QUALITY ASSURANCE (QA) OF TERMINOLOGY IN A TRANSLATION QUALITY MANAGEMENT SYSTEM (QMS) IN THE BUSINESS ENVIRONMENT

Sanja SELJAN
University of Zagreb - Faculty of Humanities and Social Sciences,
Department of Information and Communication Sciences, Croatia

ABSTRACT
The main objective of this paper is to identify the various levels of the Quality Assurance (QA) process as part of the Quality Management System (QMS), in relation to the regulatory framework and applied to the translation and terminology management process. The paper also outlines the main principles of quality management. The QMS is intended to be an integrative element which brings together aspects of the company, the core business domain and customer satisfaction, integrating people (skills, roles and responsibilities), processes (workflow, interactions), regulatory requirements, technology, risks and opportunities, with the main purpose of delivering the best product. This paper provides an insight into the importance of terminology management at a corporate level from the perspective of business communication and the benefits that it can bring to a company. Terminology management is closely interlinked with the translation process, insofar as it has a bearing on the quality of the translation product, but it can also be considered an entirely separate process, one that is carried out in line with the needs of a company or its customers.

QA focuses on the planning and documentation of all activities, with the aim of preventing errors and achieving overall quality. The QA process encompasses project management, the pre-, post- and actual production phases, and the follow-up to specific activities, phases or processes. In this paper, three steps in the optimisation of the terminology management process are presented.

QA, as part of the QMS, is analysed at the following levels: workflow (including the different activities involved), production (the use of QA tools) staff (team experts), workflow and users in a collaborative environment, software and output (the use of human and automatic evaluation metrics).

Overall, this paper presents a holistic view on the importance of QA for terminology within the framework of a translation QMS and in relation to the regulatory framework, with the overall aim of ensuring product quality in a business environment.

Key words: quality assurance (QA), quality management system (QMS), workflow, products, experts, translation, terminology, business communication

Today’s working environment, the competitive jobs market, the need for information to be conveyed in other languages and the field of business communication all necessitate quality control solutions. Terminology management, integrated within the translation process, forms part of the document production system, which require controlled and standardised QA processes.

For many agencies, industries and business stakeholders, it is desirable to ensure a consistent transfer of information and to build positive communication between customers and service/product providers. Successful communication in a global environment requires a systematic quality management process, which should be fully supported by a quality management system. Regardless of the type of text (manuals, official records, everyday documents, promotional materials, software, applications, e-mails, presentations, websites, etc.), a centralised terminology and translation process can serve to improve quality, reduce costs, and deliver faster turnarounds and more satisfied customers. Consistent terminology can also positively contribute to company image, muster new clients and expand business opportunities. Terminology management is defined as a set of strategies for storing, retrieving and updating specific terminology in order to ensure that corporate terminology is used correctly and as part of business communication. Bogaert (2008), for instance, asserts that
“Terminology Management is a fundamental part of a company’s Globalization Information Management (GIM)”.

1. TERMINOLOGY MANAGEMENT AT A CORPORATE LEVEL

Terminology management comprises data collection, the creation of databases, and reviewing and storing terminology and associated metadata in databases, but it also involves categorisation, validation, updating, conversion and other additional subtasks. Inadequate terminology management, however, can lead to integration problems, which in turn bring central storage issues at a company level (Warburton, 2006) and misunderstandings. In view of this, it is therefore necessary to define the proper protocols in order to produce clean and useful terminology databases.

The main component of the concept of terminology management is the central storage element, whereby all project members can access the database. The central storage of terminology and defined protocols require adequate support systems, including software packages and plug-ins, a monitoring system and data analytics. Terminology can be tagged with metadata and later re-used or analysed with a view to its improvement. Kudashev (2013) presents the TermFactory project, which is operated as a web service, distributed in cloud format and run as a collaborative work, and possibly added as a plug-in.

In addition to central storage, the evaluation strategies for terminology databases also need to be defined. As terminology management work is labour-intensive and time consuming, requiring a team with skills in a variety of disciplines, it is important to define at a corporate level a set of criteria for the evaluation process (e.g. time, finances, the staff involved, additional education, tools and resources, customer satisfaction, business communication). Time-consuming work and decisions to invest in certain companies should be data-driven and justified by a business case, often referred to in terms of ROI (return-on-investment) or a cost/benefit analysis (Childress, 2007).

However, it is difficult to separate terminology management from the whole document production system (Valentini, 2016) and from translation, particularly in the field of legislation (Chiocchetti et al., 2017). It also depends on the company’s infrastructure, the number of experts, the type and volume of work and number of languages, in-house processes or outsourcing, the company’s management strategy, distribution channels, etc.

Terminology management is increasingly attractive for commercial enterprises, where it can be leveraged for a variety of uses: to improve business communications (e.g. for marketing campaigns and product information), to increase productivity (Warburton, 2014; Champagne, 2004), for voice-machine communication (Seljan and Duñer, 2014), to improve readability and information transfer through consistency, or for cost reduction, faster turnaround, and improved communication within the organisation itself, or with business partners and customers (Rirdance, 2007), in addition to situations when a substantial number of documents need to be translated (e.g. EU accession, as in Seljan and Pavuna, 2006).

Warburton (2014) expounds the direct, indirect and strategic benefits of implementing a terminology management system. Direct benefits include measurable reductions (less time required for writers, translators, editors and localisers, and financial savings and the avoidance of additional costs), improvements in productivity (non-duplication of work, a consistent and more efficient performance of new tasks, faster production, less turnaround time) and the generation of valuable new information products (for practical work, further research and NLP applications, customers). Indirect benefits
include the re-use of existing resources, flexibility and the possibility to facilitate new tasks for new jobs, increased accessibility by search engines, etc. Champagne (2004) asserts that terminology is mainly assessed through intangible or indirect values through external (customer-oriented) and internal indicators and conditions of volume and effectiveness. Strategic benefits boost a company’s vision and mission, such as branding on the market, customer satisfaction through improved quality and the protection of logos and messages.

2. FROM QUALITY MANAGEMENT (QM) TO QUALITY ASSURANCE (QA)

Quality assurance (QA) focuses on the planning and documentation of all activities (guidelines, steps, rules, the selection of experts, and quality management, which comprises verification, testing activities, and procedures to check products or services in order to detect and rectify any mistakes), with the aim of preventing errors and achieving quality. A further aim of QA is to ensure that all quality requirements are met, both within the organisation and with customers. QA can be achieved by means of a regulatory framework of standards and procedures.

According to Stejskal (2006), QA in the translation process is traditionally centred on the three P’s: provider, process and product.

The QA process encompasses the project management process, pre-production, production proper (localisation engineers, translators and reviewers) and – by means of customer feedback – post-production. Seljan and Katalinić (2017) point out that investing in quality assurance (QA) of the localisation process is essential and achieved through a number of methods and storage structures, with systematic access ensuring visibility.

Ideally, the central translation management system should keep track of the provider, process and product.

Terminology management is closely interlinked with the translation process, insofar as it has a bearing on the quality of the translation product, but it can also be considered as an entirely separate process, one that is carried out in line with the needs of the company. Bogaert (2008) points out the first three steps optimising the terminology management process:

- cleaning (eliminating duplicates, inconsistencies, and unnecessary characters such as blanks, hard returns, etc.) and conversion tools for producing files in standardised formats
- creating guidelines, as part of pre-production, in order to define when to use specific corporate terminology, how to use specific terms, how to add abbreviations or explanations, etc. Guidelines in.pdf format are not very efficient because of the constant waste of time perusing them entails. Instead, the application or plug-in for terminology QA could be used, which suggests inconsistencies or flags up the need for editing and verification. The QA application should provide the possibility of using regular expressions for the automation of some parts and statistics on terminology manipulations.
- feedback on quality, conducted by means of a data analysis of the metadata of edit-terms (noise, duplicates, multiple translations)
- customer-oriented QA of corporate terminology through control verification
- Cleaning, conversion and the guideline drafting begins in the pre-production stage, while the integration of the terminology base into the translation workflow and QA takes place during production proper. Verification of the use of corporate terminology is carried out during post-production.
3. Standards

A quality management system identifies processes (what should be done), procedures (how the process should be done) and working instructions (how procedures should be performed). When applied to translation workflow and terminology management, the QMS will have to meet all ISO 9001 requirements, including documentation conditions. Some of the key components of a QMS include:

- a system that is repeatable, quantifiable and constantly improving
- documented information (such as procedures, working instructions, forms) in order to achieve the pre-defined quality
- external or internal revision to ensure compliance with requirements

Good standards, delivered through systematic quality management, are an indispensable part of the quality assurance process. One of the best-known standards is ISO 9001:2015, *Quality management systems – Requirements*, which defines the set of requirements for a QMS to help businesses to be more efficient and improve their customer relations. The QMS is intended to be an integrative element which brings together aspects of the company, the core business domain and customer satisfaction, integrating people (skills, roles, responsibilities), processes (workflow, interactions), regulatory requirements, technology, risks and opportunities, with the main purpose of delivering the best product.

ISO standard 9001:2015 is regarded as the framework for all management system standards, as it is structured to provide consistency for various quality management standards, including customer-orientation, process approach, the working environment, relationships and leadership, regulatory requirements and the final quality control of products and services. It is underpinned by the principle that quality resides in the perceptions of users and in the applications of a particular product or service.

ISO 9001 is based on seven quality management principles (Figure 1), which help the organisation or businesses to create value:

ISO 9001:2015, *Quality management systems – Requirements*, does not prescribe specific requirements for documented procedures, but leaves organisations the flexibility for their own documentation needs in order to enhance the process. As it is desirable for the quality management
process to be driven by all committed parties (language engineers, localisation experts, IT staff, translators, project managers, the sales team, HR and marketing groups, data analysts, etc.), it is important to ensure that all team members are fully involved, the workflow is efficient, the distribution of tasks is clear, and that employees are given the proper training to broader their knowledge.

ISO 9001:2015 is intended to be easily integrated with other quality management systems, such as ISO 17100:2015, *Translation services – Requirements for translation services*, which specifies the requirements for all processes in the translation workflow in the areas of management, qualifications and skills, resources and other aspects needed for a successful product delivery. As part of the “Concepts related to translation workflow and technology”, ISO 17100 has introduced ideas of machine translation and CAT (computer-aided translation) technology, concepts relating to technology and translation workflow, and, as part of the “Concepts related to people and translation services”, the role of the project manager, and of technical skills as specific skills, supporting the whole translation process.

ISO 12620:2009, *Terminology and other language and content resources*, provides, at an international level, guidelines, constraints, interchange formats and management procedures for a Data Category Registry (DCR) for a number of language resources, such as corpora, machine translation, terminology and lexicography.

4. Levels of quality assurance (QA)

Quality assurance can be carried out at several levels, at different steps or at the end of a specific phase. In any event, it is recommended that it be integrated into everyday work, and part of a routine. Chiocchetti et al. (2017) point out that QA can be performed at various levels, including: workflow, product and staff; with processes, products and people the key elements. Kudashev (2013) points out the workflows and user roles in collaborative terminology work, in contradistinction to the traditional models of content creation. The various sub-levels of quality assurance will be outlined in the following subsections.

4.1. Workflow

Quality Assurance at the level of workflow largely depends on the particular institution, needs, the level of expertise, user-orientation, etc. While some phases precede others, some phases are optional and some overlap. Although workflow quality assurance involves a range of different activities, performed by various types of specialists, they are not always carried out in practice, but nonetheless mainly include:

- the need for analysis
- defining customer needs
- defining domain and sub-domains, size, languages, main principles
- defining time-frame
- literature review on the subject and critical analysis of existing resources
- the principles of data collection
- defining workflow and priorities
- creating the framework for a terminology database, i.e. entry structure (e.g. defining the object of description, semantic properties, domain classification, expression, context, languages, types of metadata)
• programming
• consultations with the domain expert
• term extraction (using linguistic, language-independent or hybrid methods, as in Seljan et al., 2009, 2017)
• term validation
• creating objects of description (concept of fields), population in the terminology base
• translation
• reviewing
• evaluation
• conversion
• integration
• documentation
• revision
• quality analysis
• data analysis
• preparation for publication

Kudashev (2013) describes the various roles involved in this process: language or localisation engineers can develop a framework for monolingual or languageneutral ontology, terminologists can add semantic concepts and relate them, linguists can add linguistic descriptions, translators can translate, programmers can programme, and information and data specialists can perform evaluations. Because of flexibility and object-oriented description, it can be adapted to different lexical resources.

With regard to the process of creating the terminology database, different types of specialists can be involved, such as information and computer scientists, data analysts, localisation and language engineers, project managers, terminologists, linguists and translators, but also the sales and branding team, and economists. QA can also be performed at a variety of stages: to keep track of finished or ongoing tasks, at the end of a specific phase, and at the end of the whole process. This resource can serve for practical implementation as part of the translation process, for the purposes of further research and NLP (natural language processing) applications, or as a business solution for a specific customer or wider audience, where keeping track of data use can serve for future decision-making.

4.2. Production

Quality assurance at the level of production largely falls within the remit of localisation engineers, terminologists, translators, reviewers and QA evaluators, with a direct focus on the actual product. A terminology database has to fulfil criteria relating to content (correct relations), linguistics (appropriate domain, phraseology) and formal criteria (completeness).

QA during production can be performed by humans, although this method is subjective, time-consuming, inconsistent and error-prone. The use of QA tools, on the other hand, can provide great assistance for assessing translation quality and the severity of mistakes. In his description of QA, Makoushina (2007) suggests that tools can be deployed to perform rudimentary linguistic and formatting tasks, which are ideally suited to automation. Where quality assurance is the most demanding however, in terms of time, and also of the utmost importance, is at the level of words. While there are a range of different tools available offering checks on different types of errors, error categories and settings, pure quality control tools mainly focus on the following:
• segment-level checks – the detection of non-existent or incomplete translations, differing numbers of sentences,
• inconsistency checks – at segment and word level, including:
  • terminological inconsistency – i.e. the possible misuse of terminology, the use of terminology that does not comply with the database or project glossary
    - source text inconsistency – identifying different source segments which have the same translations in the target text
    - target text inconsistency – identifying identical source text segments translated in different ways
    - URL errors – URLs which are either mistransferred, inconsistent or not transferred at all
• punctuation checks – end-segment punctuation, double spaces, double punctuation marks, inconsistent use of brackets and quotation marks, i.e. inconsistent formatting or formatting incompatible with target rules, spaces before and after percentage or degree marks, corrupted characters
• orthographical errors and spelling mistakes – capitalisation in source and target text, double words
• date, number, and currency checks – verifying formats, separators, measurement unit conversion, etc.
• terminology control – compliance with terminological resources, correcting misspellings and blacklist checks – ensuring that target segments contain no undesirable translations
• tag checks – ensuring the equivalence of tags and that tagged formats such as HTML, XML, MIF, etc. are in the correct order

4.3. Staff

When it comes to staff, quality assurance requires a range of different experts for highly specialised tasks: translators, terminologists, linguists, information and data analysts, localisation engineers, project managers, domain specialists (e.g. legal, medical), computer scientists, marketing and sales teams, economists. Their roles largely depend on the project type, the customer’s demands and project stakeholders, and can be documented and visualised in order to be utilised to the full. As it is desirable for the quality management process to be driven by all committed parties, it is important to ensure that all team members are fully involved, the workflow is efficient, the distribution of tasks is clear, and that employees are given the proper training to deepen their knowledge.

4.4. Workflow and users in a collaborative environment

Kudashev (2013) suggests an evaluation of the level of workflow and the roles of users in a collaborative environment, setting out the main differences from the more traditional evaluation models:

• working environment (electronic/ virtual online environment, usually not limited to the organisation’s structure, continuous work as opposed to being time-limited, project-driven work, varying commitments to a plan, intermittent instead of more regular working methods, varying working speeds)
• participants: the number of staff working on the construction of a terminology database (potential largescale collaboration); roles (e.g. voting, content creation, term population, object definition tasks, administration), lead editor often undefined, working relationships of
employees (they may not know one another), motivation (voluntary work, self-interest),
assessment of fellow employees’ work (mutual assessment, trust in reputations, trust on the
part of the administrator)

- direct editing (online work, authentication, authorisation, information about expertise,
  constant updating, administrator, supervisor for resolving problems relating to abuse) and
  quality issues (user-generated content, often borrowed from other resources and estimated
  by users, QA depending on collaborative validation, articles edited by a number of different
  people, freedom of expression, selfcensorship, quality contingent on self-appraisal, and thus
  a systematic approach and inconsistent coverage, greater number of languages)

4.5. Software

Most translation quality tools are designed to detect compatibility and formatting errors in segments,
in individual words, and throughout an entire document.

Some of the criteria used to classify quality control tools include (Makoushina, 2007): architecture (as
a standalone tool, plug-in or integral part, compatible with various formats and performing various
functions), reliability (function performance, i.e. whether they are consistently performing according
to specifications), consistency (whether standards are used consistently in a varied multilingual
environment, and whether information is produced in a consistent manner), customisation,
learnability, data analytics, etc.

4.6. Output

Information and communication technology, including human language technologies, can, for the
purposes of transferring a message from the source to the receiver, play an important role in
transferring information to a specific user, by harnessing technology as a communication channel
(Seljan et al, 2015). Desirable outcomes in this regard include a consistent transfer of information and
building positive communication both within an organisation and with customers. It is therefore
necessary to evaluate machine-translated output, both in the form of translations and terminology
databases.

There are essentially two main types of evaluation methods: human and automatic. While human
evaluation is time-consuming, expensive and subjective, automatic quality evaluation metrics are
designed to approach and approximate human evaluation as much as possible. Hovy (2002) describes
the context of machine translation evaluation, as regards quality, purpose and context, and the
desirability of establishing a coherent approach. Brkić et al. (2011) delineate a range of automatic
metrics, but these are problematic for output in morphologically rich languages – a topic of interest
for numerous researchers and organisations. The main advantages of automatic evaluation metrics
are speed, cost and objectiveness. They always perform consistently, and in addition to being
adjustable, can provide meaningful, consistent, correct and reliable information on the level of quality
of machine translation. On the other hand, human evaluation is regarded as the “gold standard”, with
fluency and adequacy criteria of particular use when it comes to evaluating machine translation (Seljan
et al, 2015b).

The quality of automatically extracted terminology databases is also subjected to human and
automation metrics. Seljan et al. (2009) conducted research on term extraction from a monolingual
corpus created from publicly accessible English legislative documents. The research was carried out
using statistical and hybrid approaches and then evaluated by precision, recall and f-measures. In a paper presented by Seljan et al. (2014), an analysis of professional and non-professional terminology concerning patients with diabetes was evaluated and compared with three types of reference lists. Results were evaluated through precision, recall and f-measures.

Further research presented by Seljan et al. (2017) presents an automatic extraction process from monolingual text in the domain of pharmaceutical documentation, performed by three language-independent tools, but utilising different principles. The automatic extraction process was performed following digitisation and the use of OCR techniques. Results were compared with the reference terminology list created by the responsible institution and evaluated, and were also discussed with a view to possible integration into the process of digital archiving.

Gašpar et al. performed an evaluation of automatically extracted terminology in the domain of legislation using a Herfindahl-Hirshman Index (HHI), a commonly accepted measure of market concentration, but used in the paper in question to assess terminology consistency. Automatically extracted terminology is then at the contrastive level and verified in online terminology resources (IATE and EuroVoc).

5. CONCLUSION

The main objective of this paper was to identify the various levels of the Quality Assurance (QA) process as part of the Quality Management System (QMS) in relation to the regulatory framework and applied to the translation and terminology management process. The QMS of a translation workflow is intended to be an integrative element which brings together aspects of the company, the core business domain and customer satisfaction, integrating people (skills, roles and responsibilities), processes (workflow, interactions), regulatory requirements, technology, risks and opportunities, with the main purpose of delivering the best product.

In order to ensure the best possible quality of product, not only one must consider the final product itself (the translation or terminology database), but the entire process and product, as well as all providers and people involved, should be taken into account.

This paper provides an insight into the importance of terminology management at a corporate level, from the perspective of business communication and the benefits that it can deliver to a company.

Quality Assurance (QA), as part of the Quality Management System (QMS), has been analysed at the following levels: workflow, staff, product, users in a collaborative environment, software and output. This paper provides a holistic perspective on the importance of QA for the purposes of delivering product quality, and of its significance within the overall QMS, in relation to the regulatory framework, with product quality always the core aim.

References:


Seljan, S.; Dunder, I.; Gašpar, A. (2013) From Digitisation Process to Terminological Digital Resources. MIPRO.


