

Uta Seewald-Heeg, Stefanie Geldbach

Introduction

The widespread use of Computer Aided Translation (CAT) tools has revolutionized the daily work of translators and localizers. In an increasingly automated workflow the use of standardized formats provides a significant contribution to the management and quality assurance in large translation projects. Consequently, the Localization Industry Standards Association (LISA) is actively promoting the development of various standards covering the different stages of the translation workflow from job creation to archival (see Figure 1)¹. In this scenario, translation memories, term bases and machine translation (MT) lexicons are regarded as linguistic assets. Standards provide a way to protect these assets against market and technology changes since they keep users from being locked into a particular CAT tool.

While TMX (Translation Memory eXchange format) which was defined first in 1998 (www.lisa.org/standards/tmx/) has been widely adopted as standard exchange format and is nowadays supported not only by most translation memory systems but also by a growing number of MT vendors such as Systran, linguatrec, Lingenio or braintribe, the picture is less clear on the terminological and lexical side. Although various standards such as OLIF or TBX have been proposed for the exchange of terminological and lexical resources many vendors of MT or other CAT tools have not yet adopted these standards and continue to use proprietary formats only.

This LDV Forum volume contains the proceedings of an international workshop of the GLDV interest group “Machine Translation” entitled “Exchange of Lexical and Terminological Resources in Machine Translation (MT), Com-

puter-Aided Translation (CAT) and Terminology Management Systems (TMS)” which was held at Anhalt University of Applied Sciences (Hochschule Anhalt) in Köthen (Anhalt) (<http://www.inf.hs-anhalt.de>) on June 17, 2005. The workshop brought together MT developers, researchers and translators interested in the integration of lexical and terminological resources. Consequently, the issues addressed at the workshop ranged from practical problems arising during the mass export/import of terminology to fundamental conceptual differences between term bases and MT lexicons which complicate standardization.

Uta Seewald-Heeg reports on the exchange functionalities of terminology management systems (TMS) such as MultiTerm, TermStar, or SDLTermBase (to name just a few) focussing on the question whether the formats currently supported by these systems enable terminology exchange without loss of information. Wolfgang Zenk’s contribution centres on the UniTerm TMS which is developed by Acolada GmbH. Zenk discusses the database design and the various import/export formats currently used by UniTerm and elaborates on the problems of blind terminology interchange.

Stefanie Geldbach gives an overview of lexicon exchange formats currently used by commercial MT systems. Her paper also investigates whether the standardization efforts of the OLIF Consortium have actually resulted in a widespread acceptance of the OLIF2 standard.

Gregor Thurmair discusses two of the formats which are promoted by LISA, namely TBX (Term Base eXchange) and OLIF (Open Lexicon Interchange Format), which originally was

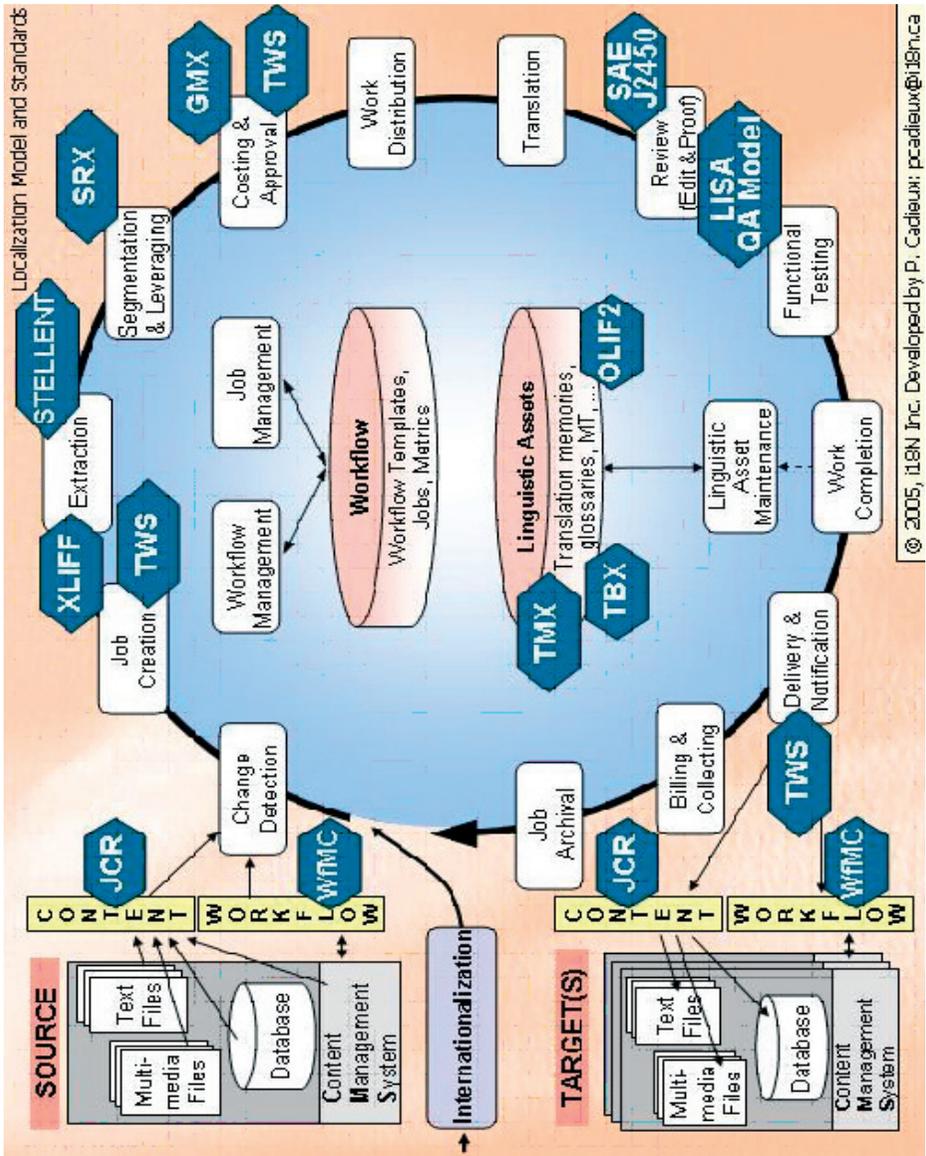


Fig. 1: Localization Model and Standards

Introduction

mainly intended for the exchange of MT lexicons. In this context, he also mentions some of the difficulties which complicate the conversion of proprietary MT lexicons from and into OLIF. As the development of OLIF converters is a non-trivial task it is not surprising that MT vendors continue to use other exchange formats ranging from simple text files to complex proprietary formats.

Monica Gavrilă, Walther von Hahn and Cristina Vertan present MANAGELEX, a generic lexicon management tool for creating, converting and merging lexicons which has been developed at Hamburg University. They outline the architecture of MANAGELEX and describe two of the modules which already have been implemented.

Georg Heeg discusses a software design approach to allow interchange of linguistic data. He focuses on the modelling of the linguistic concepts represented in the data and describes the transfer between exchange formats as a multi-tier interpretation/generation. The discussed concepts are implemented in Smalltalk, a programming environment enabling flexible conversion of data between formats supported by TMS.

Finally, Rachel Herwartz and Birgit Wöllbrink present a non-commercial internet discussion platform open to terminologists, translators and technical writers (*www.terminologieforum.de*) which was launched in January 2005.

The contributions in this issue show that the ultimate goal – blind interchange of terminological and lexicographical data – is still out of reach. Consequently, the development of suitable standards, which opens interesting perspectives for further research, is the objective of several ongoing research projects.

A list of important terminological and lexicological standards and research projects such as MILE can be found in the Appendix of this issue.

Endnote

- ¹ This model of the localization process was created by Pierre Cadieux, president of it8N Inc. (*www.it8n.ca*) and regular speaker at LISA events. The model has been used to describe and compare localization management systems and standards that apply to the localization process.