers (see Appendix E).

4.3.6 The future of translation quality management

In the last section of the survey, candidates had the opportunity to share their insights regarding the main challenges they believed would impact their companies' translation quality management processes in the near future. This question had an open-ended format, and the replies were coded and grouped to show the most recurrent topics shared by the respondents (see Figure 83 below). Some of the replies collected in this question mention the same themes present in the two previous questions, in particular, the need to automatise and integrate different tools for TQM and TQE purposes and the demand of more scalable and flexible translation quality management and evaluation methodologies.

Future challenges in translation quality management
Increase capacity and scalability to cover bigger volumes and more languages
New types of TQE metrics: user centric, holistic, post-editing effort, MT evaluation
Automation and integration of tools for TQM and TQE purposes
New text types in the TQM workflows: Transcreation, creative translations, MT evaluation, post-editing
Setting or reshaping the whole quality management strategy
Keep up with the most recent industry standards and methodologies
Translation quality management business analytics
Evaluation of different quality levels requested by clients
New tecnologies: Machine Learning, AI

Figure 83. Future challenges in translation quality management and evaluation – Translation service providers.
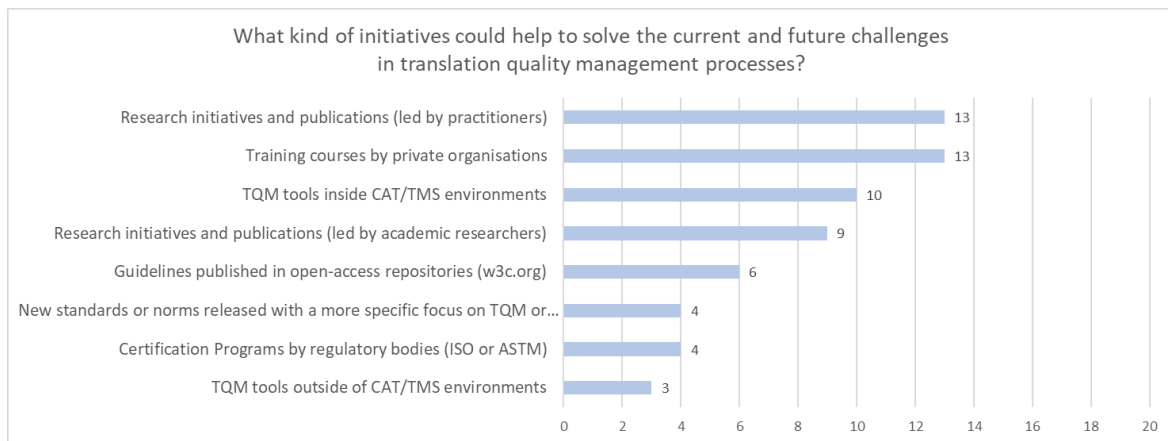
Figure 84. Useful initiatives in translation quality management and evaluation – Translation service providers.

The last closed-ended question of the survey asked the respondents to select which initiatives they believed could help the advancement of the translation quality management programs in professional environments. These are some of the main findings of this questions:

- The initiatives that were mentioned more frequently by the participants were research initiatives, publications and training courses led by practitioners or industry-related organisations.

- Respondents also favoured translation quality management solutions that would work inside their existing CAT/TMS environments instead of standalone ones.

- In the last positions of this list, we can find initiatives lead by standardisation bodies, such as new norms or certification programs.

4.3.7 General comments and feedback

Out of 20 participants on the provider's side, seven chose to fill the free text question at the end of the survey to confirm their interest in receiving a summary of the results of the survey, but none of them added any additional information or feedback.

4.4 Translation technology companies

Out of the 68 participants in this research, 4 of them worked in translation technology companies. Even though the number of participants is too small to offer a representative sample of the current practices followed by these companies to facilitate translation quality management and evaluation tasks, the qualitative data compiled can serve as a baseline from which continue exploring this research topic at a bigger scale in future research initiatives.

In the following sections, we will present the data points collected from this target group.

4.4.1 Demographics

Job title	Technology service providers
CEO	1
Product Owner (software industry)	1
Quality Strategist	1
Translation/Localisation Quality Manager	1

Figure 85. Job titles – Translation technology providers.

Given the smaller number of companies that specialised in translation technology compared to the number of companies that provide translation services (Choudhury & McConnell, 2013), it is not surprising to see a lower among participants from this target group. Two of these participants had job titles closely related to the translation quality field (Quality Strategist and Translation/Localisation Quality Manager), while the other two had more generic roles (CEO and Product Owner).
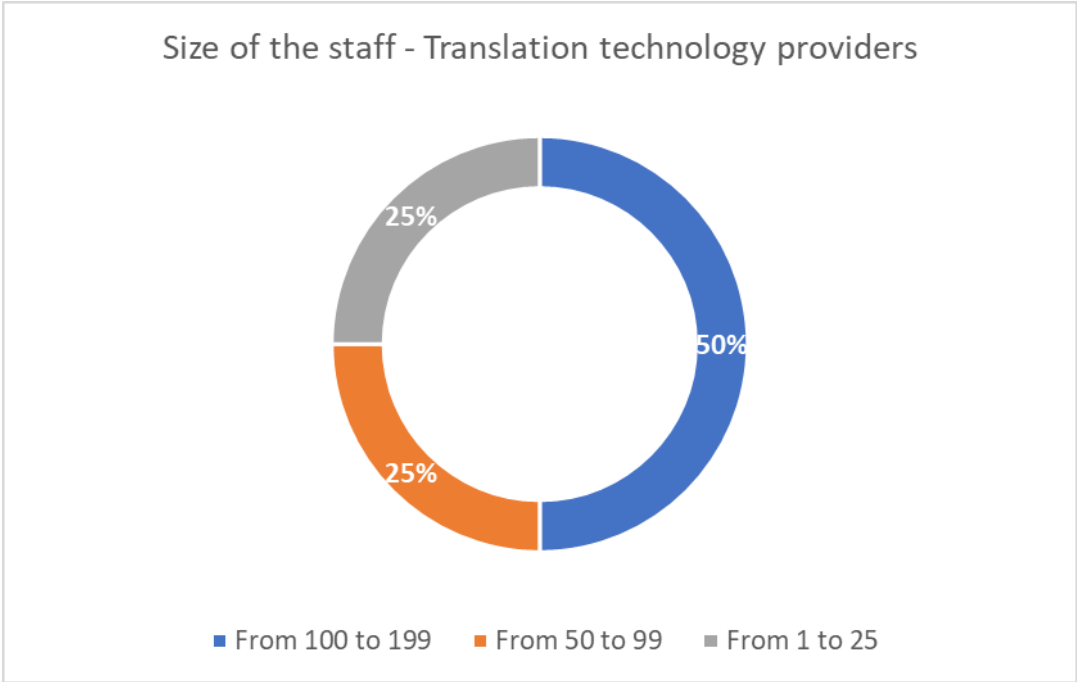
Figure 86. Size of the staff – Translation technology providers.

Half of the participants stated that they worked in translation technology providers with 100 or more employees, while the other 25% were from medium-sized companies (from 50 to 99 employees) or small-sized companies (25 or fewer employees).
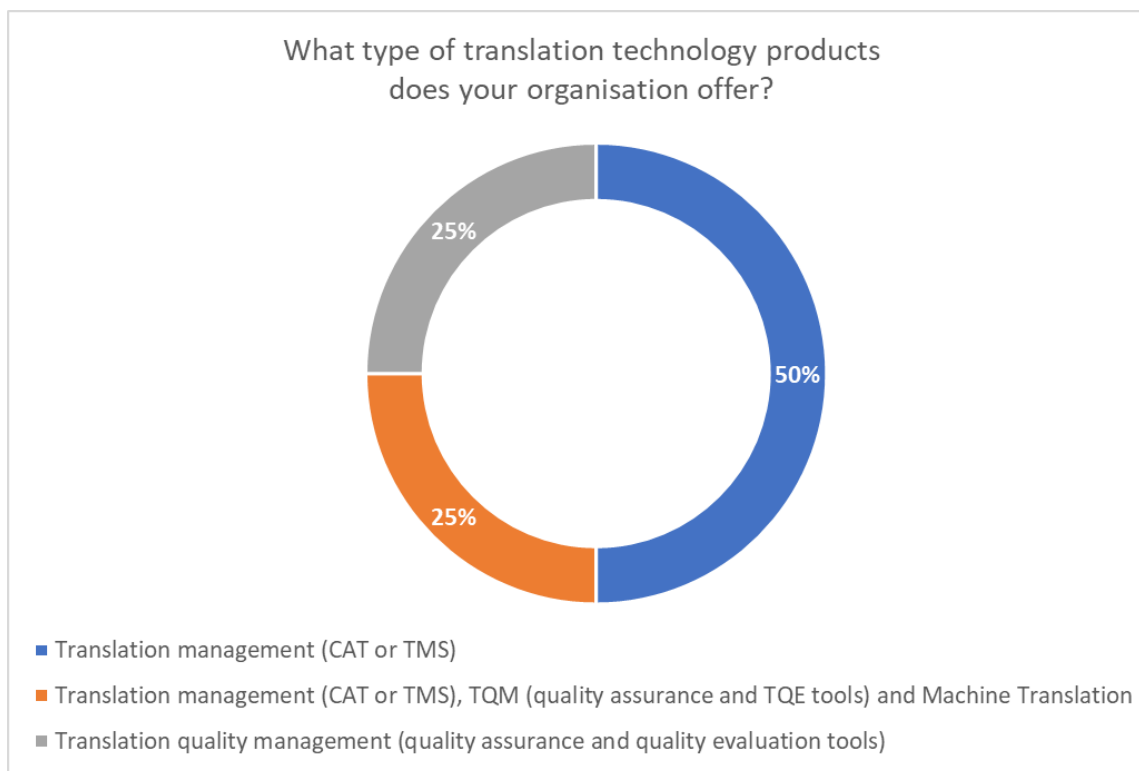
Figure 87. Translation technology products offered – Translation technology providers.

Two of the respondents reported that the translation technology providers they worked for offered mainly translation management products (CAT or TMS tools), while the others stated that they worked at a provider that offered translation quality management solutions and at a provider that had products for translation management, translation quality management and machine translation technologies.

4.4.2 Translation quality management

The second part of this version of the survey for translation technology providers covered different aspects of the translation quality management workflows that could be performed in the solutions of these providers, and it was divided into three main topics: adherence to norms and standards, translation management and evaluation workflows, and technical specifications.

4.4.2.1 Adherence to norms and standards
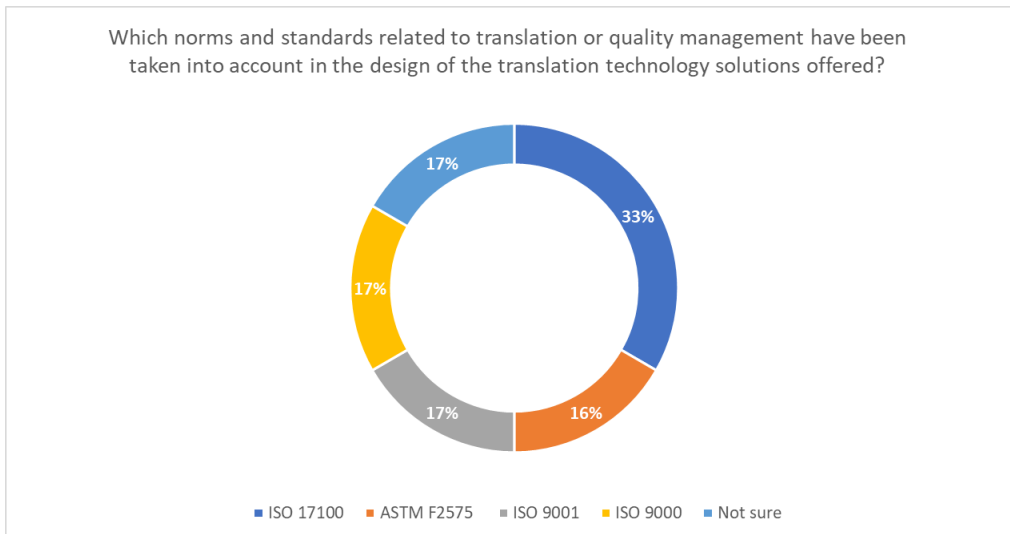
Figure 88. Adherence to norms at standards – Translation technology providers.

The majority of the participants confirmed that the technology products offered by their companies had taken into consideration at least one of the main norms and standards on translation services or quality management, mainly ISO 17100, which describes the requirements for translation services.
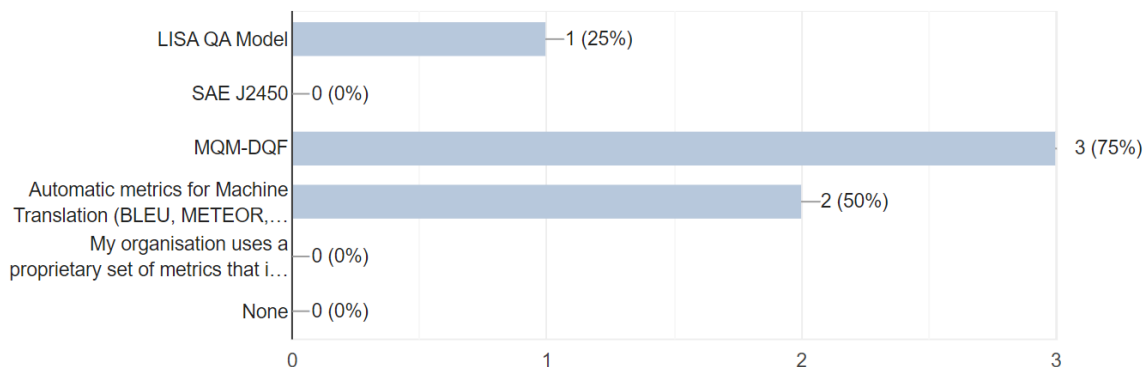
Figure 89. Adherence to standardised translation metrics – Translation technology providers.

The representatives of the translation technology providers also confirmed that their solutions allowed the use of at least one standardised translation metric, and 3 out of 4 stated their solution was compatible with MQM-DQF instead of the older LISA QA model.

131

4.4.2.2 Translation quality management and evaluation workflows

This part of the survey aimed to gain deeper insights into the features of the technology solutions in terms of translation quality management and evaluation workflows, and it was divided into three different topics: translation quality management, translation quality evaluation and user profiles.

Translation quality management

In order to have a better understanding of the potential limitations of the solutions offered by these technology providers, participants were asked if they had detected any gaps or opportunities to enhance further the features related to translation quality tasks. This question had an open-ended format, and the replies were summarised in Figure 90 below. Two of the main limitations mentioned by the participants were the lack of dedicated quality management and evaluation functionalities, and the combination of different types of metrics (human and automatic) to perform machine translation evaluation.

Nr	Reply
1	Government clients requirement are different from enterprise customers.
2	Some client uses exclusively desktop CAT tools, but we are 100% cloud-based.
3	Our tool is AI-based where the engine is being trained and can lead to very good quality and less good quality depending on the effort of the human. A BLEU score would not be sufficient but our MQM-DQF based model leads to misleading quality concerns.
4	Other tools specifically made for translation quality management and evaluation have more functionalities, and clients would like to see in our tool too.

Figure 90. Gaps and opportunities for improvement – Translation technology providers.

Translation quality evaluation
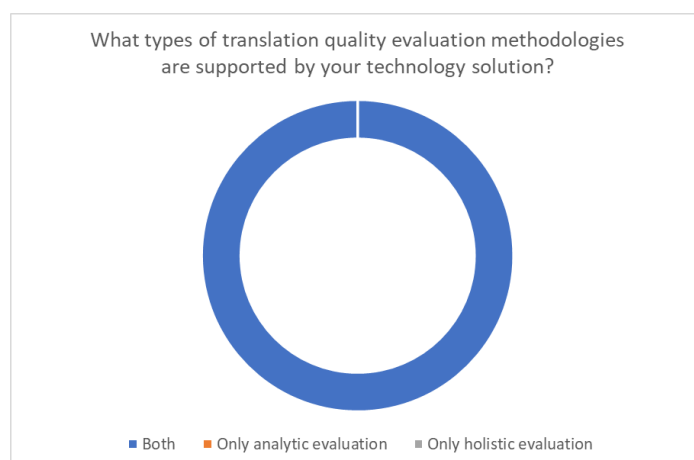
Figure 91. Types of translation quality evaluation methodologies supported – Translation technology providers.

Does your translation quality management solution allow the evaluation of whole texts or select a sample for evaluation?

4 responses

- Only whole texts
- Only samples
- Both

100%
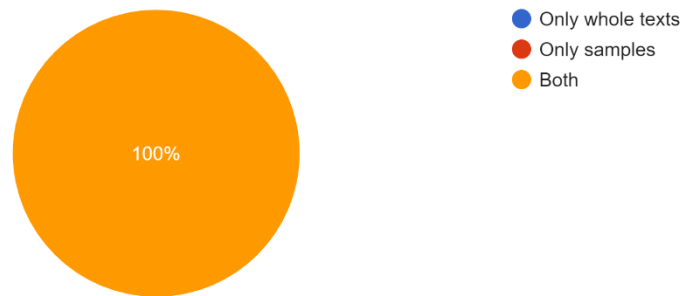
Figure 92. Sampling supported – Translation technology providers.

As can be seen in Figures 91 and 92 above, all the respondents on the translation technology provider's side declared that their companies supported analytic and holistic evaluations, as well as evaluation of whole texts or samples.

How are samples selected?

4 responses

- Samples can be selected to evaluate random, representative text passages
- Samples can be selected to focus on high-priority text passages
- Both methods
- It is not possible to select samples for evaluation purposes
- Probably both methods but it's not fully automated and requires manual work
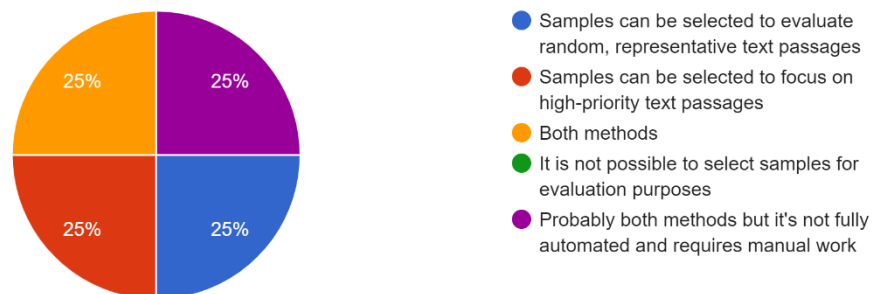
Figure 93. Sampling methods supported – Translation technology providers.

Only one of the participants stated that their technology solution allowed different sampling approaches (random versus targeted), while two of the participants mentioned that only one approach was possible.

133
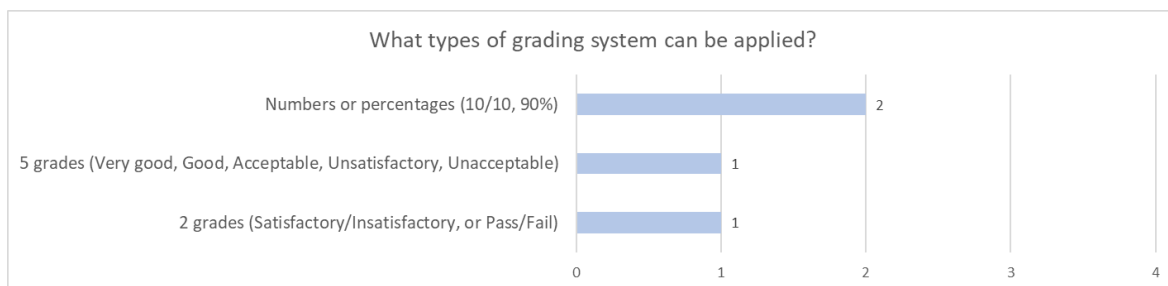
Figure 94. Grading systems supported – Translation technology providers.

Half of the participants reported that their solution allowed the use of numbers or percentages as part of the grading system, while the other half declared that their solution only offered one type of grading system (either two grades or five grades).
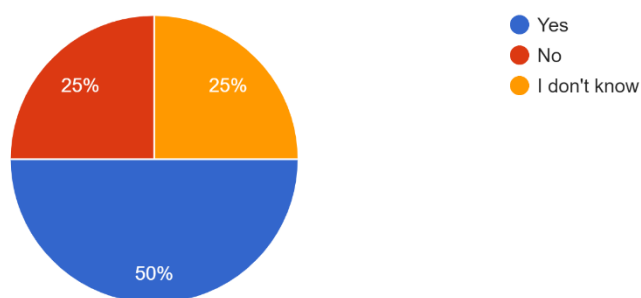
Figure 95. Uniformity of evaluation metrics in translation quality evaluation – Translation technology providers.

Two of the four respondents confirmed that their solution was designed to allow the use of different evaluation metrics to different text types or user profiles, while the other two said that it was not possible or that they were not sure.
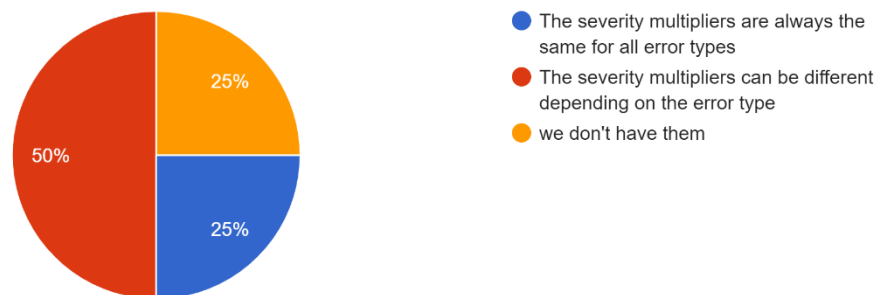
Figure 96. Uniformity of severity multipliers in translation quality evaluation – Translation technology providers.

Similar results can be observed in the following questions, as half of the participants confirmed that their solution allowed the use of nonuniform severity multiplier distributions depending on the error type, while the other half revealed that their solution did not have a severity multiplier or that the severity multipliers were the same for all error types.

Figure 97. Error typology in translation quality evaluation – Translation technology providers.

Three out of the four respondents stated that the error types used in their solutions used the same terminology of the three main error types proposed by the standardised metrics mentioned in section 2.3.7: Accuracy, Fluency and Terminology. However, some of them did not have Style, Design, Locale convention, or Verity. Additionally, some of the

135

respondents specified that their solutions used a different terminology or set of categories, such as:

- Text Specification (similar to Style and Verity)
- Technical (tag errors, link errors)
- Representation of the brand

What severity levels does it include? Please, select all the applicable options:
4 responses

Minor	3 (75%)
Major	3 (75%)
Critical	3 (75%)
Neutral/Preferential	1 (25%)
None	1 (25%)
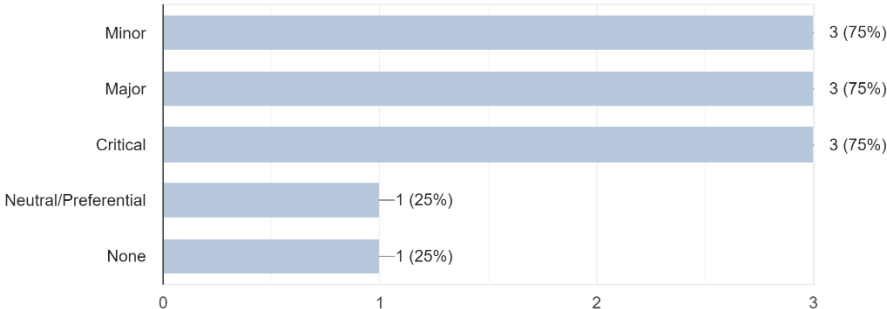
Figure 98. Error severity in translation quality evaluation – Translation technology providers.

The results were more homogeneous in the next question, which was related to the number and name of the different severity levels that could be selected in the solutions of the technology providers. Three of the four participants confirmed that they had the three categories proposed by LISA QA Model and MQM-DQF ("Minor", "Major", and "Critical"). However, only one of them had the "Minor/Preferential" category.

Does the evaluation include a pass/fail threshold?
4 responses

- Yes
- No

100%
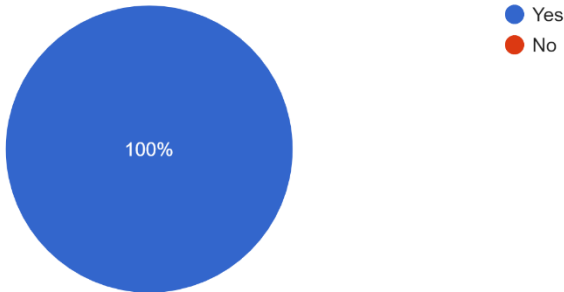
Figure 99. Use of Pass/Fail thresholds in translation quality evaluation – Translation technology providers.

All the participants in the study on the translation technology side declared that their solutions did include certain parameters to define a pass/fail threshold for translation quality evaluation purposes. In some cases, the threshold can be customised by the users of the tools, while in others, the tools have certain defined thresholds that are included as guidance.

User profiles

How many different user roles are available in the translation quality management/evaluation solution?

4 responses

Translators	3 (75%)
Reviewers/proofreaders	3 (75%)
Language leads	1 (25%)
Quality managers	1 (25%)
Evaluators	0 (0%)
Subject matter experts	0 (0%)
Others	2 (50%)
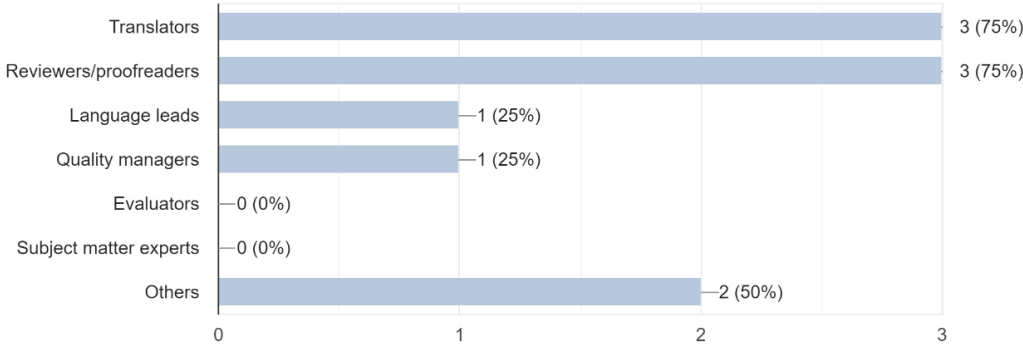
Figure 100. User roles available – Translation technology providers.

Most of the translation technology solutions (3 out 4) had differentiated user profiles for translators and reviewers/proofreaders, and two of them had other user profiles that could be used by other types of users, such as project managers. However, only one of the solutions had specific user roles for language leads or quality managers, and none of them had a specific user profile for evaluators or subject matter experts.

4.4.2.3 Technical specifications

How would you define the technical setting of the solution you provide for translation quality management purposes?

- 25% It is an API that can be connected programmatically to any system
- 75% It is a module or feature included in a CAT/TMS tool
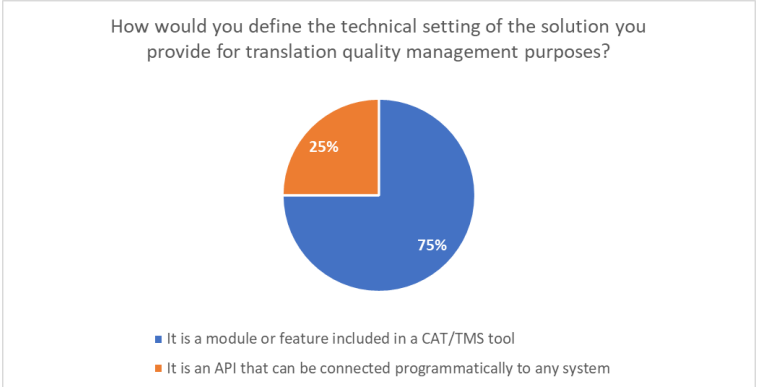
Figure 101. Translation quality management solution – Translation technology providers.

Three out of the four respondents reported that the specific solution provided for translation quality management was a module or feature included in their CAT/TMS tool, while the other participant explained that their solution was an API that could be connected to other translation systems.
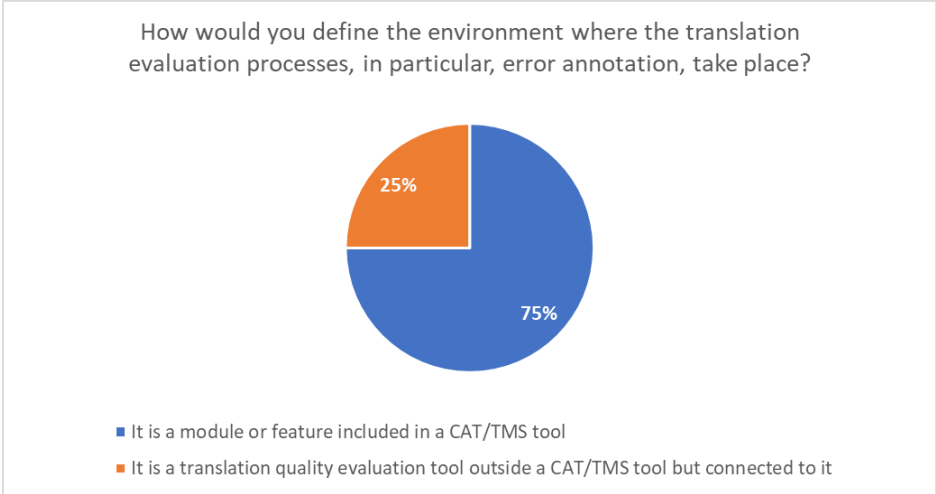
Figure 102. Translation quality evaluation solutions – Translation technology providers.

The technical setting of the solutions for error annotation and translation evaluation was also very similar, as three of the participants shared that the solutions for these particular tasks were also part of a module of feature included in their CAT/TMS, while the fourth respondent said that their solution was a standalone tool that could be connected to a CAT tool or TMS.
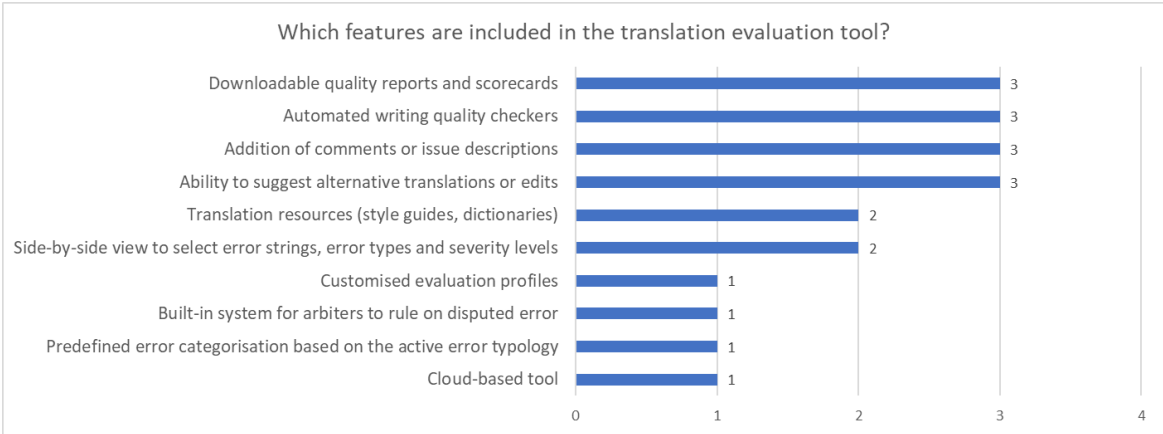
Figure 103. Translation quality evaluation solution inside CAT/TMS tools – Translation technology providers.

Those participants who confirmed their solutions allowed to perform translation quality evaluations inside a CAT/TMS were asked a follow-up question to get deeper insights into the most common features of this technical setting. As shown in Figure 104, some features

138

(such as downloadable reports, automated writing quality checkers or editing/review capabilities) were available in all the solutions, while others were available in two out of the three solutions, or only in one of them.

Features available	Features not available
Built-in system for arbiters to rule on disputed error	Built-in system for arbiters to rule on disputed error
Ability to suggest alternative translations or edits	Ability to suggest alternative translations or edits
Addition of comments or issue descriptions	Addition of comments or issue descriptions
Cloud-based tool	
Predefined error categorisation based on the active error typology	
Customised evaluation profiles	
Side-by-side view to select error strings, error types and severity levels	
Translation resources (style guides, dictionaries)	
Automated writing quality checkers	
Downloadable quality reports and scorecards	
Connection with CAT/TMS tool through an API	
Quality dashboards with information on quality scores and quality levels	

Figure 104. Translation quality evaluation solution outside CAT/TMS tools – Translation technology providers.

A similar follow-up question was shown to the other participant to get more information about the features available in the solution outside any CAT/TMS environment. As shown in Figure 104, most of the features specifically related to the evaluation of the translation itself (such as arbitration capabilities or addition of issue descriptions) were not available in this solution.

4.4.3 Current challenges and pain points

In the first question of this part of the survey, the respondents were requested to rate, on a scale of 1 to 5, the level of adequacy of the tools and solutions for translation quality management and evaluation.
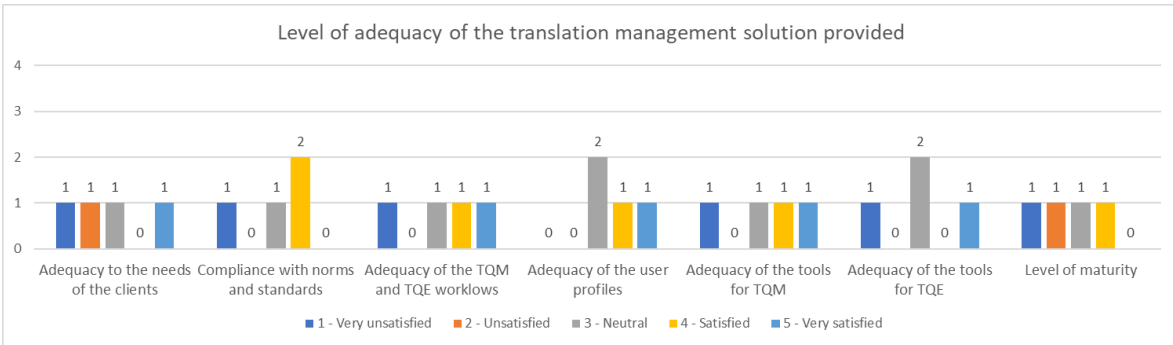
Figure 105. Level of adequacy of the solution provided – Translation technology providers.

As shown in Figure 105, the sentiment of the different participants was quite varied, so in order to better analyse the results, the answers were grouped into two categories: positive sentiment (ratings of 4 and 5) and negative sentiment (ratings of 1 and 2). According to this

classification, the categories with a higher number of replies indicating a positive sentiment (and, therefore, a higher satisfaction level) were the adequacy to the needs of the clients and the overall level of maturity of the solutions.

On the other hand, the categories with a higher number of replies indicating a negative sentiment (and, therefore, lower satisfaction level) were the level of adherence to translation quality management and evaluation norms and standards, the adequacy of the translation quality management and translation evaluation workflows of the solutions, and the adequacy of the tools for translation quality management purposes.

Current pain points in translation quality management and evaluation
Lack of automatisation
Sampling methodologies used in quality assessment
Lack of real-time quality assurance
Inflexibility and lack of customization in QA systems using style guides
Closing the feedback loop
Simplifying the data collection
Allowing business stakeholders to quickly analyze that data

Figure 106. Main pain points in translation quality management and evaluation – Translation technology providers.

The next question had an open-ended format to allow the participants to share the main challenges in translation quality management and evaluation workflows that their solutions were trying to address. The replies covered pain points such as the lack of automatization and real-time functionalities, the use of sampling methodologies, and the simplification and centralisation of data analytics.

4.4.4 Solutions

In the following part of the survey, candidates were asked to share the functionalities their companies were developing to try to solve the challenges mentioned in the previous question. The replies were added to the table shown in Figure 107 to show the correlations between pain points and solutions. As shown below, the main technical solutions developed by the participants aim to improve the lack of automation and to address the need to simplify and provide quicker quality insights throughout the quality management process.

Current pain points in translation quality management and evaluation	Technical solutions developed
	Automated feature to making data available at various phase in the process
Lack of automatisation	Automated alert to PMs when a translation has a certain number of comments and QA errors
Sampling methodologies used in quality assessment	Product development to cater to different needs
	Live QA in translation editors
Lack of real-time quality assurance	Integrate style guide rules into QA module warnings
Inflexibility and lack of customization in QA systems using style guides	Product development to cater different needs
Closing the feedback loop	Automatisation of feedback loops (reviewers, evaluators, arbiters)
Simplifying the data collection	Aggregation of data in a meaningful way
Allowing business stakeholders to quickly analyze that data	Automation of reporting features

Figure 107. Pain points and solutions developed – Translation technology providers.

4.4.5 The future of translation quality management

In this part of the research, the respondents from the translation technology side were asked to share the needs and trends in translation quality management processes that their solutions would need to be able to address in the future. This question was open-ended, and the replies were coded to show the most recurrent topics mentioned by the participants. As shown in Figure 108 below, some of the replies to this question focus on the same gaps identified in the previous question, namely, the lack of automation, and the need to provide simpler and quicker quality data at different points of the production process.

Future trends and needs in translation quality management processes
TMS that allow changes in the workflows after translation starts
MT backed up translation reviews
Predictions on translation quality coming from MT engines
QA support for low-resource locales
Morphology checks in heavily inflected languages
Consolidate quality assurance with quality assessment into one comprehensive quality management process

Figure 108. Future trends and needs in translation quality management processes – Translation technology providers.
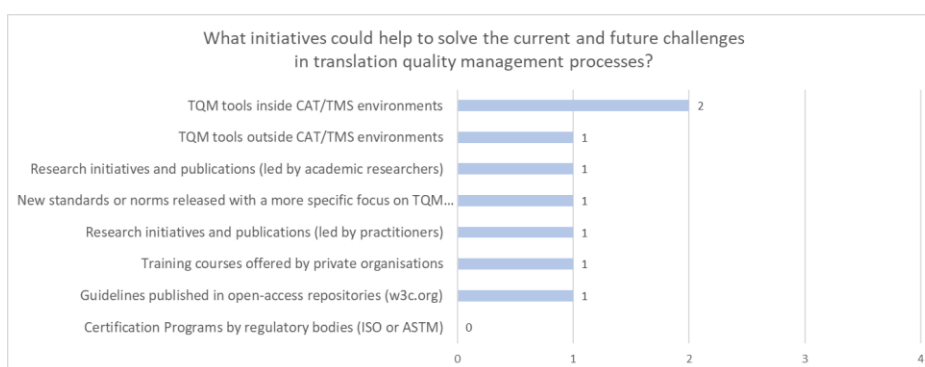
Figure 109. Future initiatives in translation quality management processes – Translation technology providers.

The next question aimed to get additional insights as to what type of initiatives would be helpful to the advancement of the translation quality management programs in professional environments.

- Half of the respondents mentioned that having TQM tools inside CAT/TMS environments was preferable to having them outside.
- The rest of the options were selected at least one, with the only exception of "Certification Programs by regulatory bodies (ISO or ASTM) that was not selected by any of the candidates.

4.4.6 General comments and feedback

One of the participants on the translation technology side shared some final thoughts on the future of translation quality:

> "My personal opinion as to translation "quality": Once the AI and MT develops further, the "accuracy" and "language error" will be less important in translation quality management (I am hoping that that [sic] very little errors of this nature will be introduced). I would feel that the "meet the purpose" criteria that focus on user experiences may start playing an important part within "quality" management." [Questionnaire respondent]

4.5 Experts in translation quality management

Out of the 68 participants, 13 were experts in translation quality management (19.2%), of which 11.8% were non-academic, and 7.4% were teachers or academic researchers. In the following sections, we will introduce the data collected in each of the six parts of the survey that was designed for this target group.

4.5.1 Demographics
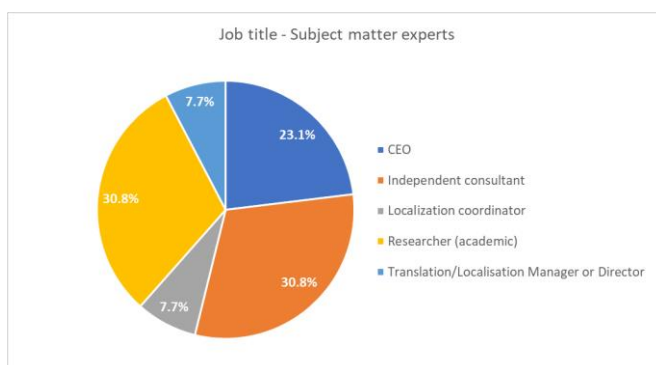
Figure 110. Job titles – Experts in translation quality management.

30.8% of the participants of this study as subject matter experts on translation quality management and evaluation stated that they worked primarily as academic researchers, while another 30.8% were independent consultants. The third category in terms of representation was the CEOs, who accounted for 23.1% of the entries. Finally, a few respondents worked as Localisation coordinators or as Translation/Localisation Managers or Directors.
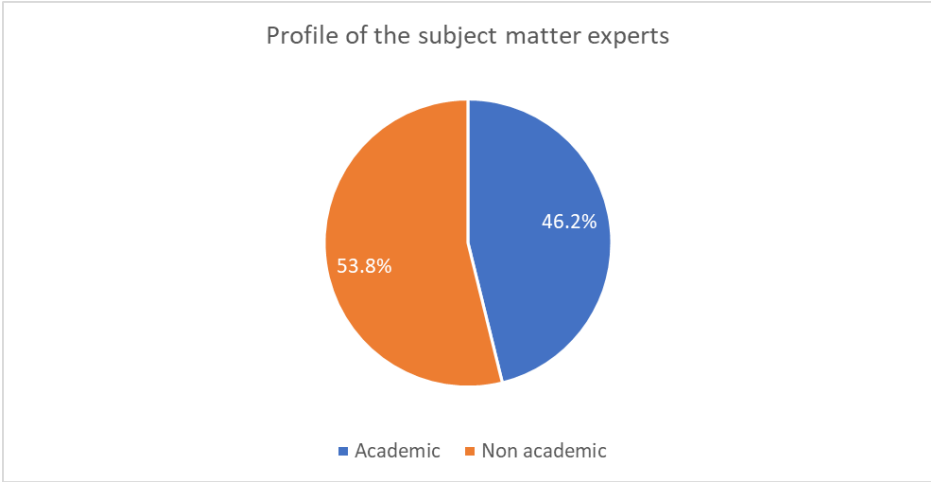
Figure 111. Profile of the participants – Experts in translation quality management.

In terms of professional background, 53.8% of the respondents reported having a non-academic background, while the other 46.2% mentioned being teachers or academic researchers.
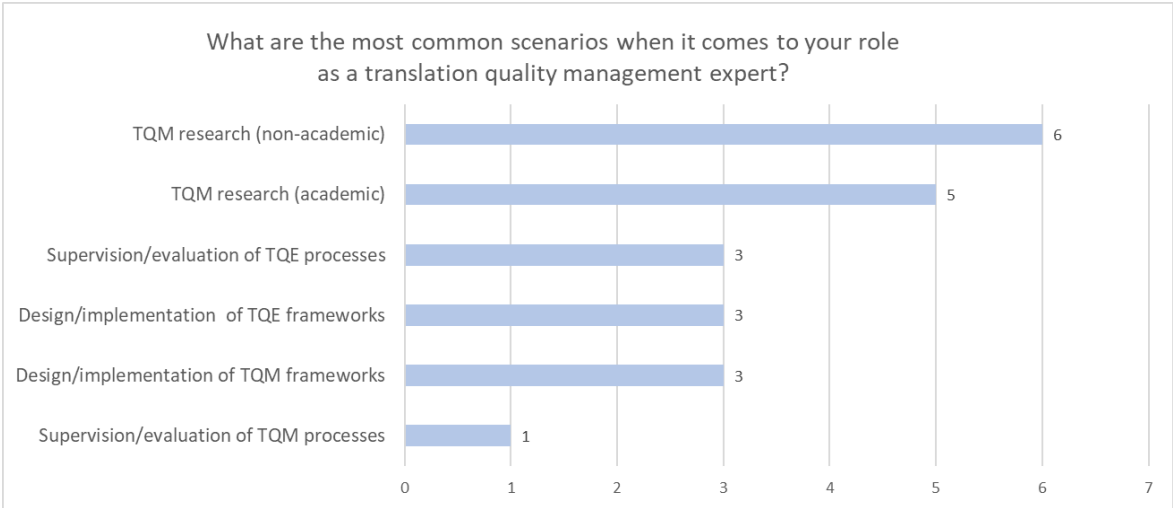
Figure 112. Professional background – Experts in translation quality management.

To the question of what were the main tasks they performed as experts in translation quality management, 46% of the participants replied that they were involved in research initiatives

143

(non-academic), while academic research was mentioned by 38% of the respondents. Other tasks reported included supervision or evaluation of translation quality evaluation processes (23%) and design or implementation of translation quality management or translation evaluation frameworks (23%).
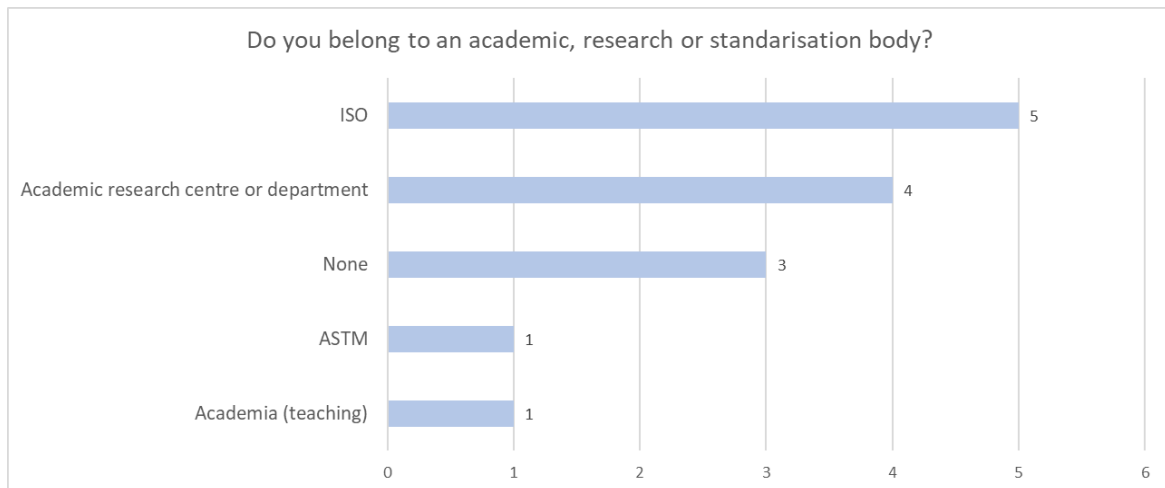
Figure 113. Professional affiliation – Experts in translation quality management.

When asked about their professional affiliations, 38% of the respondents stated that they belonged to or participated in ISO committees, and another 31% belonged to an academic research centre or department. On the other hand, 15% of the participants did not have any affiliation with any academic, research or standardisation body.

4.5.2 Translation quality management

The second part of the survey designed for the experts in translation quality management covered different aspects of the translation quality management and evaluation processes, such as norms and standards, recommendations and best practices for translation quality management and evaluation, and tools and technologies. We will see the results of each of these topics in the following sections.

4.5.2.1 Norms and standards
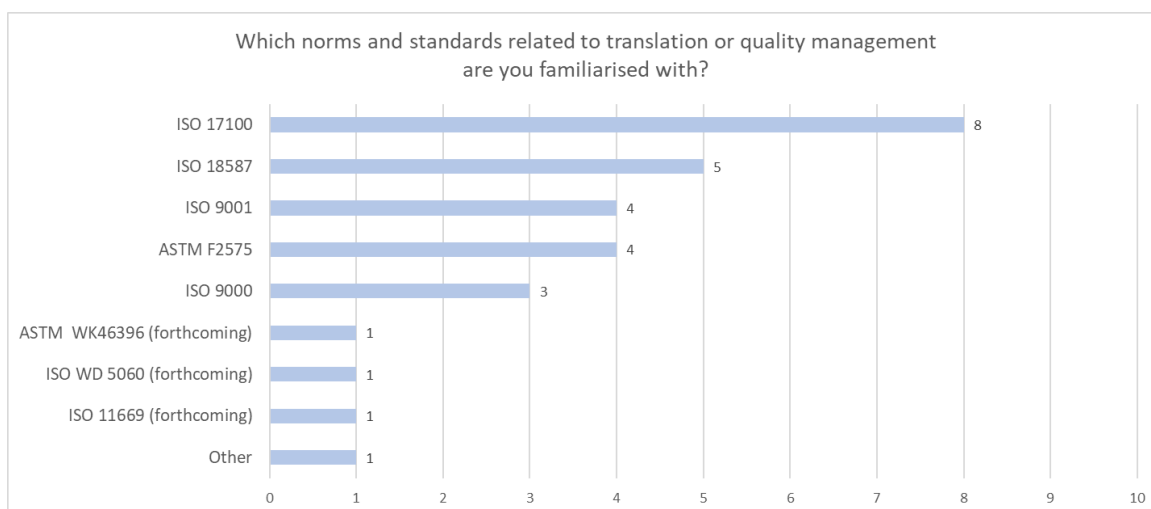
Figure 114. Knowledge of norms and standards – Experts in translation quality management.

The best-known standards were the ISO 17100 (Requirements for Translation Services), selected by 62% and ISO 18587 (Post-editing of Machine Translation Output), selected by 38% of the respondents. ISO 9001 (Quality Management Systems, Requirements) and ASTM F2575 (Standard guide for quality Assurance in Translation) seemed to be less relevant or at least less well-known, as each of them was mentioned by 31% of the participants. It is also worth noting that some of the respondents reported three additional standards that have not been published yet: ISO WD 5060, ISO 11669, and ASTM WK46396 (see section 2.4.2 for more details about these standards and guidelines).
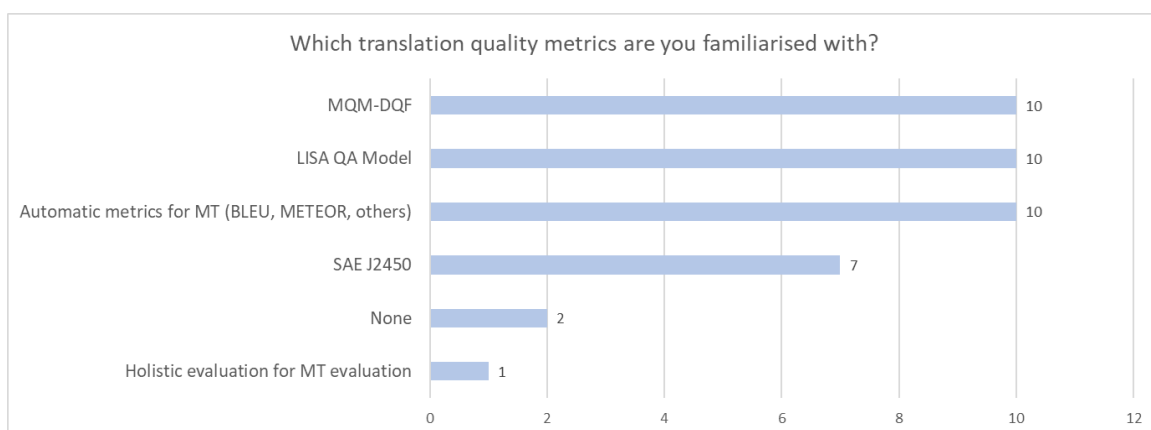
Figure 115. Knowledge of standardised metrics – Experts in translation quality management.

77% of the experts in translation quality management mentioned being familiarised with the translation quality metrics proposed by MQM-DQF, LISA QA Model, as well as different types of automatic metrics for MT evaluation such as BLEU, METEOR, and BLEURT. In

contrast, 15% of the participants stated that they were not familiar with any standardised quality metric.

4.5.2.2 Recommendations and best practices

This subsection of the research covered translation quality management and evaluation recommendations and best practices, and it was divided into three different topics: quality management, translation quality evaluation and human resources.

Translation quality management
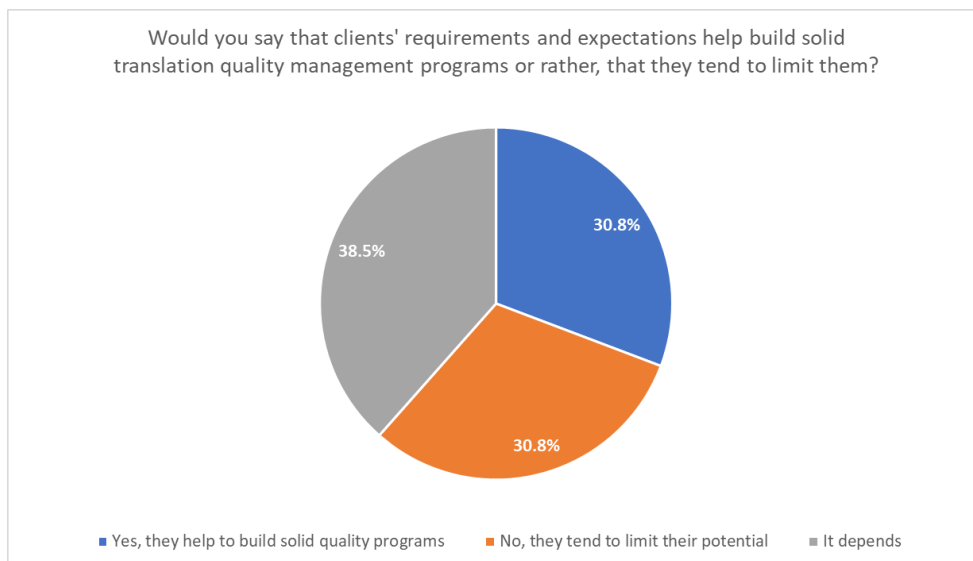
Figure 116. Clients' requirements and expectations – Experts in translation quality management.

The next question of this part of the research aimed to get deeper insights on the influence (positive, negative, or neutral) of the clients' requirements and expectations when designing a translation quality management program tailored to them. This was an open-ended question, and the replies were grouped into three categories, each of them with a very similar percentage. These are the main reasons cited by the participants to support the three different views:

Clients' requirements and expectations help build solid quality programs:

> "I think the clients' requirements shall be part of the criteria to measure quality." [Questionnaire respondent]

146

"Of course, the expectations and requirements of our customers also help a lot to improve our work as we are in the process of evolving and taking what we do more seriously." [Questionnaire respondent]

"Usually they help but deeper involvement from the translation teams would ideally be added at the beginning of the process." [Questionnaire respondent]

Clients' requirements and expectations tend to limit the development of solid quality programs:

"They often risk to limit QM programs to clients' priorities and budget to the detriment of translation priorities." [Questionnaire respondent]

"I think there is often too much focus on just error metrics, rather than using the methodology as part of a sound quality development programme." [Questionnaire respondent]

"Most clients carry on vices from previous vendors and workflows, and they want to follow the same processes, so it takes some education to implement something optimal. […] It is a challenge to be able to adapt to the constraints." [Questionnaire respondent]

The outcome varies depending on the type of client:

"It depends on the type of the client. If clients have internal localization teams, they help us improve ours or we are just expected to use theirs, but if they are companies with employees who don't have much experience in getting translation services, they are generally prone to not understanding why we have some workflows and have to do quality checks." [Questionnaire respondent]

"2 types of clients: Low maturity and high maturity. High maturity: help. Low maturity: limit (create artificial barriers). For LSPs (they have customers and internal teams, no single maturity level, depends on their clients)." [Questionnaire respondent]
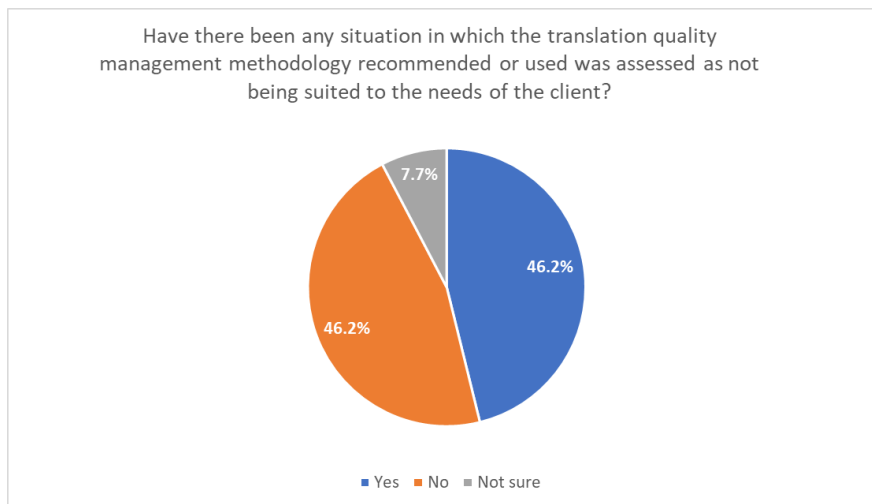
Figure 117. Translation quality management best practices – Experts in translation quality management.

When the participants were asked whether there had been any situation in which the translation quality management methodology recommended or used by them was considered as not suitable for the needs of the client, almost half the respondents (46.2%) replied they had encountered those situations. They were also requested to share some examples in a free-text form, which are presented below:

Example 1

"Yes, the client believed some categories or parameters were not relevant for them. The client re-evaluated what they wanted and they had full control of the methodology." [Questionnaire respondent]

Example 2

"Yes. We were using a "naive" approach on a new component with projects what were defined while they were created. We relied in [sic] previous translation quality assurance systems, but this project needed to be review [sic] in flight, not post-mortem, when it is too late. Their product was very established, but there was no process […]. translation quality assurance needs to be reviewed more often than established processes." [Questionnaire respondent]

Translation quality evaluation
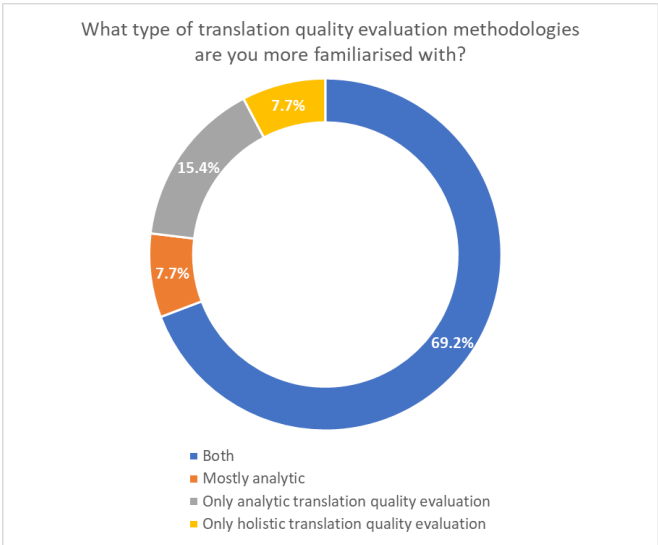
Figure 118. Types of translation quality evaluation methodologies – Experts in translation quality management.

The majority of the experts in translation quality management (69.2%) stated that they were familiar with analytic and holistic translation quality evaluation methods, and 15.4% shared they specialised in analytic evaluation methodologies, while 7.7% specialised only in holistic evaluation methodologies.
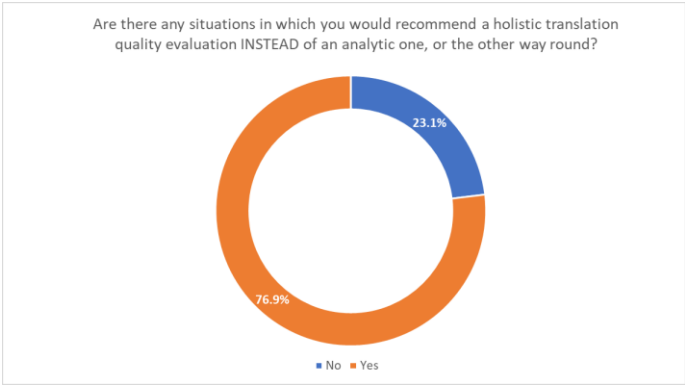
Figure 119. Use of holistic instead of analytic translation quality evaluation – Experts in translation quality management.

The majority of the experts in translation quality management (76.9) also agreed to recommend the use of holistic or anality evaluation methodologies depending on factors such as the type of text, the client's requirements, or the characteristics of the project. This question allowed the respondents to provide some specific examples, which are shown in the figure below.

Holistic methodology recommended	Analytic methodology recommended
Marketing texts, creative translation, transcreation	Translation of controlled languages (medical, technical, software)
Mid-term reviews (in flight)	Scheduled evaluations or audits
Visual checks (in app review, localisation implementation)	Translations and references available in CAT/TMS or evaluation tools
The goal is to get data quickly and cheap	The goal is to get actionable, deeper and granular insights
Evaluators have no training in linguistic review or evaluation	Trained evaluators with calibration methods in place
The goal is to get results on usability and user experience	The goal is to scrutinise the performance of a single translator/vendor

Figure 120. Use of holistic versus analytic translation quality evaluation – Experts in translation quality management.
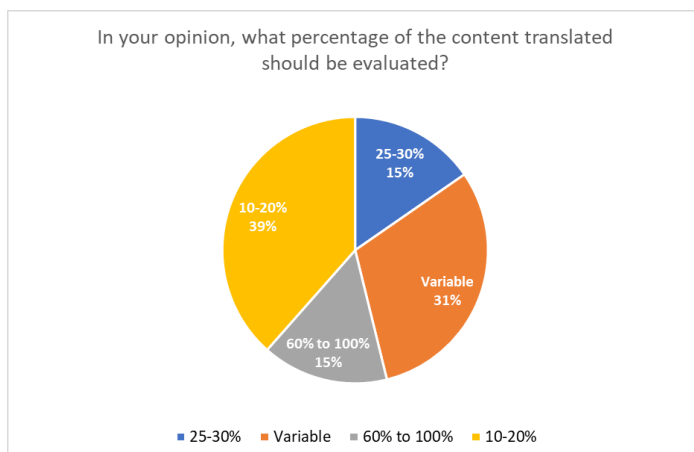
Figure 121. Percentage of content that should be included in the quality evaluation process – Experts in translation quality management.

While 39% of the experts in translation quality management considered that an evaluation of between 10% and 20% of the translated content would be advisable, 15% of them thought the percentage should be around 25-30%, and another 15% recommended increasing this percentage to 60% or more. On the other hand, 31% of the participants declared that it was not possible to recommend a generic percentage, as different factors should be taken into consideration. The main factors mentioned by the experts in translation quality management were:

- **Visibility or priority of the project:** Projects with higher visibility or priority should include a higher percentage of the translation to be evaluated.

- **New projects or vendors:** New projects or translation vendors should be evaluated more extensively until there is enough historical data and satisfactory quality results to decrease the percentage of the content evaluated.

- **Historical data:** The evaluation of projects or translation vendors with enough historical data and consistently good quality results tend to include a lower percentage of translated content. However, if there are quality concerns or a lack of maturity in translation processes that can increase the quality risks, the recommendation is to increase the percentage of the content evaluated.
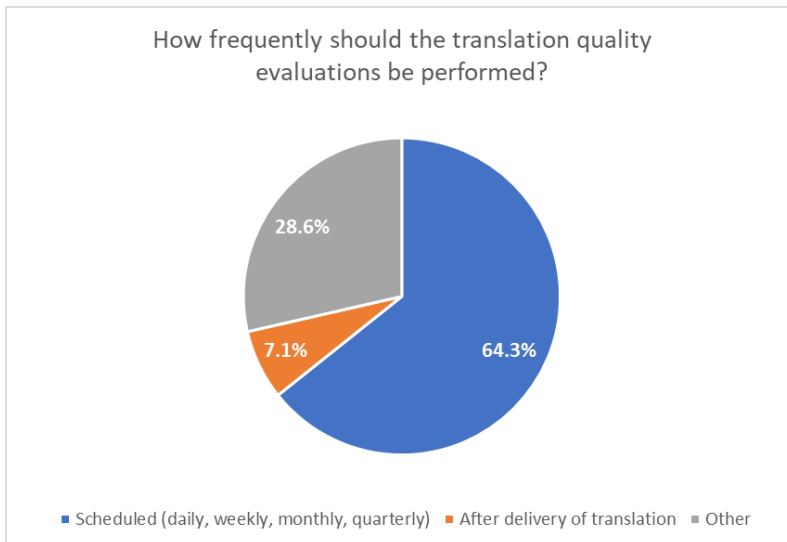
Figure 122. Frequency of translation quality evaluations – Experts in translation quality management.

More than half of the experts in translation quality management (64.3%) considered that translation quality evaluations should be performed with a scheduled frequency (daily, weekly, monthly, or quarterly), while 28.6% mentioned that the frequency could vary greatly depending on factors such as volumes, or type of content.
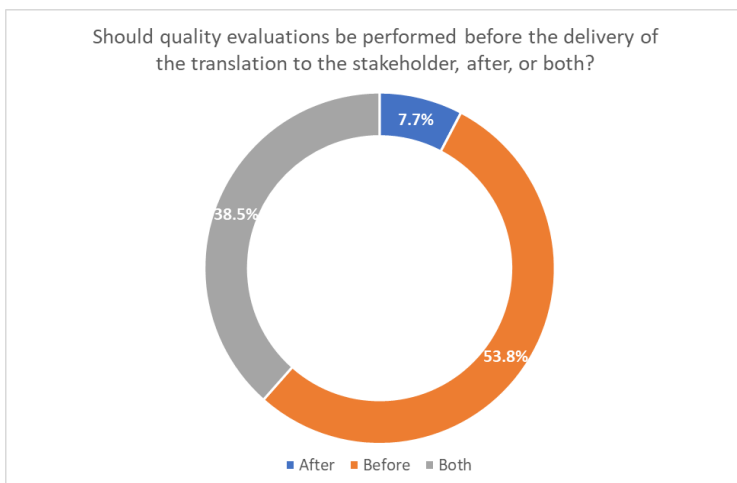
Figure 123. Timing of translation quality evaluations – Experts in translation quality management.

53.8% of the participants stated that quality evaluations should be performed, ideally, before the delivery of the translation to the stakeholder or client, while 38.5% considered that evaluations should take place before and after the delivery. This way, high risk or high priority portions of the translation project could be assessed before the delivery to detect and correct quality issues, while the rest of the translation project can be evaluated after the delivery to the client.

151

Advantages of using sampling techniques
Works better for bigger projects with low quality risks, as it provides more statistical data
Allows to spot recurrent errors and trends
More cost-effective (cheaper and faster)
It is useful for evaluations and audits performed after the projects have been delivered
It is recommended when there are time limitations, to detect mistakes in ongoing projects
It is recommended when the goal is to ensure the consistency of quality on the whole, rather than to ensure that a single project meets the quality criteria

Figure 124. Advantages of using sampling techniques in translation quality evaluation – Experts in translation quality management.

When asked whether they recommended the use of sampling techniques for translation quality evaluation, the majority of the experts in translation quality management (92%) confirmed that this method was helpful in certain contexts and provided some additional information about the advantages and recommended use of sampling techniques in a free-text form. These insights were compiled and summarised in Figure 124 above. According to the participants, sampling techniques are better suited for big projects with low quality risks, when there are cost or time limitations, or when the goal of the quality evaluation is to measure the overall consistency of the quality provided in a specific product or context.
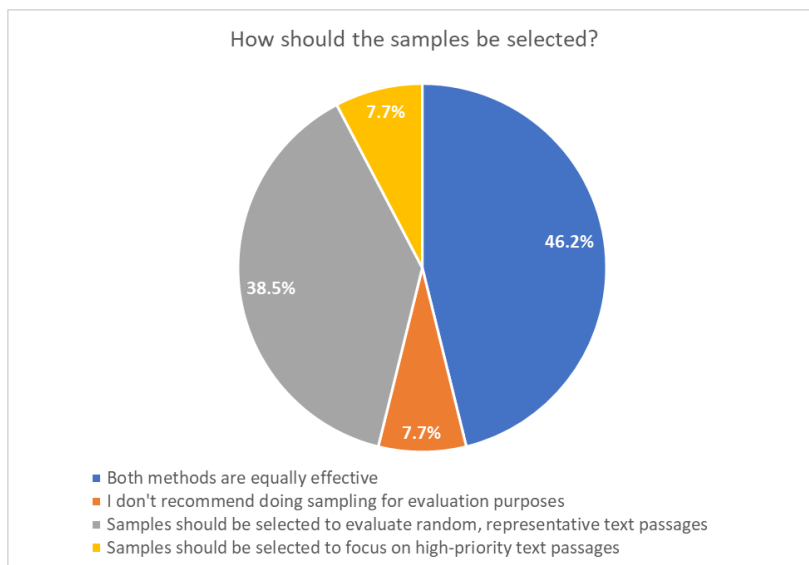
Figure 125. Sampling methods in translation quality evaluation – Experts in translation quality management.

In terms of evaluation sampling techniques, 38.5% of the participants preferred the use of random samples, while only 7.7% favoured the selection of high-priority text passages. On the other hand, 46.2% of the participants declared that, in their experience, both methods were equally effective.

152
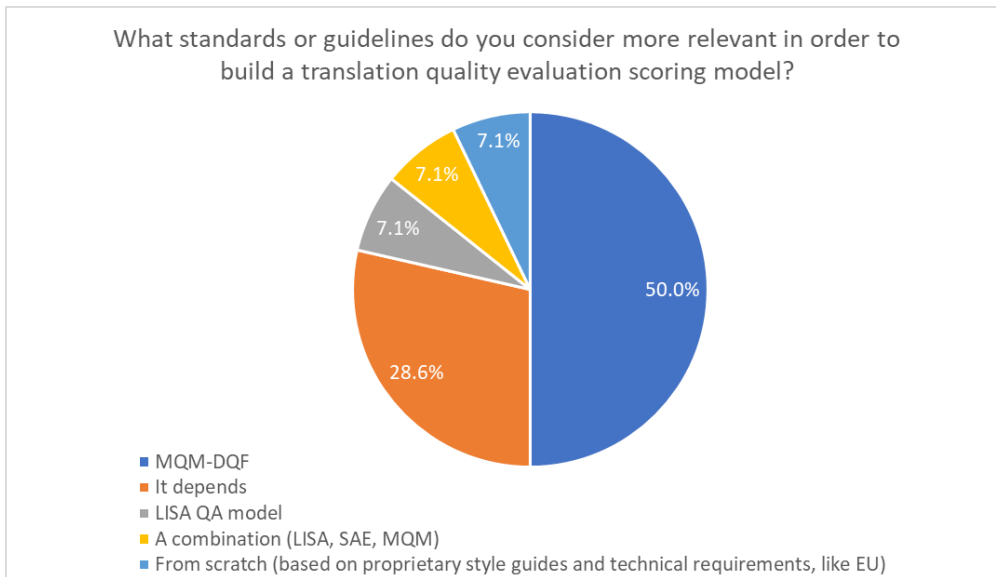
Figure 126. Guidelines for translation quality evaluation models – Experts in translation quality management.

50% of the experts in translation quality management recommended MQM-DQF as one of the most valuable guidelines to build a translation quality evaluation model. However, 28.6% also considered that different types of guidelines and considerations should be taken into account depending on the purpose of the translation quality evaluation scoring model.
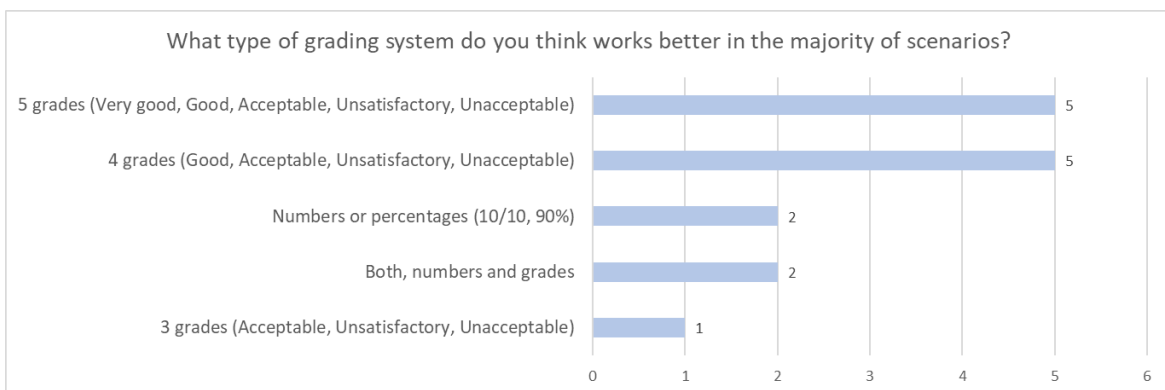
Figure 127. Grading systems recommended in translation quality evaluation – Experts in translation quality management.

The majority of the participants recommended using either 5 or 4 grades in the evaluation grading system, and many of them mentioned that, for analytic evaluation, these grades should be combined with the numbers or percentages obtained in the scoring system.

Do you think it makes sense to apply different evaluation metrics (error types, severity levels, severity multipliers) to different text types?

13 responses

Yes
No

7.7%
92.3%
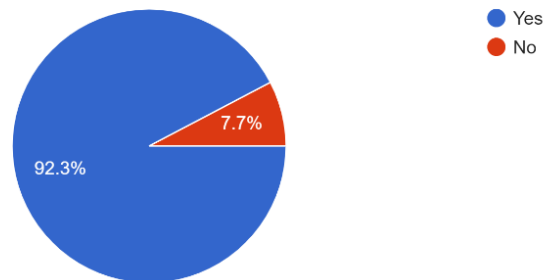
Figure 128. Uniformity of evaluation metrics in translation quality evaluation – Experts in translation quality management.

The majority of the experts in translation quality management (92.3%) also stated that evaluation metrics such as error types, severity levels and severity multipliers should be adapted to the type of text being evaluated.

Should severity multipliers be the same for all error types (uniform severity multiplier distributions), or defined for individual error types (nonuniform severity multiplier distributions)?

23.1%

■ The severity multipliers should be the same for all error types

■ The severity multipliers should NOT be the same for all error types

76.9%
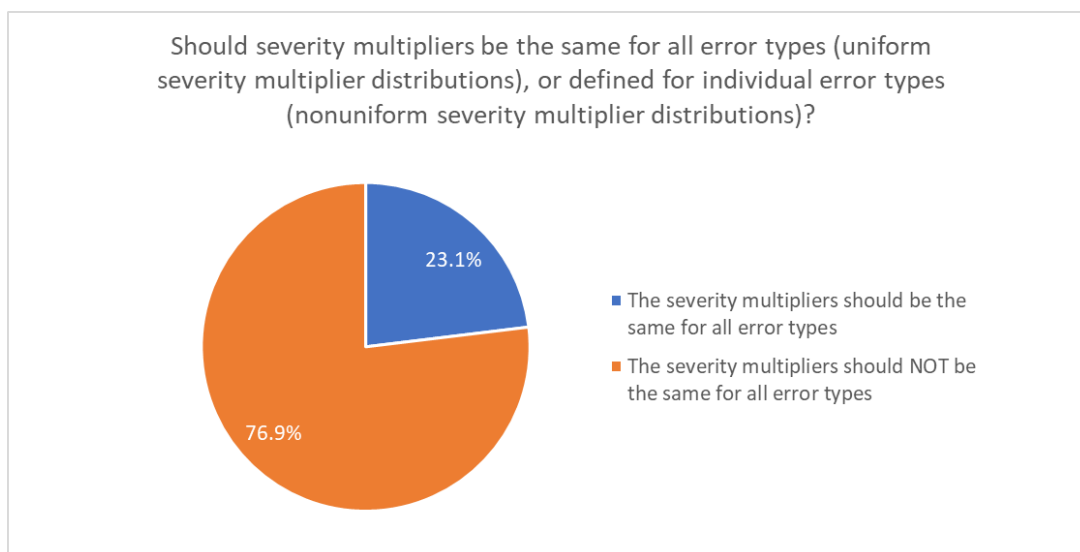
Figure 129. Uniformity of severity multipliers in translation quality evaluation – Experts in translation quality management.

The majority of the respondents (76.9%) also agreed that severity multipliers should not be the same for all error types (nonuniform severity multiplier distribution).

154
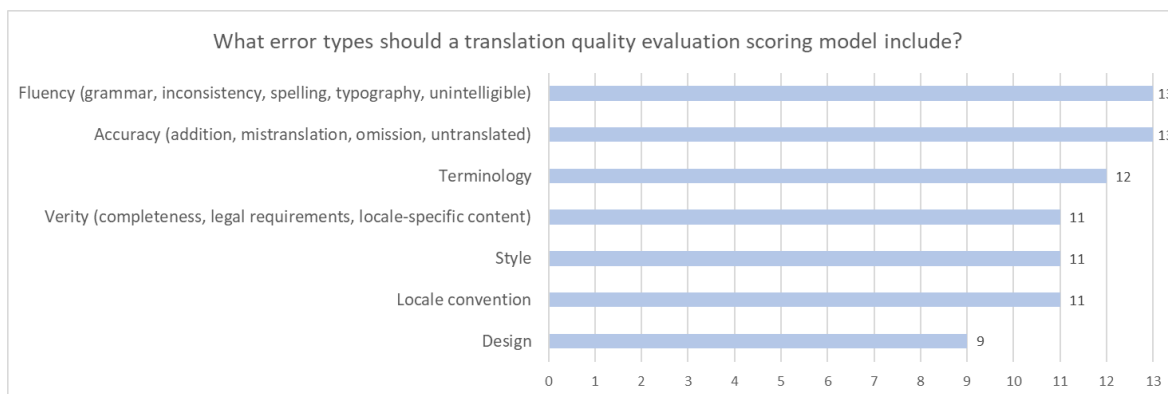
Figure 130. Error typology in translation quality evaluation – Experts in translation quality management.

All the experts in translation quality management agreed to choose Fluency and Accuracy as the main error types that should be included in a quality evaluation scoring model, with Terminology in the third place (92.3%), followed by Verity, Style, and Locale convention (84.6%) and Design (69.2%) in the last place.
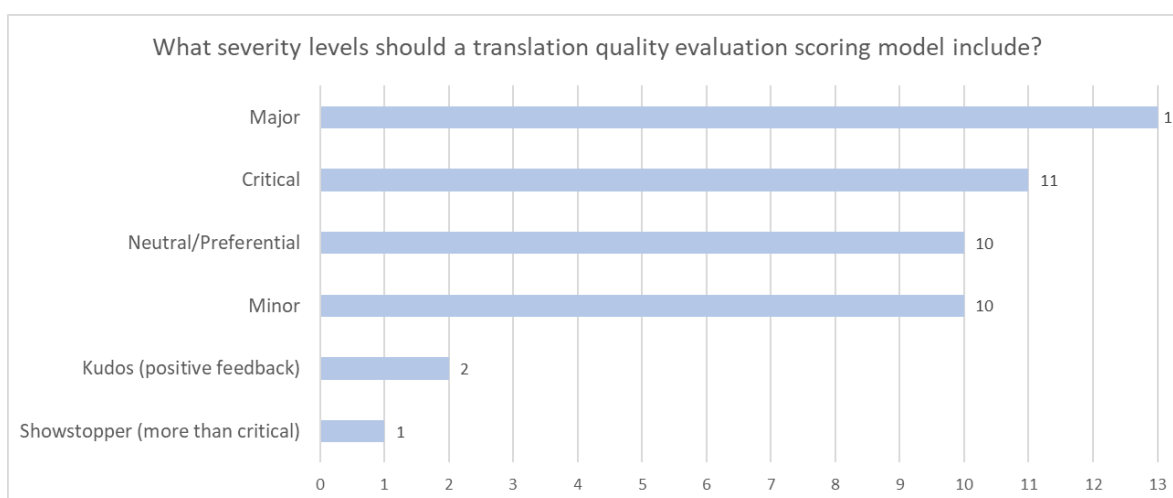
Figure 131. Error severity in translation quality evaluation – Experts in translation quality management.

Similarly to the results observed in the previous question, all the respondents followed the same nomenclature proposed by the LISA QA Model or MQM-DQF. However, only the category "Major" was selected by all of them, "Critical" was selected by 11 out of the 13 participants, and Neutral/Preferential and Minor by 10 of the experts in translation quality management. In addition, some of them also mentioned two additional categories: "Kudos" (to share positive feedback on particularly good translation solutions), and "Showstopper", which could be used in those cases in which the error detected was so severe that it required some immediate action.

155
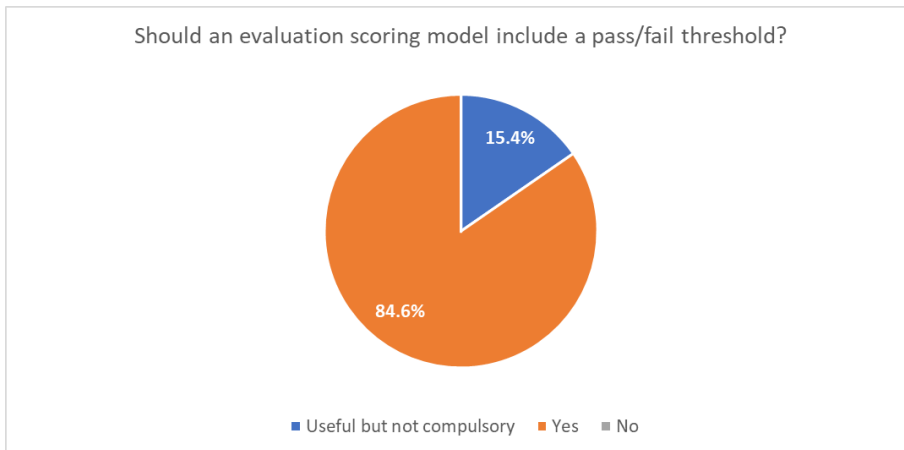
Figure 132. Use of Pass/Fail thresholds in translation quality evaluation – Experts in translation quality management.

Although 84.6% of the participants considered that an evaluation scoring model should include a pass/fail threshold, 15.4% also stated that this threshold was useful but did not need to be compulsory.
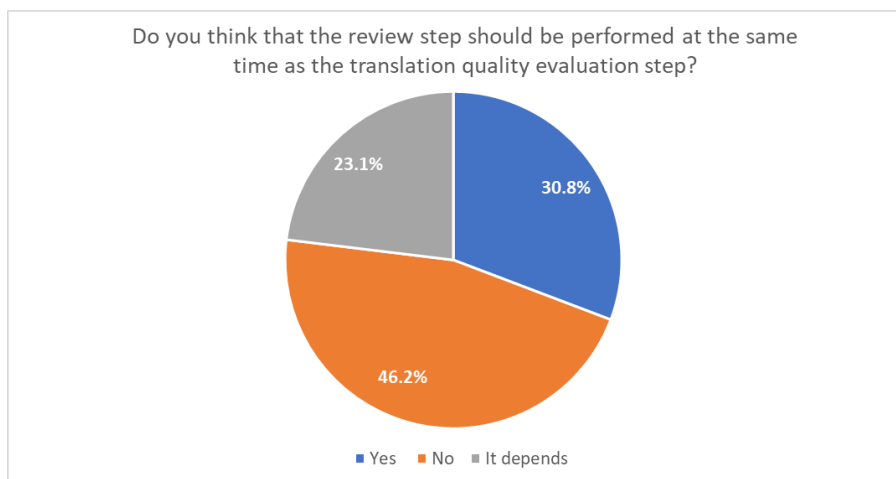
Figure 133. Review step and evaluation step in translation quality methodologies – Experts in translation quality management.

When asked whether they recommended performing the review step and the evaluation step simultaneously, the experts in translation quality management expressed three differentiated points of view. 46.2% said that each step was part of a different process and had a different goal, so they should not be performed simultaneously. 30.8% of the participants did think that both steps could be performed in one go, as it was more cost-efficient for many clients. The remaining 23.1% stated that there were cases in which it was possible or needed to combine review and evaluation in one step, for example, if there is not enough time to perform a thorough review.

156

Human resources and qualifications

What kind of resources do you think should be involved in the translation quality management process?

13 responses
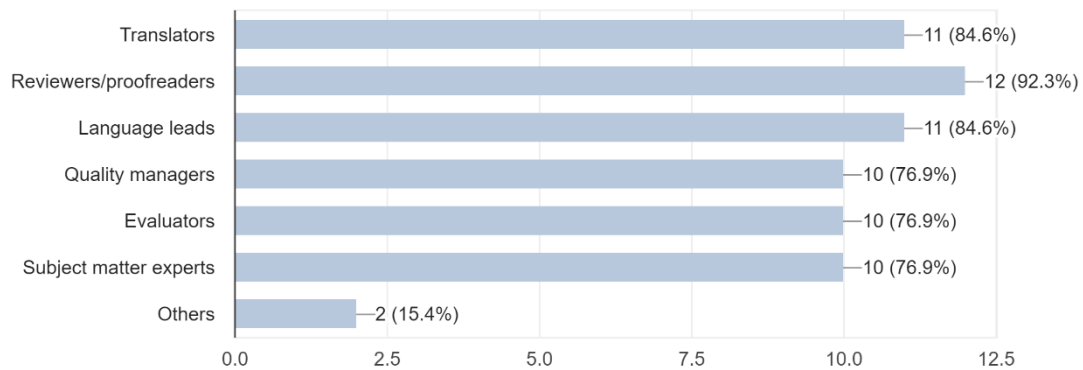
Figure 134. Human resources involved in translation quality management – Experts in translation quality management.

The type of human resources that should be involved in translation quality management processes include reviewers/proofreaders (according to 92.3% of the respondents), translators (84.6%), and language leads (84.6%). Quality managers, evaluators and subject matter experts were also mentioned by 76.9% of the participants.

Scenarios	Translation	Quality assurance 1	Quality assurance 2	Evaluation 1	Evaluation 2	Evaluation 3
Scenario 1	Translators	Reviewers/proofreaders		Evaluators	SMEs	
Scenario 2	Translators	Reviewers/proofreaders		Language leads Quality managers	SMEs	
Scenario 3	Translators	Reviewers/proofreaders		Language leads	Quality managers SMEs	
Scenario 4	Translators	Reviewers	Proofreaders	Evaluators		
Scenario 5	Translators	Reviewers	Proofreaders	Language Leads Evaluators	Quality Managers	SMEs

Figure 135. Most common scenarios in translation quality management and evaluation workflows – Experts in translation quality management.

In the following question, the participants were asked to give more details regarding the most common scenarios in translation quality management and evaluation workflows to understand better the number of steps and types of resources involved in each step. The question had an open-ended format, and the results were coded and grouped. As a result, five main scenarios were identified, with various degrees of complexity that went from two compulsory steps:

- Scenario 1 (2 compulsory steps): one quality assurance step performed by the same person, who reviews and proofreads and the same time; one quality evaluation step performed by an evaluator; and one optional quality evaluation step, performed by a subject matter expert.

- Scenario 2 (2 compulsory steps): one quality assurance step performed by the same person, who reviews and proofreads and the same time; one quality evaluation step performed by a language lead or quality manager, who acts as an evaluator; and one optional quality evaluation step, performed by a subject matter expert.

- Scenario 3 (2 compulsory steps): one quality assurance step performed by the same person, who reviews and proofreads and the same time; one quality evaluation step performed by a language lead, who acts as an evaluator; and one optional quality evaluation step, performed by a quality manager or a subject matter expert.

- Scenario 4 (3 compulsory steps): one quality assurance step performed by a reviewer, a second quality assurance step performed by a proofreader, and one quality evaluation step performed by an evaluator.

- Scenario 5 (4 compulsory steps): one quality assurance step performed by a reviewer, a second quality assurance step performed by a proofreader, one quality evaluation step performed by a language lead or evaluator, a second evaluation step performed by a quality manager, and a third quality evaluation performed by a subject matter expert.
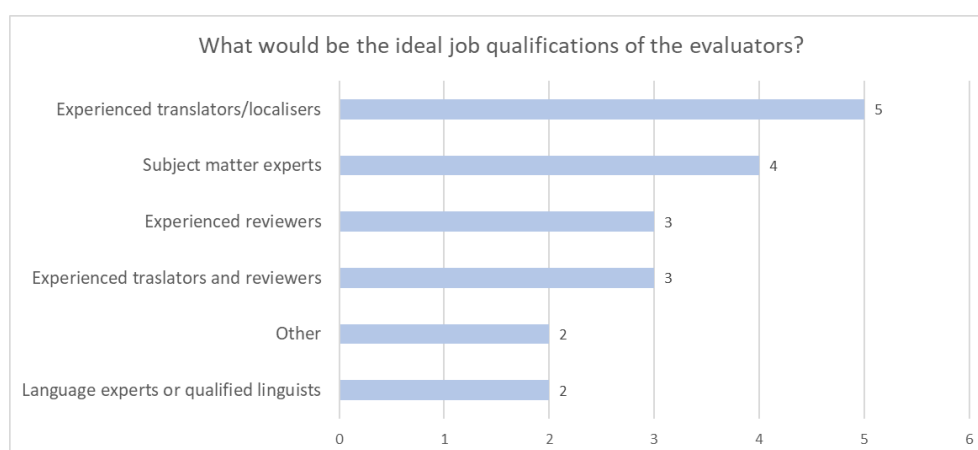
Figure 136. Skills and qualifications of evaluators – Experts in translation quality management.

38% of the participants considered that evaluators should have experience in translation or localisation, while 31% mentioned the importance of having specific subject matter

knowledge. Another qualification mentioned by 31% of the participants was experience as reviewers or combined experiences as reviewers and translators.
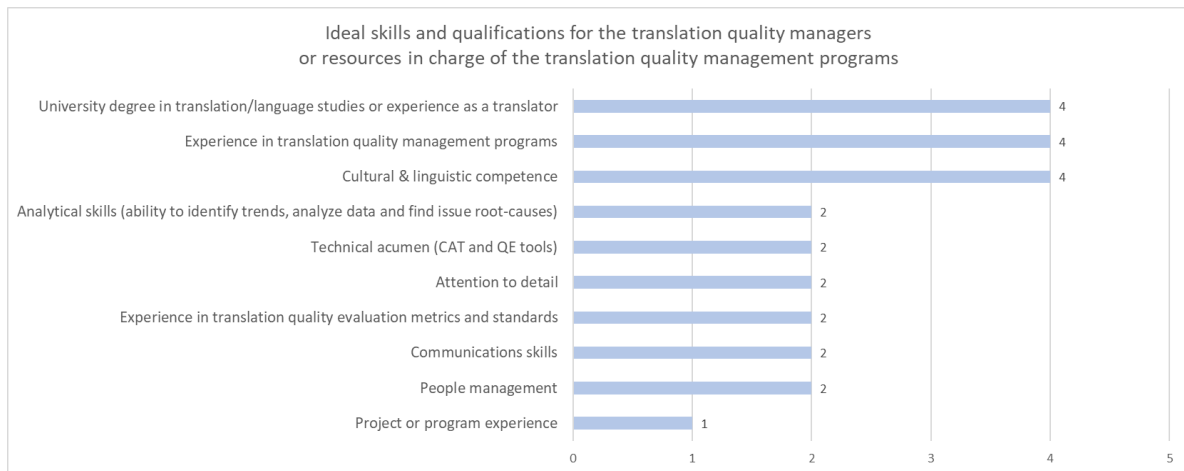
Figure 137. Skills and qualifications of translation quality managers – Experts in translation quality management.

In the following question, the experts in translation quality management were asked to share the ideal qualifications and skills of translation quality managers or other human resources in charge of translation programs. This question had an open-ended format, and the replies were coded and grouped into main themes that resulted in the 10 top qualifications shown in Figure 138. The majority of the participants highlighted the importance of providing at least two different types of training programs: generic training sessions on translation quality management metrics, error categorisation, and standards, and a specific hands-on training with internal documentation and examples of categorisation of errors.
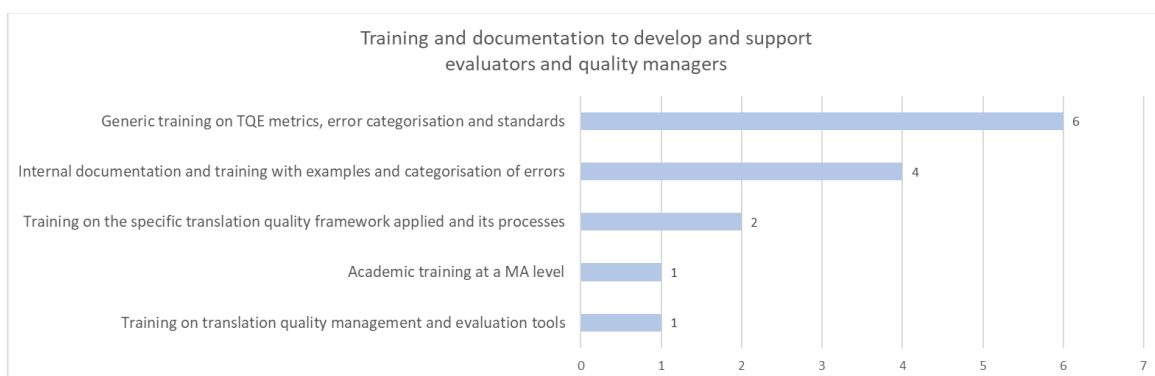
Figure 138. Training and documentation for human resources involved in translation quality management – Experts in translation quality management.

The last question of this part of the study was what training and documentation would help develop and support evaluators and quality managers in their work. This question also had

an open-ended format. Some of the most recurrent recommendations were related to general training on different translation quality methodologies and standards and internal training with example-based documentation on how error evaluation is conducted in the organisation.

4.5.2.3 Tools and technologies
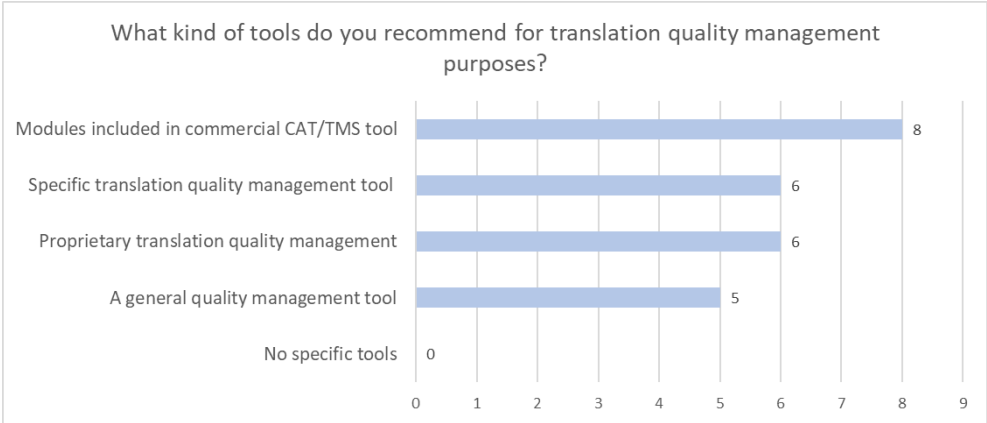
Figure 139. Tools recommended for translation quality management purposes – Experts in translation quality management.

62% of the experts in translation quality management considered that the best tools for translation quality management purposes were the quality features available in CAT tools or TMS environments such as SDL Trados or MemoQ. In comparison, 46% recommended using either proprietary translation quality management tools developed by the clients or translation service providers or specific translation quality tools such as ContentQuo, QA Distiller, Verifika, Xbench, or TQAuditor.
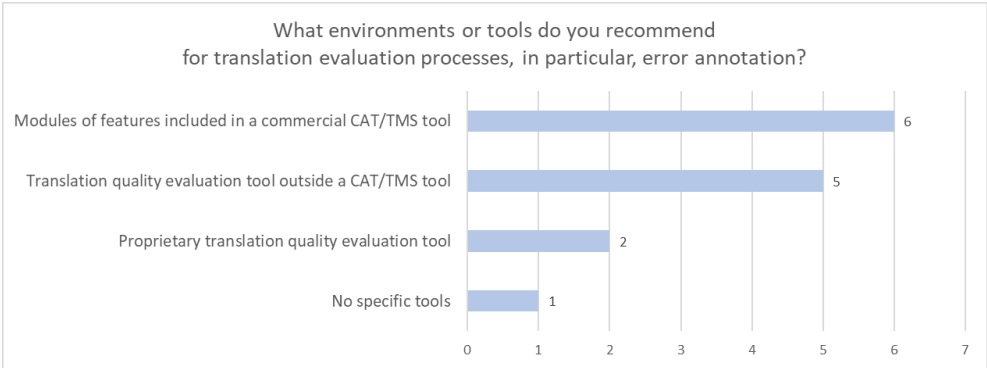
Figure 140. Tools recommended for translation quality management purposes – Experts in translation quality management.

Similarly to what was reported in the previous question, 48% of the experts in translation quality management also recommended the use of modules or features included in commercial CAT tools or TMS to perform error annotation, and a 38% suggested using the

same translation quality tools mentioned above for translation evaluation and error annotation.

4.5.3 Current challenges in translation quality management and evaluation

In this part of the survey designed for experts in translation quality management, the participants were asked to rate, on a scale of 1 to 5, the level of maturity of the current translation management frameworks used in the industry according to different paraments, and to share the main pain points on translation quality management and evaluation they have observed.
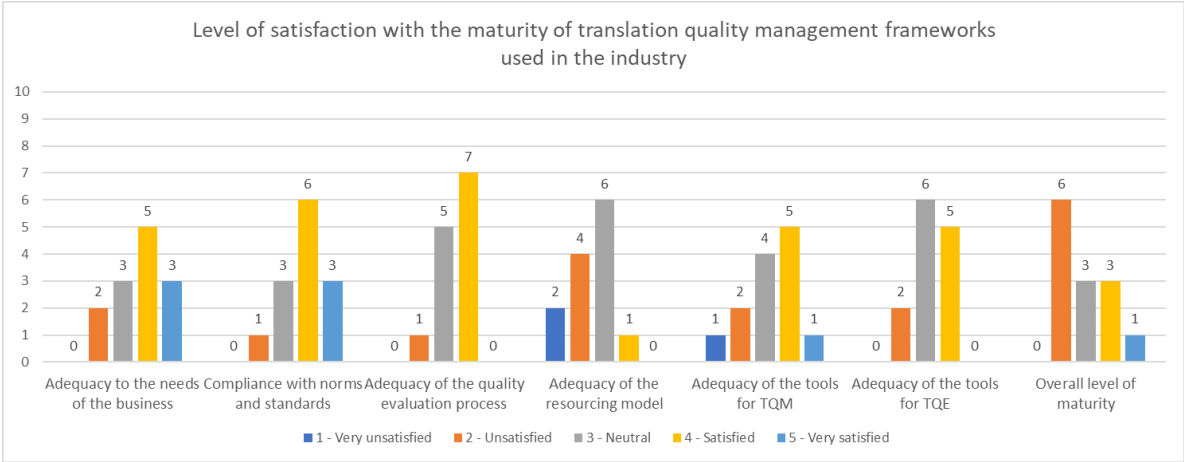
Figure 141. Level of satisfaction with the maturity of TQM frameworks used in the industry – Experts in translation quality management.

The two aspects with a higher level of satisfaction reported by the experts in translation quality management were the adequacy to the needs of the business and the level of compliance with norms and standards. On the other hand, the two categories with the lowest level of satisfaction were the adequacy of the resourcing model and the overall level of maturity of the frameworks.

Main pain points in translation quality management and evaluation	Nr of replies
TQM: Lack of centralised data or dashboards with visibility on trends	2
TQM: Financial limitations to invest in tools or dedicated resources	5
TQM: Lack of cost-effective, scalable and flexible TQM methodologies	5
TQM: Lack of training and knowledge	3
TQM: Lack of access to practical information on TQM methodologies	1
Clients: Unrealistic, inconsistent or unclear quality expectations	1
Clients: Aligning and educating clients on quality management best practices	3
Clients: Client-based priorities over translation priorities	1
Quality improvement: Time constrains	4
Tools: Lack of advanced tools for TQM and TQE	4
TQE: Appropriate training and calibration among different reviewers	4
TQE: High costs, time-consuming processes and rigidity of the evaluation model	4
TQE: Use of unstructured holistic models without objective grading	1

Figure 142. Main pain points in translation quality management and evaluation – Experts in translation quality management.

The second question of this part of the survey had an open-ended format, and the replies were coded and grouped into five main topics: challenges in translation quality management, challenges related to the collaboration with clients, challenges related to quality improvement, challenges related to the tools used, and specific challenges in translation quality evaluation. The main findings are summarised in Figure 142.

4.5.4 Solutions

In this part of the research, experts in translation quality management were asked to share their insights as to the potential solutions to try to solve the pain points they had mentioned in the previous questions. The replies to this open-ended question were coded and added to the table with the compilation of pain points to show both sets of information side to side. As shown in the figure below, the main suggestions of this group of participants revolve around three main topics:

- Development of more advanced tools a better integration solutions.
- Use of automated processes and technologies.
- Training and hiring efforts to attract and develop evaluators and quality managers.

Main pain points in translation quality management and evaluation	Suggested solutions
TQM: Lack of centralised data or dashboards with visibility on trends	Develop APIs and connections
TQM: Financial limitations to invest in tools or dedicated resources	Educating clients to better use of MT to reduce costs
TQM: Lack of cost-effective, scalable and flexible TQM methodologies	Develop TQM processes focused on total quality control, TQE should be just one component of it
TQM: Lack of training and knowledge	Bring in a qualified quality manager
TQM: Lack of access to practical information on TQM methodologies	Standards: Have a more flexible approach that is easier to apply
Clients: Unrealistic, inconsistent or unclear quality expectations	Educate clients to be more flexible in their expectations and tools used
Clients: Aligning and educating clients on quality management best practices	Educate clients to be more flexible in their expectations and tools used
Clients: Client-based priorities over translation priorities	TSP should be the gatekeepers of quality, they should be trusted by clients
Quality improvement: Time constrains	Automated quality evaluation platform or computer assisted evaluation
Tools: Lack of advanced tools for TQM and TQE	Develop APIs, connections, propose other tools
	Move quality evaluation from external to internal in CAT tools
TQE: Appropriate training and calibration among different reviewers	National regulations and investments that allow companies to train and hire enough competent evaluators
	Closer collaboration and partnership with reviewers to improve calibration
TQE: High costs, time-consuming processes and rigidity of the evaluation mode	Automated quality evaluation platform or computer assisted evaluation
	Mid-term review of the TQE process
TQE: Use of unstructured holistic models without objective grading	Bring in a qualified quality manager

Figure 143. Translation quality management and evaluation pain points and recommended solutions – Experts in translation quality management (see Appendix E).

4.5.5 The future of translation quality management

The goal of the last section of this version of the survey was to gather the insights of the experts in translation quality management regarding the future challenges they believe would impact the translation quality management and evaluation frameworks and what kind of initiatives would be helpful for the advancement of this discipline in the translation industry. The first question had an open-ended format, and the replies were coded and grouped to show the most recurrent topics mentioned by the participants (see Figure 144 below). Similarly to the information provided by the buyers of translation services and the translation service providers, the experts on translation quality considered that some of the biggest challenges in translation quality management will be the need to increase the capacity, flexibility and scalability of the quality management programs to cover bigger languages, more languages, and faster, the rise of new technologies (such as Machine Learning and AI), and the development of new translation quality management workflows and metrics.

Future challenges in translation quality management
Increase capacity and scalability to cover bigger volumes, more languages, and faster
New types of TQE metrics: user-centric, holistic, post-editing effort, automatic metrics, quality estimation
Automation and integration of tools for TQM and TQE purposes
New text types in the TQM workflows: Transcreation, MT evaluation, multimedia formats such us web streaming, voice to text
Keep up with the most recent methodologies and approaches to avoid corporate silos
Evaluation of different quality levels requested by clients
New tecnologies: Machine Learning, AI
Insufficient training and remuneration of reviewers and evaluators, labor intrusion
Increased demand of qualified evaluators and quality managers

Figure 144. Future challenges in translation quality management and evaluation – Experts in translation quality management.
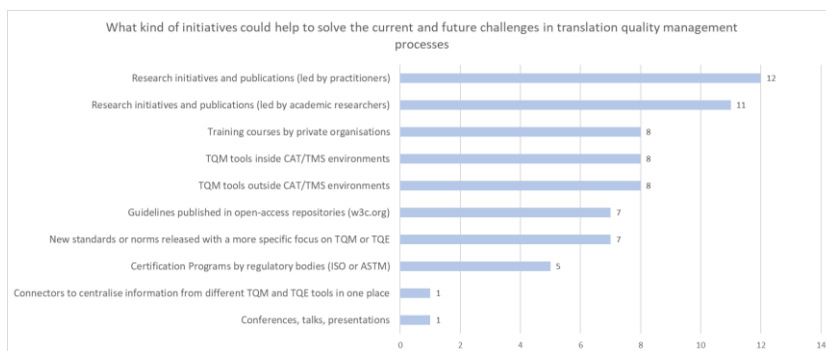
Figure 145. Useful initiatives in translation quality management and evaluation – Experts in translation quality management.

The last closed-ended question of the survey asked the experts in translation quality management to select which initiatives they considered could help the advancement of the translation quality management programs in professional environments. These are some of the main findings of this question:

- The initiatives that were mentioned more frequently by the participants were research initiatives, led either by practitioners and industry-related organisations or by universities and research centres.

- Participants also favoured training courses organised by private organisations and advanced translation quality management tools, regardless of whether they were integrated into existing CAT/TMS environments or standalone ones.

- Some of the respondents added two new initiatives that, in their opinion, were worth considering: the organisation of conferences talks and presentations around translation quality management and evaluation to facilitate knowledge sharing among different sectors of the industry, and the development of connectors that could help to centralise the information from different translation quality management and evaluation tools in one place for data analysis.

4.5.6 General comments and feedback

One of the experts in translation quality management that participated in the survey chose to fill the free text question at the end of the questionnaire to share an additional remark on this topic:

"Do localization engineers have any influence in [sic] the translation quality management processes? If so, they could help to have the tools and technologies needed to automatise translation quality management". [Questionnaire respondent]

5 Discussion and summary

The following sections of this dissertation provide a summary and discussion of the key findings (KFs) of the survey per target group (buyers of translation services, translation service providers, translation technology companies, experts in translation quality management) and theme. The themes correspond to the research questions introduced in section 1.3:

- What is the demographic and professional profile of the companies and participants that carry out translation quality management and evaluation tasks?
- What is the level of maturity in terms of translation processes and analytics?
- What is the level of maturity in terms of translation quality management processes?
    - What is the overall level of maturity in translation quality management?
    - What is the overall level of knowledge and use of translation quality management norms and standards?
    - What are the documented processes and human resources used in translation quality management programs?
        - How does the translation quality management program work?
        - What are the translation quality evaluation methodologies used?
        - How does the translation quality evaluation process work?
        - What kind of human resources are involved in translation quality management and evaluation processes?
- What are the main pain points and limitations in translation quality management and evaluation processes?
- What solutions and workarounds are applied in the industry to overcome the current pain points and limitations?
- What are the future challenges and trends in translation quality management and what kind of initiatives would help the advancement of the translation quality management practices in professional environments?

5.1 Buyers of translation services

45.6% of the participants in this research stated they worked in internal translation or localisation teams of companies that required translation services. These participants provided information about these themes:

- Profile of the buyers of translation services that carry out translation quality management and evaluation tasks
- Level of maturity in terms of translation processes and analytics
- Level of maturity in terms of translation quality management processes
    - Overall level of maturity in translation quality management
    - Overall level of knowledge and use of translation quality management norms and standards
    - Documented processes and human resources used in translation quality management programs
        - Translation quality management methodologies
        - Translation quality evaluation methodologies
        - Human resources involved
- Main tools and technologies used in translation quality management and evaluation processes
- Main pain points and limitations in translation quality management and evaluation processes
- Solutions and workarounds applied to overcome the current pain points and limitations
- Future challenges and trends in translation quality management

5.1.1 Profile of the buyers of translation services that carry out translation quality management and evaluation tasks (KF1)

The results of this research reveal that the vast majority of the companies that employ human resources that carry out translation quality management tasks tend to have a medium to big size if we consider three parameters:

- Overall staff (71.0% reported having more than 200 employees).

- Annual translation volumes (51.7% revealed having an annual translation volume of 10M words or more).

- Number of languages managed or outsourced (77.4% stated that their organisations needed translation for more than ten languages).

5.1.2 Level of maturity in terms of translation processes and analytics (KF2)

These high translation volumes, as well as the number of languages managed and outsourced by the majority of the companies that participated in the study, correlated with a high maturity level of these organisations in terms of translation and localisation processes, as 80.6% of the respondents stated that their organisations were either at level 4 (Optimised translation processes) or level 5 (Transparent translation processes).

This high level of maturity can also be seen in how the participants deal with translation analytics, although here the percentage of companies at the highest levels was not so high. The majority of the respondents (51.6%) considered that their companies were either at level 3 (Diagnostic analytics) or level 4 (Predictive analytics), and none of them thought that their companies had reached the highest level (Prescriptive analytics).

5.1.3 Level of maturity in terms of translation quality management processes

In order to get more information about the average level of maturity of the buyers of translation services regarding their translation quality management processes, the study covered four different aspects: level of maturity of the translation quality management program, level of knowledge and use of norms and standards, documented processes and human resources, and tools and technologies.

5.1.3.1 Level of maturity of the translation quality management program (KF3, KF4)

The high level of maturity in translation analytics and the predominance of levels 3 and 4 can also be seen when it comes to the maturity of the translation quality management program. The same percentage of respondents (51.6%) considered that their companies were either at level 3 (Agile quality management) or level 4 (Centralised quality management), and none of them thought that their companies had reached the highest level (Expert quality management).

5.1.3.2 Level of knowledge and use of translation quality management norms and standards (KF5, KF6)

Even though the leading translation services and quality management norms and standards are explicitly aimed to guide companies that want to build quality management systems, more than half of the translation service buyers do not follow or take into consideration any norm or standard to build their translation quality management program, either because they considered that they were not flexible or not customisable enough, or because they were not aware of the existence of these standards.

5.1.3.3 Documented processes and human resources used in translation quality management programs

5.1.3.3.1 Translation quality management methodologies (KF7)

The types of content that are more frequently included in translation quality management process of translation service buyers are marketing texts, websites, and mobile apps, as well as technical documents and external documentation, as they tend to have a bigger impact on the business and a higher priority in the translation management processes.

More than 50% of the translation service buyers include at least 80% of the translated content in their translation quality management processes, and 58% of the participants confirmed that the percentage applied is static and does not vary.

5.1.3.3.2 Translation quality evaluation methodologies (KF8, KF9, KF10, KF11)

Even though almost half of the translation service buyers use analytic and holistic translation quality evaluation methods, there is a clear preference for the analytic method over the holistic one. Participants also mentioned that the reason for using both methods is that they complement each other. The analytic method is used as a baseline, and the holistic method helps to add an additional layer of information to capture data regarding sentiment, influence of the quality in the user experience, and level of relevance and effectivity of the translation in local markets.

45% of the translation service buyers include a maximum of 50% of the translated content in the translation evaluation workflows, 54.8% of them conduct evaluations with a scheduled frequency, and 40% follow a hybrid approach in which some translations are evaluated before the delivery to the internal stakeholders, and others, after the delivery. This hybrid

approach is probably one of the causes why a very high number of translation service buyers (48.4%) use whole texts or samples to perform translation evaluations depending on the needs instead of just one of these two methods.

In terms of grading and scoring systems used, the majority of the translation service buyers use a combination of numbers or percentages that represents the score obtained in the evaluation, a threshold that defines whether the evaluated text met the required level of quality or not, and a grading system with only two grades. The evaluation metrics (error types, severity levels or severity multipliers) can vary depending on the type of text, or the quality requirements, which indicates a higher level of sophistication. However, the severity multipliers are always the same for all error types.

Given that the majority of the translation service buyers stated that their quality metrics were based on the MQM-DQF model, it is not surprising to see that their error typology has the same error categories (Accuracy, Fluency, Terminology and Style) and error levels proposed by this model (Minor, Major, Critical).

The vast majority of the participants on the buyer side reported feeling confident about the level of adequacy of the quality evaluation methodology used, although almost 50% of them were only moderately confident, and none of them recognised being extremely confident.

5.1.3.3.3 Human resources involved (KF12)

The types of roles most commonly involved in the translation quality management programs of the translation service buyers are mainly reviewers and proofreaders, followed by translators, quality managers and language leads. Only 38.7% of the participants mentioned having evaluators as a specific role, and there was no particular trend as to whether these resources were employed directly by the company or were external.

In terms of qualifications and skills, the majority of the participants stated that evaluators were required to have previous experience as translators and reviewers, while for quality managers, other skills were considered more or at least as important, for example, cultural and linguistic competence and experience in translation quality management programs.

5.1.3.4 Tools and technologies (KF13)

Almost half the translation service buyers (48%) use the modules available in commercial CAT/TMS tools such as Memsource, Smartling and XTM for translation quality

management purposes and 23% have a proprietary translation quality management tool, while 45% use these same features for quality evaluation and error annotation purposes, and 27% prefer other tools such as ContentQuo or Excel.

5.1.4 Current challenges in translation quality management (KF14)

Despite (or maybe because of) their reliance on CAT/TMS tools for translation quality management and evaluation, translation service buyers pointed out that the two aspects of their translation quality management that they were less satisfied with were precisely the adequacy of the tools used for translation quality management and the tools used for translation quality evaluation. This can be corroborated with the pain points shared by the participants, as they included topics such as the lack of automation and the lack of centralised data or dashboards with information about quality. Other common pain points mentioned by the participants were the lack of dedicated human resources, the difficulty to establish communication loops and feedback channels for quality improvement, and the lack of clear, scalable, and customisable evaluation metrics and thresholds that can cover both analytic and holistic methodologies.

5.1.5 Solutions and workarounds applied (KF15)

The solutions to the challenges currently faced by the translation service buyers include the migration to new CAT/TMS or the use of internal localisation engineering teams to design connectors to automatise and centralise quality data, the collaboration with other teams to have access to content audits and other repositories of holistic quality data provided by internal subject matter experts or users, and the partnership with translation service or translation technology vendors to learn from their expertise in translation quality management.

5.1.6 Future challenges and trends in translation quality management (KF16, KF17)

In the opinion of the participants from the translation service buyers, some of the future trends will be the increase in volumes, languages and text types covered by translation quality programs, the addition of new types of translation quality management workflows and translation quality metrics that cater for a more holistic, scalable, and agile approach to translation quality, and the increase in the use of tools and technologies that can automatise processes and provide a greater integration.

Finally, the translation service buyers considered that the most valuable initiatives to contribute to the advance of their translation quality management programs were the development of more advanced translation quality management features inside the current CAT/TMS tools already used, and research initiatives on translation quality management practices led by practitioners and industry-related organisations.

5.2 Translation service providers

29.4% of the total of the participants in this research worked for translation service providers. The participants provided information about these themes:

- Profile of the translation service providers
- Level of maturity in terms of translation processes and analytics
- Level of maturity in terms of translation quality management processes
    - Overall level of maturity in translation quality management
    - Overall level of knowledge and use of translation quality management norms and standards
    - Documented processes and human resources used in translation quality management programs
        - Translation quality management methodologies
        - Translation quality evaluation methodologies
        - Human resources involved
- Main tools and technologies used in translation quality management and evaluation processes
- Main pain points and limitations in translation quality management and evaluation processes
- Solutions and workarounds applied to overcome the current pain points and limitations
- Future challenges and trends in translation quality management

5.2.1 Profile of the translation service providers (KF18)

Unlike the average size of the translation service buyers, which tend to have more than 200 employees, the size of the staff of the translation service providers was much more varied,

as 45% of them have 25 employees, 30% have from 26 to up to 100 employees, and 35% have more than 100 employees.

This distribution correlates with the annual volumes reported by these companies, given that 30% have an annual translation volume of over 100M words, 40% translate between 500k and 100M words per year, and 20% translate less than 500k words per year. Furthermore, these three balanced tiers in terms of size of the staff and annual translation volumes can also be seen in the number of languages translated by these companies, as 30% provide translation services for three languages or less, 30% work with more than ten languages, and 40% offer translation services in 4 to 10 languages.

5.2.2 Level of maturity in terms of translation processes and analytics (KF19)

Similarly to the results observed in the translation service buyers, the majority of the respondents (45%) considered that their companies were either at level 3 (Diagnostic analytics) or level 4 (Predictive analytics), and none of them thought that their companies had reached the highest level (Prescriptive analytics).

5.2.3 Level of maturity in terms of translation quality management processes

In order to get more information about the average level of maturity of the providers of translation services regarding their translation quality management processes, the study covered four different aspects: level of maturity of the translation quality management program, level of knowledge and use of norms and standards, documented processes and human resources, and tools and technologies.

5.2.3.1 Level of maturity of the translation quality management program (KF20)

Interestingly enough, even though 70% of the respondents had reported that their companies were in the intermediate level in terms of translation analytics, in this question, the replies were concentrated either on the lower levels of maturity (60% of companies are either at level 1, manual quality management, or 2, automated quality management) or at the highest levels of maturity (40% of the companies consider they are in level 4, centralised quality management, or 5, expert translation quality management).

5.2.3.2 Level of knowledge and use of translation quality management norms and standards (KF21, KF22)

Many translation service buyers (especially in the public sector) use the certification on certain quality management and translation services standards as a requisite to select new translation providers. Therefore, it is not surprising to see that 70% of the translation service providers use or follow one or more of these standards. However, it is interesting to note a preference for translation services specific standards, such as ISO 17100, over quality management standards such as 9001, and the prevalence of the ISO standards over the ASTM one.

45% of the participants also reported that their companies use either MQM-DQF or LISA QA model as the baseline standards to design their translation metrics, and 21% of the translation service providers do not use any standardised metric.

5.2.3.3 Documented processes and human resources used in translation quality management programs

5.2.3.3.1 Translation quality management methodologies (KF23, KF24)

70% of the translation service providers include all the translated content in their translation quality management processes, while the other 30% revealed that there were exceptions, mainly due to two main scenarios: clients that prefer not to pay for translation quality assurance steps or projects that only require unedited machine translation.

60% of the translation providers consider that the quality requirements and expectations from their clients help them build a solid quality management program tailored to their needs, while 35% think that their client's requirements and expectations do not affect the development of the translation quality management, or that they can limit or hinder it.

5.2.3.3.2 Translation quality evaluation methodologies (KF25, KF26, KF27, KF28)

In contrast to the data compiled from the translation service buyers, who favour a combined holistic and analytic evaluation approach, only 10% of the respondents from the provider's side reported using both holistic and analytic evaluation methods. Instead, each method is chosen as the preferred one by half of the remaining participants (45%).

Additionally, 35% of the translation service providers include all the translated content in the translation evaluation workflows as a norm, while another 30% follow a more granular approach and tailor this percentage depending on different factors such as the reliability of the translation resources or the quality risks identified during the quality planning phase.

Similarly to the information provided by the translation service buyers, 55% of the translation service providers follow a hybrid approach in which some translations are evaluated before the delivery to the internal stakeholders, and others, after the delivery. This hybrid approach is probably one of the causes why a very high number of translation service buyers (65%) use whole texts or samples to perform translation evaluations depending on the needs instead.

In terms of grading and scoring systems used, the majority of the translation service providers use a combination of numbers or percentages that represents the score obtained in the evaluation, a grading system with three to five grades, and a defined threshold to indicate whether the translation meets the requirements or not. A slight majority of the translation service providers (55%) use non-uniform evaluation metrics depending on the type of text or the quality requirements, and half of the participants stated their companies use non-uniform severity multiplier distributions, while the other half reported that the severity multipliers are always the same for all error types.

The majority of the translation service providers confirmed that their quality metrics were based on the MQM-DQF or the LISA QA model, so it is not surprising to see that their error typology has the same error categories (Accuracy, Fluency, Terminology and Style) and error levels proposed by these two models (Minor, Major, Critical).

90% of the participants on the provider's side reported feeling confident about the level of adequacy of the quality evaluation methodology used, and more than 60% of them were very confident or extremely confident.

5.2.3.3.3 Human resources involved (KF29)

The types of roles most commonly involved in translation service providers' translation quality management programs are mainly the reviewers and proofreaders, followed by the language leads, the translators, and the quality managers. Only 35.7% of the participants mentioned having evaluators as a specific role.

There were two distinctive trends as to whether these resources were employed directly by the company or were external: Either the resourcing model of the translation service provider favoured the externalisation of these roles (45% of the candidates revealed that 10% or less of these resources were internal), or it adopted a flexible approach in which translators and reviewers were usually external, but language leads and evaluators were in-house workers (16% of the candidates mentioned this approach).

In terms of qualifications and skills of the evaluators, the majority of the participants stated that these resources were required to have previous experience as translators or reviewers, follow a specific training path provided by the company and have specific knowledge of the subject matter or the client's account.

On the other hand, quality managers should ideally have academic qualifications or experience as translators, possess proven analytical skills and have experience in translation quality management programs.

5.2.3.4 Tools and technologies (KF30)

35% of the translation service providers use the modules available in commercial CAT/TMS tools (such as Memsource, XTM, Trados Studio, Matecat, MemoQ, Wordfast and XTRF) for translation quality management purposes, 27% do not have any specific tools, 19% uses standalone translation quality tools such as ContentQuo, Xbench and Verifika, and 15% have proprietary tools for translation quality management purposes.

The situation is quite similar in translation quality evaluation, as 39% use these same features for quality evaluation and error annotation purposes, 22% have proprietary tools, 13% use other types of tools, and 9% do not use any specific tools for error annotation.

5.2.4 Current challenges in translation quality management (KF31)

The two parameters with a higher level of satisfaction stated by the translation service provides were the adequacy to the needs of the business and the compliance with norms and standards. This could be because translation service providers need to be able to fulfil the needs of their clients and adapt to the requirements of the translation industry in general. On the other hand, the two categories with the lowest level of satisfaction were the adequacy of the tools used for translation quality evaluation and the adequacy of the tools used for translation quality management.

This lack of adequacy of the tools used can also be seen in the pain points shared by the participants, as they included topics such as the lack of automation and the need for better integration of translation quality management tools and translation quality evaluation tools in their current TMS environments. Other common pain points mentioned by the participants were the lack of cost-effective, scalable, and flexible translation quality management methodologies and time constraints to perform translation quality management and evaluation tasks.

5.2.5 Solutions and workarounds applied (KF32)

The solutions to the challenges currently faced by the translation service providers include the incorporation of experienced quality managers or the use of consultancy services specialised in translation quality management, the investment in advanced automation tools, and benchmarking initiatives to compare new tools to perform translation quality management or evaluation.

5.2.6 Future challenges and trends in translation quality management (KF33, KF34)

According to the participants from the translation service providers, some of the future trends will be the increase in volumes, languages and text types covered by translation quality programs, the need for a more scalable, faster, and cost-effective approach to translation quality, and the influence of new technologies such as machine learning and AI.

The participants also shared their views on the most valuable initiatives that could help the development of their translation quality management programs. The ones mentioned more frequently were research initiatives, publications and training courses led by practitioners or industry-related organisations, and the development of more advanced translation quality management features inside CAT/TMS tools.

5.3 Experts in translation quality management

19.2% of the participants in this research were either non-academic subject matter experts in translation quality management or academic researchers or teachers. While the surveys designed for buyers of translation services and translation service providers were aimed at discovering the most common practices in the industry (descriptive approach), the goal of the surveys designed for the experts in translation quality management was to get insights as

to what the best practices should be. For this reason, the following themes were covered mainly from a prescriptive approach:

- Profile of the experts in translation quality management
- Recommendations and best practices
    - Translation quality management
    - Translation quality evaluation
    - Human resources and qualifications
    - Tools and technologies
- Current challenges in translation quality management and evaluation
- Solutions recommended
- Future challenges and trends in translation quality management

5.3.1 Profile of the experts in translation quality management (KF35)

30.8% of the experts on translation quality management and evaluation stated that they worked primarily as academic researchers, while another 30.8% were independent consultants. In terms of professional background, 53.8% of the respondents reported having a non-academic background, while the other 46.2% mentioned being teachers or academic researchers. 38% of the respondents declared that they participate in ISO committees, while 31% belonged to an academic research centre or department.

Probably due to this slight prominence of non-academic translation quality researchers with participation in ISO committees, the standards that were more popular among this research group were the ISO norms specifically related to translation, such as ISO 17100 (Requirements for Translation Services), mentioned by 62% of the participants, and ISO 18587 (Post-editing of Machine Translation Output), selected by 38% of the experts in translation quality management.

77% of the participants were also familiarised with the translation quality metrics proposed by MQM-DQF and LISA QA Model and different types of automatic metrics for MT evaluation such as BLEU, METEOR, and BLEURT.

5.3.2 Recommendations and best practices

This part of the research was divided into four topics: quality management, translation quality evaluation, human resources, and tools and technologies. The following sections will discuss the results of each of these topics.

5.3.2.1 Translation quality management (KF36, KF37)

There was not a clear consensus among the experts in translation quality management as to whether the requirements and expectations from their clients help them build a solid quality management program or hinder it. A third of the participants considered that these expectations were important and useful to tailor the quality of these programs to the needs of the clients, the other third stated that some of the client's requirements and expectations presented specific challenges that needed to be overcome to ensure the development of a solid quality program, and 38.5% of the participants believed that the outcome depended entirely on the level of translation maturity of the client.

Almost half the experts in translation quality management (46.2%) also reported encountering situations in which the translation quality management methodology used by them was considered as not suitable for the needs of the client and explained that it was critical to review the translation quality assurance methodologies frequently to make sure they could adapt to different needs and new types of translation projects.

The experts consulted also defined five common scenarios in translation quality management and evaluation workflows with various degrees of complexity. These scenarios went from two compulsory steps (Quality assurance 1 and Evaluation 1) plus one optional step (Evaluation 2) to four compulsory steps (Quality assurance 1, Quality assurance 2, Evaluation 2, Evaluation 2) plus one optional step (Evaluation 3).

5.3.2.2 Translation quality evaluation (KF38, KF39, KF40, KF41, KF42)

The majority of the experts in translation quality management (69.2%) are familiarised with both analytic and holistic translation quality evaluation methods, with a slight preference for the analytic method over the holistic one. Similarly to the information shared by the translation service buyers, the experts in translation quality management consider that both methods can complement each other and provide some examples as to what type of texts, goals, or circumstances determine which method is more appropriate than the other. For

example, holistic translation quality evaluation is more effective to assess creative texts when there is a need to get quality data quickly and cheaply or when the goal is to get quality results related to usability and local relevancy of the translated text.

39% of the translation quality management experts recommend including between 10% and 20% of the translated content in the translation evaluation workflows, while 31% of them prefer a flexible approach to determine the most adequate percentage depending on factors such as the visibility or priority of the project, and the level of confidence in the resources used. More than half of the participants (64.3%) also consider that translation quality evaluations should be performed with a scheduled frequency, and 53.8% think that these evaluations should ideally take place before the delivery of the translation to the stakeholder or client. 92% of them agree that sampling techniques should be used in specific scenarios such as projects with very big volumes but low-quality risks, evaluations and audits conducted after the delivery of the translation or when there are time constraints.

In terms of grading and scoring systems used, the majority of the translation quality experts recommend using a threshold that defines whether the evaluated text met the required level of quality or not, and a grading system with four or five grades. In addition, between 92.3% and 76.9% of them consider that evaluation metrics (error types, severity levels or severity multipliers) and severity multipliers should not be uniform. Instead, evaluation metrics should be adapted to the type of text being evaluated, and severity multipliers should not be the same for all error types.

Given their familiarity with quality evaluation metrics such as MQM-DQF and the LISA QA model, it is not surprising to see that the majority of the experts (between 84.6% and 92.3%) recommend using six of the main error categories defined in these models (Accuracy, Fluency, Terminology, Verity, Style, and Locale conventions). In addition, they also recommend using the three error levels (Minor, Major, Critical) plus Neutral/Preferential.

When asked whether they recommended performing the review step at the same time as the evaluation step, almost half of the participants (46.2%) said they should not be performed at the same time, while 30.8% of the translation quality management experts did think that both steps could be performed in one go, as it was more cost-efficient for many clients.

179

5.3.2.3 Human resources and qualifications (KF43, KF44)

According to the translation quality experts, the roles that should be involved in the translation quality management programs are mainly the reviewers and proofreaders, followed by translators and language leads. In addition, 76.9% of the participants mentioned quality managers, evaluators, and subject matter experts.

In terms of qualifications and skills of the evaluators, 61% of the participants stated that these resources should possess previous experience as translators or reviewers and have specific knowledge of the subject matter or the client's account.

On the other hand, quality managers should ideally be in possession of academic qualifications or be experienced translators, have experience in translation quality management programs and possess cultural and linguist competencies.

The majority of the participants also mentioned that the two most valuable types of training to support evaluators and quality managers in their job were generic training on translation quality metrics, error categorisation and standards; and company-specific training with practical examples of the application of the evaluation methodology (in particular, error categorisation and error severities).

5.3.2.4 Tools and technologies (KF45)

65% of the translation quality experts consider that the best tools for translation quality management purposes are the quality features available in CAT tools or TMS environments such as SDL Trados or MemoQ, while 46% recommend using proprietary translation quality management tools or specific translation quality tools such as ContentQuo, QA Distiller, Verifika, Xbench, or TQAuditor.

The situation is quite similar for translation quality evaluation tasks, and 48% of the experts recommend the use of modules or features included in commercial CAT tools or TMS to perform error annotation, while 38% suggest the same translation quality tools mentioned above.

5.3.3 Current challenges in translation quality management and evaluation (KF46)

According to the translation quality experts, the two aspects of the translation management frameworks used in the industry that show a higher level of maturity are their adequacy to

the needs of the business and their level of compliance with norms and standards. On the other hand, the two categories with the lowest level of maturity were the adequacy of the resourcing model and the overall level of maturity of the frameworks.

The lack of adequacy of the resourcing model can also be seen in the pain points shared by the participants, as they included topics such as the lack of training and knowledge, the financial limitations to invest in tools or dedicated resources, and the need for appropriate training and calibration among reviewers. Other pain points mentioned by the experts and that are related to the overall low level of maturity of the translation quality management frameworks are the need to have cost-effective, scalable, and flexible translation quality management methodologies and the fact that the current evaluation models are time-consuming, expensive, and rigid.

5.3.4 Solutions and workarounds applied (KF47)

The solutions to the challenges detected by translation quality experts include the development of translation quality management processes focused on a holistic view of translation quality, leveraging the use of new technologies such as machine learning and AI to perform computer-assisted evaluations, and the publication of national or international regulations and guidelines to hire and train qualified evaluators

5.3.5 Future challenges and trends in translation quality management (KF48, KF49)

According to the translation quality management experts, some of the future trends will be the need for a more scalable, faster, and cost-effective approach to translation quality, the influence of new technologies such as machine learning and AI, the increased demand for qualified evaluators and quality managers, and the development of new types of translation quality metrics (user-centric, holistic, automated, quality estimations) and quality levels.

The participants also shared their views on the most valuable initiatives that could help the advancement of translation quality management frameworks in the industry. The ones mentioned more frequently were research initiatives, led either by practitioners and industry-related organisations or by universities and research centres; training courses organised by private organisations; and advanced translation quality management tools, regardless of whether they are integrated into existing CAT/TMS environments or standalone ones.

5.4 Translation technology companies

The fourth target group of this research were the translation technology companies, which represented 5.9% of the participants. The rationale behind the inclusion of this target group in this study is that, even though technology has been an integral part of the translation profession for at least 20 years, there has not been much research on specific aspects of the technology used for translation, particularly on the tools and mechanisms used for quality management and evaluation tasks (Chesterman & Williams, 2002; Doherty, 2016; Petrova, 2019; Vela-Valido, 2021). For this reason, the survey designed for translation technology companies aimed to get insights into the functionalities, limitations and future capabilities of the translation quality management and evaluation solutions offered by different types of translation technology companies. Therefore, the survey was structured into five main topics:

- Profile of the translation technology companies
- Functionalities and limitations
  - Adherence to norms and standards
  - Translation quality management solutions
  - Translation quality evaluation solutions
  - Technical specifications
- Current challenges and pain points
- Solutions
- The future of translation quality management

5.4.1 Profile of the translation technology companies (KF50)

Unlike the average size of the translation service buyers, which generally had more than 200 employees, the size of the staff of the translation technology companies mainly was between 100 and 199 employees (as stated by 50% of the participants), and there were also smaller companies that had between 50 and 99 employees (25%), or less than 25 employees (25%). Half of the respondents reported that they worked for companies that offered mainly translation management products (CAT tools or TMS), the other participants worked at technology companies specialised in translation quality solutions or technology providers with different products for translation management, translation quality management and machine translation technologies

5.4.2 Functionalities and limitations

The part of the survey for translation technology providers was divided into three main topics: adherence to norms and standards, translation management and evaluation workflows, and technical specifications. We will discuss each of them in the following sections.

5.4.2.1 Adherence to norms and standards (KF51)

87% of the participants confirmed that their translation technology products had taken into consideration at least one of the main norms and standards on translation services or quality management, mainly ISO 17100, which describes the requirements for translation services. 75% of the translation technology providers also shared that their solutions allow the use of MQM-DQF evaluation metrics.

5.4.2.2 Translation quality management solutions (KF52)

The main gaps and opportunities for further development of the features related to translation quality tasks detected by the translation technology providers were:

- Different requirements and needs depending on the size and type of clients.
- Integration of automated metrics for machine translation quality management.

The participants also mentioned that those tools that have specifically developed for translation quality management and evaluation have more functionalities than modules or features included in CAT tools and TMS environments.

5.4.2.3 Translation quality evaluation solutions (KF53, KF54, KF55)

All the translation technology providers confirmed that their companies supported analytic and holistic evaluations and evaluations of whole texts or samples.

In terms of grading and scoring systems used, half of the translation technology providers confirmed that their solutions allow the use of numbers or percentages as part of the grading system, while the other half declared that their solution only offered one type of grading system.

Half of the respondents also confirmed that their solution allows different evaluation metrics and severity multipliers distributions (non-uniform approach), while the other half said it was not possible or that they were not sure.

75% of the technology service providers reported that their solutions use the same terminology and at least three of the main error types proposed by the standardised metrics mentioned in section 2.3.7 (Accuracy, Fluency and Terminology) and the three error levels (Minor, Major, Critical). However, the rest of the error types, such as Design, Locale Convention, or Style, were supported by only one of the solutions.

Most of the translation technology solutions (75%) have different user profiles for translators and reviewers/proofreaders. However, only one of the solutions had specific user roles for language leads or quality managers, and none of them had a specific user profile for evaluators or subject matter experts.

5.4.2.4 Technical specifications (KF56)

75% of the participants reported that the specific solution provided for translation quality management and error annotation is a module or feature included in their CAT/TMS tool, while the other solution is an API that can be connected to other translation systems. All the solutions also have some standard features such as downloadable reports, automated writing quality checkers or editing/review capabilities, while other features that are typically related to the evaluation of the translation itself (such as arbitration capabilities or addition of issue descriptions) were available only on some of the tools.

5.4.3 Current challenges and pain points (KF57)

In this part of the survey, the participants were requested to rate the level of adequacy of their tools and solutions according to different parameters.

According to the technology service providers, the two aspects that show a higher level are the adequacy of the tools to the needs of the business and the overall level of maturity of the translation technology. On the other hand, the categories that scored the lowest were the level of adherence to translation quality management and evaluation norms and standards, the adequacy of the translation quality management and translation evaluation workflows of the solutions, and the adequacy of the tools for translation quality management purposes.

This lack of adequacy of the translation quality workflows offered by these solutions, as well as the lack of adequacy of these tools for specific translation quality management purposes, also appears in the pain points shared by the participants, as they mentioned topics such as the need to improve automation, real-time quality assurance and sampling methodologies, as well as simplifying data collection and improving data analysis capabilities.

5.4.4 Solutions (KF58)

The solutions to the technological challenges in translation quality management shared by translation technology companies include the development of different automated quality assurance, evaluation, and business analytic features and reinforcing product development to cater to different quality management and evaluation needs.

5.4.5 Future challenges and trends in translation quality management (KF59, KF60)

According to translation technology companies that participated in this research, some of the future trends will be the use of machine learning and AI to provide automated translation quality management and evaluation features and the consolidation of quality assurance and quality evaluation workflows into one comprehensive quality management process.

The respondents also shared their insights on the most valuable initiatives that could help the advancement of translation quality management frameworks in the industry. Most of the participants considered especially important to have more advanced translation quality management tools integrated into existing CAT/TMS environments.

5.5 Summary of key findings (KFs)

Key findings of this research per target group are summarised below.

5.5.1 Buyers of translation services

- **KF1:** Buyers with bigger volumes and number of languages tend to have a more significant need for quality management and evaluation programs.
- **KF2**: Buyers with higher maturity levels in terms of translation quality processes and analytics are more involved in translation quality management and evaluation programs.

- **KF3**: There seems to be a relationship between the level of maturity in translation analytics and the level of maturity in translation quality management.

- **KF4:** Translation service buyers gravitate between very high levels of translation quality maturity (45.1%) and very low levels of translation quality management maturity (45.2%).

- **KF5:** More than 50% of the translation service buyers do not take into consideration any standard on translation services or quality management to build their translation quality management programs.

- **KF6:** Almost 50% of the translation service buyers use MQM-DQF to design their metrics, 29% have a proprietary set of metrics, and 23% do not use any metric.

- **KF7:** Translation service buyers prioritise marketing texts, websites, mobile content, technical documents, and external documentation in their translation quality programs, and they include 80% of the total content that is translated in their translation quality management programs.

- **KF8:** The average translation service buyer includes a maximum of 50% of the translated content in its translation evaluation workflow, conducts evaluations with a scheduled frequency, and combines predelivery evaluations with postdelivery evaluations and sampling evaluation methods with non-sampling evaluation methods.

- **KF9:** The average translation service buyer has a grading system with two grades (Pass/Fail) and defined thresholds and a numerical scoring system. The evaluation metrics are non-uniform, and the severity multipliers are uniform.

- **KF10:** The average translation service buyer uses the MQM-DQF model as a baseline for this quality metrics methodology, which includes four main error categories (Accuracy, Fluency, Terminology and Style) and the three error levels (Minor, Major, Critical).

- **KF11:** The average translation service buyer feels moderately confident about the level of adequacy of the quality evaluation methodology used.

- **KF12:** Translation service buyers demand evaluators with experience as translators and reviewers and consider that translation quality managers should have cultural and linguist competencies, experience or certification in language studies, and experience in translation quality management programs.

- **KF13:** The average translation service buyer uses the modules available in commercial CAT/TMS tools such as Memsource, Smartling and XTM for translation quality management and evaluation purposes.

- **KF14:** The main pain points in translation quality management suffered by translation service buyers are the lack of adequacy of the tools used for translation quality management and evaluation, the lack of dedicated human resources, and the need to establish clear, scalable, and customisable evaluation metrics and thresholds that can be used for analytic and holistic evaluations.

- **KF15:** The solutions to translation quality management challenges adopted by translation service buyers are using more advanced tools (developed either internally or externally), adopting a cross-functional and cross-team approach to translation quality, and learning best practices from translation service or translation technology vendors.

- **KF16:** The most important trends in translation quality management identified by translation service buyers are the increase in volumes, languages and text types, the advent of new types of translation quality management workflows and translation quality metrics, and the increase in the use of tools and technologies.

- **KF17:** The most valuable initiatives in translation quality management identified by the translation service buyers are the development of more advanced translation quality management features inside CAT/TMS tools and research initiatives led by practitioners and industry-related organisations.

5.5.2 Translation service providers

- **KF18:** The demographics of the translation service providers that participated in the study show an equal distribution of small, medium, and big companies in terms of size of the staff, annual translation volumes and number of languages translated.

- **KF19:** Translation service providers show, on average, the same level of maturity in terms of translation analytics as the translation service buyers that have translation quality management programs.

- **KF20:** Translation service providers gravitate between high levels of translation quality management maturity (40%) and low levels of translation quality management maturity (60%).

- **KF21:** The average translation service provider is certified or follows one or more quality management and translation services standards, in particular ISO 17100 (translation services) and 9001 (quality management).

- **KF22:** Almost 50% of the translation service providers use the MQM-DQF or LISA QA model to design their metrics, 21% does not use any metric, and 17% have a proprietary set of metrics.

- **KF23:** 70% of translation service providers include 100% of the translated content in their translation quality management programs unless a different strategy has been specifically requested by the client.

- **KF24:** 60% of the translation service providers consider that their clients' quality requirements and expectations help them build a solid quality management program.

- **KF25:** The average translation service provider tailors the percentage of the translated content that is included in the translation evaluation workflow depending on different factors, combines predelivery evaluations with postdelivery evaluations, and sampling evaluation methods with non-sampling evaluation methods.

- **KF26:** The average translation service provider has a grading system with between three and five grades, defined pass/fail thresholds, and a numerical scoring system.

- **KF27:** Similarly to the translation service buyers studied in this research, the average translation service provider uses the MQM-DQF or LISA QA models as a baseline for its quality metrics methodology, which includes four main error categories (Accuracy, Fluency, Terminology and Style) and the three error levels (Minor, Major, Critical).

- **KF28:** The average translation service buyer feels very confident about the level of adequacy of the quality evaluation methodology used.

- **KF29:** Translation service providers select evaluators with experience as translators or reviewers and tend to provide specific training. They also consider that translation quality managers should have academic qualifications or experience as translators, proven analytical skills, and experience in translation quality management programs.

- **KF30:** The average translation service buyer uses the modules available in commercial CAT/TMS tools such as Memsource, XTM, Trados Studio, Matecat, MemoQ, Wordfast and XTRF for translation quality management and evaluation purposes. The biggest translation service providers tend to use standalone translation

quality management and evaluation tools or develop proprietary tools, while the smallest providers do not use any tools or rely on spreadsheets to perform these tasks.

- **KF31:** The main pain points in translation quality management suffered by translation service providers are the lack of adequacy of the tools used for translation quality management and evaluation, the need to establish cost-effective and flexible translation quality management methodologies and the time constraints they encounter to perform translation quality management and evaluation tasks.

- **KF32:** The solutions to translation service providers' current translation quality management challenges are the use of human resources specialised in translation quality management (either internally or externally), the investment in automation tools, and the research of new tools to perform translation quality management or evaluation.

- **KF33:** The most important trends in translation quality management identified by translation service buyers are the increase in volumes, languages and text types, the need for a more scalable, faster, and cost-effective approach to translation quality, and the influence of new technologies such as machine learning and AI.

- **KF34:** The most valuable initiatives in translation quality management identified by the translation service providers are the development of more advanced translation quality management features inside CAT/TMS tools and research initiatives, publications and training courses led by practitioners and industry-related organisations.

5.5.3 Experts in translation quality management

- **KF35**: 53.8% of the experts in translation quality management have a non-academic background, 62% are mostly familiarised with ISO norms such as ISO 17100 and ISO 18587, and 77% are used to working with evaluation models such as MQM-DQF and LISA QA Model for evaluation of human translation, and BLEU and METEOR for evaluation of machine translation outputs.

- **KF36:** 38.5% of the experts in translation quality management consider that the level of translation maturity of the organisations is a determining factor in whether their needs and expectations can help them develop a solid quality program or limit its potential.

- **KF37:** Translation quality experts consider five main scenarios in translation quality management and evaluation workflows with various degrees of complexity. The simplest scenarios have two compulsory steps, while the most complex ones have four compulsory steps.

- **KF38:** The majority of experts in translation quality management recommend both the analytic and the holistic translation quality evaluation methods, and they consider that holistic translation quality evaluation is more effective to assess creative texts when there is a need to get quality data quickly and cheaply or when the goal is to get quality results related to usability and local relevancy of the translated text.

- **KF39:** Translation quality experts recommend either including between 10% and 20% of the translated content in the translation evaluation workflows or adapting the percentage depending on factors such as the visibility or priority of the project and the level of confidence in the resources used. They also consider that evaluations should take place before the delivery of the translation to the stakeholder and that sampling evaluation methods are advisable when there are projects with big translation volumes, low-quality risks, or time constraints.

- **KF40:** Translation quality experts recommend a grading system with four or five grades and defined pass/fail thresholds. The evaluation metrics and severity multipliers should not be uniform for all text types and error types.

- **KF41:** Translation quality experts recommend using the MQM-DQF model as a baseline to establish quality evaluation metrics with six main error categories (Accuracy, Fluency, Terminology, Verity, Style, and Locale conventions) and three error levels (Minor, Major, Critical), plus a specific level for Minor/Preferential changes.

- **KF42:** Almost half of the translation quality experts do not recommend performing review and evaluation steps simultaneously, while 30.8% recognise that this practice can be adopted to save time and improve the efficiency of the overall quality management framework.

- **KF43:** Translation quality experts consider that evaluators should have experience as translators or reviewers and have specific knowledge of the subject matter. They also think that translation quality managers should ideally have academic qualifications or experience as translators, experience in translation quality management programs and cultural and linguist competencies.

- **KF44:** Translation quality experts consider that evaluators and quality managers should have generic training sessions on translation quality metrics, error categorisation and standards, and company-specific training with practical examples of the application of the evaluation methodology.

- **KF45:** The majority of translation quality experts recommend using modules or features included in commercial CAT tools or TMS to perform translation quality management and evaluation, although there is also an important number of experts that prefer using proprietary translation quality management tools or specific translation quality tools such as ContentQuo, QA Distiller, Verifika, Xbench, or TQAuditor.

- **KF46:** The main pain points in translation quality management pointed out by the translation quality experts are the need to establish cost-effective and flexible translation quality management methodologies and the fact that the current evaluation models are time-consuming, expensive, and rigid.

- **KF47:** The solutions to the current translation quality management challenges pointed out by translation quality experts are the development of more holistic translation quality management frameworks, the use of technologies to perform computer-assisted evaluations, and the publication of regulations or guidelines to hire and train qualified evaluators.

- **KF48:** The most important trends in translation quality management identified by translation quality experts are the need for a more scalable, faster, and cost-effective approach to translation quality, the influence of new technologies such as machine learning and AI, the increased demand for qualified evaluators and quality managers, and the development of new types of translation quality metrics and quality levels.

- **KF49:** The most valuable initiatives in translation quality management identified by translation quality experts are research initiatives (both academic and non-academic), training courses, and more advanced translation quality management features inside or outside CAT/TMS tools.

5.5.4 Translation technology companies

- **KF50:** The demographics of the translation technology companies that participated in the study show a prevalence of medium-size providers specialised in translation management products (CAT tools or TMS).

- **KF51:** 87% of the translation technology providers have considered norms such as ISO 17110 in the design of their translation solutions, and 75% allow the use of MQM-DQF evaluation metrics.

- **KF52:** Translation technology providers understand there are specific gaps in the solutions provided for translation quality management, such as the lack of flexibility to adapt to the quality management needs of different types of clients, the need to integrate automated metrics for machine translation quality management, and the difference in the number of functionalities available between tools that have been designed explicitly for translation quality management and the quality modules or features that can be found in CAT tools or TMS.

- **KF53:** All the translation technology providers support analytic and holistic evaluations and evaluation of whole texts or samples, although only 50% support non-uniform evaluation metrics and severity multipliers distributions.

- **KF54:** 75% of the technology service solutions allow a maximum of three error categories out of the seven proposed by models such as MQM-QDF (Accuracy, Fluency and Terminology), and they have three error levels (Minor, Major, Critical).

- **KF55:** 75% of the technology service solutions have user profiles for translators and reviewers/proofreaders, but only one of them offers specific user roles for language leads or quality managers.

- **KF56:** Most translation technology solutions work as a module or feature in a CAT/TMS tool. These modules usually have some standard functionalities for translation quality management and review, but some do not have specific evaluation functionalities such as arbitration or the addition of comments or issue descriptions.

- **KF57:** The main current challenges pointed out by technology service providers are the need to optimise and automatise the translation quality management and translation evaluation workflows inside the translation management solutions and the overall level of adequacy of these tools for translation quality management purposes.

- **KF58:** The solutions to the technological challenges shared by translation technology companies include the development of automated quality assurance, evaluation, and business analytic features and enhancing product development to cater to different quality management needs.

- **KF59:** Two of the most important trends in translation quality management identified by translation technology providers are the use of machine learning and AI

in translation quality management and evaluation, and the consolidation of quality assurance and quality evaluation workflows into one single process.

- **KF60:** According to the translation technology companies, the most helpful initiative in translation quality management is the development of more advanced translation quality management features inside existing CAT/TMS tools.

6 Conclusions

This empirical study set out to investigate the current quality management and evaluation practices in the translation industry by asking the following six research questions:

1. What is the demographic and professional profile of the companies and participants that carry out translation quality management and evaluation tasks?
2. What is the level of maturity in terms of translation processes and analytics?
3. What is the level of maturity in terms of translation quality management processes?
    3.1. What is the overall level of maturity in translation quality management?
    3.2. What is the overall level of knowledge and use of translation quality management norms and standards?
    3.3. What are the documented processes and human resources used in translation quality management programs?
        3.3.1. How does the translation quality management program work?
        3.3.2. What are the translation quality evaluation methodologies used?
        3.3.3. How does the translation quality evaluation process work?
        3.3.4. What kind of human resources are involved in translation quality management and evaluation processes?
    3.4. What are the main tools and technologies used in translation quality management and evaluation processes?
4. What are the main pain points and limitations in translation quality management and evaluation processes?
5. What solutions and workarounds are applied in the industry to overcome the current pain points and limitations?
6. What are the future challenges and trends in translation quality management, and what kind of initiatives would help the advancement of the translation quality management practices in professional environments?

After providing the background for this research by reviewing the literature available on academic, professional and regulatory approaches to translation quality management, this research proceeded to solve these questions using a mixed methods approach that encompassed the design of a survey consisting of questionnaires and semi-structured interviews targeting four different audiences: buyers of translation services with internal

translation quality management teams, translation service providers, translation technology companies and experts in translation quality management and evaluation. By including four types of target groups, the aim was to be able to observe and analyse the current practices in translation quality management from two different angles (prescriptive and descriptive) and through different lenses.

6.1 Original contributions of the research

The results of this survey were analysed and discussed in chapters 4 and 5, and the key findings for each target group have been summarised in section 5.5. Based on these key findings, the following answers to the research questions are set forth:

Answer to research question 1: What is the demographic and professional profile of the companies and participants that carry out translation quality management and evaluation tasks?

The findings of the survey conducted for this study suggest that overall, international companies with big translation volumes and several languages tend to have a more significant need for quality management and evaluation programs and that the type of resources most commonly involved in translation quality management are Translation/Localisation Managers or Directors, Translation/Localisation Quality Managers, and Translation/Localisation Program Managers.

Answer to research question 2: What is the maturity level in translation processes and analytics?

According to the data analysed, translation buyers and providers with higher maturity levels in terms of translation quality processes and analytics are more involved in translation quality management and evaluation programs. Both types of companies have, on average, the same level of maturity.

Answer to research question 3.1: What is the overall level of maturity in translation quality management?

The findings of the questions regarding this topic suggest that translation service buyers gravitate between very high levels of translation quality maturity (levels 4 or 5) and very low levels of translation quality management maturity (levels 1 or 2). In contrast, translation

service providers gravitate between high levels of translation quality management maturity (level 4) and low levels of translation quality management maturity (level 2).

Answer to research question 3.2: What is the overall level of knowledge and use of translation quality management norms and standards?

The data compiled from translation buyers, translation providers, experts in translation quality management and translation technology providers suggest that more than 50% of the translation service buyers do not take into consideration any standard on translation services or quality management to build their translation quality management programs. This percentage goes up to 70% for translation service providers, 87% for translation technology providers, and 100% for translation quality management experts.

Similar results have been observed when it comes to the use of standardised metrics, as less than 50% of the translation buyers use any existing model to design their metrics, and 23% of them do not use any metric. On the other hand, almost 50% of the translation service providers use standardised metrics such as MQM-DQF or LISA QA, and 21% do not use any metric. In contrast, 77% of the translation quality management experts are familiar with norms (particularly ISO norms) and evaluation metrics such as MQM-DQF or LISA QA, and 75% of the translation technology solutions are designed to be compatible with the standardised evaluation metrics mentioned above.

Answer to research question 3.3.1: How does the translation quality management methodology work?

According to the replies analysed, translation service buyers prioritise marketing texts, websites, mobile content, technical documents, and external documentation in their translation quality programs, and they include 80% of the total content that is translated in their translation quality management programs. In addition, the average translation service buyer includes a maximum of 50% of the translated content in its translation evaluation workflow, conducts evaluations with a scheduled frequency, and combines predelivery evaluations with postdelivery evaluations and sampling evaluation methods with non-sampling evaluation methods.

On the other hand, translation service providers tend to include 100% of the translated content in their translation quality management programs unless a different strategy has been specifically requested by the client. They also tailor the percentage of the translated content

197

that is included in the translation evaluation workflow depending on different factors and combine predelivery evaluations with postdelivery evaluations and sampling evaluation methods with non-sampling evaluation methods.

Translation quality management experts recommend both the analytic and the holistic translation quality evaluation methods, and they consider that holistic translation quality evaluation is more effective to assess creative texts when there is a need to get quality data quickly and cheaply or when the goal is to get quality results related to usability and local relevancy of the translated text. These experts also recommend either including between 10% and 20% of the translated content in the translation evaluation workflows or adapting the percentage depending on factors such as the visibility or priority of the project and level of confidence in the resources used. They also consider that evaluations should take place before the delivery of the translation to the stakeholder and that sampling evaluation methods are advisable when there are projects with big translation volumes, low-quality risks, or time constraints.

Answer to research question 3.3.2: What translation quality evaluation methodologies are used?

The findings of the survey conducted suggest that translation service buyers tend to have a grading system with two grades (Pass/Fail), defined thresholds, and a numerical scoring system. The evaluation metrics used are non-uniform, but the severity multipliers are uniform. They use the MQM-DQF model as a baseline for their evaluation metrics methodology, which includes four main error categories (Accuracy, Fluency, Terminology and Style) and three error levels (Minor, Major, Critical). Translation service buyers feel moderately confident about the level of adequacy of the quality evaluation methodology used.

Similarly to the translation service buyers studied in this research, translation service providers use the MQM-DQF or LISA QA models as a baseline for their quality metrics methodology, which includes four main error categories (Accuracy, Fluency, Terminology and Style) and the three error levels (Minor, Major, Critical). However, their grading system is more complex because it uses between three and five grades, defined pass/fail thresholds, and a numerical scoring system. Translation service providers feel very confident about the level of adequacy of the quality evaluation methodology used.

The level of complexity of the translation quality methodology increases if we analyse the data obtained from the translation quality management experts. They recommend a grading system with four or five grades and defined pass/fail thresholds and non-uniform evaluation metrics and severity multipliers. Moreover, they also suggest using the MQM-DQF model as a baseline to establish quality evaluation metrics, and to use all the error categories (Accuracy, Fluency, Terminology, Verity, Style, and Locale conventions), three error levels (Minor, Major, Critical), plus a specific level for Minor/Preferential changes. Interestingly enough, even though half of the translation quality experts do not recommend performing review and evaluation steps simultaneously, 30.8% of them agree that this practice can be adopted to save time and improve the efficiency of the overall quality management framework.

All the translation technology providers that participated in the research support analytic and holistic evaluations, as well as evaluation of whole texts or samples; however, only 50% of them support non-uniform evaluation metrics and severity multipliers distributions, and 75% of them allow to use a maximum of three error categories (Accuracy, Fluency and Terminology), out of the six proposed by MQM-QDF as well as three error levels (Minor, Major, Critical).

Answer to research question 3.3.3: What kind of human resources are involved in translation quality management and evaluation processes?

According to the data analysed, the types of roles most commonly involved in translation service buyers' translation quality management programs are mainly the reviewers and proofreaders, followed by the translators, quality managers, and language leads. Only 38.7% of the participants mentioned having evaluators as a specific role, and there was no particular trend as to whether these resources were employed directly by the company or were external. In terms of training and qualifications, translation service buyers prefer evaluators with experience as translators and reviewers, and they consider that translation quality managers should have cultural and linguist competencies, experience or certification in language studies, and experience in translation quality management programs.

Translation service providers follow a similar approach, as the type of roles most commonly involved in the translation quality management programs on their side are mainly the reviewers and proofreaders, followed by the language leads, the translators, and the quality

managers. Only 35.7% of the participants mentioned having evaluators as a specific role. In terms of training and qualifications, translation service providers prefer evaluators with experience as translators or reviewers. In addition, they also tend to provide specific training to train new evaluators who have previously worked for them as translators or reviewers. Translation providers also consider that translation quality managers should have academic qualifications or experience as translators, proven analytical skills, and experience in translation quality management programs.

Translation quality experts are pretty aligned with both translation buyers and providers, and they believe evaluators should have experience as translators or reviewers and possess specific knowledge of the subject matter. They also think that translation quality managers should ideally have academic qualifications or experience as translators, experience in translation quality management programs and cultural and linguist competencies. Translation quality experts also consider that evaluators and quality managers should have generic training sessions on translation quality metrics, error categorisation and standards, and company-specific training with practical examples of the application of the evaluation methodology.

Finally, even though the majority of the translation technology solutions studied have user profiles for translators and reviewers/proofreaders, only one of them offers specific user roles for language leads or quality managers.

Answer to research question 3.4: What are the main tools and technologies used in translation quality management and evaluation processes?

The data compiled in this study suggests that most translation service buyers and providers use the modules available in commercial CAT/TMS tools for translation for both quality management and evaluation purposes. However, the biggest translation service providers tend to use standalone translation quality management and evaluation tools or develop proprietary tools, while the smallest translation service providers do not use any tools or rely on spreadsheets to perform these tasks.

This practice is confirmed by the majority of the translation quality experts that participated in the study. However, an important number of experts also recommend using proprietary translation quality management tools (whenever that is feasible) or specific translation quality tools such as ContentQuo, QA Distiller, Verifika, Xbench, or TQAuditor.

Most technology service solutions studied work as a module or feature in a CAT/TMS tool. Even though these modules have some basic functionalities for translation quality management and review, some do not have with specific evaluation functionalities such as arbitration or adding comments or issue descriptions. Moreover, translation technology providers acknowledge that there are certain gaps in the solutions provided for translation quality management, such as the lack of flexibility to adapt to the quality management needs of different types of clients, the need to integrate automated metrics for machine translation quality management, and the difference in the number of functionalities available between tools that have been designed explicitly for translation quality management and the quality modules of features that can be found in CAT tools or TMS.

Answer to research question 4: What are the main pain points and limitations in translation quality management and evaluation processes?

The main pain points in translation quality management shared by translation service buyers are the lack of adequacy of the tools used for translation quality management and evaluation, the lack of dedicated human resources, and the need to establish clear, scalable, and customisable evaluation metrics and thresholds that can be used for analytic and holistic evaluations.

For translation service providers, the main pain points in translation quality management are, again, the lack of adequacy of the tools used for translation quality management and evaluation, the need to establish cost-effective and flexible translation quality management methodologies and the time-constraints they encounter to perform translation quality management and evaluation tasks.

From the point of view of the translation quality management experts, the main pain points they have observed are the need to establish cost-effective and flexible translation quality management methodologies and the fact that the current evaluation models are time-consuming, expensive, and rigid.

Finally, the main challenges pointed out by the technology service providers are the need to optimise and automatise the translation quality management and translation evaluation workflows inside the translation management solutions, as well as the overall level of adequacy of these tools for translation quality management purposes.

Answer to research question 5: What are the solutions and workarounds applied in the industry to overcome the current pain points and limitations?

According to the data analysed, the main solutions adopted by translation service buyers to overcome the translation quality management challenges mentioned are the use of more advanced tools (developed either internally or externally), the adoption of a cross-functional and cross-team approach to translation quality, and taking the opportunity to learn best practices from translation service providers or translation technology vendors.

Translation service providers propose the increase of human resources specialised in translation quality management (either internally or externally), the investment in automation tools, and the research of new tools to perform translation quality management or evaluation.

The solutions pointed out by translation quality experts include the development of more holistic translation quality management frameworks, the use of technologies to perform computer-assisted evaluations, and the publication of regulations or guidelines to hire and train qualified evaluators.

For translation technology companies, the main solutions to the technological challenges in translation quality management processes are the development of automated quality assurance, evaluation, and business analytic features and improving their product to cater to different quality management needs.

Answer to research question 6: 2. What are the future challenges and trends in translation quality management, and what kind of initiatives would help the advancement of the translation quality management practices in professional environments?

According to the data analysed, the most important trends in translation quality management identified by the translation service buyers are the increase in volumes, languages and text types, the advent of new types of translation quality management workflows and translation quality metrics, and the increase in the use of tools and technologies. This target group also considers that the most valuable initiatives in translation quality management are the development of more advanced translation quality management features inside CAT/TMS tools and research initiatives led by practitioners and industry-related organisations.

Translation service providers also consider that one of the most important trends is the increase in volumes, languages and text types, the need for a more scalable, faster, and cost-effective approach to translation quality, and the influence of new technologies such as machine learning and AI. Similarly to what was reported by the translation service buyers, the most valuable initiatives in translation quality management identified by the translation service providers include the development of more advanced translation quality management features inside CAT/TMS tools, as well as research initiatives, publications and training courses led by practitioners and industry-related organisations.

Translation quality management experts have a very similar view on this topic, as they consider that the most important trends in translation quality management are, yet again, the need for a more scalable, faster, and cost-effective approach to translation quality and the influence of new technologies such as machine learning and AI. In addition, this target group also mentions two more trends that should be considered: the increased demand for qualified evaluators and quality managers and the development of new types of translation quality metrics and quality levels. Finally, this target group considers that the most valuable initiatives in translation quality management are research initiatives (both academic and non-academic), training courses and more advanced translation quality management features inside or outside CAT/TMS tools.

6.2 Limitations and future research

Even though this dissertation provides several original contributions to the research of the current practices in translation quality management and evolution in professional environments, it also has a set of limitations that need to be considered.

Firstly, the researcher acknowledges that the findings of this research are limited by the sampling approach chosen and the overall size of the sample. For this reason, these key findings should be considered as an exploratory baseline for further and more profound research on this topic rather than a generalisable view of the practices followed and recommended by the wider translation industry.

Secondly, while the aim of this research was to take a broad approach and select different types of target groups, it certainly does not cover all the possible perspectives on translation quality management, as it favours the experiences and opinions of those who perform translation quality management and evaluation, are directly involved in these processes, or

belong to the translation industry. Consequently, translation service buyers without internal translation quality teams, translators and reviewers that are the recipients of the processes described, and the end-users that consume the final translations are missing from this research.

In addition, and as explained in the introduction of this dissertation, the scope of this research is the human quality management and evaluation of the translation of written texts into written texts without any audiovisual elements. Therefore, this study does not include audiovisual translation scenarios (such as subtitling, voice-over, or dubbing), nor does it cover automatic translation quality management and evaluation processes or tools.

Finally, this dissertation is, to the best of the researcher's knowledge, the first empirical research initiative that describes and analyses the current practices in translation quality management and evaluation from the point of view of translation buyers and translation service providers; provides an overview of the relation between these practices and different technological solutions from the point of view of some translation technology providers, and compiles a set of recommendations and best practices from translation quality management experts.

7 References

Ackland, R. (2013). *Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age*. London: SAGE Publications Ltd. https://doi.org/10.4135/9781446270011

al-Qinai, J. (2002). Translation Quality Assessment. Strategies, Parametres and Procedures. *Meta*, *45*(3), 497–519. https://doi.org/10.7202/001878ar

ASTM. (2014). *ASTM F2575 - 14 Standard Guide for Quality Assurance in Translation*. Standard Guide for Quality Assurance in Translation. https://doi.org/10.1520/F2575-14

ASTM. (2021a). *WK46396: Standard Practice for Analytic Translation Quality Evaluation*. ASTM. https://www.astm.org/WORKITEM-WK46396.htm

ASTM. (2021b). *WK46396 - New Practice for Development of Translation Quality Metrics*. ASTM. https://www.astm.org/DATABASE.CART/WORKITEMS/WK46396.htm

ASTM. (2021c). *WK54884 - New Practice for Holistic Quality Evaluation System for Translation*. ASTM. https://www.astm.org/DATABASE.CART/WORKITEMS/WK54884.htm

ATA. (2021). *Certification – American Translators Association*. https://www.atanet.org/certification/

Bangalore, S., Rambow, O., & Whittaker, S. (2000). Evaluation metrics for generation. *INLG '00: Proceedings of the First International Conference on Natural Language Generation*, vol 14, June 2000, 1-8 https://doi.org/10.3115/1118253.1118255

Beste, K. (2020). What content quality really means. In J.-M. Dalla-Zuanna & C. Kurz (Eds.), *Translation Quality in the Age of Digital Transformation* (pp. 410-423). Berlin: BDÜ Weiterbildung- und Fachverlagsgesellschaft

Birch, A., Abend, O., Bojar, O., & Haddow, B. (2016). Hume: Human UCCA-based evaluation of machine translation. *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 1264–1274. https://doi.org/10.18653/v1/d16-1134

Bojar, O., Buck, C., Callison-Burch, C., Federmann, C., Haddow, B., Koehn, P., Monz, C., Post, M., Google, R. S., & Specia, L. (2013). Findings of the 2013 Workshop on Statistical Machine Translation. *WMT13 Conference on Machine Translation (WMT)*, 1–44. http://www.statmt.org/wmt13/pdf/WMT01.pdf

Bojar, O., Federmann, C., Haddow, B., Koehn, P., Post, M., & Specia, L. (2016). Ten Years of WMT Evaluation Campaigns: Lessons Learnt. *Proceedings of the LREC 2016 Workshop "Translation Evaluation – From Fragmented Tools and Data Sets to an Integrated Ecosystem."*, 27–34. http://www.cracking-the-language-barrier.eu/wp-content/uploads/Bojar-Federmann-etal.pdf

Cadieux, P. (2004). *Globalization: People, Process and Technology*. Translation Directory. https://www.translationdirectory.com/article386.htm

Calvo, E. (2018). From translation briefs to quality standards: Functionalist theories in today s translation processes. *Translation and Interpreting*, *10*(1), 18–32. https://doi.org/10.12807/ti.110201.2018.a02

Cao, D., & Zhao, X. (2008). Translation at the united nations as specialized translation. *Journal of Specialised Translation*, *9*(9), 39–54. https://www.jostrans.org/issue09/art_cao.pdf

Carroll, J. B. (1966). An Experiment in Evaluating the Quality of Translations. *Mechanical Translation and Computational Linguistics*, *9*(3–4), 67–75. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.231.769&rep=rep1&type=pdf

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018a). Approaches to Human and Machine Translation Quality Assessment. In S. Castilho, S. Doherty, F. Gaspari, & J. Moorkens (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications* (pp. 9–38). Cham: Springer. https://doi.org/10.1007/978-3-319-91241-7_2

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018b). Translation quality assessment: from principles to practice. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Machine Translation: Technologies and Applications* (Vol. 1). Cham: Springer. https://doi.org/10.1007/978-3-319-91241-7

CCHI. (2021). *Certification Commission for Healthcare Interpreters*. https://cchicertification.org/

CEN. (2006). *EN 15038 Translation services - Service requirements*. www.din.de

Chesterman, A., & Williams, J. (2002). *The Map: A Beginner's Guide to Doing Research in Translation Studies*. Manchester: Taylor & Francis Group

Chriss, R. (2006). *Translation as a Profession*. Morrisville: Lulu.com

Church, K. W., & Hovy, E. H. (1993). Good applications for crummy machine translation. *Machine Translation*, *8*(4), 239–258. https://doi.org/10.1007/BF00981759

Colina, S. (2008). Translation quality evaluation: empirical evidence for a functionalist approach. *The Translator*, *14*(1), 97–134. https://doi.org/10.1080/13556509.2008.10799251

Creswell, J., Clark, V., Gutmann, M., & Hanson, W. (2003). Advance Mixed methods Research Designs. In *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage

CSA Research. (2021a). *Localization Maturity Assessment*. https://csa-research.com/Featured-Content/For-Global-Enterprises/Localization-Maturity-Assessment

CSA Research. (2021b). *Pro-Bono Survey Data*. https://csa-research.com/Featured-Content/For-LSPs/COVID-19-Industry-Data-Research

De Vaus, D. (2013). Surveys In Social Research. In *Surveys In Social Research* (5th ed.). New York: Routledge. https://doi.org/10.4324/9780203501054

DePalma, D. A., & Pielmeier, H. (2013). *The price-quality link*. http://www.commonsenseadvisory.com/AbstractView.aspx?ArticleID=5494

DFKI. (2015). *Multidimensional Quality Metrics Definition*. http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#dqf-mapping

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication*, *10*, 947–969. https://ijoc.org/index.php/ijoc/article/viewFile/3499/1573

Doherty, S., & Gaspari, F. (2013). *Understanding and Implementing Effective Translation Quality Evaluation Techniques*. https://www.qt21.eu/launchpad/sites/default/files/QTLP%20GALA%20Webinar%20 3.pdf

Doherty, S., Moorkens, J., Gaspari, F., & Castilho, S. (2018). On Education and Training in Translation Quality Assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications* (Vol 1, pp. 95–106). Cham: Springer. https://doi.org/10.1007/978-3-319-91241-7_5

Drugan, J. (2013). Quality in Professional Translation: Assessment and Improvement. In *Quality in Professional Translation: Assessment and Improvement*. London: Bloomsbury. https://doi.org/10.5040/9781472542014

Drugan, J., Strandvik, I., & Vuorinen, E. (2018). Translation Quality, Quality Management and Agency: Principles and Practice in the European Union Institutions. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications* (pp. 39–68). Cham: Springer. https://doi.org/10.1007/978-3-319-91241-7_3

Dzeguze, D., Melby, A., & Lommel, A. (2018, October). *Next Generation Quality: Assessing Translation Quality with ASTM WK46396*. https://locworld.com/wp-content/uploads/2018/10/TS2.pdf

ELIA, EMT, EUATC, FIT Europe, GALA, LIND, & Women in Localization. (2021). *ELIS 2021 - European Language Industry Survey* (p. 87). https://ec.europa.eu/info/sites/default/files/about_the_european_commission/service_s tandards_and_principles/documents/elis_2021_european_language_industry_survey.p df

Esselink, B. (2019). Multinational language service provider as user. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology* (pp. 109–126). London: Routledge. https://doi.org/10.4324/9781315311258-7

Fields, P., Hague, D. R., Koby, G. S., Lommel, A., & Melby, A. (2014). What Is Quality? A Management Discipline and the Translation Industry Get Acquainted. *Tradumàtica: Tecnologies de La Traducció*, *12*, 404. https://doi.org/10.5565/rev/tradumatica.75

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). *Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation*. http://arxiv.org/abs/2104.14478

Garvin, D. A. (1984). What Does "Product Quality" Really Mean? *Sloan Management Review*, *26*(1).

Gladkoff, S., Sorokina, I., Han, L., & Alekseeva, A. (2021). *Measuring Uncertainty in Translation Quality Evaluation (TQE)*. http://arxiv.org/abs/2111.07699

Görög, A. (2017). *The 8 most used standards and metrics for Translation Quality Evaluation*. TAUS Blog. https://blog.taus.net/the-8-most-used-standards-and-metrics-for-translation-quality-evaluation

Graham, Y., Baldwin, T., Moffat, A., & Zobel, J. (2013). Continuous measurement scales in human evaluation of machine translation. *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, 33–41. https://aclanthology.org/W13-2305.pdf

Han, L., Jones, G. J. F., & Smeaton, A. F. (2021). *Translation Quality Assessment: A Brief Survey on Manual and Automatic Methods*. https://arxiv.org/abs/2105.03311v1

Hatim, B., & Mason, I. (2014). Discourse and the translator. London: Routledge. https://doi.org/10.4324/9781315846583

Heyn, M. (2016). *Measuring translation quality in SDL Trados Studio 2015 with TQA*. https://www.trados.com/blog/measuring-translation-quality-in-sdl-trados-studio-2015-with-tqa.html

Holmes, J. (1988). The name and Nature of Translation Studies. In *Translated! Papers in Literary Translation and Translation Studies* (pp. 67–80). Leiden: Brill. https://doi.org/10.1163/9789004486669_008

Horguelin, P. A., & Brunette, L. (1998). Pratique de la révision, 3e édition revue et augmentée. Québec: Brossard

Horton, D. (1998). Translation assessment: Notes on the interlingual transfer of an advertising text. *IRAL - International Review of Applied Linguistics in Language Teaching*, *36*(2). https://doi.org/10.1515/iral.1998.36.2.95

House, J. (1981). A model for translation quality assessment. *Tubinger Beitrage Zur Linguistik*, *88*, 1-344. Tübingen: Narr

House, J. (1997). Translation quality assessment. A model revisited. Tübingen: Narr. https://books.google.co.uk/books?id=D16aYuTCBJ0C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

House, J. (2002). Translation Quality Assessment: Linguistic Description versus Social Evaluation. *Meta*, *46*(2), 243–257. https://doi.org/10.7202/003141ar

House, J. (2015). *Translation quality assessment: past and present*. London: Routledge

ISO. (2012). *ISO/TS 11669:2012 Translation projects - General guidance*. https://www.iso.org/standard/50687.html

ISO. (2015a). Quality management systems - Fundamentals and vocabulary. In *English* (Vol. ISO9000). https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en

ISO. (2015b). *Translation services — Requirements for translation services. ISO 17100:2015(E)*. *17100*(17100), 27. https://www.iso.org/obp/ui/#iso:std:iso:17100:ed-1:v1:en

ISO. (2017). *ISO 18587:2017 (en), Translation services — Post-editing of machine translation output — Requirements*. https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:en

ISO. (2019). *ISO 20539:2019 (en), Translation, interpreting and related technology — Vocabulary*. ISO. https://www.iso.org/obp/ui/#iso:std:iso:20539:ed-1:v1:en

ISO. (2021). *ISO/CD 5060 - Translation services — Evaluation of translation output — General guidance*. ISO. https://www.iso.org/standard/80701.html

ISO, & TMB. (2004). *Policy on Global Relevance*. https://share.ansi.org/ISOT/Other%20ISOT%20Resources/ISO%20Global%20Relevance%20Policy.pdf

Jiménez-Crespo, M. A. (2009). The evaluation of pragmatic and functionalist aspects in localization. *The Journal of Internationalization and Localization*, *1*, 60–93. https://doi.org/10.1075/jial.1.03jim

Johnson, T. (2021). Experiential Localization Quality Assessment. *Multilingual*. https://multilingual.com/issues/september-october-2021/experiential-localization-quality-assessment-xlqa-and-user-centricity-illustrated/

King, M., Popescu-Belis, A., & Hovy, E. (2003). FEMTI: creating and using a framework for MT evaluation. *Proceedings of Machine Translation Summit IX*. https://aclanthology.org/2003.mtsummit-papers.30.pdf

Koby, G. S., Fields, P., Hague, D. R., Lommel, A., & Melby, A. (2014). Defining Translation Quality. *Tradumàtica: Tecnologies de La Traducció*, *12*, 413. https://doi.org/10.5565/rev/tradumatica.76

Koehn, P., & Monz, C. (2006). Manual and automatic evaluation of machine translation between European languages. *Proceedings of the Workshop on Statistical Machine Translation (StatMT '06).* Association for Computational Linguistics, USA, 102–121. https://dl.acm.org/doi/10.5555/1654650.1654666

Koskinen, K. (2014). Translating Institutions: An ethnographic study of EU translation. London: Routledge. https://doi.org/10.4324/9781315759760

Kurz, C. (2020). Translation quality, translation evaluation, error categories and score sheets. In J.-M. Dalla-Zuanna & C. Kurz (Eds.), *Translation Quality in the Age of Digital Transformation* (pp. 38–68). Berlin: BDÜ Weiterbildung- und Fachverlagsgesellschaft

Larose, R. (1989). *Théories contemporaines de la traduction.* (2nd ed.). Québec: Presses de l'Université du Québec

Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., & Toral, A. (2020). A set of recommendations for assessing human-machine parity in language translation. In *Journal of Artificial Intelligence Research* (Vol. 67, pp. 653–672). AI Access Foundation. https://doi.org/10.1613/JAIR.1.11371

Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet? *Translator*, *6*(2), 149–168. https://doi.org/10.1080/13556509.2000.10799063

Lavie, A., & Agarwal, A. (2007). METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. *Proceedings of the Second Workshop on Statistical Machine Translation*, *0*(June), 228–231. https://www.aclweb.org/anthology/W07-0734.pdf

Lommel, A. (2018). *Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies* (pp. 109–127). https://doi.org/10.1007/978-3-319-91241-7_6

Lommel, A., & Melby, A. K. (2018). *MQM-DQF: A Good Marriage (Translation Quality for the 21st Century)*. www.conference.amtaweb.org

Makoushina, J. (2007). Translation Quality Assurance Tools: Current State and Future Approaches. *Translating and the Computer*, *29*. https://aclanthology.org/2007.tc-1.8.pdf

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, *50*(1), 79–104. https://doi.org/10.1177/147078530805000107

Mateo, R. M. (2014). A deeper look into metrics for translation quality assessment (TQA): A case study. *Miscelanea*, *49*(49), 73–93. https://www.researchgate.net/publication/283399796_A_deeper_look_into_metrics_for_translation_quality_assessment_TQA_A_case_study

Melby, A., Fields, P., Hague, D. R., Koby, G. S., & Lommel, A. (2014). Defining the Landscape of Translation. *Tradumàtica: Tecnologies de La Traducció*, *12*, 392. https://doi.org/10.5565/rev/tradumatica.74

Mellinger, C. D. (2018). Re-thinking translation quality. *Target. International Journal of Translation Studies*, *30*(2), 310–331. https://doi.org/10.1075/target.16104.mel

Memsource. (2021). *Memsource Launches Comprehensive Language Quality Assessment (LQA) Features*. https://www.memsource.com/blog/memsource-launches-comprehensive-language-quality-assessment-features/

Nida, E. (1964). *Toward a Science of Translation*. Leiden: Brill

Nimdzi. (2021). *Nimdzi 100 2021: Ranking of Top Language Service Companies*. https://www.nimdzi.com/nimdzi-100-top-lsp/

Nishino, R. (2020). Translation Quality and JTF's Evaluation Guidelines. In J.-M. Dalla-Zuanna & D. C. Kurz (Eds.), *Translation Quality in the Age of Digital Transformation* (pp. 154–164). Berlin: BDÜ Weiterbildung- und Fachverlagsgesellschaft

Nord, C. (1991). *Textanalyse und Übersetzen: theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*. Heidelberg: Groos

Nord, C. (1996). Text type and translation method: An objective approach to translation criticism. *Translator*, *2*(1). https://doi.org/10.1080/13556509.1996.10798965

Nord, C. (1997). Translating as a Purposeful Activity. Functionalist Approaches Explained (Translation Theories Explained 1). Manchester: St. Jerome

O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *Journal of Specialised Translation*. https://www.jostrans.org/issue17/art_obrien.pdf

Oates, B. J. (2006). *Researching Information Systems and Computing*. London: SAGE Publications Ltd.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). BLEU: a method for automatic evaluation of machine translation. *ACL*, 311–318. https://doi.org/10.3115/1073083.1073135

Parra-Galiano, S. (2005). *La revisión de traducciones en traductología: aproximación a la práctica de la revisión en el ámbito profesional mediante el estudio de casos y propuestas de investigación*. http://hdl.handle.net/10481/660

Petrova, V. (2019). Translation Quality Assessment Tools and Processes in Relation to CAT Tools. *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT) 2019*, 89–97. https://doi.org/10.26615/issn.2683-0078.2019_011

Pym, A. (2009). Exploring translation theories. London: Routledge https://doi.org/10.4324/9780203869291

QT21 Consortium. (2015). *QT21 – Quality Translation 21*. https://www.qt21.eu/

Reiss, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. Munich: Hueber

Reiss, K. (2014). Translation criticism-potentials and limitations: Categories and criteria for translation quality assessment. In *Translation Criticism - Potentials and Limitations: Categories and Criteria for Translation Quality Assessment*. London: Routledge. https://doi.org/10.4324/9781315760407

Reiss, K., & Vermeer, H. J. (1984). *Grundlegung einer Allgemeinen Translationstheorie*. Tübingen: Niemeyer

SAE. (2001). SAE. The Engineering Society for Advancing Mobility Land Sea Air and Space. https://www.sae.org/

Salmons, J. (2010). *Online interviews in real time*. London: SAGE Publications Ltd.

Schäffner, C. (1997). From "Good" to "Functionally Appropriate": Assessing Translation Quality. *Current Issues In Language and Society*, *4*(1), 1–5. https://doi.org/10.1080/13520529709615476

Schäffner, C. (2011). Theory of translatorial action. In G. Yves & Van D. Luc (Eds.), *Handbook of Translation Studies 2* (pp. 157–162). Amsterdam: John Benjamins. https://doi.org/10.1075/hts.2.the1

Secară, A. (2005). Translation Evaluation - a State of the Art Survey. *Proceedings of the ECoLoRe/MeLLANGE Workshop, Leeds*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.3654&rep=rep1&type=pdf

Slator (2021). *Localization Industry and Translation Research for 2020*. https://slator.com/data-research/

Snow, T. A. (2015). *Establishing the Viability of the Multidimensional Quality Metrics Framework*. Provo: Brigham Young University. https://scholarsarchive.byu.edu/etd/5593

Specia, L., Raj, D., & Turchi, M. (2010). Machine translation evaluation versus quality estimation. *Machine Translation*, *24*(1), 39–50. https://doi.org/10.1007/s10590-010-9077-2

Specia, L., & Shah, K. (2018). Machine Translation Quality Estimation: Applications and Future Perspectives. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: from principles to practice* (pp. 201–235). Cham: Springer. https://doi.org/10.1007/978-3-319-91241-7_10

Stejskal, J. (2009). Quality Assessment in Translation. *MRA's Alert! Magazine*, 18–22. https://www.cetra.com/wp-content/uploads/Files/Alert_05_07.pdf

Strandvik, I. (2017). Evaluation of outsourced translations. State of play in the European Commission's Directorate-General for Translation (DGT). In T. Svoboda, Ł. Biel, & K. Łoboda (Eds.), *Quality aspects in institutional translation* (pp. 123–137). Berlin: Language Science Press. https://doi.org/10.5281/zenodo.1048172

Sun, S., Guzmán, F., & Specia, L. (2020). Are we Estimating or Guesstimating Translation Quality? *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6262–6267. https://doi.org/10.18653/v1/2020.acl-main.558

Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electronic Journal*, 5. https://doi.org/10.2139/ssrn.3205035

TAUS. (2015). *TAUS*. Harmonized DQF-MQM Error Typology. https://www.taus.net/qt21-project#error-typology-template

Thelen, M. (2008). Translation Quality Assessment or Quality Management & Quality Control of Translation? *Translation and Meaning*, *Part 8*(January 2008), 411–424. https://www.researchgate.net/publication/316460849_TRANSLATION_QUALITY_ASSESSMENT_OR_QUALITY_MANAGEMENT_QUALITY_CONTROL_OF_T RANSLATION

Toral, A., Castilho, S., Hu, K., & Way, A. (2018). *Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation*. *1*, 113–123. https://doi.org/10.18653/v1/w18-6312

Tranquality. (2021). *Tranquality | Translation Quality*. https://www.tranquality.info/

Vela-Valido, J. (2021). Translation Quality Management in the AI Age: New technologies to Perform Translation Quality Management Operations. *Revista Tradumàtica. Tecnologies de La Traducció*, *18*, 128–146. https://doi.org/https://doi.org/10.5565/rev/tradumatica.285

Vermeer, H. J. (1996). *A Skopos Theory of Translation (Some arguments for and against)*. Heidelberg: TEXTconTEXT

Vilar, D., Leusch, G., Ney, H., & Banchs, R. E. (2007). Human Evaluation of Machine Translation Through Binary System Comparisons. *Proceedings of the Second Workshop on Statistical Machine Translation*, 96–103. https://aclanthology.org/W07-0713

W3C. (2021a). *Multidimensional Quality Metrics (MQM) Community Group*. W3C. https://www.w3.org/community/mqmcg/

W3C. (2021b). *Multidimensional Quality Metrics (MQM) Terminology*. W3C. https://www.w3.org/community/mqmcg/mqm-terminology/

Way, A. (2018). Quality Expectations of Machine Translation. In Moorkens J., S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment. Machine Translation: Technologies and Applications* (1st ed., pp. 159–178). Cham: Springer. https://doi.org/10.1007/978-3-319-91241-7_8

White, J. S., Connell, T. O., & Mara, F. O. (1994). The ARPA MT Evaluation Methodologies: Evolution, Lessons and Future Approaches. *Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA 1994)*, 193–205. https://aclanthology.org/1994.amta-1.25.pdf

Williams, M. (1989). The Assessment of Professional Translation Quality: Creating Credibility out of Chaos. *TTR : Traduction, Terminologie, Rédaction*, *2*(2), 13. https://doi.org/10.7202/037044ar

Williams, M. (2001). The Application of Argumentation Theory to Translation Quality Assessment. *Meta*, *46*(2), 326–344. https://doi.org/10.7202/004605ar

Williams, M. (2004). Translation quality assessment: An argumentation-centred approach. In *Translation Quality Assessment: An Argumentation-Centred Approach*. Ottawa: University of Ottawa Press

XTM International. (2021). *TAUS Dynamic Quality Framework - XTM International*. https://xtm.cloud/taus/

Ziganshina, L. E., Yudina, E. V., Gabdrakhmanov, A. I., & Ried, J. (2021). Assessing human post-editing efforts to compare the performance of three machine translation engines for english to Russian translation of cochrane plain language health information: Results of a randomised comparison. *Informatics*, *8*(1), 9. https://doi.org/10.3390/informatics8010009

8 Appendices

Appendix A

This appendix contains a list of publications and presentations where parts of this research have previously been made public.

Title: "New Methodologies Applied to the Linguistic Quality Assurance in Multilingual Corporations"

Authors: Jennifer Vela Valido

Venue: LocWorld40 Conference, "Go Global, Be Global" in Lisbon, Portugal.

Date presented: June 13th, 2019.

Title: "Translation Quality Management in the AI Age New technologies to Perform Translation Quality Management Operations"

Authors: Jennifer Vela Valido

Date presented: July 1st 13th, 2021

Publication: Vela-Valido, J. (2021). Translation Quality Management in the AI Age: New technologies to Perform Translation Quality Management Operations. Revista Tradumàtica. Tecnologies de La Traducció, 18, 128–146. https://doi.org/https://doi.org/10.5565/rev/tradumatica.285.

Appendix B

This appendix contains the participant information sheet and consent form used for the online questionnaires and interviews of this research.

Translation Quality Management methodologies and technologies

Welcome to the 2021 Translation Quality Management methodologies and technologies study. The goal of this academic research initiative is to have an overview of the different approaches, tools, and processes adopted by companies, translation service providers and subject matter experts to manage and evaluate translation quality. The results will be shared with the participants and the wider public to contribute to the advancement of this discipline in the translation and localisation industry.

If you are involved in translation quality management processes, we would love to hear from you! We are looking for established professionals from all over the world, as well as consultants and researchers.

This survey is divided into 5 parts and will take 30 minutes to complete. Alternatively, you can schedule an online interview through this link.

This survey will close on the 15th of October 2021.

To thank you for your time, we will send you a complimentary summary of the results BEFORE the publication of the research in a public repository. You will find the instructions to receive the summary at the end of this survey.

We also welcome you to contact us if you have any questions: jennifer.vela101@alu.ulpgc.es.

<div align="center">ONLINE CONSENT FORM</div>

INFORMATION REGARDING THE STUDY AND INFORMED CONSENT *

Thank you for your participation in this academic study. This questionnaire is part of the doctoral study of the researcher Jennifer Vela Valido, a PhD candidate at the University of Las Palmas de Gran Canaria (Spain), under the supervision of Dr. Laura Cruz (ULGPC) and Dr. Jorge Díaz Cintas (University College London). Your information is strictly confidential and will be only used for academic and research purposes. This research will not store or publish personally identifiable information such as personal names or company names. All the responses will be anonymised and aggregated. No risks are associated with this study. You can withdraw your answers at any point within the time that the questionnaire is online; without providing any explanations nor bearing any consequences. By clicking "Yes" you agree that you have read and understood the above and you will proceed to the following step of this study.

- Yes
- No

Appendix C

This appendix contains the design of each questionnaire and the research questions each section aims to address. Open-ended questions appear marked as (O), while closed-ended questions have (C) at the end.

Questionnaire 1: Buyers of translation services

Research questions	Questionnaire design
1 Understand the profile of the respondent and the demographics (generic category for all groups)	1.1 What is your job title? (C) 1.2 Which description best fits you? (C) 1.3 What is the size of the staff (FTE) of your company? (C) 1.4 In which sector does your company operate? (C) 1.5 What is your company's annual revenue? (C) 1.6 What are the annual translation volumes of your organisation? (O)
2 Understand the level of maturity in terms of localization and analytics	2.1 How would you score the actual level of maturity of your organization in terms of translation and localization processes? (C) 2.2 How would you score the actual level of maturity of your organization in terms of translation and localization analytics? (C)
3 Understand the level of maturity in terms of translation quality management processes	
3.1 Overall level of maturity in translation quality management	3.1.1 How would you score the actual level of maturity of your organization in terms of translation and localization quality management processes? (C)
3.2 Knowledge and use of norms and certifications	3.2.1 Which ones of the following norms and standards related to translation or quality management has your organisation incorporated in their translation quality management program (completely or partially)? (C) 3.2.2 Which ones of the following translation quality metrics has your organisation incorporated in their translation quality management program? (C) 3.2.3 What is the reason your organisation decided to use a proprietary or customised set of metrics? (C)
3.3 Documented processes and resources	**Translation Quality Management Processes** 3.3.1 What happens from the moment a translation request is received until the translation quality management workflow of that request is considered closed? If there are different quality management workflows depending on the context,

the type of content or the type of translation, please briefly describe the 3 most common scenarios. (O)

3.3.2 What types of content are included in the translation quality management process? (C)

3.3.3 What percentage of the translated content is included in the translation quality management workflow? (O)

3.3.4 Do you consider different percentages depending on the context of the translation request, the type of the text or other factors? If so, how are these percentages distributed? (O)

Translation Quality Evaluation methodologies

3.3.5 What translation quality evaluation methodologies are used in your organisation? (C)

3.3.6 Does your organisation sometimes perform holistic translation quality evaluation instead of analytic translation quality evaluation, or the other way round? If yes, under what circumstances? (O)

3.3.7 Does your organisation perform holistic translation quality evaluation on the same texts that are submitted for analytic translation quality evaluation? If yes, under what circumstances? (O)

3.3.8 Does your organisation assess the reliability of the translation quality evaluation used? If yes, how? (O)

3.3.9 How would you rate the current level of confidence in the quality scores? (C)

3.3.10 Are there any situations in which the translation quality evaluation system your organisation has in place is assessed as not being applicable or reliable? Which ones? (O)

Translation Quality Evaluation Processes

3.3.11 What percentage of the translated content is evaluated and how frequently are translation quality evaluations performed?

3.3.12 Do you perform quality evaluations before the delivery of the translation test to the stakeholder, after, or both? If both, what is the percentage of each type of evaluation and why?

3.3.13 Do you evaluate whole texts or select a sample for evaluation?

3.3.14 If you do sampling, what sampling criteria do you use? For example, text size, text types, end-user profiles, type of resource used, type of translation quality requirement.

3.3.15 Are samples selected to evaluate random, representative text passages, or are they selected to focus on high-priority text passages? Or both?

3.3.16 Do you consider other evaluation scenarios depending on the context of the translation request, the type of the text

	or other factors? If so, what are these different evaluation workflows?
	3.3.17 Are translation requirements designed to be aligned with translation grades or translation tiers? If so, what are the translation grades or tiers used?
	3.3.18 Are different translation grades or tiers associated with different text types? If so, which ones?
	3.3.19 Are different evaluation metrics (error types, severity levels, severity multipliers) applied to different text types?
	3.3.20 What does the translation quality evaluation scoring model look like? What was its origin?
	3.3.21 What error types does it include? For example: Grammar, Terminology, Mistranslation...
	3.3.22 What severity levels does it include? For example: Minor, major, critical.
	3.3.23 Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)?
	3.3.24 Does the evaluation include a pass/fail rating? How is that determined? What are the quality thresholds in pass/fail assessments?
	3.3.25 Does the scoring model have one uniform metric or are there different metrics for different content types?
	Resources
	3.3.26 What kind of resources are involved in the translation quality management process and what is their role? Translators, quality managers, evaluators, subject matter experts, others?
	3.3.27 How many of these resources are internal and how many are external? Is there any specific reason for this decision?
	3.3.28 Taking into account the roles mentioned above, whose translations are being evaluated by whom? Are there any differences depending on whether the resources are internal or external?
	3.3.29 What are the job qualifications of the evaluators? For example, are they required to be qualified as experienced translators and/or revisers?
	3.3.30 What are the most relevant job qualifications of the translation quality managers or the resources in charge of the translation quality management program?
	3.3.31 What kind of training and documentation are provided to develop and support evaluators and/or quality managers?
3.4 Tools and technologies	3.4.1 What kind of tools does your organisation use for translation quality management purposes? (C)

	3.4.2 What is the environment or tool where the translation evaluation processes, in particular, error annotation, take place? (C) 3.4.3 How are evaluations of translation samples processed through the CAT/TMS tool or translation quality evaluation tool? (C) 3.4.4 Please, reply to this question if your organisation performs translation quality evaluation inside a CAT/TMS tool. Select which ones of the following features are included in the tool used for translation quality evaluation? (C) 3.4.5 Please, reply if your organisation performs translation quality evaluation using a dedicated quality tool outside CAT/TMS tool. Select which ones of the following features are included in the tool used for translation quality evaluation? (C)
4 Understand the main pain-points that still remain, their impact and priority	4.1 On a scale of 1 to 5, where 1 is completely unsatisfied and 5 is completely satisfied, how would you rate the level of adequacy of the current translation management process in your organisation according to the following parameters (C) 4.2 What are the 3 main pain points that your organisation is currently facing when it comes to translation quality management? (O)
5 Understand the workarounds and solutions used to address the remaining pain-points	5.1 What are the partial solutions or workarounds that your organisation has applied to the three pain points mentioned before? (O)
6 Understand the relevance and gaps of the current norms, certifications and practices used for TQM	6.1 What are the 3 main challenges that your organisation will probably have to face in the near future when it comes to translation quality management processes? (O) 6.2 Are there any tools, initiatives or innovation initiatives your organisation is planning to adopt to enhance their translation quality management program? Which ones? 6.3 What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes? Please, tick all the ones that apply:(C)

Questionnaire 2: Translation service providers

GOAL	QUESTIONS
1 Understand the profile of the respondent and the demographics (generic category for all groups)	1.1 What is your job title? (C) 1.2 Which description best fits you? (C) 1.3 What is the size of the staff (FTE) of your company? (C) 1.4 In which sector does your company operate? (C) 1.5 What is your company's annual revenue? (C) 1.6 What are the annual translation volumes of your organisation? (O)
2 Understand the level of maturity in terms of localization and analytics	2.1 How would you score the actual level of maturity of your organization in terms of translation analytics? (C)
3 Understand the level of maturity in terms of translation quality management processes	
3.1 Overall level of maturity in translation quality management	3.1.1 How would you score the actual level of maturity of your organization in terms of translation and localization quality management processes? (C)
3.2 Knowledge and use of norms and certifications	3.2.1 Which ones of the following norms and standards related to translation or quality management has your organisation incorporated in their translation quality management program. Please, indicate if your company is certified in any of them (C) 3.2.2 Which ones of the following translation quality metrics has your organisation incorporated in their translation quality management program? (C) 3.2.3 What is the reason your organisation decided to use a proprietary or customised set of metrics? (C)
3.3 Documented processes and resources	**Translation Quality Management Processes** 3.3.1 What happens from the moment a translation request is received until the translation quality management workflow of that request is considered closed? If there are different quality management workflows depending on the context, the type of content or the type of translation, please briefly describe the 3 most common scenarios. (O) 3.3.2 What types of content are included in the translation quality management process? (C) 3.3.3 What percentage of the translated content is included in the translation quality management workflow? If the percentage varies depending on the client or other factors, please briefly describe the 3 most common scenarios. (O)

	3.3.4 Do you consider different percentages depending on the context of the translation request, the type of the text or other factors? If so, how are these percentages distributed? (O)
	3.3.5 Do you have a general translation quality program that applies to all your clients or do you adapt it to each client? If so, what percentage of clients require a specific translation quality management? (O)
	3.3.6 In general, would you say that your clients' requirements and expectations help you build your translation quality management program or rather limit it?
	Translation Quality Evaluation methodologies
	3.3.7 What translation quality evaluation methodologies are used in your organisation? (C)
	3.3.8 Does your organisation sometimes perform holistic translation quality evaluation instead of analytic translation quality evaluation, or the other way round? If yes, under what circumstances? (O)
	3.3.9 Does your organisation perform holistic translation quality evaluation on the same texts that are submitted for analytic translation quality evaluation? If yes, under what circumstances? (O)
	3.3.10 Does your organisation assess the reliability of the translation quality evaluation method used? If yes, how? (O)
	3.3.11 How would you rate the current level of confidence in the quality scores of the quality evaluation method used? (C)
	3.3.12 Are there any situations where the translation quality evaluation system your organisation has in place is assessed as not being applicable or reliable? Which ones? (O)
	Translation Quality Evaluation Processes
	3.3.13 What percentage of the translated content is evaluated and how frequently are translation quality evaluations performed?
	3.3.14 Do you perform quality evaluations before the delivery of the translation test to the stakeholder, after, or both? If both, what is the percentage of each type of evaluation?
	3.3.15 Do you evaluate whole texts or select a sample for evaluation?
	3.3.16 If you do sampling, what sampling criteria do you use? For example, text size, text types, end user profiles, type or resource used, type of translation quality requirement.
	3.3.17 Are samples selected to evaluate random, representative text passages, or are they selected to focus on high-priority text passages? Or both?

	3.3.18 Are translation requirements designed to be aligned with translation grades or translation tiers? If so, what are the translation grades or tiers used? 3.3.19 Are different translation grades or tiers associated with different text types or clients? If so, which ones? 3.3.20 Are different evaluation metrics (error types, severity levels, severity multipliers) applied to different text types or clients? 3.3.21 What does the translation quality evaluation scoring model look like? What was its origin? 3.3.22 What error types does it include? For example: Grammar, Terminology, Mistranslation... 3.3.23 What severity levels does it include? For example: Minor, major, critical. 3.3.24 Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)? 3.3.25 Does the evaluation include a pass/fail rating? How is that determined? What are the quality thresholds in pass/fail assessments? 3.3.26 Does the scoring model have one uniform metric, or are there different metrics for different content types? **Resources** 3.3.27 What kind of resources are involved in the translation quality management process and what is their role? Translators, quality managers, evaluators, subject matter experts, others? 3.3.28 What percentage of these resources are internal vs external? Is there any specific reason for this decision? 3.3.29 Taking into account the roles mentioned above, whose translations are being evaluated by whom? Are there any differences depending on whether the resources are internal or external? 3.3.30 What are the job qualifications of the evaluators? For example, are they required to be qualified as experienced translators and/or revisers? 3.3.31 What are the most relevant job qualifications of the translation quality managers or the resources in charge of the translation quality management program? 3.3.32 What kind of training and documentation are provided to develop and support evaluators and/or quality managers?
3.4 Tools and technologies	3.4.1 What kind of tools does your organisation use for translation quality management purposes? (C) 3.4.2 What is the environment or tool where the translation evaluation processes, in particular, error annotation, take place? (C)

	3.4.3 How are evaluations of translation samples processed through the CAT/TMS tool or translation quality evaluation tool? (C) 3.4.4 Please, reply to this question if your organisation performs translation quality evaluation inside a CAT/TMS tool. Select which ones of the following features are included in the tool used for translation quality evaluation? (C) 3.4.5 Please, reply if your organisation performs translation quality evaluation using a dedicated quality tool outside CAT/TMS tool. Select which ones of the following features are included in the tool used for translation quality evaluation? (C)
4 Understand the main pain-points that still remain, their impact and priority	4.1 On scale of 1 to 5, where 1 is completely unsatisfied and 5 is completely satisfied, how would you rate the level of adequacy of the current translation management process in your organisation according to the following parameters (C) 4.2 What are the 3 main pain-points that your organisation is currently facing when it comes to translation quality management? (O)
5 Understand the workarounds and solutions used to address the remaining pain-points	5.1 What are the partial solutions or workarounds that your organisation has applied to the three pain-points mentioned before? (O)
6 Understand the relevance and gaps of the current norms, certifications and practices used for TQM	6.1 What are 3 main challenges that your organisation will probably have to face in the near future when it comes to translation quality management processes? (O) 6.2 Are there any tools, initiatives or innovation initiatives your organisation is planning to adopt to enhance their translation quality management program? Which ones? 6.3 What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes mentioned above? Please, tick all the ones that apply: (C)

Questionnaire 3: Translation technology companies

GOAL	QUESTIONS
1 Understand the profile of the respondent and the demographics (generic category for all groups)	1.1 What is your job title? (C) 1.2 Which description best fits you? (C) 1.3 What is the size of the staff (FTE) of your company? (C) 1.4 In which sector does your company operate? (C) 1.5 What is your company's annual revenue? (C) 1.6 What are the annual translation volumes of your organisation? (O)
2 Translation quality management processes	
2.1 Knowledge and use of norms and certifications	2.1.1 Which ones of the following norms and standards related to translation or quality management have been taken into account in the design of the translation technology solutions offered by your company (either completely or partially)? (C) 2.1.2 Which ones of the following translation quality metrics have been taken into account or incorporated in the design of the translation technology solutions offered by your company (either completely or partially)? (C)
2.2 Documented processes and resources	**Translation Quality Evaluation methodologies** 2.2.1 What translation quality evaluation methodologies are supported by your technology solution? (C) 2.2.2 Are there any situations in which the translation quality management or the translation quality tools provided by your company have been assessed as not being suited to the needs of your clients, or not reliable enough? If yes, can you please explain why? (O) **Translation Quality Evaluation Processes** 2.2.3 Does your translation quality management solution allow the evaluation of whole texts or select a sample for evaluation? 2.2.4 If your tool allows sampling, what sampling criteria can be used? For example, text size, text types, end-user profiles, type of resource used, type of translation quality requirement. 2.2.5 Is it possible to apply different evaluation metrics (error types, severity levels, severity multipliers) to different text types or user profiles? 2.2.6 What does the translation quality evaluation scoring model look like? Is it based on certain standards or guidelines?

	2.2.7 What error types does it include? For example: Grammar, Terminology, Mistranslation... 2.2.8 What severity levels does it include? For example: Minor, major, critical. 2.2.9 Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)? 2.2.10 Does the evaluation include a pass/fail rating? How is that determined? What are the quality thresholds in pass/fail assessments? **Resources** 2.2.11 How many different user roles are available in the translation quality management solution? Translators, quality managers, evaluators, subject matter experts, others?
2.3 Tools and technologies	2.3.1 How would you define the technical setting of the solution you provide for translation quality management purposes? (C) 2.3.2 How would you define the environment or tool where the translation evaluation processes, in particular, error annotation, take place? (C) 2.3.3 Please, reply to this question if your technology solution allows performing a translation quality evaluation inside a CAT/TMS tool. Select which ones of the following features are included in the tool? (C) 2.3.4 Please, reply if your technology solution allows performing a translation quality evaluation outside a CAT/TMS tool. Select which ones of the following features are included in the tool: (C)
3 Understand the main pain-points that still remain, their impact and priority	3.1 On a scale of 1 to 5, where 1 is completely unsatisfied and 5 is completely satisfied, how would you rate the level of maturity of the translation management solution provided by your company, according to the following parameters: (C) 3.2 What are the 3 main pain points that your technology solution is currently working on to address when it comes to translation quality management processes? (O)
4 Understand the workarounds and solutions used to address the remaining pain-points	4.1 What are the partial solutions or workarounds that your technology solution has applied to the main 3 pain points mentioned above? (O)
5 Understand the relevance and gaps of the current	5.1 What are the 3 main challenges that your technology solution will probably have to face in the near future

norms, certifications and practices used for TQM	when it comes to translation quality management processes and technologies? (O) 5.2 Are there any technologies or innovation initiatives your company is planning to adopt to enhance your technology solution? Which ones? (O) 5.3 What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes? Please, tick all the ones that apply(C)

Questionnaire 4: Experts in translation quality management

GOAL	QUESTIONS
1 Understand the profile of the respondent and the demographics (generic category for all groups)	1.1 What is your job title? (C) 1.2 Which description best fits you? (C)
2 Translation quality management processes	
2.1 Knowledge and use of norms and certifications	2.1.1 Which ones of the following norms and standards related to translation or quality management are you familiarised with? (C) 2.1.2 Which ones of the following translation quality metrics are you familiarised with? (C)
2.2 Documented processes and resources	**Translation Quality Management Processes** 2.2.1 What are the most common scenarios when it comes to your role as a translation quality management expert? Please, select a maximum of 3. (C) 2.2.2 In your experience, would you say that the clients' requirements and expectations help build sound translation quality management programs or rather limit them? (O) **Translation Quality Evaluation methodologies** 2.2.3 What type of translation quality evaluation methodologies are you familiarised with? (C) 2.2.4 Are there any situations in which you would recommend or perform a holistic translation quality evaluation instead of analytic translation quality evaluation, or the other way round? If yes, what would be those situations? 2.2.5 Are there any situations in which you would recommend or perform a holistic translation quality evaluation and an analytic translation quality evaluation on the same texts? If yes, what would be those situations? 2.2.6 Do you or the companies you work with assess the reliability of the translation quality evaluation used? If yes, how? 2.2.7 Have there been any situations in which the translation quality management methodology recommended or used by you has been assessed as not being suited to the needs of your clients, or not reliable enough? If yes, can you please explain why? **Translation Quality Evaluation Processes**

	2.2.8 In your opinion, what percentage of the content translated by a company or a language service should be evaluated and how frequently should the translation quality evaluations be performed?
	2.2.9 In your opinion, should quality evaluations be performed before the delivery of the translation to the stakeholder, after, or both? If both, what would be the ideal percentage of each type of evaluation?
	2.2.10 Do you think that translation quality assurance (also called "review step") should be or could be performed at the same time as the translation quality evaluation step? If yes, in which situations?
	2.2.11 In which situations do you think it would be advisable to use sampling techniques rather than evaluate whole tests?
	2.2.12 If you recommend sampling, what criteria is more effective in your opinion? For example, text size, text types, end-user profiles, type of resource used, type of translation quality requirement...
	2.2.13 What standards or guidelines do you consider more relevant in order to build a translation quality evaluation scoring model?
	Resources
	2.2. 14 What kind of resources do you think should be involved in the translation quality management process and in what capacity? For example translators, quality managers, evaluators, subject matter experts, others?
	2.2.15 Taking into account the roles mentioned above, whose translations should be evaluated by whom? Are there any differences depending on whether the resources are internal or external?
	2.2.16 What would be the ideal job qualifications of the evaluators? For example, should they be qualified as experienced translators and/or revisers?
	2.2.17 In your opinion, what are the most relevant job qualifications of the translation quality managers or the resources in charge of the translation quality management program?
	In your opinion, what kind of training and documentation should be provided to develop and support evaluators and/or quality managers?
2.3 Tools and technologies	2.3.1 What kind of tools do you use/recommend for translation quality management purposes? (C)
	2.3.2 What environment or tool do you use/recommend for translation evaluation processes, in particular, error annotation? (C)

	2.3.3 What method do you use/recommend to perform evaluations of translation samples processed through a CAT/TMS tool or translation quality evaluation tool?(C) 2.3.4 Please, reply to this question if you perform or recommend performing translation quality evaluation inside a CAT/TMS tool. Select which ones of the following features are included in the tool recommended for translation quality evaluation:(C) 2.3.5 Please, reply to this question if you perform or recommend performing translation quality evaluation outside a CAT/TMS tool. Select which ones of the following features are included in the tool recommended for translation quality evaluation:(C)
3 Understand the main pain-points that still remain, their impact and priority	3.1 On a scale of 1 to 5, where 1 is completely unsatisfied and 5 is completely satisfied, how would you rate the level of maturity of the current translation management frameworks used in the industry, according to the following parameters: (C) 3.2 In your opinion, what are the 3 main pain points when it comes to translation quality management? (O)
4 Understand the workarounds and solutions used to address the remaining pain-points	4.1 What are the partial solutions or workarounds to the main 3 pain-points mentioned above that you have applied or recommended? (O)
5 Understand the relevance and gaps of the current norms, certifications and practices used for TQM	5.1 In your opinion, what are the 3 main challenges that the industry will be facing in the near future when it comes to translation quality management processes and technologies? (O) 5.2 Are there any technologies, tools or innovation initiatives you are researching or are planning to adopt to enhance the current translation quality management processes? Which ones? (O) 5.3 What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes? Please, tick all the ones that apply: (C)

Appendix D

This appendix contains the graphic design and the format of the different types of open and close ended questions used in the survey part of this research.

Demographics (common section for all groups)

Part 1: Demographics

What is your job title? *

- ○ Translation/Localisation Quality Manager
- ○ Translation/Localisation Manager or Director
- ○ Language services Manager or Director
- ○ Translation/Localisation Program Manager
- ○ CEO
- ○ Researcher (academic)
- ○ Researcher (non-academic)
- ○ Independent consultant
- ○ Other: _____

What is the size of the staff of your organisation? *

- ○ 1-25
- ○ 26-49
- ○ 50-99
- ○ 100-199
- ○ 200+
- ○ Not sure/Not applicable

In which sector(s) does your organisation operate? *

- ☐ Aerospace & Defense
- ☐ Engineering & Manufacturing
- ☐ Financial Services
- ☐ Gaming
- ☐ Legal
- ☐ Life Sciences
- ☐ Media
- ☐ Professional Services
- ☐ Public Sector
- ☐ Technology
- ☐ Travel & Retail
- ☐ Not sure/Not applicable
- ☐ Academia
- ☐ Other: _____

What is your organisation's annual revenue? *

◯ Under $1 million

◯ $1 million - $4.9 million

◯ $5 million - $9.9 million

◯ $10 million -$25 million

◯ Over $26 million

◯ Not sure/Not applicable

What are the annual translation volumes of your organisation? *

◯ Under 500k words

◯ Between 500k and 10M words

◯ Between 10M and 100M words

◯ Over 100M words

◯ Not sure/Not applicable

Which description best fits you? *

If you hold multiple jobs within the language services industry, answer from the perspective of the role where you are more involved with translation quality management processes and methodologies. Please, answer the rest of the survey with that role in mind.

◯ I am part of an Internal Corporate Translation/Localisation Team (client or buyer)

◯ I work for a Translation or Language Services provider (services provider)

◯ I work for a Language Technology Company (technology provider)

◯ I offer consulting or services in translation quality management

◯ I work as a teacher or researcher

Back   Next   Clear form

Questionnaire 1: Buyers of translation services

How many languages/locales does your organisation work with or outsource? *

○ Less than 5

○ Between 5 and 10

○ More than 10

Part 2: Level of maturity in translation processes and analytics

How would you score the actual level of maturity of your organisation in terms of translation and localisation processes? *

○ 1 - Reactive: My organisation responds to multicultural opportunities on an ad hoc basis (no formal processes, technology, or staff in place)

○ 2 - Repeatable: My organisation has some processes for core localisation tasks and recognises that someone needs to be in charge

○ 3 - Managed: My organisation has a translation team (internal or external), a consistent use of basic processes and some specialized language tools

○ 4 - Optimised: My organisation has deployed shared language services enterprise-wide, and the localisation function is centralised and supported by translation management systems

○ 5 - Transparent: My organisation treats localisation as just another business process to be managed and continuously improved. Global customers and their requirements are integrated into all strategic and operational planning

---

How would you score the actual level of maturity of your organisation in terms of translation analytics? *

○ 0 - Inactive: My organisation does not track translation activity

○ 1 - Reactive: My organisation has historical data reports focused on costs, compiled manually

○ 2 - Descriptive: My organisation has operational reports on costs and performance based on historical data with manual consolidation data

○ 3 - Diagnostic: My organisation has dashboards indicating errors or problem conditions, there is some loose or manual integration of multiple data sources

○ 4 - Predictive: My organisation has fully automated predictive analytics that indicate what is likely to happen, there is integration with multiple data sources (business analytics, reporting and visual dashboards)

○ 5 - Prescriptive: My organisation has real-time analysis that shapes immediate and future actions, and it may count with data acquired from external sources

Part 3: Level of maturity in translation quality management processes

3.1 Overall maturity in translation quality management

How would you score the actual level of maturity of your organisation in terms of translation and localisation quality management processes? *

○ 1 - Manual: My organisation uses manual and/or legacy processes

○ 2 - Automated: My organisation has started to integrate some automation in the quality management workflow

○ 3 - Agile: My organisation has automated and scalable quality management processes across multiple languages

○ 4 - Centralised: My organisation practices strategic and centralised translation management processes across multiple channels and languages

○ 5 - Expert: My organisation benefits from translation as a revenue center and is using translation quality management metrics for growth

3.2 Norms, standards and metrics

Which ones of the following norms and standards related to translation or quality management has your organisation incorporated in their translation quality management program (completely or partially)? *

☐ ISO 17100 (Translation services — Requirements for translation services)

☐ ISO 18587 (Translation services — Post-editing of machine translation output — Requirements)

☐ ISO 9001 (Quality management systems — Requirements)

☐ ISO 9000 (Quality management systems — Fundamentals and vocabulary)

☐ ASTM F2575 (Standard Guide for Quality Assurance in Translation)

☐ None

☐ Other:

Please, reply to this question if you selected "None" above. What is the reason your organisation hasn't incorporated any of the norms or standards mentioned in their translation quality program? *

☐ My organisation is not aware of the existence of those standards

☐ The standards are not flexible or customisable enough

☐ The standards are too complex or difficult to apply

☐ None of the standards satisfied the specific needs of my organisation

☐ Not applicable

☐ Other:

Which ones of the following translation quality metrics has your organisation incorporated in their translation quality management program? *

☐ LISA QA Model

☐ SAE J2450

☐ MQM-DQF

☐ Automatic metrics for Machine Translation (BLEU, METEOR, others)

☐ My organisation uses a proprietary set of metrics that is not based on any specific standard

☐ None

☐ Other: _____

Please, reply to this question if you selected "None" above. What is the reason your organisation doesn't use quality metrics or decided to use a proprietary or customised set of metrics? *

☐ My organisation is not aware of the existence of standarised metrics

☐ The standarised metrics are not flexible or customisable enough

☐ The standarised metrics are too complex or difficult to apply

☐ None of the standarised metrics satisfy the specific needs of my organisation

☐ Not applicable

☐ Other: _____

3.3 Translation quality management processes

Are there different quality management workflows depending on factors such as the type of content or the type of translation? If yes, please, mention the most common scenarios. *

Your answer

What types of content are included in the translation quality management process? *

☐ Internal-facing material (HR, Legal, Training, KB)

☐ Marketing materials

☐ Technical documents and external documentation

☐ Helpdesk and customers service content

☐ Websites and mobile apps

☐ Social media

☐ Chat and user-generated content

☐ Other: _____

240

What percentage of the translated content is included in the translation quality management workflow? *

Your answer

Do you consider different percentages depending on the context of the translation request, the type of the text or other factors? If yes, how are these percentages distributed? *

Your answer

What type of translation quality evaluation methodology is used in your organisation? *

○ Only analytic translation quality evaluation (see definition in https://www.w3.org/community/mqmcg/mqm-terminology/)

○ Only holistic translation quality evaluation (see definition in https://www.w3.org/community/mqmcg/mqm-terminology/)

○ Both

○ None

○ Other:

Does your organisation sometimes perform holistic translation quality evaluation INSTEAD of analytic translation quality evaluation, or the other way round? If yes, under what circumstances? *

Your answer

Does your organisation perform holistic translation quality evaluation on the SAME TEXTS that are submitted for analytic translation quality evaluation? If yes, under what circumstances? *

Your answer

How would you rate the current level of confidence in the adequacy of the quality scores of the quality evaluation methodology used? *

	Not confident	Slightly confident	Moderately confident	Very confident	Extremely confident
Level of confidence in the adequacy of the quality scores	○	○	○	○	○

241

Are there situations in which the translation quality evaluation methodology your organisation has in place is assessed as not being applicable or reliable? Please, give an example. *

Your answer

3.4 Translation quality evaluation processes

What percentage of the translated content is evaluated? *

Your answer

How frequently are translation quality evaluations performed? *

Your answer

Does your organisation perform quality evaluations before the delivery of the translation to the stakeholder, after, or both? *

☐ Before

☐ After

☐ Both (please, specify the percentage for each in "Other")

☐ Other:

Does your organisation evaluate whole texts or select a sample for evaluation? *

○ Only whole texts

○ Only samples

○ Both

If your organisation does sampling, what sampling criteria is used? Please, select all the options that would be applicable. *

☐ Text size

☐ Text type

☐ End-user profile

☐ Type of resources used

☐ Level of translation quality required

☐ We don't don't select samples for evaluation purposes

☐ Other:

242

How are samples selected? *

○ Samples are selected to evaluate random, representative text passages

○ Samples are selected to focus on high-priority text passages

○ Both methods

○ We don't don't select samples for evaluation purposes

○ Other: _____

Does your organisation consider different evaluation workflows depending on factors such as the context of the translation request, the type of the text or other criteria? *

○ Yes

○ No

○ Other: _____

Are your translation requirements designed to be aligned with translation grades or translation tiers? *

○ Yes

○ No

○ Other: _____

Is the translation quality evaluation grading based on certain standards or guidelines? If yes, which ones? *

Your answer

What type of grading system is used? *

☐ 2 grades (for example: Satisfactory/Insatisfactory, or Pass/Fail)

☐ 3 grades (for example: Acceptable, Unsatisfactory, Unacceptable)

☐ 4 grades (for example: Good, Acceptable, Unsatisfactory, Unacceptable)

☐ 5 grades (for example, Very good, Good, Acceptable, Unsatisfactory, Unacceptable)

☐ We don't use a grading system

☐ Numbers or percentages (for example: 10/10, 90%)

☐ Other: _____

Are different evaluation metrics (error types, severity levels, severity multipliers) applied to different text types or quality requirements? *

○ Yes

○ No

○ Other: _____

What error types does it include? Please, select all the applicable options. *

☐ Accuracy (addition, mistranslation, omission, untranslated)

☐ Design

☐ Fluency (grammar, inconsistency, spelling, typography, unintelligible)

☐ Locale convention

☐ Style

☐ Terminology

☐ Verity (completeness, legal requirements, locale-specific content)

☐ Other: _____

What severity levels does it include? Please, select all the applicable options: *

☐ Minor

☐ Major

☐ Critical

☐ Neutral/Preferential

☐ Other: _____

What severity levels does it include? Please, select all the applicable options: *

☐ Minor

☐ Major

☐ Critical

☐ Neutral/Preferential

☐ Other: _____

Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)? *

○ The severity multipliers are the same for all error types

○ The severity multipliers are NOT the same for all error types

○ Other: _____

Does the evaluation include a pass/fail threshold? *

○ Yes

○ No

○ Other: _____

How are the quality thresholds used to determine the pass/fail assessments? *

Your answer

Is the quality threshold uniform, or there are there different thresholds for different content types or quality criteria? *

○ Same threshold (uniform)

○ Different thresholds depending on the content type or quality criteria (non-uniform)

○ Other: _____

3.5 Resources and qualifications

What kind of resources are involved in the translation quality management process? *

☐ Translators

☐ Reviewers/proofreaders

☐ Language leads

☐ Quality managers

☐ Evaluators

☐ Subject matter experts

☐ Others

What percentage of these resources are internal versus external? *

Your answer

Is there any specific reason for this distribution of internal vs external? *

Your answer

Taking into account the roles mentioned above, whose translations are being evaluated by whom? *

Your answer

Are there any differences depending on whether the translation or the evaluation resources are internal or external? *

Your answer

What are the job qualifications of the evaluators? For example, are they required to be qualified as experienced translators and/or reviewers? *

Your answer

In your experience, what are the most relevant job qualifications of the translation quality managers or the resources in charge of the translation quality management program? *

Your answer

What kind of training and documentation are provided in your organisation to develop and support evaluators and/or quality managers? *

Your answer

In your experience, what are the most important competencies of an evaluator? Please, rate the following competencies with 1-5, where 1 means "Not important" and 5 "Extremely important". *

	1 - Not important	2 - Slightly important	3 - Moderately important	4 - Very Important	5 - Extremely important
Cultural competence in both the source language and the target language.	○	○	○	○	○
Domain competence in both the source language and the target language	○	○	○	○	○
Evaluation competence: ability to classify and annotate any errors found during the evaluation process	○	○	○	○	○
Linguistic and textual competence in both the source language and the target language	○	○	○	○	○
Revision competence: ability to identify potential errors and suggest or make corrections	○	○	○	○	○
Translation competence: ability to translate content.	○	○	○	○	○
Technical competence	○	○	○	○	○

3.6 Tools and technologies

What kind of tools does your organisation use for translation quality management purposes? *

☐ Modules and features already included in a commercial CAT/TMS tool. Please, specify the name of the tool in "Other"

☐ A specific translation quality management tool outside a CAT/TMS tool. Please, specify the name of the tool in "Other"

☐ A general quality management tool. Please, specify the name of the tool in "Other"

☐ A proprietary translation quality management tool

☐ No specific tools

☐ Other:

What is the environment or tool where the translation evaluation processes, in particular, error annotation, take place? *

☐ Modules or features already included in a commercial CAT/TMS tool. Please, specify the name of the tool in "Other"

☐ A specific translation quality evaluation tool outside a CAT/TMS tool but connected to it. Please, specify the name of the tool in "Other"

☐ A specific translation quality evaluation tool outside a CAT/TMS tool and disconnected from it. Please, specify the name of the tool in "Other"

☐ A proprietary translation quality evaluation tool developed by my organisation

☐ No specific tools

☐ Other:

How is the evaluation of translation samples processed through the CAT/TMS tool or translation quality evaluation tool? *

○ The samples are created outside the translation quality evaluation system with only the sample text being fed into the system

○ The selected sample segments are identified inside the translation quality evaluation system (CAT tool or other tools) and the evaluation is limited to the sampled segments flagged as being in the sample.

○ Not applicable

○ Other:

248

Part 4 Challenges and compromises

On a scale of 1 to 5, where 1 is "Very unsatisfied" and 5 is "Very satisfied", how would you rate the level of adequacy of the current translation management process in your organisation according to the following parameters? *

	1 - Very unsatisfied	2 - Unsatisfied	3 - Neutral	4 - Satisfied	5 - Very satisfied
Adequacy to the needs to the business and the stakeholders	◯	◯	◯	◯	◯
Compliance with the main norms and standards related to translation quality management	◯	◯	◯	◯	◯
Adequacy of the quality evaluation process and methodology	◯	◯	◯	◯	◯
Adequacy of the resourcing and training model	◯	◯	◯	◯	◯
Adequacy of the tools used for translation quality management	◯	◯	◯	◯	◯
Adequacy of the tools used for translation quality evaluation	◯	◯	◯	◯	◯
Overall level of maturity of the whole translation quality management process	◯	◯	◯	◯	◯

What are the 3 main pain points that your organisation is currently facing when it comes to translation quality management and what is their impact in your organisation? *

Your answer

What are the partial solutions or workarounds that your organisation has applied to the 3 pain points mentioned above? *

Your answer

249

Part 5 Looking into the future

What are 3 main challenges that your organisation will probably have to face in the near future when it comes to translation quality management processes? *

Your answer

Are there any tools, initiatives or innovation initiatives your organisation is planning to adopt to enhance their translation quality management program? Which ones? *

Your answer

What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes mentioned above? Please, tick all the ones that apply: *

☐ New standards or norms released by organisations such as ISO and ASTM with a more specific focus in translation quality management or translation quality evaluation

☐ Guidelines and recommendations published in open-source repositories such as w3c.org

☐ Certification and Declaration Programs organised by regulatory bodies such as ISO or ASTM

☐ Training courses offered by private organisations

☐ Research initiatives and publications led by some of the main actors in the translation and localisation industry (language service providers and organisations)

☐ Research initiatives and publications led by universities in collaboration with some of the main actors in the translation and localization industry (language service providers and organisations)

☐ Development of advanced translation quality management tools outside of CAT/TMS environments

☐ Development of advanced translation quality management tools inside CAT/TMS environments

☐ Other:

Questionnaire 2: Translation service providers

How many languages does your organisation work regularly with? *

○ Between 1 and 3

○ Between 4 and 10

○ More than 10

Part 2: Level of maturity in translation processes and analytics

How would you score the actual level of maturity of your organisation in terms of translation analytics? *

○ 0 - Inactive: My organisation does not track translation activity

○ 1 - Reactive: My organisation has historical data reports focused on costs, compiled manually

○ 2 - Descriptive: My organisation has operational reports on costs and performance based on historical data with manual consolidation data

○ 3 - Diagnostic: My organisation has dashboards indicating errors or problem conditions, there is some loose or manual integration of multiple data sources

○ 4 - Predictive: My organisation has fully automated predictive analytics that indicate what is likely to happen, there is integration with multiple data sources (business analytics, reporting and visual dashboards)

○ 5 - Prescriptive: My organisation has real-time analysis that shapes immediate and future actions, and it may count with data acquired from external sources

---

Part 3: Level of maturity in translation quality management processes

3.1 Overall maturity in translation quality management

How would you score the actual level of maturity of your organisation in terms of translation and localisation quality management processes? *

○ 1 - Manual: My organisation uses manual and/or legacy processes

○ 2 - Automated: My organisation has started to integrate some automation in the quality management workflow

○ 3 - Agile: My organisation has automated and scalable quality management processes across multiple languages

○ 4 - Centralised: My organisation practices strategic and centralised translation management processes across multiple channels and languages

○ 5 - Expert: My organisation benefits from translation as a revenue center and is using translation quality management metrics for growth

3.2 Norms, standards and metrics

Which ones of the following norms and standards related to translation or quality management has your organisation incorporated in their translation quality management program? *

- ☐ ISO 17100 (Translation services — Requirements for translation services)
- ☐ ISO 18587 (Translation services — Post-editing of machine translation output — Requirements)
- ☐ ISO 9001 (Quality management systems — Requirements)
- ☐ ISO 9000 (Quality management systems — Fundamentals and vocabulary)
- ☐ ASTM F2575 (Standard Guide for Quality Assurance in Translation)
- ☐ None
- ☐ Other:

Please, reply to this question if you selected "None" above. What is the reason your organisation hasn't incorporated any of the norms or standards mentioned in their translation quality program? *

- ☐ My organisation is not aware of the existence of those standards
- ☐ The standards are not flexible or customisable enough
- ☐ The standards are too complex or difficult to apply
- ☐ None of the standards satisfied the specific needs of my organisation
- ☐ Not applicable
- ☐ Other:

Which ones of the following translation quality metrics has your organisation incorporated in their translation quality management program? *

- ☐ LISA QA Model
- ☐ SAE J2450
- ☐ MQM-DQF
- ☐ Automatic metrics for Machine Translation (BLEU, METEOR, others)
- ☐ My organisation uses a proprietary set of metrics that is not based on any specific standard
- ☐ None
- ☐ Other:

Please, reply to this question if you selected "None" above. What is the reason your organisation doesn't use quality metrics or decided to use a proprietary or customised set of metrics? *

- ☐ My organisation is not aware of the existence of standarised metrics
- ☐ The standarised metrics are not flexible or customisable enough
- ☐ The standarised metrics are too complex or difficult to apply
- ☐ None of the standarised metrics satisfy the specific needs of my organisation
- ☐ Not applicable
- ☐ Other:

252

3.3 Translation quality management processes

Are there different quality management workflows depending on the type of client? What are the most common scenarios. *

Your answer

What types of translated texts are included in the translation quality management process? *

○ All the translated texts

○ Some of them (please, explain which ones are excluded in "Other")

○ Other: _____

What percentage of the translated content is included in the translation quality management workflow? If the percentage varies depending on the client's specifications or requirements, please briefly describe the 3 most common scenarios. *

Your answer

Do you apply different percentages depending on the context of the translation request, the type of the text or other factors? If yes, how are these percentages distributed? *

Your answer

Do you have a general translation quality program that applies to all your clients, or do you adapt it to each client? *

○ General translation quality program for all our clients

○ Different translation quality program depending on the client

○ Other: _____

What percentage of your clients require a specific translation quality management program? *

Your answer

What type of translation quality evaluation methodology is more common in your organisation? *

○ Analytic translation quality evaluation (see definition in https://www.w3.org/community/mgmcg/mqm-terminology/)

○ Holistic translation quality evaluation (see definition in https://www.w3.org/community/mgmcg/mqm-terminology/)

○ Both

○ Other: _____

Does your organisation sometimes perform holistic translation quality evaluation INSTEAD of analytic translation quality evaluation, or the other way round? If yes, under what circumstances? *

Your answer

Does your organisation perform holistic translation quality evaluation on the SAME TEXTS that are submitted for analytic translation quality evaluation? If yes, under what circumstances? *

Your answer

How would you rate the current level of confidence in the adequacy of the quality scores of the quality evaluation methodology used? *

	Not confident	Slightly confident	Moderately confident	Very confident	Extremely confident
Level of confidence in the adequacy of the quality scores	○	○	○	○	○

Are there situations in which the translation quality evaluation methodology your organisation has in place is assessed as not being applicable or reliable? Please, give an example. *

Your answer

In general, would you say that your clients' requirements and expectations help you build your translation quality management program or rather limit it? *

○ My clients' requirements and expectations help my organisation to build a solid quality management program.

○ My clients' requirements and expectations limit the ability of my organisation to build a solid quality management program.

○ My clients' requirements and expectations don't affect the ability of my organisation to build a solid quality management program.

○ Other: _____

254

3.4 Translation quality evaluation processes

If your organisation follows different quality evaluation processes, please only mention the most common scenario for each question.

What percentage of the translated content is evaluated? *

Your answer

How frequently are translation quality evaluations performed? *

Your answer

Does your organisation perform quality evaluations before the delivery of the translation to the client, after, or both? *

○ Before

○ After

○ Both (please, specify the percentage for each in "Other")

○ Other: _____

Does your organisation evaluate whole texts or select a sample for evaluation? *

○ Only whole texts

○ Only samples

○ Both

○ Other: _____

If your organisation does sampling, what sampling criteria is used? Please, select all the options that would be applicable. *

☐ Text size

☐ Text type

☐ End-user profile

☐ Type of resources used

☐ Level of translation quality required

☐ We don't don't select samples for evaluation purposes

☐ Other: _____

255

How are the samples selected? *

○ Samples are selected to evaluate random, representative text passages

○ Samples are selected to focus on high-priority text passages

○ Both methods

○ We don't don't select samples for evaluation purposes

○ Other: _____

Are the translation quality management requirements designed to be aligned with translation grades or translation tiers? *

○ Yes

○ No

○ Other: _____

Is the translation quality evaluation grading model based on certain standards or guidelines? Does it vary depending on the client or type of translation? *

Your answer

What type of grading system is used? If it varies depending on the client, please select a maximum of 3. *

☐ 2 grades (for example: Satisfactory/Insatisfactory, or Pass/Fail)

☐ 3 grades (for example: Acceptable, Unsatisfactory, Unacceptable)

☐ 4 grades (for example: Good, Acceptable, Unsatisfactory, Unacceptable)

☐ 5 grades (for example, Very good, Good, Acceptable, Unsatisfactory, Unacceptable)

☐ We don't use a grading system

☐ Numbers or percentages (for example: 10/10, 90%)

☐ Other: _____

Are different evaluation metrics (error types, severity levels, severity multipliers) applied to different text types? *

○ Yes

○ No

○ Other: _____

What error types does it include? Please, select all the applicable options. *

☐ Accuracy (addition, mistranslation, omission, untranslated)

☐ Design

☐ Fluency (grammar, inconsistency, spelling, typography, unintelligible)

☐ Locale convention

☐ Style

☐ Terminology

☐ Verity (completeness, legal requirements, locale-specific content)

☐ Other: _____

What severity levels does it include? Please, select all the applicable options: *

☐ Minor

☐ Major

☐ Critical

☐ Neutral/Preferential

☐ Other: _____

Are severity multipliers the same for all error types (uniform severity multiplier distributions), or are they defined for individual error types (nonuniform severity multiplier distributions)? *

⦿ The severity multipliers are the same for all error types

◯ The severity multipliers are NOT the same for all error types

◯ Other: _____

Does the evaluation include a pass/fail threshold? *

◯ Yes

◯ No

◯ Other: _____

How are the quality thresholds used to determine the pass/fail assessments? *

Your answer

Is the quality threshold uniform, or there are there different thresholds for different content types or quality criteria? *

◯ Same threshold (uniform)

◯ Different thresholds depending on the content type or quality criteria (non-uniform)

◯ Other: _____

3.5 Resources and qualifications

What kind of resources are involved in the translation quality management process? *

☐ Translators

☐ Reviewers/proofreaders

☐ Language leads

☐ Quality managers

☐ Evaluators

☐ Subject matter experts

☐ Others

What percentage of these resources are internal versus external? *

Your answer

Is there any specific reason for this distribution of internal vs external? *

Your answer

Taking into account the roles mentioned above, whose translations are being evaluated by whom? *

Your answer

Are there any differences depending on whether the translation or the evaluation resources are internal or external? *

Your answer

What are the job qualifications of the evaluators? For example, are they required to be qualified as experienced translators and/or reviewers? *

Your answer

In your experience, what are the most relevant job qualifications of the translation quality managers or the resources in charge of the translation quality management program? *

Your answer

What kind of training and documentation are provided in your organisation to develop and support evaluators and/or quality managers? *

Your answer

In your experience, what are the most important competencies of an evaluator? Please, rate the following competencies with 1-5, where 1 means "Not important" and 5 "Extremely important". *

	1 - Not important	2 - Slightly important	3 - Moderately important	4 - Very Important	5 - Extremely important
Cultural competence in both the source language and the target language.	○	○	○	○	○
Domain competence in both the source language and the target language	○	○	○	○	○
Evaluation competence: ability to classify and anotate any errors found during the evaluation process	○	○	○	○	○
Linguistic and textual competence in both the source language and the target language	○	○	○	○	○
Revision competence: ability to identify potential errors and suggest or make corrections	○	○	○	○	○
Translation competence: ability to translate content.	○	○	○	○	○
Technical competence	○	○	○	○	○

259

3.6 Tools and technologies

What kind of tools does your organisation use for translation quality management purposes? *

- [ ] Modules and features already included in a commercial CAT/TMS tool. Please, specify the name of the tool in "Other"
- [ ] A specific translation quality management tool outside a CAT/TMS tool. Please, specify the name of the tool in "Other"
- [ ] A general quality management tool. Please, specify the name of the tool in "Other"
- [ ] A proprietary translation quality management tool
- [ ] No specific tools
- [ ] Other: _____

What is the environment or tool where the translation evaluation processes, in particular, error annotation, take place? *

- [ ] Modules or features already included in a commercial CAT/TMS tool. Please, specify the name of the tool in "Other"
- [ ] A specific translation quality evaluation tool outside a CAT/TMS tool but connected to it. Please, specify the name of the tool in "Other"
- [ ] A specific translation quality evaluation tool outside a CAT/TMS tool and disconnected from it. Please, specify the name of the tool in "Other"
- [ ] A proprietary translation quality evaluation tool developed by my organisation
- [ ] No specific tools
- [ ] Other: _____

How is the evaluation of translation samples processed through the CAT/TMS tool or translation quality evaluation tool? *

- ( ) The samples are created outside the translation quality evaluation system with only the sample text being fed into the system
- ( ) The selected sample segments are identified inside the translation quality evaluation system (CAT tool or other tools) and the evaluation is limited to the sampled segments flagged as being in the sample.
- ( ) Not applicable
- ( ) Other: _____

260

Part 4 Challenges and compromises

On a scale of 1 to 5, where 1 is "Very unsatisfied" and 5 is "Very satisfied", how would you rate the level of adequacy of the current translation management process in your organisation according to the following parameters? *

	1 - Very unsatisfied	2 - Unsatisfied	3 - Neutral	4 - Satisfied	5 - Very satisfied
Adequacy to the needs to the business and the stakeholders	◯	◯	◯	◯	◯
Compliance with the main norms and standards related to translation quality management	◯	◯	◯	◯	◯
Adequacy of the quality evaluation process and methodology	◯	◯	◯	◯	◯
Adequacy of the resourcing and training model	◯	◯	◯	◯	◯
Adequacy of the tools used for translation quality management	◯	◯	◯	◯	◯
Adequacy of the tools used for translation quality evaluation	◯	◯	◯	◯	◯
Overall level of maturity of the whole translation quality management process	◯	◯	◯	◯	◯

What are the 3 main pain points that your organisation is currently facing when it comes to translation quality management and what is their impact in your organisation? *

Your answer

What are the partial solutions or workarounds that your organisation has applied to the 3 pain points mentioned above? *

Your answer

261

Part 5 Looking into the future

What are 3 main challenges that your organisation will probably have to face in the near future when it comes to translation quality management processes? *

Your answer

Are there any tools, initiatives or innovation initiatives your organisation is planning to adopt to enhance their translation quality management program? Which ones? *

Your answer

What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes mentioned above? Please, tick all the ones that apply: *

☐ New standards or norms released by organisations such as ISO and ASTM with a more specific focus in translation quality management or translation quality evaluation

☐ Guidelines and recommendations published in open-source repositories such as w3c.org

☐ Certification and Declaration Programs organised by regulatory bodies such as ISO or ASTM

☐ Training courses offered by private organisations

☐ Research initiatives and publications led by some of the main actors in the translation and localisation industry (language service providers and organisations)

☐ Research initiatives and publications led by universities in collaboration with some of the main actors in the translation and localization industry (language service providers and organisations)

☐ Development of advanced translation quality management tools outside of CAT/TMS environments

☐ Development of advanced translation quality management tools inside CAT/TMS environments

☐ Other:

Questionnaire 3: Translation technology companies

What type of language technology products does your organisation offer? *

☐ Translation management (CAT or TMS)

☐ Translation quality management (quality assurance and quality evaluation tools)

☐ Other: _____

Part 2: Translation quality management methodologies

2.1 Norms, standards and metrics

Which ones of the following norms and standards related to translation or quality management have been taken into account in the design of the translation technology solutions offered by your company (either completely or partially)? *

☐ ISO 17100 (Translation services — Requirements for translation services)

☐ ISO 18587 (Translation services — Post-editing of machine translation output — Requirements)

☐ ISO 9001 (Quality management systems — Requirements)

☐ ISO 9000 (Quality management systems — Fundamentals and vocabulary)

☐ ASTM F2575 (Standard Guide for Quality Assurance in Translation)

☐ None

☐ Other: _____

Which ones of the following translation quality metrics have been taken into account or incorporated in the design of the translation technology solutions offered by your company (either completely or partially)? *

☐ LISA QA Model

☐ SAE J2450

☐ MQM-DQF

☐ Automatic metrics for Machine Translation (BLEU, METEOR, others)

☐ My organisation uses a proprietary set of metrics that is not based on any specific standard

☐ None

☐ Other: _____

2.2 Translation quality management and evaluation processes

What types of translation quality evaluation methodologies are supported by your technology solution? *

○ Only analytic translation quality evaluation (see definition in https://www.w3.org/community/mqmcg/mqm-terminology/)

○ Only holistic translation quality evaluation (see definition in https://www.w3.org/community/mqmcg/mqm-terminology/)

○ Both

○ None

○ Other: _____

Has there been situations in which the translation quality management technology or the translation quality tools provided by your company have been assessed as not being suited to the needs of your clients? If yes, can you please explain why? *

Your answer

Does your translation quality management solution allow the evaluation of whole texts or select a sample for evaluation? *

○ Only whole texts

○ Only samples

○ Both

○ Other: _____

If your solution does sampling, what sampling criteria can be used? Please, select all the options that would be applicable. *

☐ Text size

☐ Text type

☐ End-user profile

☐ Type of resources used

☐ Level of translation quality required

☐ It is not possible to select samples for evaluation purposes

☐ Other: _____

264

How are samples selected? *

○ Samples can be selected to evaluate random, representative text passages

○ Samples can be selected to focus on high-priority text passages

○ Both methods

○ It is not possible to select samples for evaluation purposes

○ Other:

Is the translation quality evaluation grading model based on certain standards or guidelines? *

Your answer

What types of grading system can be applied? *

☐ 2 grades (for example: Satisfactory/Insatisfactory, or Pass/Fail)

☐ 3 grades (for example: Acceptable, Unsatisfactory, Unacceptable)

☐ 4 grades (for example: Good, Acceptable, Unsatisfactory, Unacceptable)

☐ 5 grades (for example, Very good, Good, Acceptable, Unsatisfactory, Unacceptable)

☐ We don't use a grading system

☐ Numbers or percentages (for example: 10/10, 90%)

☐ Other:

Is it possible to apply different evaluation metrics (error types, severity levels, severity multipliers) to different text types or user profiles? *

○ Yes

○ No

○ Other:

What error types does it include? Please, select all the applicable options. *

☐ Accuracy (addition, mistranslation, omission, untranslated)

☐ Design

☐ Fluency (grammar, inconsistency, spelling, typography, unintelligible)

☐ Locale convention

☐ Style

☐ Terminology

☐ Verity (completeness, legal requirements, locale-specific content)

☐ Other:

265

What severity levels does it include? Please, select all the applicable options: *

☐ Minor

☐ Major

☐ Critical

☐ Neutral/Preferential

☐ Other:

Are severity multipliers the same for all error types (uniform severity multiplier distributions), or can they be defined for individual error types (nonuniform severity multiplier distributions)? *

○ The severity multipliers are always the same for all error types

○ The severity multipliers can be different depending on the error type

○ Other:

Does the evaluation include a pass/fail threshold? *

○ Yes

○ No

○ Other:

How are the quality thresholds used to determine the pass/fail assessments? *

Your answer

2.3 Resources

How many different user roles are available in the translation quality management/evaluation solution? *

☐ Translators

☐ Reviewers/proofreaders

☐ Language leads

☐ Quality managers

☐ Evaluators

☐ Subject matter experts

☐ Others

2.4 Tools and technologies

How would you define the technical setting of the solution you provide for translation quality management purposes? *

○ It is a module or feature included in a CAT/TMS tool

○ It is a specific translation quality management tool outside a CAT/TMS tool

○ It is a generic quality management tool

○ Other: _____

How would you define the environment or tool where the translation evaluation processes, in particular, error annotation, take place? *

○ It is a module or feature included in a CAT/TMS tool

○ It is a translation quality evaluation tool outside a CAT/TMS tool but connected to it

○ It is a translation quality evaluation tool outside a CAT/TMS tool and disconnected from it

○ Other: _____

How is the evaluation of translation samples processed through the translation quality evaluation tool? *

○ The samples are created outside the translation quality evaluation system with only the sample text being fed into the system

○ The selected sample segments are identified inside the translation quality evaluation system (CAT tool or other tools) and the evaluation is limited to the sampled segments flagged as being in the sample.

○ Not applicable

○ Other: _____

Please, reply to this question if your technology solution allows performing a translation quality evaluation inside a CAT/TMS tool. Select which ones of the following features are included in the tool: *

☐ Cloud-based tool

☐ User interface with side-by-side column view to select error strings and mark error types and severity levels in the same step

☐ Predefined error categorisation based on the active error typology used in the evaluation

☐ Ability to suggest alternative translations or edits for translation error

☐ Addition of comments or issue descriptions that can be included in the final evaluation scorecard

☐ Translation resources such as style guides or monolingual dictionaries

☐ Automated writing quality checkers (spell checkers, grammar checkers)

☐ Built-in system for arbiters to rule on disputed errors in favor of the translator or the evaluator

☐ Customised evaluation profiles

☐ Downloadable quality reports and scorecards

☐ Quality dashboards with information such as quality scores and quality levels to support business intelligence

☐ Not applicable

☐ Other:

Please, reply if your technology solution allows performing a translation quality evaluation outside a CAT/TMS tool. Select which ones of the following features are included in the tool: *

☐ Cloud-based tool

☐ Connection with CAT/TMS tool through an API

☐ User interface with side-by-side column view to select error strings and mark error types and severity levels in the same step

☐ Predefined error categorisation based on the active error typology used in the evaluation

☐ Ability to suggest alternative translations or edits for translation error

☐ Addition of comments or issue descriptions that can be included in the final evaluation scorecard

☐ Integrated terminology resources such as glossaries and termbases (ready-only)

☐ Integrated terminology resources such as glossaries and termbases (read-write)

☐ Automated writing quality checkers (spell checkers, grammar checkers)

☐ Translation resources such as style guides or monolingual dictionaries

☐ Built-in system for arbiters to rule on disputed errors in favor of the translator or the evaluator

☐ Customised evaluation profiles

☐ Downloadable quality reports and scorecards

☐ Quality dashboards with information such as quality scores and quality levels to support business intelligence.

☐ Not applicable

☐ Other:

Part 3: Challenges and compromises

On a scale of 1 to 5, where 1 is "Very unsatisfied" and 5 is "Very satisfied", how would you rate the level of maturity of the translation management solution provided by your company, according to the following parameters? *

	1 - Very unsatisfied	2 - Unsatisfied	3 - Neutral	4 - Satisfied	5 - Very satisfied
Adequacy to the needs to the business and the stakeholders	◯	◯	◯	◯	◯
Compliance with the main norms and standards related to translation quality management	◯	◯	◯	◯	◯
Adequacy of the quality evaluation process and methodology	◯	◯	◯	◯	◯
Adequacy of the resourcing and training model	◯	◯	◯	◯	◯
Adequacy of the tools used for translation quality management	◯	◯	◯	◯	◯
Adequacy of the tools used for translation quality evaluation	◯	◯	◯	◯	◯
Overall level of maturity of the whole translation quality management process	◯	◯	◯	◯	◯

What are the 3 main pain points that your technology solution is currently working on to address when it comes to translation quality management processes? *

Your answer

What are the solutions that your technology solution proposes to overcome the 3 pain points mentioned above? *

Your answer

269

Part 4: Looking into the future

What are 3 main challenges that your technology solution would like to tackle in the near future when it comes to translation quality management processes and technologies? *

Your answer

Are there any technologies or innovation initiatives your company is planning to adopt to enhance your technology solution? Which ones? *

Your answer

What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes mentioned above? Please, tick all the ones that apply: *

☐ New standards or norms released by organisations such as ISO and ASTM with a more specific focus in translation quality management or translation quality evaluation

☐ Guidelines and recommendations published in open-source repositories such as w3c.org

☐ Certification and Declaration Programs organised by regulatory bodies such as ISO or ASTM

☐ Training courses offered by private organisations

☐ Research initiatives and publications led by some of the main actors in the translation and localisation industry (language service providers and organisations)

☐ Research initiatives and publications led by universities in collaboration with some of the main actors in the translation and localization industry (language service providers and organisations)

☐ Development of advanced translation quality management tools outside of CAT/TMS environments

☐ Development of advanced translation quality management tools inside CAT/TMS environments

☐ Other:

Questionnaire 4: Experts in translation quality management

Do you belong or participate in any of the following research or standarisation bodies? *

☐ ISO

☐ ASTM

☐ Academic research centre or department (please, specify in "Other")

☐ Professional research centre or department (please, specify in "Other")

☐ Other: _____

Part 2: Translation quality management methodologies

2.1 Norms, standards and metrics

Which ones of the following norms and standards related to translation or quality management are you familiarised with? *

☐ ISO 17100 (Translation services — Requirements for translation services)

☐ ISO 18587 (Translation services — Post-editing of machine translation output — Requirements)

☐ ISO 9001 (Quality management systems — Requirements)

☐ ISO 9000 (Quality management systems — Fundamentals and vocabulary)

☐ ASTM F2575 (Standard Guide for Quality Assurance in Translation)

☐ None

☐ Other: _____

Which ones of the following translation quality metrics are you familiarised with? *

☐ LISA QA Model

☐ SAE J2450

☐ MQM-DQF

☐ Automatic metrics for Machine Translation (BLEU, METEOR, others)

☐ None

☐ Other: _____

2.2 Translation quality management processes

What are the most common scenarios when it comes to your role as a translation quality management expert? Please, select a maximum of 3. *

☐ I design/implement end-to-end TQM methodologies for different companies

☐ I supervise or evaluate end-to-end TQM methodologies for different companies

☐ I design/implement translation quality evaluation methodologies for different companies

☐ I supervise or evaluate translation quality evaluation methodologies for different companies

☐ I research translation quality methodologies and processes (academic)

☐ I research translation quality methodologies and processes (non-academic)

What type of translation quality evaluation methodologies are you more familiarised with? *

○ Only analytic translation quality evaluation (see definition in https://www.w3.org/community/mqmcg/mqm-terminology/)

○ Only holistic translation quality evaluation (see definition in https://www.w3.org/community/mqmcg/mqm-terminology/)

○ Both

○ Other:

In your experience, would you say that the clients' requirements and expectations help build sound translation quality management programs or rather limit them? *

Your answer

2.3 Translation quality management and evaluation processes

If your organisation follows different quality evaluation processes, please briefly mention the most common scenario for each question.

Are there any situations in which you would recommend or perform a holistic translation quality evaluation INSTEAD of analytic translation quality evaluation, or the other way round? If yes, what would those situations be? *

Your answer

Are there any situations in which you would recommend or perform a holistic translation quality evaluation AND an analytic translation quality evaluation on the same text? If yes, what would those situations be? *

Your answer

Do you or the companies you work with assess the reliability of the translation quality evaluation used? If yes, how? *

Your answer

272

Have there been any situations in which the translation quality management methodology recommended or used by you has been assessed as not being suited to the needs of the client? If yes, can you please explain why? *

Your answer

In your opinion, what percentage of the content translated by a company or a language service provider should be evaluated? *

Your answer

How frequently should the translation quality evaluations be performed? *

Your answer

Should quality evaluations be performed before the delivery of the translation to the stakeholder, after, or both? *

○ Before

○ After

○ Both (please, specify the ideal percentage for each in "Other")

○ Other:

Do you think that translation quality assurance (also called "review step") should be performed at the same time as the translation quality evaluation step? If yes, in which situations? *

Your answer

In which situations do you think it would be advisable to use sampling techniques rather than evaluate whole texts? *

Your answer

If you recommend sampling, what sampling criteria are more effective in your opinion? *

☐ Text size

☐ Text type

☐ End-user profile

☐ Type of resources used

☐ Level of translation quality required

☐ I don't recommend doing sampling for evaluation purposes

☐ Other:

273

How should the samples be selected? *

○ Samples should be selected to evaluate random, representative text passages

○ Samples should be selected to focus on high-priority text passages

○ Both methods are equally effective

○ I don't recommend doing sampling for evaluation purposes

○ Other: _____

What standards or guidelines do you consider more relevant in order to build a translation quality evaluation scoring model? *

Your answer

What type of grading system do you think works better in the majority of scenarios? *

○ 2 grades (for example: Satisfactory/Insatisfactory, or Pass/Fail)

○ 3 grades (for example: Acceptable, Unsatisfactory, Unacceptable)

○ 4 grades (for example: Good, Acceptable, Unsatisfactory, Unacceptable)

○ 5 grades (for example, Very good, Good, Acceptable, Unsatisfactory, Unacceptable)

○ Numbers or percentages (for example: 10/10, 90%)

○ Other: _____

Do you think it makes sense to apply different evaluation metrics (error types, severity levels, severity multipliers) to different text types? *

○ Yes

○ No

○ Other: _____

What error types should a translation quality evaluation scoring model include? Please, select all the applicable options. *

☐ Accuracy (addition, mistranslation, omission, untranslated)

☐ Design

☐ Fluency (grammar, inconsistency, spelling, typography, unintelligible)

☐ Locale convention

☐ Style

☐ Terminology

☐ Verity (completeness, legal requirements, locale-specific content)

☐ Other: _____

274

What severity levels should a translation quality evaluation scoring model include? Please, select all the applicable options: *

☐ Minor

☐ Major

☐ Critical

☐ Neutral/Preferential

☐ Other:

Should the severity multipliers be the same for all error types (uniform severity multiplier distributions), or defined for individual error types (nonuniform severity multiplier distributions)? *

○ The severity multipliers should be the same for all error types

○ The severity multipliers should NOT be the same for all error types

○ Other:

Should an evaluation scoring model include a pass/fail threshold? *

○ Yes

○ No

○ Other:

How should the quality thresholds be used to determine the pass/fail assessments? *

Your answer

2.4 Resources and qualifications

What kind of resources do you think should be involved in the translation quality management process? *

☐ Translators

☐ Reviewers/proofreaders

☐ Language leads

☐ Quality managers

☐ Evaluators

☐ Subject matter experts

☐ Others

Taking into account the roles mentioned above, whose translations are being evaluated by whom? *

Your answer

Are there any differences depending on whether the translation or the evaluation resources are internal or external? *

Your answer

What would be the ideal job qualifications of the evaluators? For example, should they be qualified as experienced translators and/or revisers? *

Your answer

In your experience, what are the most relevant job qualifications of the translation quality managers or the resources in charge of the translation quality management program? *

Your answer

In your opinion, what kind of training and documentation should be provided to develop and support evaluators and/or quality managers? *

Your answer

276

In your experience, what are the most important competencies of an evaluator? Please, rate the following competencies with 1-5, where 1 means "Not important" and 5 "Extremely important". *

	1 - Not important	2 - Slightly important	3 - Moderately important	4 - Very Important	5 - Extremely important
Cultural competence in both the source language and the target language.	○	○	○	○	○
Domain competence in both the source language and the target language	○	○	○	○	○
Evaluation competence: ability to classify and annotate any errors found during the evaluation process	○	○	○	○	○
Linguistic and textual competence in both the source language and the target language	○	○	○	○	○
Revision competence: ability to identify potential errors and suggest or make corrections	○	○	○	○	○
Translation competence: ability to translate content.	○	○	○	○	○
Technical competence	○	○	○	○	○

2.5 Tools and technologies

What kind of tools do you use/recommend for translation quality management purposes? *

☐ Modules and features already included in a commercial CAT/TMS tool. Please, specify the name of the tool in "Other"

☐ A specific translation quality management tool outside a CAT/TMS tool. Please, specify the name of the tool in "Other"

☐ A general quality management tool. Please, specify the name of the tool in "Other"

☐ A proprietary translation quality management tool developed by the client

☐ No specific tools

☐ Other: _____

What environment or tool do you use/recommend for translation evaluation processes, in particular, error annotation? *

☐ Modules or features already included in a commercial CAT/TMS tool. Please, specify the name of the tool in "Other"

☐ A specific translation quality evaluation tool outside a CAT/TMS tool but connected to it. Please, specify the name of the tool in "Other"

☐ A specific translation quality evaluation tool outside a CAT/TMS tool and disconnected from it. Please, specify the name of the tool in "Other"

☐ A proprietary translation quality evaluation tool developed by the client

☐ No specific tools

☐ Other: _____

What method do you use/recommend to perform evaluations of translation samples processed through a CAT/TMS tool or translation quality evaluation tool? *

○ The samples should be created outside the translation quality evaluation system with only the sample text being fed into the system

○ The selected sample segments should be identified inside the translation quality evaluation system (CAT tool or other tools) and the evaluation should be limited to the sampled segments flagged as being in the sample

○ Other: _____

278

Part 3: Challenges and compromises

On a scale of 1 to 5, where 1 is "Very unsatisfied" and 5 is "Very satisfied", how would you rate the level of maturity of the current translation management frameworks used in the industry, according to the following parameters? *

	1 - Very unsatisfied	2 - Unsatisfied	3 - Neutral	4 - Satisfied	5 - Very satisfied
Adequacy to the needs to the business and the stakeholders	○	○	○	○	○
Compliance with the main norms and standards related to translation quality management	○	○	○	○	○
Adequacy of the quality evaluation process and methodology	○	○	○	○	○
Adequacy of the resourcing and training model	○	○	○	○	○
Adequacy of the tools used for translation quality management	○	○	○	○	○
Adequacy of the tools used for translation quality evaluation	○	○	○	○	○
Overall level of maturity of the whole translation quality management process	○	○	○	○	○

In your opinion, what are the 3 main pain points when it comes to translation quality management? *

Your answer

What are the partial solutions or compromises to the main 3 pain points mentioned above that could help to overcome them in the short term? *

Your answer

279

Part 4: Looking into the future

In your opinion, what are the 3 main challenges that the industry will be facing in the near future when it comes to translation quality management processes and technologies? *

Your answer

Are there any technologies, tools or innovation initiatives you are researching or are planning to adopt to enhance the current translation quality management processes? Which ones? *

Your answer

What kind of initiatives do you think could help to solve the current and future challenges in translation quality management processes mentioned above? Please, tick all the ones that apply: *

☐ New standards or norms released by organisations such as ISO and ASTM with a more specific focus in translation quality management or translation quality evaluation

☐ Guidelines and recommendations published in open-source repositories such as w3c.org

☐ Certification and Declaration Programs organised by regulatory bodies such as ISO or ASTM

☐ Training courses offered by private organisations

☐ Research initiatives and publications led by some of the main actors in the translation and localisation industry (language service providers and organisations)

☐ Research initiatives and publications led by universities in collaboration with some of the main actors in the translation and localization industry (language service providers and organisations)

☐ Development of advanced translation quality management tools outside of CAT/TMS environments

☐ Development of advanced translation quality management tools inside CAT/TMS environments

☐ Other:

Closing questions (common section for all groups)

Thank you!

Thank you for taking the time to complete this survey. We truly value the information you have provided. Your responses will contribute to our research and to a better understanding of the current state of the translation quality management methodologies in the translation industry.

If you have any comments about the survey or the project, you can reach out us at jennifer.vela101@alu.ulpgc.es.

Many thanks!
Jennifer Vela-Valido
PhD candidate

Is there anything else you would like to share with us?

Your answer

Please, indicate whether you would like to receive a summary of the findings from this survey and/or be contacted for a follow-up interview.

☐ I would like to receive a summary of the findings from this survey

☐ I agree to be contacted to arrange a follow-up interview

Please, enter your email address below if you wish to receive the findings of this survey or be contacted for a follow-up interview:

Your answer

Appendix E

Main pain points in translation quality management and evaluation	Applied solutions and workarounds
Insourced TQM: Lack of centralised data or dashboards with visibility on trends	Automatise as much as possible (macros, proprietary tools) Weekly meetings to review data Moving to a TMS that has dashboard
Insourced TQM: Lack of clearly defined TQM process	Team is working on finding solutions to these gaps
Insourced TQM: Lack of dedicated human resources	Translators help with evaluations Content audits with different type of contents and quality levels per content
Insourced TQM: Lack of support or interest from internal stakeholders	
Insourced TQM: Lack of technical resources to implement advanced tools	Internal automatic checks with TMS
Insourced TQM: Lack of training and competences	Learning from our translation vendors
Insourced TQM: Lack ot automatisation	Migrating to a new TMS to gain flexibility Regular updates of excel dashboards to make sure the data is available at any given time
Insourced TQM: Level of maturity of the translation/localisation team	Translators help with evaluations Introducing DQF model
Insourced TQM: Missing requirements or guidelines from internal stakeholder	
Outsourced TQM: Lack of visibility on how the QM process is performed	Quality management policy with KPIs to assess service received
Outsourced TQM: Maintaining a stable quality (especially freelancers)	Metrics, monitoring, autoQA
Quality improvement: Communication loops and feedback channels	Dedicated email and ticketing system Slack discussions, informally, between linguists peer to peer
Quality improvement: Time constrains	Automatise as much as possible (macros, proprietary tools)
Tools: Ability to review content easily in context (app and web content)	Creating manual workaround with the third-party reviewer vendor's help Working with TQE tool that can display same content that is sent to translators
Tools: More integration required into TMS (TQM and TQE)	Localisation engineering providing interim solutions with CAT tool exports Planning to integrate quality reporting into TMS
Tools: No tools allow a macro LQA view	
Tools: Use of proprietary tools that cannot easily be updated or improved	Using scorecards to issue quality reports and assess quality in a more granular way
TQE: Alignment and calibration among different reviewers	SLAs with reviewers, survey based model rather than out of the box Calibration sessions with all the senior localisation especialists
TQE: High costs, time-consuming processes and rigidity of the evaluation model	Crowd-based testing by non-linguists, linguistic testing on a bigger scale only with indicator
TQE: Lack of clear, scalable and customisable evaluation metrics and thresholds	Introducing DQF model
TQE: Use of very unstructured holistic models without any grading or sampling	Currently working on a process improvement plan

Figure 46. Pain points and solutions applied in translation quality management and evaluation – Translation service buyers.

Main pain points in translation quality management and evaluation	Applied solutions and workarounds
TQM: Lack of centralised data or dashboards with visibility on trends	Spreadsheet to register the results of each evaluation
TQM: Financial limitations to invest in tools or dedicated resources	Researching affordable tools that would mean an improvement to the TQM processes
TQM: Lack of cost-effective, scalable and flexible TQM methodologies	Hiring a Quality Manager
TQM: Lack of training and knowledge	Hiring consultants to help us optimise our TQM processes
TQM: Lack of automatisation	Investing in new automatisation tools
TQM: Lack of access to practical information on TQM methodologies	Learning from colleagues and similar companies
Clients: Lack of visibility of upcoming work	Forecast requests, scalable teams
Clients: Unrealistic, inconsistent or unclear quality expectations	Setting up specific processes per client to have clear instructions
Clients: Aligning and educating clients on quality management best practices	Proactive communication towards clients concerning expectation management
	Helping less experienced clients understand the benefits of analytical quality monitoring and to correctly interpret data
Clients: Lack of content strategy	Helping clients create a Content Matrix, define quality expectations with their stakeholders
Quality improvement: Time constrains	
Tools: Ability to review content easily in context (app and web content)	
Tools: More integration required into TMS (TQM and TQE)	Research on available TQM and TQE tools that can be plugged into our TMS
Tools: Excel still used as main tool for TQM and TQE	Research on specific translation quality management tool to replace Excel
Tools: Lack of advanced tools for TQM and TQE	Research on available TQM and TQE tools
TQE: Alignment and calibration among different reviewers	Hiring a Quality Manager
TQE: High costs, time-consuming processes, rigidity of the evaluation model	Using alternative approaches (such as holistic approach or, fit for purpose evaluation)
TQE: MT evaluation (raw and post-edited)	Trying to harmonize MTPE evaluation instructions
	Going to conferences to try to learn better ways to do MT evaluation
TQE: Use of unstructured holistic models without objective grading	Hiring a Quality Manager

Figure 82. Pain points and solutions applied in translation quality management and evaluation – Translation service providers.

Main pain points in translation quality management and evaluation	Suggested solutions
TQM: Lack of centralised data or dashboards with visibility on trends	Develop APIs and connections
TQM: Financial limitations to invest in tools or dedicated resources	Educating clients to better use of MT to reduce costs
TQM: Lack of cost-effective, scalable and flexible TQM methodologies	Develop TQM processes focused on total quality control, TQE should be just one component of it
TQM: Lack of training and knowledge	Bring in a qualified quality manager
TQM: Lack of access to practical information on TQM methodologies	Standards: Have a more flexible approach that is easier to apply
Clients: Unrealistic, inconsistent or unclear quality expectations	Educate clients to be more flexible in their expectations and tools used
Clients: Aligning and educating clients on quality management best practices	Educate clients to be more flexible in their expectations and tools used
Clients: Client-based priorities over translation priorities	TSP should be the gatekeepers of quality, they should be trusted by clients
Quality improvement: Time constrains	Automated quality evaluation platform or computer assisted evaluation
Tools: Lack of advanced tools for TQM and TQE	Develop APIs, connections, propose other tools
	Move quality evaluation from external to internal in CAT tools
TQE: Appropriate training and calibration among different reviewers	National regulations and investments that allow companies to train and hire enough competent evaluators
	Closer collaboration and partnership with reviewers to improve calibration
TQE: High costs, time-consuming processes and rigidity of the evaluation model	Automated quality evaluation platform or computer assisted evaluation
	Mid-term review of the TQE process
TQE: Use of unstructured holistic models without objective grading	Bring in a qualified quality manager

Figure 143. Translation quality management and evaluation pain points and recommended solutions – Experts in translation quality management