



Owning MT: Lionbridge and SDL as in-house MT users

TAUS Report

April 2009

| T | A | U | S |
Translation Automation User Society



Owning MT: Lionbridge and SDL as in-house MT users

April 2009

Authored by:

Andrew Joscelyne

Key words:

Machine Translation

LSP

Business Relevance:

This report looks at how two of the world's leading language service providers envision and use their own machine translation systems.



Table of Contents

1. SUMMARY: MT IN LSPS	2
2. LIONBRIDGE: MACHINE TRANSLATION BELOW THE LINE.....	3
<i>Why own MT?.....</i>	<i>3</i>
<i>How MT is marketed</i>	<i>3</i>
<i>The technology stack.....</i>	<i>4</i>
<i>Is SMT a competitor?.....</i>	<i>4</i>
<i>Workflow</i>	<i>4</i>
<i>Source control.....</i>	<i>4</i>
<i>Post-editing</i>	<i>5</i>
<i>Language pairs</i>	<i>5</i>
<i>Pricing.....</i>	<i>5</i>
<i>Roadmap</i>	<i>5</i>
3. SDL: TRANSLATION AUTOMATION AS PRODUCT AND SERVICE	6
<i>The technology stack.....</i>	<i>6</i>
<i>What about SMT?.....</i>	<i>6</i>
<i>Using MT for customers.....</i>	<i>6</i>
<i>Business arguments.....</i>	<i>7</i>
<i>Language pairs</i>	<i>7</i>
<i>Pricing.....</i>	<i>7</i>
<i>Roadmap</i>	<i>7</i>
4. COMMONALITIES AND DIFFERENCES	8
5. CONCLUSIONS	9
<i>Information sources.....</i>	<i>10</i>



1. Summary: MT in LSPs

According to the results of the March 2009 TAUS market survey, only 14% of the LSP (language service provider) respondents state that they will never use MT, 40% already use it today. This provides strong evidence that MT is moving into the mainstream among LSPs. To launch a series of reports on LSP deployment, this report focuses on the two major LSPs which own and develop their own MT systems.

The world's second and third largest language service provider companies **Lionbridge Technologies** (US) (2008 revenues \$419.0 million 3,820 employees, 48 offices) and **SDL International** (UK) (2008 revenues \$174.5 million, 1,700 employees, over 50 offices), are also two of the very few LSPs to have acquired, developed and use their own machine translation systems.

Other LSPs who have some experience in deploying MT systems include CLS Communications (DE, CH), which uses an in-house version of the BrainTribe MT system (now developed and marketed under the Lucy Software umbrella), and WorldLingo (US), which uses MT technology as both an advertising gimmick and a production tool. Compared with them, however, SDL and Lionbridge maintain substantial development teams for their wholly-owned systems, but until very recently, have tended to use the technology as a discreet asset rather than as a major plank in their strategic or marketing development.

However there are clear differences in the two companies' MT agendas. Unlike Lionbridge, SDL also sells its translation technology, alongside its services. Having first acquired Trados (translation memory and terminology tools) and then Idiom (GMS solution), with their existing broad customer bases in the translation industry, SDL announced in October 2008 that it was marketing a new range of translation technology products which includes its "automated" translation solution. This brings SDL out of the MT as a production option space and into the MT *vendor* market where systems are sold either under license or as off the shelf products. It is therefore competing either head-on or indirectly with players such as Language Weaver (US), Systran, and PROMT (RU), as well as with the dozen or so much smaller technology providers working across restricted language pairs, (e.g. AppTek (US) and LINGUATEC(DE)). Like some other vendors, it also offers a free machine translation service online (<http://www.freetranslation.com/>) for the automatic translation of websites and short text.

This report looks at the MT strategies of these two emblematic LSP competitors at a time when the translation industry is experiencing a bout of radical innovation in technology, business and publishing models, along with workforce organization.

2. Lionbridge: Machine Translation below the Line

Machine Translation (MT) at Lionbridge is conceived as one of a range of in-house tools that enhance translation quality and lower translation costs for customers. It is not promoted as an individual tool or special asset in the company's marketing material, and Lionbridge does not position itself as a player in the MT vendor marketplace.

In addition to its unbranded MT system (which we shall refer to by its original name of *Barcelona*), Lionbridge provides customers with the Logoport Translation Memory server – a linguistic asset management tool – and its Linguistic Toolbox -- a set of software tools to automate various types of quality checks on documents.

The company uses MT selectively as a translation automation tool on the basis of specific customer needs. It will also develop partnerships to promote MT adoption and expansion, rather than increase its own ability to compete in the MT market directly.

The Lionbridge MT team comprises of 10 permanent members, responsible for throughputting an average of 20 million words a year using MT for a relatively small line-up of large-volume clients.

Why own MT?

The use of the company's MT engine is limited to the commercial demands of Lionbridge's customers, where MT is part of a total delivery mechanism. It is not touted as an asset or apparently used in stealth mode as a production enhancer for translation jobs where its use is not disclosed to customers. Lionbridge therefore espouses its customers' own image of the potential 'dangers' or 'limits' of MT technology, rather than trying to educate the market to any advantages it might find in MT, or leveraging early-adopter advantage from its own technology.

For Lionbridge, the key to owning MT is to have full control over the system. For an RBMT system, proprietary power is located in the rule sets for the various languages covered. These are engineered and maintained by computational linguists and represent the system's key investment.

When purchasing RBMT systems, users can usually only customize the dictionaries – i.e., add their own domain terms. Any other customization involves expensive support work by the product owner. Lionbridge controls each stage of the MT process (including the grammar rules) so it can optimally adapt the output to a given customer's style, terminology, 'do not translate' terms, and so on.

Rather than using its MT engine as a competitive advantage, Lionbridge believes that partnering with best-of-breed (hybrid) solutions is also a good direction for the company. This is achieved by providing connectivity to its SaaS platform solutions, which are open to any web services compliant MT engine. The company is therefore poised to benefit from any increase in MT deployment from its customer base, but has not itself attempted to invest more deeply in MT as a competitive service.

How MT is marketed

Lionbridge only uses MT in strict agreement with a given customer. It does not market or compete within the MT space as a commercial engine. The company claims that it evangelizes MT as a relevant solution for customers who have high volumes of technical source material, associated with a clear domain, as part of an overall approach to global content. Where it proves commercially viable, Lionbridge will use its engine for its own production requirements. Where commercially intelligent, it will connect through its SaaS-based platform to best-of-breed solutions from other companies. Standalone MT is not a competitive focus for Lionbridge.

The technology stack

When it acquired Bowne Global Solutions in 2000 -- after the meltdown of the Lernout & Hauspie 1990s scheme to consolidate language and translation technologies -- Lionbridge inherited an MT engine called *Barcelona*. The engine was originally developed as the core engine for Globalink, a US translation automation company that flourished in the mid-1990s. *Barcelona* was bought by Bowne when Globalink was wound down.

Barcelona is a rules-based MT system (RBMT) with modules that analyze the sentence structure of the source, look up words in lexicons, and convert the content to the target language by applying mainstream transfer rules.

In 2002, Lionbridge rebuilt *Barcelona* to separate the linguistic modules (dictionary and grammar rules) from the underlying software engine. The purpose of this overhaul was to enable linguists (i.e., staff members with a knowledge of language structure) to enhance the *translation quality* of the engine, without having to be trained in computational linguistics to interact with the machine's software.

As a result, the system is easier and faster to customize for specific instances of customer terminology and text style. In addition, this tactical division of labor has lowered Lionbridge's development cost for using MT as a professional productivity tool.

Is SMT a competitor?

Lionbridge contrasts MT for high quality professional translation with MT for "gisting" and believes that SMT is better suited to the latter activity.

They contend that SMT engines need massive amounts of data for each language pair, and need to be trained for each language pair. While they provide effective near-term delivery for gisting, it has proved much harder to produce quality that can feed an effective post editing stage (despite the initial buzz). Ultimately, Lionbridge has found that SMT-only approaches tend to drive up costs more significantly for the publication quality translation demands the company focuses on.

Although RBMT systems are more expensive to customize than SMT engines (due to the human work needed to craft the system's rules), they can be customized in a more controlled manner, and this makes an RBMT system more manageable in a quality publishing environment.

Workflow

Lionbridge is pushing for the general adoption of Localization 2.0, an approach that will effectively deal with the wide-ranging market changes taking place in content creation and delivery. As part of this overall strategy, Lionbridge integrated *Barcelona* in 2003 into its technology platform, which also combines Translation Memory resources and builds out to an array of support tools including terminology mining, named entity extraction, Edit Distance quality tests, and sentence complexity analyzers.

Source control

More and more of Lionbridge's clients are looking at how controlled authoring (CL) can deliver benefits in a content translation environment, even though CL deployment is by no means easy. To address these kinds of concern, Lionbridge generates translatability reports on source material to identify areas for linguistic/formatting improvement, and notes that even quite *small* changes in writing behavior can have wide-ranging positive effects in the translation process. Being able to continually and instantly effect and monitor changes within dynamic language assets, and combining them with MT, is a fundamental part of extracting total possible value from a combined MT solution.



Post-editing

Given Lionbridge's own focus on MT as part of a content creation process, post-editing is seen as a vital step in delivering viable final quality. Post-editing requires careful and specific training, and it is not something that any linguist can be asked to do. Lionbridge uses Edit Distance and other Linguistic Toolbox tools for "Computer-Aided Revision" to maximize the efficiency of human post-editing. Experience has shown that some post-editing patterns can benefit from automation.

Language pairs

For Lionbridge's own engine, the language pairs are English <=> FR, IT, DE, SP and PT (both Brazilian and Iberian Portuguese), plus cross-language translation between some of these pairs. This represents a fairly basic deck, and leaves out strategic Tier 1 languages such as Japanese, Simplified and Traditional Chinese, which are currently handled using only Translation Memory leverage. Lionbridge looks to partners and MT specialists to fill the gaps left by its own engine.

Pricing

Lionbridge prices MT words as if they were post-edited words. Lionbridge does not commercially compete within the specific MT engine market on prices.

Roadmap

Lionbridge believes that its MT usage will increase. As such, the company plans to introduce new language pairs as appropriate through partnership. As MT usage expands through the growing acceptance of "good enough" and lower price/quality expectations, Lionbridge will use its ability to connect to best-of-breed solutions, and will continue to develop its own MT capabilities for use with high quality translation and "good enough" approaches. There are no plans to take MT to a direct, standalone commercial offering.

3. SDL: translation automation as product and service

With the launch of SDL's Automated Translation Solutions in late 2008, SDL both leveraged its own in-house R&D in machine translation, and re-positioned itself as an MT vendor on the world market, incidentally offering an accessible plan for LSPs to get more involved in MT as a logical extension of TMs.

This new product portfolio integrated a number of existing MT-driven services marketed by SDL as a comprehensive tool box. This includes the instant translation of web pages, and the integration of its SDL Trados translation memory suite and/or its SDL TMS tool into an MT solution built around its SDL Knowledge-based Translation System (SDL KbTS), together with APIs for various business applications.

At the same time, however, SDL will continue with its two track policy by also using KbTS to deliver *services* for its customers. As the new product has only just reached the market, the focus here is on how SDL conceives of its MT *service* offering.

The technology stack

Over last decade years, SDL has developed a comprehensive translation automation system whose roots go back to the earliest days of *localization* technology (as opposed to MT as such) when the Weidner Multi-Lingual Word Processing System was launched in 1981.

This underlying technology was also acquired by Intergraph in 1989, renamed Transcend, and in 1996 sold to Transparent Language for use as a web-based and enterprise translation service. SDL acquired this technology in 2001, radically upgraded and extended the system and three years later launched its Knowledge-based Translation System which was used by one of its clients. Today SDL has more than 20 customers using this MT system in various types of application.

By knowledge based system, SDL does not mean that it is driven by an AI-type, symbolic knowledge processing approach, but simply that the MT process draws on "linguistic knowledge" embodied in software such as parsers, dictionaries, and data stored in TMs. In other words, it is a standard "hybrid" system that either combines data recycling from TMs with RBMT on new sentences, or as a gisting system that raw-translates the input for rapid understanding by humans.

The SDL MT workforce now has over 100 people, including engineers and post-editing staff.


What about SMT?

SDL currently feels that there is very little documented evidence on the clear superiority of SMT over RBMT, and notes that improvement in SMT will come more through the addition of language knowledge than the brute force of data. It agrees that Google Translate can offer useful gisting support, but doubts that any LSPs or end users would wish to invest the time in post-editing Google output to publishing quality.

Using MT for customers

SDL never uses MT without telling its clients, and finds that the very concept of MT still scares off certain people. The only way to get benefits from MT is to invest in extensive customization to reach the appropriate quality.

Customers that use SDL's MT services include: Chrysler (owner manuals shipped with cars), Microsoft (virus alerts, MSDN alerts), RS (component catalog); Hewlett Packard (support content and product



master), Renault (service manuals for cars), Dell (support content), SolidWorks (support content), CNH (support content) and Best Western (web site/services).

A potential MT growth area will be Customer Support Services, where there is considerable untranslated content in knowledge bases.

Business arguments

The main reasons for deploying MT are to lower the cost of translation, provide flexible solutions to customers needs, and above all to provide translation where the alternative would have been no translation.

SDL claims 35% reduction in time to market for certain MT applications (e.g. translating large product catalogs), and 27% savings in costs compared to all-manual translation methods.

Language pairs

SDL offers 11 core languages, and 15 language pairs, set to grow to 13/17 by the beginning of 2009.

- English ↔ French/Canadian French
- English ↔ Italian
- English ↔ Spanish/LA Spanish/ International Spanish
- English ↔ German
- English ↔ Portuguese/Brazilian Portuguese
- English ↔ Dutch
- English → S. Chinese
- English → T. Chinese
- English → Norwegian (Q2 2009)
- English → Swedish (Q1 2009)
- English → Finnish (Q1 2009)
- English → Danish (Q2 2009)

Rather than investing deeply in new language development at every new step, SDL finds that its clients can derive substantial benefits from the existing language deck. By using MT for the available languages, a company can then use the savings gained to open up *new* markets with human translation of newer languages.

Pricing

SDL prices MT *at discounted word rates* when KbTS is used.

Roadmap

In its September 2008 study of the potential market for MT, carried out with MT organizations, SDL polled global companies such as Cisco, Dell, HP, Motorola, Oracle, Panasonic, Philips, Siemens, Sun and Vodafone about their intentions to use MT. From the companies questioned, it found a “substantial industry shift in perception” about the usability of the technology. This study was then used to prepare mindsets for the launch of the company’s new automated translation package, and its Common Enterprise Application Framework (CEAF), the new “open” architecture underlying its Global Information Management Platform.

4. Commonalities and Differences

- ✓ Both SDL and Lionbridge have RBMT systems.

Although this is a purely historical accident, SDL and Lionbridge both say that the *controllability* and *output consistency* of this type of MT probably makes it easier to manage in a low-investment environment – users only need to learn how to populate the dictionaries, rather than improve the rules.

SMT, though initially cheaper to acquire due to the existence of open source version, demands extensive domain data management and training time. And there is little practical evidence of large-scale savings from using SMT in a publication quality production environment.

- ✓ Both only use their MT for a small set of clients, even if their volumes can be large
- ✓ Neither uses it systematically as a resource for their translator population as a web service

Lionbridge is prepared to negotiate “partnerships” with other MT developers to leverage its web service model, although cannot say if it has done so. But there appear to be no immediate plans to open up MT output for more general production throughput.

SDL is committed to helping customers, but apparently not to using MT as an aid to translation tasks in general in partnership with its translator base.

- ✓ Development Teams

Lionbridge: 10 in R&D / language engineering,

SDL: over a hundred in R&D / language engineering and other MT-related production tasks

- ✓ Profitability:

SDL claims its MT service has been profitable for 4 years,

Lionbridge tends to emphasize the manpower and investment cost of MT

- ✓ Pricing

Both apply the mainstream word-based approach to MT pricing.



5. Conclusions

There is probably only a slight chance today of another LSP investing deeply in MT technology as a development option, more because of the very high cost of skilled human resources and the development/training process than the actual technology itself in a market that is under increasing price pressure.

However, TAUS has collected evidence of the strong likelihood that LSPs will engage much more closely with MT, either by licensing some form of technology, working in partnership with an MT vendor (as in the PROMT's LSP Partnership Program) or working more closely as a supplier of post-editing services with large-account customers who may decide to use the technology in-house.

It is also more likely for language-centric people such as translators to opt for RBMT rather than SMT systems. Rule-based systems embody declarative "linguistic knowledge", whereas data-driven language-independent SMT systems take a more "engineering" approach to the process.

However, a number of LSPs are beginning to experiment with the Moses open-source MT engine, which means that their principle costs will be in people rather than software licences. It may well be possible for even smaller LSPs to train and tune very finely-targeted SMT engines, once they have access to large collections of language data (e.g. via TDA), and build up capabilities in SMT deployment. Examples of LSPs going down this route are Translated.net and Pangeanic.

Lionbridge and SDL are therefore likely to remain in a special category as LSPs due to their capacity to develop MT technology as well as service clients with production solutions. At the same time, they are pursuing quite different strategies:

SDL continues to act as technology supplier, both at the level of infrastructure with its newly Common Enterprise Application Framework, and in terms of selling tools (MT system, authoring system, GMS, translation memory), much to dismay of many LSPs who feel locked into these products while SDL is competing against them for translation business. It will be interesting to see how many LSPs interested in deploying MT as a partial solution to their service requirements benefit from SDL's MT offering, rather than using pure plays such as PROMT or Systran, for example. LSPs who act as subcontractors to SDL already have access to the service as an option, and a small subset of these also license MT software from MT system vendors, playing off different quality levels and language pairs as they search for the optimum cost-time-quality equation.

Lionbridge, on the other hand, appears to be playing the workflow card, by opening up its translation facilities via Logoport in such a way that third party MT could be deployed via an interface in tandem with the company's TM cloud.



Information sources

Many thanks to TAUS members and everyone else who provided input to this report.

COLOPHON

About TAUS

TAUS is a networking community for users and practitioners of authoring, translation and localization services and technologies. By sharing user cases, good practices and intelligence in cross-industry meetings and online forums the TAUS group aims at advancing the adoption of translation automation technologies.

TAUS reports cover:

- **Technology review.** Introductions to key areas of translation automation.
- **Best practices.** Overview of best practices in applying technologies.
- **User cases.** Analyses of processes in member and non-member companies.
- **Meeting reports.** Reports on TAUS Executive Forums and Summits.

For more information on TAUS, see: www.translationautomation.com

Replies, questions and comments to Jaap van der Meer (Director) at jaap@translationautomation.com

Address: Oosteinde 11, 1483 AB De Rijp, The Netherlands.

© 2009, by TAUS B.V. All rights reserved.

TAUS reports are published by TAUS B.V. exclusively for its subscribers. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means.