

Linport as a Standard for Interoperability Between Translation Systems

Alan K. Melby¹, Tyler A. Snow²

[1]Department of Linguistics & English Language

[2]Translation Research Group

Brigham Young University, Utah, United States

akmtrg@byu.edu, tylerasnow@byu.edu

Abstract

In the translation industry there are a plethora of emerging and evolving technology standards, and a great need for interoperability among them. Some interoperability standards include XLIFF, for translatable bi-texts; TIPP, which adds a package format to coordinate files including XLIFF for a single translation task; and finally Linport, which contains translation data at the project level. This paper introduces and analyzes the viability of Linport, a solution to translation tool incompatibility issues. Linport and XLIFF are complementary interoperability standards.

Keywords: *standards, interoperability, linport, XLIFF, bi-texts, translation data*

1. The Need for Linport

The modern world shows ever increasing opportunities for connection and communication with others. Thanks to modern technologies such as satellites, cell phones, and the Internet, instant global communication is possible, a phenomenon unimaginable even a few decades ago. This level of global interoperability, or the capacity to work with others to accomplish tasks quickly and easily, is a defining achievement of our era. The great advancement of interoperability in the age of globalization will continue to make life easier. However, increased interconnectivity presents problems of its own. Being able to talk to someone on the other side of the world does not guarantee *communication* can be achieved. This is why the translation industry exists. There is also a need for technological communication within this industry. This suggests that new standards should be developed in order to promote interoperability.

A classic example of interoperability breakdown comes from the shipping industry. For thousands of years countries and companies would ship products internationally, attempting to maximize trade profits by cutting costs and travel time. Attempts to streamline these processes revealed that cargo ship containers did not fit onto the trains and boats that carried them to their final destinations. Container types and sizes varied greatly. Consequently, trucking and train companies found it difficult to plan for the

movement of incoming goods. The trade goods would have to be unpacked from the shipping container and then repacked into containers which would fit on a truck or train, wasting valuable time and money in the process. The industry needed a standard. Ultimately, the international community agreed upon a standard shipping container size. These standardized shipping containers could be easily moved from ship to truck without removing the contents as both the ship and the trucks were made to handle the exact dimensions of the pre-specified containers. These improved measures of interoperability enhanced the shipping industry's productivity on a global scale.

As mentioned earlier, this need for standards applies not only to the shipping industry but to many other fields as well, including translation technologies. Translation companies are constantly vying for their translation software to be used and recognized in the world of commercial translation. Some of these companies include SDL, LingoTek, MultiLing, Kilgray, SYSTRAN, and XTM-Intl, among many others. Each translation software system has a different interface that handles translation projects in different ways, often creating interoperability difficulties.

For example, if one company starts a translation project using SDL tools and then subcontracts out of house to a freelance translator who uses Kilgray, the same problem ensues as found in the shipping container example. The various project

components, such as source text, specifications, and file types, all have to be “unpacked” from the SDL format and “re-packed” or reformatted in the Kilgray-style format in order for the freelance translator to do their allotted portion of the project. Finally, the completed translation then has to return to SDL format. Similar to the shipping industry’s need for standardized containers, the translation industry needs a standard “container” of its own, to allow for interoperability between the numerous translation software tools now available to professional translators and companies worldwide. Fortunately, a standardized translation project container will soon be available.

2. Linport: A Standardized Container for the Translation Industry.

Limport (Language Interoperability Portfolio), a complete and interoperable container solution for all translation processes and projects, is already under development. A Limport container documents the details of an entire translation project, and carries each of the individual components that comprise the various tasks relative to a translation project. These tasks could include the initial translation of a text from one language to another, a translation revision, a review, or a proofreading task. Each of these tasks would be accessible to any participant in the project who, upon task completion, would be able to “pack” their goods into the Limport container for further use. Therefore, keeping with the shipping example, a Limport portfolio represents an overall project view much like a shipping container but also would be able to define particular translation tasks within the project. These are comparable to standardized boxes that are shipped in the larger container and could be represented by any number of translation formats, including XLIFF and TIPP.

2.1 Elements of Limport

The current implementation of Limport is represented by a directory structure format containing two sub-folders. The first is the *portInfo folder*, similar to an HTML header element. It contains information about the portfolio as a whole, such as specifications and support files that apply to all relative subtasks, and universal identifiers to facilitate breakdown and reintegration of the portfolio.

The second element is the *payload folder*, which contains translated or un-translated documents for

review, as well as the supporting resources needed for translation or any other required task, such as revision or review. Examples of these resources could include translation memory files, textual references, terminology files, and style guides, among others.

Limport can contain almost any file format, as long as it fits into the predefined directory structure. To facilitate this methodology, the payload folder is divided into language folders, such as the “en” folder for English or “es” folder for Spanish. These folders then contain document folders which house exactly one document in a “doc” folder and its supporting files, such as glossaries and translation memories, along with the document’s specifications, in a support folder. In this manner, translation tools know how each file correlates to another and can handle them appropriately. The directory structure outline can be found at: <http://dragoman.org/limport/ldm.txt>.

2.2 Structured Translation Specifications (STS)

Structured Translation Specifications (STS) enhance Limport’s ability to store and transfer the information necessary for translation tasks. It allows companies or translation project managers to specify important metadata about the translation itself, such as the target audience and intended use of the translation. An STS file includes 21 important translation specifications that should be considered during, or even before initiating a translation task or project. The 21 specifications are provided in Table 1 and are also available at <http://tts.org/specs>. Table 1 was made by the Globalization and Localization Association (GALA) and the Localization/Translation and Authoring Consortium (LTAC).

3. History of Limport

Interestingly enough, Limport itself is an example of a conglomeration of companies and organizations working together. The project comes from three main project streams. In March 2011 many of the organizations that participated in the former LISA standards organization agreed that a container type format was needed in the translation industry. The Globalization and Localization Association (GALA) and the Localization/Translation and Authoring Consortium (LTAC) began work on what was named the Container Project. The first presentation of their work was given a month later in Torino,

A. Linguistic [1–13]	
Source-content information [1–5]	
[1]	textual characteristics <ul style="list-style-type: none"> a) source language b) text type c) audience d) purpose
[2]	specialized language <ul style="list-style-type: none"> a) subject field b) terminology
[3]	volume
[4]	complexity
[5]	origin
Target content information [6–13]	
[6]	target language information <ul style="list-style-type: none"> a) target language b) target terminology
[7]	audience
[8]	purpose
[9]	content correspondence
[10]	register
[11]	file format
[12]	style <ul style="list-style-type: none"> a) style guide b) style relevance
[13]	layout
B. Production tasks [14–15]	
[14]	typical production tasks <ul style="list-style-type: none"> a) preparation b) initial translation c) in-process quality assurance
[15]	additional tasks
C. Environment [16–18]	
[16]	technology
[17]	reference materials
[18]	workplace requirements
D. Relationships [19–21]	
[19]	permissions <ul style="list-style-type: none"> a) copyright b) recognition c) restrictions
[20]	submissions <ul style="list-style-type: none"> a) qualifications b) deliverables c) delivery deadline
[21]	expectations <ul style="list-style-type: none"> a) compensation b) communication

Table 1. Structured Translation Specifications

Italy at the JIAMCATT translation technology conference (JIAMCATT). After the presentation, a representative of the European Commission’s Directorate General for Translation (DGT) indicated that their organization already had started on a similar project, known as the Multilingual Electronic Dossier (MED) project. In MED, they aimed to represent an entire translation project in what they called a translation “dossier.” After a series of discussions, the Container project and the MED project were merged to form the Linport project in July 2011, hosted by LTAC Global a non-profit organization. It was decided that the Linport container would be called a portfolio and would contain all data pertaining to a translation project, be it an authoring, translation, or publication project.

3.1 Work by Interoperability Now!

In 2010, unbeknownst to the Linport project, an initiative company called Interoperability Now!, or IN!, had already started work on yet another similar project. The participants in the Linport project and those involved in the IN! project became aware of each other and then held a series of discussions. IN! agreed to integrate its “container” format into the Linport project.

3.2 IN!’s Translation Interoperability Protocol Package (TIPP)

IN!’s primary contribution to the Linport project is the Translation Interoperability Protocol Package (TIPP). TIPP represents a single translation task to be performed using exactly two languages in a translation workflow. A Linport portfolio by contrast contains the whole translation project, which could potentially involve many languages and tasks. By design then, a Linport portfolio could theoretically be broken down into multiple TIPP task packages which could be accessed, completed, and then reintegrated back into the portfolio for transportation to another translation tool. (History by Melby et al. 2012, *Multilingual Magazine*).

TIPP was designed with XLIFF in mind. Information about the TIPP format, including the parser tool, can be found online at: <http://code.google.com/p/interoperability-now/>.

4. How does Linport work and where does XLIFF fit in?

Linport portfolios are designed to efficiently move translation data between translation environment

tools. XLIFF users will find that a Linport portfolio incorporates XLIFF files at its very heart. A Linport portfolio can contain any number of translatable and/or previously translated documents. Although any bi-text, monolingual, or multilingual document can be contained within a Linport portfolio, it is anticipated that XLIFF will be the most common format used.

An XLIFF document will often be accompanied by several non-XLIFF supporting files, including: terminology files (e.g. .tbx), translation memory files (e.g. .tmx), translation metadata (e.g. an STS file), among countless others. All of these files can be optionally grouped together within a TIPP package or directly into the payload folder of a Linport portfolio. In this way, several XLIFF documents with their support files can easily be packaged together into one Linport portfolio, as shown in the diagram below.

-Linport provides project level interoperability for both XLIFF and non-XLIFF translation projects.

Whether you are working with a pure XLIFF-style project in house, or multiple freelance translators, Linport provides a standardized way to move project data between all stages of the translation workflow.

-Linport can handle multiple extension types.

Linport allows XLIFF to easily associate with non-XLIFF file types. Translation file formats for source texts range from XLIFF to DOCX to PDF. Glossary formats are equally diverse. Linport’s standardized file structure will help machines and humans quickly associate all of the various parts of a translation project. Future Linport-aware XLIFF tools will be able to convert translation data to and

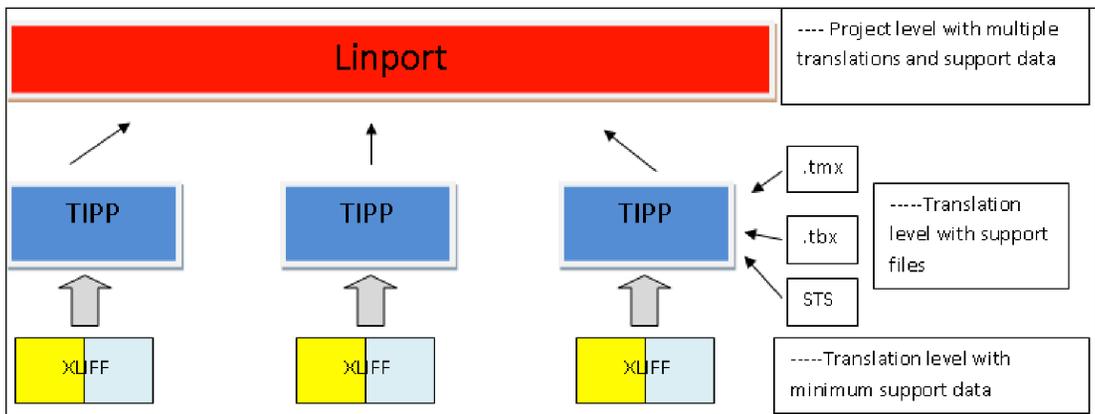


Figure 1: Linport portfolio organisation

5. Why should I use Linport? Isn't XLIFF enough?

Linport does not replace XLIFF. The two formats work together to promote and enhance organization and interoperability in the translation workflow.

Linport adds a new level of abstraction to the XLIFF format.

XLIFF is designed to organize the translation of source text to target text, whereas Linport is designed to organize multiple translations before and after the actual translation. Entire translation projects can be efficiently organized, broken down into translation tasks including XLIFF, and finally reintegrated back into a project portfolio for a reviewer or final publication.

from XLIFF format with ease.

-Linport incorporates ISO 11669.

ISO 11669 defines the Structured Translation Specifications (STS) a set of 21 translation parameters that enhance translation quality by providing additional information to the translators and reviewers of a specific translation project. More information can be found at <http://ttd.org/specs>

-Linport allows for translation quality assessment metrics such as QTLP.

Linport can easily incorporate any quality assessment metric, thus allowing enhanced interaction between translators and reviewers in order to produce higher quality translations. QTLP

is an emerging translation quality assessment metric format which is customizable for different projects or documents. QTLP or any other metric format can be contained in a Linport portfolio.

-Limport is an immediate solution that is easy to implement.

Although no builder tool currently exists to create or validate a Linport portfolio, though online Linport portfolio builders and validation systems are under development, all that is really needed to build one is an operating system with a directory structure. All existing Linport style portfolios to date have been made by hand in five minutes or less. As long as Linport guidelines are understood, any tech savvy or non tech-savvy person can create a viable Linport portfolio.

-Limport is free.

Limport is non-profit and non-proprietary. Any translation company or individual can use Linport royalty free.

6. Current and future development

Limport has much ongoing and future work to be done. The portfolio data model needs to be refined and formalized then submitted to a standards body, eventually becoming an ISO standard. Various software projects are being developed such as an online portfolio builder, a split website that breaks a portfolio into TIPPs, a merger website that integrates TIPP responses back into a portfolio, and a Linport validation system and schema, among others.

7. How to get involved

There are many ways to get involved in the ongoing development of Linport. You can join the Linport community and participate in monthly Linport conference calls by going to <http://www.limport.org> or joining the GALA Linport community group at <http://www.gala-global.org> (Search for Linport in their search box.)

Contact the authors at: tylerasnow@byu.edu or Alan Melby: akmtrg@byu.edu to contribute real project data that can be used to test the Linport portfolio model. All contributed data must be non-confidential. You can also contribute by testing apps developed for Linport, developing your own Linport

applications, or introducing Linport into your company's translation workflow as an early adopter. The Linport community is open and thankful for any support you and your company are willing to provide.

References

International Annual Meeting on Computer-Assisted Translation and Terminology (JIAMCATT). (2012) Retrieved from <http://jiamcatt.org/>.

Limport: The Language Interoperability Portfolio Project. (2012) Retrieved from <http://limport.org>.

Melby, Alan. (2011) "The Seoul of Standards and You." *The ATA Chronicle*. Oct: 12-16.

Melby, Alan. Chandler, Brian. Lommel, Arle. (2012) "Limport addresses translation package compatibility." *Multilingual Magazine*. July/Aug: 45-47.

Melby, A., Lommel, A., Rasmussen, N., & Housley, J. (2012) "The Language Interoperability Portfolio (Limport) Project: Towards an Open, Nonproprietary Format for Packaging Translation Materials." (Unpublished article submitted to *The Journal of Internationalisation and Localisation*).

Structured Specifications and Translation Parameters. (2012) Retrieved from www.ttt.org/specs.

Glossary of Acronyms Used

DGT – Directorate General for Translation
 ›A part of the EC; JIAMCATT Partner
 ›ec.europa.eu/dgs/translation

EC – European Commission

ETSI – European Telecommunications Standards Institute
 ›www.etsi.org

IN! – Interoperability Now!

›A group working to improve the interoperability of tools and technology within the localization industry
 ›code.google.com/p/interoperability-now

ISO – International Standards Organization
 ›JIAMCATT Partner

>www.iso.org

Linport – The Language **I**nter-operability **P**ortfolio Project

>www.linport.org

LISA - Localization Industry Standards Association

>*Ceased to exist March 2011*

OASIS - Organization for the Advancement of Structured Information Standards

>www.oasis-open.org

QTLP or **QTL**aunchPad - Quality Translation Launch Pad.

EC-CORDIS-PF7-LT project 296347

cordis.europa.eu/projects/rcn/103949_en.html

(2012-07-01 to 2014-06-30)

>www.qt21.uk

TAUS – Translation Automation

>www.translationautomation.com

TIPP – Translation Interoperability Protocol Package – an **IN!** project

XLIFF – XML Localisation Interchange File Format

>XLIFF 1.2

>XLIFF 2.0

>XLIFF 1.2: docs.