Translation Packet Overview

By Arle R. Lommel and Alan K. Melby • Version 0.2 (March 24, 2011) • DRAFT: Not to be used for implementation purposes.

The latest version of this document will be maintained for download at *http://www.ttt.org/specs*.

The Translation Packet

One of the consistent problems faced by the translation and localization industry is that material to be translated can be transmitted in many different fashions (FTP, email, CD-ROM, etc.), often with incomplete or inadequate instructions for how the material is to be translated. As a result, typical providers of translation and localization (T&L) providers (often also called LSPs or TSPs) spend a significant amount of manual effort in manipulating files, clarifying instructions, and verifying that files move from place to place correctly and are translated according to client expectation. One large T&L provider recently reported that addressing these issues cost it in excess of $3 million/year.

One solution to part of this problem is the use of a standard format for transmitting content, such as XLIFF, which allows localizable material to be extracted and transmitted in a standard fashion. When XLIFF is used consistently it simplifies the process of obtaining localizable material and ensuring its completeness, but it does not ensure that specifications for the translation process are correctly transmitted nor does XLIFF address the needs of all T&L tasks (e.g., graphics localization may go beyond the translation of strings and DTP tasks are frequently carried out in the T&L process). So while XLIFF is a tremendous benefit for users, it will not eliminate many of the manual issues associated with the T&L process.

The *translation packet*, by contrast, embraces existing standards to provide a standard mechanism for the transmission of translatable content together with other resources needed to facilitate the T&L process. The intention is that a complete and valid packet should contain (or reference) all of the materials and project data needed to fully process a T&L transaction while minimizing the need for manual intervention or negotiation between the T&L supplier and requester after start of the transaction. The use of the translation packet specification would apply regardless of the tools to be used. While it will not eliminate all manual processing or management steps, it will reduce them greatly by providing a standard way for translation tools to interact with the resources in the packet.

The translation packet is intended as a scalable format for the transmission of all materials needed in the T&L process. Its structure is suitable for translation requests of just a few words or for those with hundreds of thousands of words, all using the same generic control structure. Because the structure is flexible but controlled, tools that use it will know how to interpret the contents and should require manual intervention only when it is actually needed. (For example, the packet might contain instructions on how to obtain materials at a secure facility, an inherently manual task, but these instructions would persist in the container, eliminating the need to pass separate emails or messages with these instructions.)

At the heart of the packet is a *structured translation specification set (STSS)*, a set of metadata concerning the T&L transaction that explain how it is to take place. It explains customer expectations and requirements for the product of the T&L process. By using the STSS, many of the causes of conflict or rework will be eliminated from the T&L process and the reasons for any breakdown will be easier to identify. The STSS

also serves as a contract for the work to be done and can assist in the T&L procurement process because it identifies the variables that are likely to affect project costs in advance and requires requesters of T&L services to be clear about what it is they want from service providers.

Structure

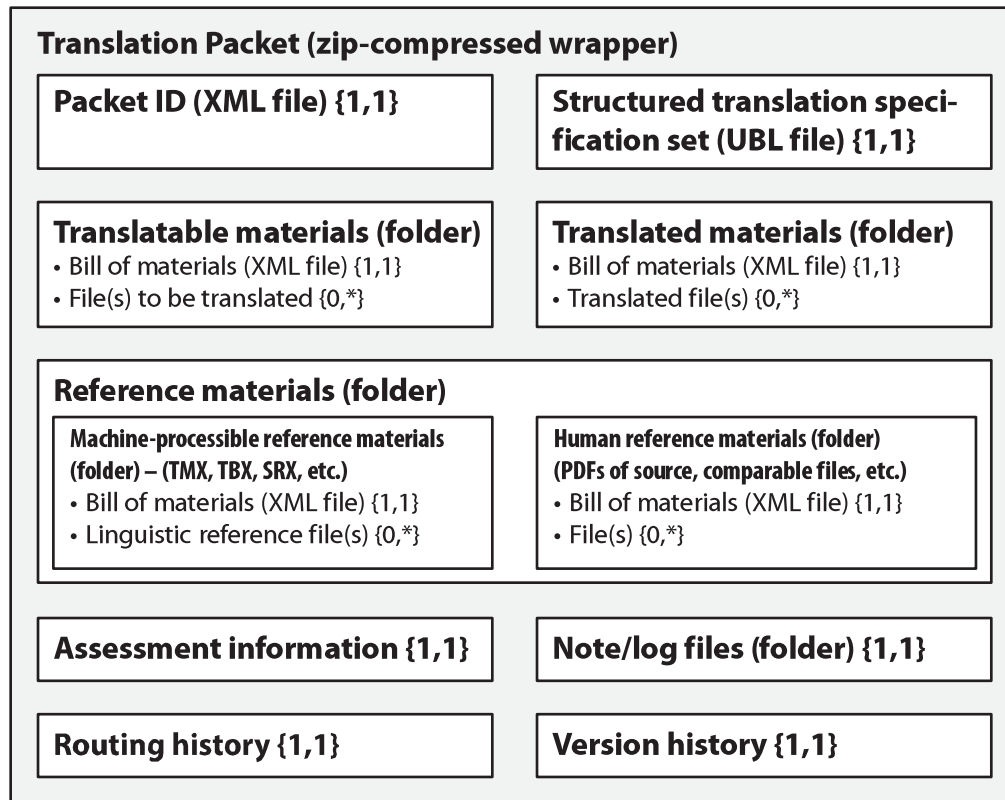A high-level overview of the structure is presented as **Figure 1**.

Translation Packet (zip-compressed wrapper)

Packet ID (XML file) {1,1}

Structured translation specification set (UBL file) {1,1}

Translatable materials (folder)
- Bill of materials (XML file) {1,1}
- File(s) to be translated {0,*}

Translated materials (folder)
- Bill of materials (XML file) {1,1}
- Translated file(s) {0,*}

Reference materials (folder)

Machine-processible reference materials (folder) – (TMX, TBX, SRX, etc.)
- Bill of materials (XML file) {1,1}
- Linguistic reference file(s) {0,*}

Human reference materials (folder) (PDFs of source, comparable files, etc.)
- Bill of materials (XML file) {1,1}
- File(s) {0,*}

Assessment information {1,1}

Note/log files (folder) {1,1}

Routing history {1,1}

Version history {1,1}

Figure 1. *High-level structure of a translation packet*

The packet itself consists of a zip-compressed directory (with optional encryption) containing a pre-defined directory structure. The contents of this directory are as follows:

1. **Packet ID** (XML file). This is a lightweight XML file that specifies the unique ID of the packet. It also contains an optional reference to a user-defined project or tracking ID that specifies for accounting and tracking purposes what project, purchase order, etc. the packet belong to. All other XML files in the packet must reference the unique ID of the packet to help ensure the integrity of the packet. It may also contain contact information for the requester.
2. **Structured translation specification file** (XML file). This UBL file contains parameters for the translation of the packet. (The UBL representation is under development, but a description of the parameters is presented starting on page 5.) For a packet to be considered valid, the specification must be filled out entirely; incomplete specifications could result in rejection of a packet or require additional manual steps that delay the T&L process. **In most**

cases it is anticipated that the specification would be filled out using pre-defined templates that would pre-specify most of the options.

3. **Translatable materials** (folder). This folder contains (or references) the actual materials that are to be translated in the T&L process. The folder contains the following materials:

    a. **Bill of materials** (XML file). This file specifies what materials are included or referenced in the directory. Each file to be translated *must* be referenced by either a local URI (that resolves to a file in the *Translatable materials* folder), or a valid Internet URI that specifies where the files may be found. In the case of materials that cannot be included or directly referenced (e.g., files stored at a secure facility or on a server requiring login validation), it may contain instructions on how to locate/obtain the files. (Note that if files are not available via URI some automated processing functions may not be available.)

    b. **File(s) to be translated**. If files that are to be translated are stored locally in the packet, they are included here. Each file *must* have a corresponding reference in the *Bill of materials.* XLIFF is the preferred format for these materials, although other formats are allowed. (Note that in many cases there may be no files to be translated in a packet since they may reside external to the packet, in which case no files will be included here.) The files to be translated may reside in a directory structure within the *Translatable materials* folder.

4. **Translated materials** (folder). This folder contains (or references) the actual materials that were translated in the T&L process. It is expected that in most cases this directory will initially be empty (except for an empty *Bill of materials*). The folder contains he following materials:

    a. **Bill of materials** (XML file). This file specifies what materials are included or referenced in the directory. Each translated file *must* be referenced by either a local URI (that resolves to a file in the *Translated materials* folder), or a valid Internet URI that specifies where the files may be found. In the case of materials that cannot be included or directly referenced (e.g., files stored at a secure facility or on a server requiring login validation), it may contain instructions on how to locate/obtain the files.

    b. **Translated files**. If translated files are stored locally in the packet, they are included here. Each file *must* have a corresponding reference in the *Bill of materials.* (Note that in many cases there may be no files to be translated in a packet since they may reside external to the packet, in which case no files will be included here.) All translated files included in the packet *must* be stored in subdirectories named with fully specified locale identifiers using CLDR locale-identifier conventions. (For example, if a project is translated in French for Canada and German for Germany, it would contain directories named "fr_CA" and "de_DE" respectively.)

5. **Reference materials** (folder). This folder contains or provides directions to access reference materials to be used in the T&L process. It contains two subdirectories:

    a. **Machine-processable reference materials** (folder). This folder contains or references linguistic reference materials intended to be used in machine processing of the file. These materials may include translation memory (TMX or application-specific files), terminology resources (TBX or application-specific terminology files), segmentation rules (SRX), etc.

        i. **Bill of materials** (XML file). This file specifies what materials are included or referenced in the directory. (Similar guidelines apply as to other *Bills of materials*.) It may also contain instructions to use existing resources (e.g., to use TM data that the T&L provider already has in-house).

        ii. **Linguistic reference files**. If any linguistic reference files are included in the packet, they are included here.

      b. **Human reference materials** (folder). This folder contains or references any materials intended for the use of human translators *that are not intended to be processed in a CAT tool* (i.e., they are for informational purposes). Examples may include PDFs of source files, samples of similar documents, or background material that may be useful to the translation.

           i. **Bill of materials** (XML file). This file specifies what materials are included or referenced in the directory. (Similar guidelines apply as to other *Bills of materials*.) It may include notes on general resources that are to be used.

          ii. **Files**. Any local files for human use are to be stored in this directory.

6. **Assessment information**. (Note: This section is not yet defined). Contains information on quality assessment requirements and tasks.

7. **Note/log files** (folder). This folder contains a log of all transactions involving packet contents. It may also include notes added by humans during processing or use of the packet.

8. **Routing history**. This file provides an audit trail for the packet. Each tool that accesses the file must log its access in the routing history, even if it does not intentionally otherwise change the packet. This history is intended for use in automatic processing as a verification step that the packet has followed proper processes.

9. **Version history**. This file contains a summary of changes made with each processing step. (The nature of this function is yet to be defined.)

Remaining tasks

The following tasks remain:

- **Creation of sample packets (March 2011)**. The initial version of the packet will be released in late March. At this point we need non-confidential/sanitized sample data from real projects (ideally in standardized formats such as XLIFF, TMX, and TBX, although these are not required) to put in example packets. The initial representation will include plain-text representations of the specification document while the full UBL representation is created.
- **Development of UBL structured translation specifications set (STSS) format (April 2011)**. We are currently working on development of the UBL representation of the STSS format. When complete it will provide a standard, machine-processable format for this component.
- **Testing (Q2, 2011)**. After the STSS format receives preliminary definition, we will request testing of the packets and comment on the format from the perspective of a production environment.
- **Submission to a standards body for further development (Q3, 2011)**. When we have a stable proposal, we will work with the community to submit the project to the appropriate standards body as a proposal for the standardization process.

Conclusion

As can be seen from this overview, the Translation packet provides significant functionality that goes beyond what can be supported by any single file format (e.g., XLIFF) to provide full support for the entire T&L process. It builds upon existing standards (TMX, TBX, XLIFF, OAXAL, etc.) to create a new infrastructure for improved translation processes.

List of Translation Parameters

The following is a list of the parameters to be addressed in a Structured Translation Specifications Set (STSS). Full descriptions follow on page 6.)

A. Linguistic [1–13]
source content information (not dependent on target language)
- [1]   content type, audience, and purpose
- [2]   subject field(s) and terms
- [3]   volume
- [4]   complexity and obstacles
- [5]   origin

target language specific information
- [6]   language(s) and region(s) (including accuracy and fluency)
- [7]   audience(s)
- [8]   purpose(s)
- [9]   content correspondence
- [10]   usage register
- [11]   file format
- [12]   style
- [13]   layout details

B. Production tasks [14–15]
- [14]   typical production tasks:
  - a)   pre-translation tasks
  - b)   initial translation
  - c)   in-process quality assurance and control tasks
    - i.   self-checking / post-editing
    - ii.   revision
    - iii.   review
    - iv.   final formatting or compilation
    - v.   final reading
- [15]   additional tasks:
  - a)   completeness check
  - b)   terminology check
  - c)   functional testing
  - d)   back translation
  - e)   random sampling

C. Translation environment [16–18]
- [16]   technology
- [17]   reference materials
- [18]   requirements provider must satisfy

D. Deliverables, deadlines, and restrictions (expected by the requester) [19–21]
- [19]   terms and contents of delivery
- [20]   legal and ethical considerations

E. Compensation and other items expected by the provider
- [21]   compensation and clarification

Descriptions of Translation Parameters

A.1 General

A structured set of 21 translation parameters can be divided into the five sections listed below. Many of these specifications are included in national and regional translation quality standards although they may not be found all in one place with exactly the same names as in this Technical Specification.

A.1.1 Specification Types

a) **Linguistic [1–13]**: information about the source content (its document type, language, and intended audience and purpose) as well as target language specific information.

b) **Production tasks [14–15]:** tasks which are to be performed during the production phase, which begins once the final specifications have been approved by all parties.

c) **Environment [16–18]:** requirements concerning the environment in which the translation takes place

d) **Deliverables, deadlines, and restrictions (expected by the requester) [19–20]:** all terms concerning delivery, including, but not limited to, items beyond that of the translated content, deadlines, and restrictions

e) **Compensation and other items expected by the provider [21]:** Terms of payment (form and timeline) in exchange for deliverables; other expectations of the requester, such as providing clarifications about the source content as need by the provider

These specifications form a package that defines and guides a translation project and allows the entire translation project to be evaluated. Sometimes a formal agreement or contract will accompany the specifications. For many projects, especially for those arranged between clients and TSPs who have previously worked together, some of the specifications are well implicitly understood and need not be laid out in writing.

The first five specification types are useful in developing initial project specifications. These types [1–5] are highly relevant to pre-production activities, especially when one of the stakeholders (e.g., purchasing agent representing the client) does not know the source content language or when the source content is confidential and only a few people in the production chain are allowed to see it. In these cases, a description of the source content must be made explicit by examination of the source content. Furthermore, selecting an adequate translator cannot be made without knowing the content type and subject field(s). To estimate the cost of a project obviously requires knowledge of the volume of the source content. Cost estimates can also be influenced by the complexity of the source document. For example, the effort required to translate the text in a graphic (e.g., images, diagrams, or even Flash presentations) depends on whether the text is editable or whether the client has provided at least a text-less version of the graphic so that the translated text may be added to the graphic as a separate layer. Such source text obstacles may dramatically affect the degree of difficulty in a translation.

The rest of the specification types [6–16] are relevant to quality (i.e., conformance to specifications that have been agreed upon by all stakeholders). Conformance to some specifications cannot be determined

solely by examining the target text. For example, an otherwise good translation that is inappropriately divulged to a third party or that is delivered late would not be considered a job well done.

A.1.2 Parameter vs. Specification

In this Technical Specification, a distinction is made between parameter and specification.

A parameter is a question. A specification is an answer.

EXAMPLE   Parameter: content type.   Specification: patent.

EXAMPLE   Parameter: *subject matter*.   Specification: *plasma physics*.

EXAMPLE   Parameter: volume.   Specification: five thousand words.

Parameters can also be called *specification types*. Parameters form a framework for creating structured specifications. Without this framework, the names, descriptions, and order of the specifications in a package might vary widely.

Project specifications (that is, the parameters values for a particular project) are relevant during all phases of a translation project. Below are descriptions of the various specification types (i.e., parameters) and their relevance to project and product assessment (i.e., assessment of the translation as a product of the project). The relevance of specifications to all aspects of the production phase is so direct that it is not reiterated for each specification type.

At first glance these five initial specifications types may seem irrelevant to project evaluations and product assessments. Nevertheless, the content type, subject field, and volume are crucial factors in choosing an appropriate person or persons to assess the translation. Also, surprises in content complexity and obstacles can be key to understanding why a cost estimate was significantly lower than the actual cost.

A.2 Linguistic [1–13]

A.2.1.1 Source content information (not dependent on target language)

A.2.1.2 [1] Content Description

What is the language of the source content? What is the regional variation (e.g. British vs. American English)? What was the audience for the source content? What was the purpose of the source content when it was created?

Is the source content a patent, a user manual for a high-tech device, the user interface to a medical device, a letter from the president of the corporation, or some other type of content? Much has been written on content type, that is, text types and genres.

Note: A separate document is being prepared that reports on a content type in terms of text type and text genre.

A.2.1.3 [2] Subject Fields(s) and terms

Is the source text about chemical engineering or political science or any of thousands of other subject fields? A subject field is also called a domain. Often, a translator specializes in particular subject fields, rather than translating only general-language documents. It is crucial to select a translator who is familiar with—or

willing and able to invest the time needed to become familiar with—any subject fields that are crucial to fully understanding the source content.

Any terms that are expected to be translated in a particular way, such as according to a corporate terminology database, should be identified by the requester in the source content.

A.2.1.4 [3] Volume

Volume can be measured in many ways (including characters, words, lines, and pages). Although billing is sometimes based on target-text volume, analysis before beginning the translation must obviously include the volume of the source text. Sometimes there is disagreement over volume since two software applications might provide differing word counts for the same document. A standard for measuring and representing volume, known as GMX-V, should eventually help in this regard when it becomes widely implemented.

A.2.1.5 [4] Complexity and Obstacles

There is a lack of formal studies of complexity. Yet, complexity is a factor that influences the difficulty of translating a text. For example, if there are graphics in a source document, is the text within the graphics editable? Is the source text more than plain text? For example, does the document have footnotes or endnotes?

Sometimes the source text contains obstacles to translation such as internally inconsistent terminology, a lack of cohesion or grammaticality, spelling, or even factual errors, difficult discourse structure, or inconsistent formatting. Such obstacles influence the translation. When faced with these source text obstacles, should the translator take on the additional task of improving the source text?

Another example of especially challenging obstacles is plays on words.

When a source text with multiple obstacles is intended for translation into several target languages, the authors may want to consider eliminating such obstacles by optimizing the source text for clarity, conciseness, and consistency before giving the text to translators. These measures will make the translation process more efficient and less costly.

A.2.1.6 [5] Origin

Although not as obviously crucial as subject field(s) and volume specification types, the origin of the source text can be highly relevant to a translation project, especially in pre-production planning.

For example, in one case a French text was commissioned and translated into English before the stakeholders had even realized that an English version existed and that the French "source content" was actually a translation of the English original. If an inquiry had been made as to the origin of the source text, it would have revealed that the source content was itself a translation. An appropriate step would have been to obtain the original text rather than to translate a translation. Thus, failure to determine the origin of source content can influence the assessment of a project as a whole and can remove the need to evaluate the translation.

Another relevant example of information important to understanding the origin of the source text would be an awareness of the author's native language. For example, if the source text author does not possess near native proficiency in the source language, interference from the native language of the author can influence the source text; knowing the source of this language interference may help in understanding any unnatural expressions.

A.2.2 Target language specific information [6–13]

A.2.2.1 [6] Language(s) and Region(s) (including accuracy and fluency)

Knowing the language of the source text and the language to which it will be translated is not enough for adequate translations. The regional variant of the *target* language (sometimes called a locale) is particularly important. A translation done for a Quebec audience may not be acceptable in France. Additionally, the regional variant of the *source* content language is also important. There are many differences between regional variants of English, such as US vs. UK, which could influence understanding on the part of the translator.

Specification type 6 is obviously relevant to **pre-production**. It is generally the first question asked about a translation project: What language do you want this translated into? And the next question—unless the answer is obvious—should be: what regional variant of the language is needed? Language and region are also relevant to **project assessment**. Although a project seldom goes so disastrously wrong that the source content is translated into the wrong language, it may more frequently be translated into the wrong regional variation.

It is usually implicit that the translation should be both accurate and fluent. Accuracy refers to the connection between the source text and the target text. If a dog becomes a tiger in a translation, that would usually be considered a lack of accuracy. In other words, an accurate translation is one free of translation errors.

Fluency, refers to the readability of the target text. It consists primarily of mechanical characteristics such as spelling, punctuation, and grammar. It can be assessed without reference to the source text. A translation can be highly fluent without being at all accurate. If the requester does not expect a fully accurate and fluent translation, this should be indicated in the language specification. For example, a translation may be needed into rough German very quickly and fluency is not essential. An important aspect of fluency is textual cohesion.

A.2.2.2 [7] Audience and [8] Purpose(s)

The audience and purpose of a translation may be the same or different as that of the source text. It is crucial to know whether a shift in audience and purpose has been requested for the translation.

In some sense, the audience of a translation will always differ because it is intended for speakers of a different language, but sometimes this shift in audience goes beyond language. For example, a source text written for specialists might be translated for a more general audience. Though some would say that this is no longer a translation, a broad view considers translation to be a production of a text in the target language that corresponds to a text in the source language and which meets a certain set of specifications; thus a project with substantial differences between the source and target text specifications may result in a text which is not a translation according to this narrow view. In many cases, the purpose of a translation is often the same as the purpose of the source text, but any shift in purpose is highly important for the translator to be aware of and must be documented. Thus, the audience and purpose parameters are crucial to a translation project.

Many other specification types are related to types 8 and 9. Indeed, it might be theoretically possible to provide an extensive description of the audience and purpose of a translation relative to the audience and purpose of the source content and include sufficient information in that description to make many other specification types redundant or at least inferable. However, it is more practical and useful to give a short

description of the audience and purpose of a translation and then distribute other information through other specification types unless the audience of the translation is not obvious or the purpose of the translation differs from the purpose of the source content.

Another complication in this specification type is that there can be multiple purposes for a translation. This must be carefully documented and can influence both production and assessment. The assessment question here is whether the translation achieves all intended purposes.

A.2.2.3 [9] Content Correspondence

The history of translation is full of debates about the content-correspondence parameter. How faithful to the source text should a translation be? Historically, this debate was couched in the difference between a literal translation and a free translation. The debate has been more recently cast as a question about whether a translation should be *foreignized* or domesticated. That is, should it reveal the "foreign" nature of its basis (the source text), or should it read like a text authored in the target language with no trace of the source text? Essentially, the same contrast is captured in the terms *overt* and *covert*. A **covert translation** pretends to be an original text. An **overt translation** overtly discloses that it is a translation.

Localization is a process that includes adaptation of non-textual material and a covert translation of textual material. The objective of localization is to create a target-locale version of a product or service that gives the appearance of having been created in the target language and geographic region. Localization began with software products but now applies to any product or service. However, localization is not usually applied to the translation of a text that is not part of larger system. There are two ways to think of a non-localized text, as 1) overt translation or as 2) untranslated text (source content never translated to the target language).

Sometimes it is thought that the covert/overt distinction only applies to literary translation. This is not so. Not all commercial, government, or non-profit translation is localized, that is, made covert. There are multiple reasons for not localizing. One reason is to produce a translation that can be used in multiple locales. This is sometimes called a generalized translation. Another is to reduce cost. Localization is expensive. For the audience and purpose, will a non-localized translation be sufficient? There are many questions that need to be answered once localization is chosen. How are proper nouns to be handled? Are they translated or left in the source language? How much liberty can be taken with the information in the source content, so long as the purpose of the translation is maintained? Can target-culture-appropriate objects be substituted for source-culture-specific objects? What if cultural substitution causes ripple effects in the translation?

In addition, it should be pointed out that covert vs. overt does not constitute the whole picture for this specification type. It is only one dimension of content correspondence.

Another dimension is whether the translation should be idiomatic. That is, should it conform exactly to the word order and collocational conventions of the target language? It is often assumed that all translations, whether covert or overt, must be idiomatic. However, some reasons exist for specifying a non-idiomatic translation. If the purpose of a translation is to study a text in its original cultural context, it may prove desirable to create a non-idiomatic translation that sticks very close to the original. In the extreme, such a non-idiomatic translation is called a "gloss." If the purpose of a translation is to allow the reader to get the gist of a dynamically created source text, in a context where human translation is not feasible (such as translating a texting messages between speakers to two different languages) then a non-idiomatic, raw machine translation that does not conform to the norms of the target language may be preferable to no translation at all (sometimes called zero translation).

Thus we have two dimensions of context correspondence: idiomatic and non-idiomatic. For an idiomatic translation, there is the additional dimension of whether the translation is overt or covert.

A third dimension of content correspondence is whether the translation is a full translation or a *summary translation*. Summary translation may sometimes be disallowed, or in other words not formally included under the definition of translation, but it is a growing part of the world of commercial translation.

Clearly, content correspondence is relevant to assessment of the product and thus the entire project. Were content correspondence specifications followed? A brilliantly localized translation for Mexico is not a good translation, and its project is not a success, if the specifications were a generalized Spanish translation for all of South America. A very literal, that is non-idiomatic translation, is not a good product if a covert translation was specified.

The default specifications associated with the content correspondence parameter are idiomatic, covert, and full. However, there are reasons for specifying departures from this default in all three dimensions of content correspondence.

A.2.2.4 [10] Usage Register

Typically, a translation is expected to use a register that corresponds to the register of the source text. That is, an informal source text would be translated into informal language, and a formal source text into formal language. However, sometimes a shift in register is expected. Register is a complex topic, and a formal/informal contrast is a vast oversimplification of this parameter.

The desired register must be agreed on during pre-production and is relevant to assessment. Has the translation reached the agreed upon register? Expensive translations have sometimes been rejected because their register did not match the intended audience.

A.2.2.5 [11] File Format

Suppose the source text is delivered as a PDF file. This does not mean the target text will be delivered as a PDF file. It is important to know the file format of the source text (e.g., HTML file, word processing document, etc.) and the type of file format that the target text should be delivered in. An extensive list of file formats is therefore necessary in the description of this parameter, even though a truly exhaustive list is not feasible. In addition, a discussion of the impact of content management systems is needed.

A.2.2.6 [12] Style

Style is a complex aspect of a translation project. Three important aspects of this specification type are 1) style guides, 2) general style, and 3) specific style of terminology variations.

a) **Style guides**. If a style guide is to be followed in the translation, it must obviously be made available to the translator. Otherwise, someone may have to make many wasteful changes during the revision and review process. If the style guide that was used in creating the source text is known, it may be helpful to make it available to the translator. Certainly, the target-language style guide and a summary of differences with the source-language style guide would be useful.

b) **General style**. For some translations style matters very much. The translator must not only understand the source content, be proficient in the specified variant of the target language, and be able to make source and target content correspond (i.e., transfer between the source and target languages and cultures), but must also be an excellent writer in the target language.

One cannot expect to translate with more style than one can write in his or her native tongue. For many translations, style is not as important as accuracy of information in the target text.

c) **Terminology style**. In scientific, technical, and commercial texts, terminology is often a crucial element. It is not sufficient to be consistent using a target language term for a given source language term when the domain-specific concept is the same. It is essential to conform to a particular terminology database (also called a termbase) or glossary provided to or developed by the TSP. A terminology resource (termbase or glossary), may be specific to an organization or even a project or product within an organization. It would be bad style to depart from the specified terminology resource and could also make a translation unusable until corrected.

A.2.2.7 [13] Layout Details

Layout details such as margins and point size may or may not be included in a style guide. Anything relevant to the exact layout of the translation which is not found in the style guide should be specified as the value of this parameter. Often, the translator is not asked to be concerned with the final layout. However, someone on the translation team will need to be aware of the requirements for the layout of the final product to be delivered to the client.

A.3 Production

A.3.1 General

Project specifications should identify the tasks that will be performed during the production phase. There are two parameters for production tasks. The first lists the typical tasks that are part of the production phase in the order they are typically performed. The second production parameter lists additional tasks that may be required in some projects at some point in the production process.

A.3.2 [14] Typical production tasks

The following are production tasks typical to translation projects.

a) **Pre-translation tasks:** this includes both identifying terms in the source text and deciding on corresponding target-language terms to be used in this project. Ideally, source-content terms were already identified in the pre-production phase.

Sometimes the terminology task is skipped as a separate task in a translation project, and terms are dealt with as they are encountered. The consequences of this may vary, depending on such factors as the volume of the source text and whether there are multiple translators involved in the project.

b) **Initial translation:** this may be human or machine translation.

Some projects involve assessing and correcting a completed translation, so even this task is not universal.

Once any preliminary terminology work is completed and the initial translation is available, then quality assurance and quality control tasks begin.

c) **In-process quality assurance and control**

In-process quality assurance often consists of sub-tasks, such as

1. **self-checking / post-editing:** the translator checks his or her own work or post-edits machine translation.
2. **revision:** the source and target texts are compared for content correspondence; also called bilingual editing in the ASTM standard.
3. **review:** the target text is evaluated by a subject matter expert; typically corresponds to monolingual editing in the ASTM standard.
4. **final formatting or compilation:** the target text is formatted for final reading or the target context, such as user interface strings, are compiled for final reading
5. **final reading:** the target text is proofread on its own and a final global comparison is made with the source, possibly requiring a clearance note.

NOTE: Sometimes, proofreading is not done by the TSP but by the client. Therefore, it is crucial in the specifications to make it clear who is doing it, so that it does not get left undone.

Pre-production planning should always involve consideration of how the initial translation will be done and who will perform the five in-process quality assurance sub-tasks. In the CEN standard, revision must be done by someone other than the translator who produced the initial translation, unless the client formally agrees to an exception to this rule. The ASTM standard is not so rigid. This guide suggests that if the client is interested the number of people involved in performing these tasks should be divulged. In a very simple one-page translation project, both the initial translation and all five in-process quality assurance and quality control sub-tasks may be performed by one person. The in-process specification types are relevant to project and product assessment, especially when tracking down the source of a problem in a faulty product delivered to a client.

A.3.3 [15] Additional production tasks

In some cases, some or all of the following additional tasks need to be performed:

a) **Completeness check:** Does every segment of source text that should be translated have a translation? Are "do not translate" segments left in the source language?

b) **Terminology check:** Necessary to verify the internal terminological consistency of a target text to an external termbase.

c) **Functional testing:** Necessary for translation and localization projects for software and hardware and websites.

d) **Back translation:** Translation of the target text back into the source language, as a means of checking the quality of the translation, much hated by most translators but sometimes required by the client.

e) **Random sampling** is not a separate task but rather a method of implementing any of the tasks in 14 and 15. On occasion, a client or provider will randomly select small samples of a large translation, especially in the case of high ongoing volume, and have them checked by someone else using a quantitative measure of translation quality. Results of random sampling conducted by a client should be at least summarized and made available to the provider at regular intervals as part of a continuous-improvement plan.

Third party review is not mentioned here, because in this document, third-party review refers to review conducted by the client after the product is delivered without involving the TSP. If the TSP is involved, then this review is part of the specifications derived from specification type 14.

A.4   Translation Environment [16–18]

A.4.1   [16] Technology

What technology will be used during the project? Sometimes the client cares—and sometimes not; but the project manager always cares. This specification type clearly affects pre-production planning. It can be relevant to post-production troubleshooting if the specified technology is not used at all or is not used appropriately.

A.4.2   [17] Reference materials

What resources are available to the translation team? This team includes not only the translator but also any other people involved in the tasks to be performed for the current project. Typical reference materials include related documents from the client company (in source and target languages), glossaries or termbases, and translation memories created in related projects. Gathering reference materials is, of course, part of pre-production, and reference materials may be consulted when a disagreement arises about the use of any term.

A.4.3   [18] Requirements provider must satisfy

Are there any restrictions on where the work is to be performed? In some government translation projects, all work must be done in a secure facility. This is one extreme. At the other extreme, work can be done anywhere, even at home, over the Internet. How confidential is the material in question? What procedures are to be followed to maintain confidentiality? This is relevant to pre-production planning and may be the most important project assessment factor if it is violated. What other requirements are imposed on the provider? This parameter can be used for material associated with an RFP (Request for Proposals), which is a Call for Tender in British English. For example, providers that reply must be certified according to EN 5038 or must be incorporated in Canada.

A.5   Deliverables, deadlines, and restrictions

A.5.1   [19] Deliverables and deadlines

What is to be delivered? Is it just the translation, or is it translation and an updated translation memory? What about new terms identified during the translation process? Are they to be entered into a termbase and delivered with the translation? What is the due date? What are the consequences of being late? In some cases, the target language and the due date of the translation are among the most obvious specifications for a project, from a client's perspective, and may override all other factors in assessing the success of a project. These are expected of the provider.

A.5.2   [20] Legal and ethical considerations

Who owns the translation? Will the translator's name appear on it? If so, will the translator be allowed to proofread the translation after all other changes are made by other parties? Who owns the translation memory associated with the project? Every project has legal and ethical considerations. Neglecting to follow this type of specification can result in loss of business or even lawsuits.

A.6   [21] What the provider expects of the requester

How much will the TSP be paid for the translation? How long, once the invoice is sent, will it take for payment to be made? In some situations, such as the in-house translation department, there may be no formal invoice or exchange of money.

What is expected of the requester during the production phase? Is the provider allowed to ask questions about the source text?