[*Translating and the Computer 22*: Proceedings of the Twenty-second international conference…16-17 November 2000. (London: Aslib, 2000)]

The EU LE4 TransRouter Project

Reinhard Schäler
Localisation Research Centre (LRC)
Department of Computer Science and Information Systems
University of Limerick
Limerick - Ireland

Reinhard.Schaler@ul.ie

1 Abstract

This article describes the European TransRouter project which produced a prototype application designed to help users identify the best possible way to translate a particular project. As a typical knowledge society application, it is designed to facilitate easy and almost instant access to knowledge by evaluating highly complex information based on user-definable criteria. The interest shown by potential investors augurs well for the commercial exploitation and further development of this prototype.

2 Introduction

In late 1996, Microsoft's vendor manager in Dublin invited a small group of translation technology experts from the localisation industry to discuss the implications of the success of translation memory (TM) systems for the localisation process. Translation memory systems, piloted since 1995 by no more than a few highly innovative localisation companies and successfully used first by Softrans-Berlitz and Oracle in a major localisation project, were now being introduced on a large scale. This meant that the knowledge - some call it, maybe more appropriately, *gut feeling* - accumulated by the few translation memory experts around at the tune, had to be made available easily and efficiently to the non-technical translation mangers.

Therefore, the main questions discussed around the table at this meeting were:

- What questions are asked and which criteria are used by the *experts* when they decide whether or not a translation memory system should be used for a particular translation project?

- Can an automated system based on these questions and criteria be developed to make this expert knowledge available to the translation managers thus helping them to decide whether they should use human translators or a translation technology application?

3 Approach

Following the development of an initial prototype at the LRC, called *ETAT*, a European consortium was set-up. This consortium successfully proposed the TransRouter project under the 4th call for proposals of the Language Engineering sector of the EU's 4th Framework Programme. In addition to *ETAT*, findings from the *EAGLES Case Study* of work in the Commission Translation Services (SdT) also influenced the development of the proposal.

The consortium was lead by Berlitz, the world's largest provider of translation services. The Localisation Research Centre (LRC) at the University of Limerick was the technical manager. Other partners in the consortium were: ISSCO, the Centre for Language Technology (CST), the universities of Edinburgh and Regensburg, and L&H Language Technologies.

A user group, the TransRouter User Advisory Group (TUAG), was also established. It consisted of a wide range of software publishers, translation technology developers, localisation vendors and translation service providers providing feedback and support to the consortium.[2]

3.1 Project aim

At the tune of the development of the project proposal and the TransRouter Technical Annex in 1997, the overall aim of the project was to develop a decision support tool for project managers who had to decide whether a specific project should be translated by a human translator, by translation memory, by MT or a combination of all three.

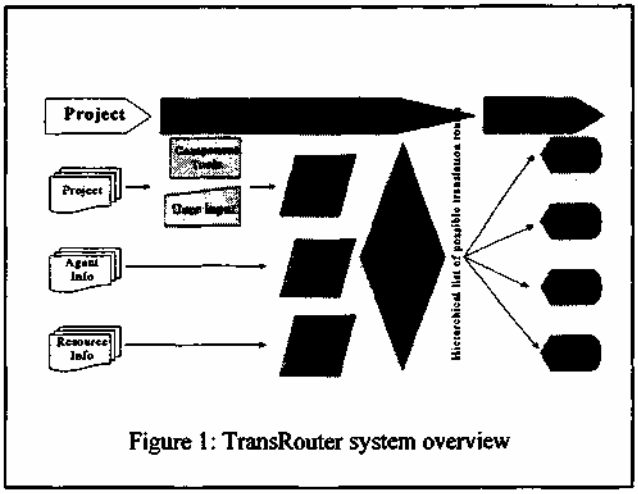

Figure 1: TransRouter system overview

TransRouter was to present a number of possible routes to translation managers, along with the consequences in terms of time, cost and quality of taking each route .

[1] The EU TransRouter project (LE.4-8345) was funded by the European Commission over a period of 28 months (January 1998 to April 2000).

[2] Among the members of the TransRouter user advisory group were: Alpnet Corporation, Corel Corporation Ltd., CTS Teoranta, EGT, Filenet Company Ireland Ltd., Gateway 2000, Lingtech A/S, LioNBRIDGE Technologies, Logos GmbH, Novell Ireland Software Ltd., Oracle WPTG, Oversaetterhuset A/S, Praetorius, STAR Deutschland GmbH, Symantec Ireland Ltd., TRADOS, Translation Experts Ltd., Translation Service of the European Commission, VistaTEC.

The suggestions proposed by TransRouter's decision support kernel were to be based on information stored in three types of profiles:

- *Project profiles* containing detailed data on specific projects, such as project size, source and target languages, file formats, time and budget available etc. This information was to be acquired by user input and the automatic analysis of the project using the TransRouter *component tools.*

- *Agent profiles* containing detailed information on the features of translation technology applications, such as file formats and language combinations a system is able to handle, its integration options with other systems, the level of adaptability to specific user requirements etc. This information was to be updated as necessary by translation technology experts or translation technology vendors.

- *Agent resources profiles* containing information on the resources available to agents such as terminology databases, translation memories, MT terminology databases etc. This information was to be maintained by the in-house translation technology expert at the user site.

While it was intended that users would eventually be able to 'connect' their own *component tools* to TransRouter, the consortium also decided to supply some essential component tools with the prototype:

- Word counter and sentence length estimator
- Translation memory/previously translated texts coverage calculator
- Version comparison
- Repetitiveness detector
- Unknown term detector
- Sentence simplicity checker

3.2 Scope

While TransRouter took the initial *ETAT* prototype as one of its central starting points, its scope, target group and range of subjects expanded substantially to include a wider range of translation technology applications in comparison to the translation memory centred approach of the *ETAT* prototype.

Although localisation scenarios are still being used as a test bed for TransRouter, the benefits of the tool will outreach this industry alone. The TransRouter consortium carried out detailed studies of translation practices in different environments to ensure that the requirements of the broadest possible range of users will be catered for.

3.3 A moving target: user requirements

It was evident from the outset that user requirements would change between the time that the project was proposed and the time of its completion, a period covering effectively three years.

Therefore, the consortium had to ensure that important developments with an impact on user requirements were carefully monitored. Users had to be consulted and kept informed of project progress. Their feedback had to be taken into consideration for design and development decisions during the different stages of the prototype development.

Over the lifetime of the project, two important developments took place which had a significant impact on the requirements of users and, therefore, on the development of the TransRouter prototypes:

- Developments in the use of translation technology applications
- New strategic approaches to the business of translation

We will consider each of these in turn.

3.3.1 Developments in the use of translation technology applications

At the tune of the project proposal development in 1997, translation technology applications such as translation memory systems were cutting edge technology with only few expert users.

This situation had dramatically changed — at least in some important user domains — at the tune that the design of the first TransRouter prototype was presented to the TransRouter User Advisory Group (TUAG).

The users consulted made it clear that they no longer just needed advice on whether to use a translation memory application or not - this was no longer an issue. However, what they now needed was a knowledge-based, easy to use and accessible *advisor.*

This tool should analyse large amounts of data and produce a report on a wider range of options and *routes* than those described in the original Technical Annex. This wider range of options should also cover items such as pricing information in relation to particular vendors, specific languages and the use of translation technology and human translation.

3.3.2 New strategic approaches to the business of translation

Since 1999, a new approach to the business of translation has rapidly been gaming momentum. *Translation portals* are completing the service offering of translation service providers by moving translation projects in their entirety to the Internet.

e-Translation offered by these portals allow each step in the translation process from request for quote (RFQ) to delivery of the translated source material and billing to be processed entirely through web-based, electronic interfaces and media.

Some translation portals connect their clients with several hundred translation service providers and many thousand freelance translators. In addition, they offer the use translation technology applications and their corresponding linguistic resources.

Clients looking for translation services on a translation portal can now choose among hundreds of service providers and thousands of freelance translators - compiling their charges based on a complex matrix, including at least some of the following factors:

- translation technology applications used
- source and target languages
- type and domain of source text
- quality of the product in the target language
- quality and relevance of existing reference material

Individual service providers such as Berlitz GlobalNet and UNISCAPE have already developed tools to support the users of their translation portals. *BerlitzIT,* for example, allows users to compute a quote for small-scale translation projects, online and with immediate results.

It was evident that TransRouter could be the perfect fit for online translation service providers trying to help clients to identify the best possible solution for their translation requirements.

4 Prototypes

The overall development of TransRouter took place in two distinct phases. During each of these phases one prototype was developed. This phased approach allowed the identification and correction of issues in relation to:

- the integration of the TransRouter components
- the functionality of the decision support kernel
- the design and usability of the user interface

The first TransRouter prototype followed in its approach the original assumption by the consortium that translation managers would mainly be interested in suggestions (from a decision support tool) on whether or not to use a particular translation technology application for a specific project.

The *second* prototype developed this original approach further but followed the suggestions and feedback from the TransRouter user advisory group and other potential TransRouter users.

They had suggested that at least in the localisation industry the use of translation technology applications, and specifically translation memory applications, had become almost a standard, i.e. the decision whether or not to use a translation memory application was not an issue anymore.

A detailed analysis of the implications on time, cost and quality when choosing different translation routes *with different translation service providers* (in-house and third party) had become a definite requirement. In addition, the result of this complex analysis - based

on the information stored in its profiles and based on the results generated by the decision support kernel and component tools - had to be made available in an easy-to-read and user-configurable report.

TransRouter developed into a typical knowledge society application – designed to facilitate easy and almost instant access to knowledge by evaluating highly complex information using user definable criteria.

5 Architecture

The architecture and design of the TransRouter prototype followed an early decision made by the consortium to use industry-standard designs, structures, platforms and applications running under MS Windows, in particular:

- MS Visual C++
- MS ACCESS
- TRADOS Translator's Workbench API
- MS WORD API

TransRouter is written in C++, using Microsoft Visual Studio, Version 6, the standard tool for developing Windows applications. The application design is object orientated, based on Microsoft Foundation Classes (MFC).

One of the most important design principles of TransRouter is the provision of a common interface and platform for a variety of tools and utilities all forming part of the overall TransRouter prototype.[3]

The following sections will describe both the TransRouter component tools and the decision kernel components.

5.1 Component Tools

The TransRouter platform offers a number of component tools for the automatic profiling of text.

5.1.1 Word Counter

Using industry-standard algorithms, this tool reports the number of words in the source text. To calculate the cost and time for a "straight-forward" translation, the number of words is multiplied by the rate for a particular language pair.

5.1.2 Translation Memory comparison tool

This tool compares the source text with an existing translation memory and provides the decision kernel with a detailed report on matches found. The decision kernel in turn is then in a position to adjust the rates applied according to the settings specified by the user. For example, companies might want to specify different ratios (time and money) for

[3] All TransRouter component tools can either be launched individually from within TransRouter or automatically as part of the overall analysis of a project. Decision kernel components cannot be launched individually.

un-matched, fuzzy-matched and fully-matched text segments, applying discounts to the rate for fuzzy-matched and fully-matched segments.

5.1.3 Previous version comparison tool

Even if not stored in a translation memory, previous translations can be used to speed-up the translation and increase the quality of similar new translation projects. This tool assesses the similarity between a new source text and a previously translated source text and reports this to the decision kernel. The decision kernel in turn can then report to the user on the impact of the creation and use of an initial translation memory (through the alignment and subsequent import of the old source and old target text into a translation memory) on overall, user-adjustable project factors such as cost and time.

5.1.4 Internal repetition checker

Certain types of text have shown an internal repetitiveness rate of 20% and higher. Even without an existing translation memory, translation memory applications can make use of these internal repetitions as translation of the source progresses. The internal repetition checker reports the internal repetition rate to the decision support kernel which in turn uses it to allow the user to adjust the rates for a translation project according to the amount and the type of internal matches.

5.1.5 Unknown term detector

Coding unknown terminology is one of the major cost factors for the automatic translation of text. The unknown term detector (UTD) compares the terms in a source text against a specified terminology database and reports the number of terms not found in the database to the decision support kernel. This figure is used by the decision support kernel to adjust the rates (tune and money) for automatic translation based on data supplied by the user.

5.1.6 Sentence simplicity checker

The sentence simplicity checker attempts to calculate the grade of complexity of a source text and then supplies this grade to the decision support kernel. Based on this grade of complexity, the decision support kernel can either eliminate highly complex source text as a suitable candidate for automatic translation, or alert the user to the fact that automatic translation will produce a low quality translation because of the complexity of the source.

5.2 Decision kernel components

The TransRouter kernel itself integrates a number of utilities and kernel component tools.

5.2.1 Router

Based on (a) the TransRouter internal, pre-defined set of possible translation routes and (b) external, project and user dependent data generated by the component tools and supplied by user input, the router matches the resources available in a given infrastructure with the requirements of a specific project and selects the possible translation routes.

5.2.2 Profiler

All TransRouter profiles are stored, maintained and accessed through a custom-designed relational database application. The architecture of this application ensures the most cost-effective and secure environment for the TransRouter profiles.

5.2.3 Cost calculator

The calculation of the cost for the translation of a project is based on a complex matrix. It allows the user to adjust the basic formula of {unit price x number of units} with factors for each match type, the source and target languages and the translation routes.[4]

5.2.4 Time calculator

The estimated time required for the completion of a project is calculated using the basic formula {unit time x number of units}. This formula is modified by the time calculator to reflect different translation routes and match types identified. The factors used for this modification are user-definable.

5.2.5 Quality estimator

The quality of the translation of the source text depends on a variety of factors, among them (a) the proficiency and experience of the translators and editors, (b) the quality and complexity of the source text, (c) the procedures used by and the infrastructure available to the translation service provider, and (d) the time and financial resource allocated. The current implementation of the quality estimator in TransRouter provides an estimate of the translation quality to be expected for each pre-defined translation route. This limitation, however, is largely overcome by features implemented in the ranker (see below).

5.2.6 Report generator

The report generator composes the TransRouter analysis report taking as input (a) the description of the possible routes for a particular project from the router and (b) the corresponding data on cost, time and quality from the cost, time and quality modules. The report shows the relevant data sorted in a hierarchical order by language, route and vendor. While the report, for obvious reasons, will always be generated by language, other factors impacting on the way the report is presented can be set by the user in the ranker.

5.2.7 Ranker

The ranker allows users to examine available translation routes under different scenarios, asking what-if questions and

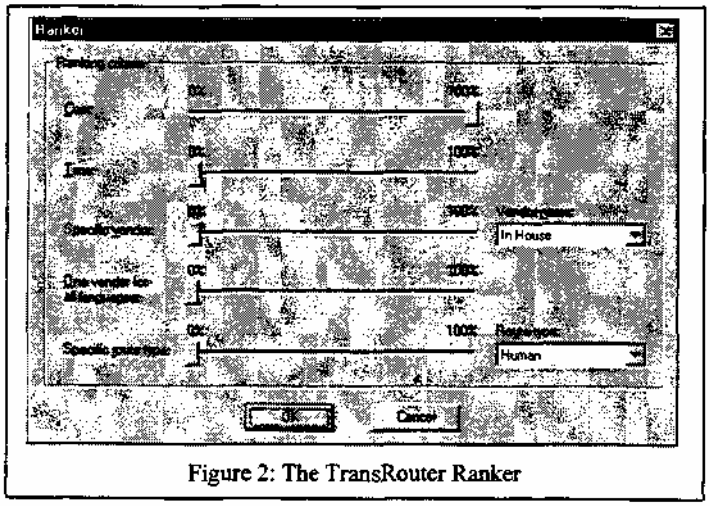

Figure 2: The TransRouter Ranker

[4] The current implementation of the cost calculator fixes the unit price in one currency (ECUs) and the type of unit as words.

modeling the data display based on their own preferences and quality expectations. Users can specify their relative preference for the criteria used to display and sort the information in the TransRouter report, among them cost, time, vendor (incl. in-house), single language/multi-language solution, and route type (human/TM/MT).

5.2.8 Report printing and saving utilities

This utility allows the user to save, print or edit the report generated by TransRouter using industry-standard word-processing applications.

6 Routes

TransRouter analyses all files, for all the languages specified, and determines the cost and time, and, to a lesser degree, the level of quality that can be expected out of a particular route using a particular vendor and for a particular language pair.

Routes in TransRouter are a limited number of possible, predefined ways of translating a particular project. Eight routes are currently defined in TransRouter.

6.1.1 Human translation

A translator translates the files without using any electronic translation tool (TM tool or terminology tool, or MT tool). Word-processors or similar text processing tools can, of course, be used.

6.1.2 Translation memory tool, with TM resource

A translator uses a translation memory application with an available translation memory resource, i.e. a specific translation memory.

6.1.3 Translation memory tool, without using an existing TM resource

A translator uses a translation memory application without a translation memory resource, i.e. a specific translation memory.

This method of translation is considered because a TM tool, even without a TM resource, can improve the quality (e.g. consistency) of a translation, especially in the case of projects with a high degree of internal repetition.

In addition, using a TM application means that at the end of the project a TM resource will have been created which may be useful if a similar project is to be translated in the future.

6.1.4 Translation memory tool, with initial TM

No translation memory is available.[5] However, possible savings from the alignment of a previously translated source with its translation and the subsequent import of the resulting aligned file into a new translation memory are considered.

[5] An initial translation memory (TM) is a memory not developed during the use of a translation memory application, but based on the alignment of previously translated source and target material and its subsequent import into an empty translation memory.

This is particularly useful if a very similar project was translated before, but did not use a TM application to build up translation memories. By aligning previous source and target files 'initial translation memories' can be created which will immediately give users the full benefit of translation memories.

6.1.5 Plain Machine Translation (MT)

This is an MT only option, i.e. without TM. The price/time will be calculated based on the plain MT matrix.

6.1.6 MT followed by post editing

This involves a proofread by a human post-editor who will perform a quick edit of the MT output. The objective is to increase the readability of the target text. The quick edit does not involve terminology research. It involves:

- correction of obvious syntactic mistakes;
- translation of untranslated terms (only general terminology, i.e. the proofread will not involve the translation or correction of highly domain specific and very specialised terminology);
- untranslate words that should not have been translated (only in obvious cases with the objective to increase the readability of the target text).

6.1.7 MT with pre- and post-editing

This route involves pre-editing of the source text and post-editing of the target text. The objective is to increase the quality of the MT output and to produce a text that can be published.

6.1.8 Combination of MT and TM

Here, TM is used for 100% matches and MT is used to translate all the fuzzy/no matches.

7 Workflow

A user would like to translate a project. His requirement is to identify the most suitable translation route for this project.

7.1 Project profile

A project is made up of a number of files, which may be of a number of different types (doc, rtf, html, etc.). These files have to be translated into a number of languages.

The user specifies the directory where the files are located. TransRouter determines the type of the files in question (based on the file extension).

Additional information about the project is collected during the automatic analysis. This analysis uses the TransRouter component tools to determine:

- the absolute amount of words to be translated
- the relative amount of words to be translated, taking into account matching phrases in either existing translation memories, previously translated text or internal repetition

- the number of unknown terms
- the degree of complexity of the source text

The result of this analysis is stored in the project profile.

7.2 Agent and agent resources profile

The information on available agents and agent resources is stored in the corresponding profiles and maintained by TransRouter expert users, e.g. translation technology experts within a company or consultants specifically employed for this purpose. A stand-alone database application and interfaces have been developed for this purpose.

TransRouter users would generally not modify this information.

They might wish, however, to select or de-select specific agent resources according to their requirements or the perceived relevance of particular resources for a given project. This is possible from within the main TransRouter interface.

7.3 Translation service information

A translation service can be provided by anybody – from in-house translators, to large multinational companies, to freelancers working from home.

Each translation service provider has translation resources (agent and agent resources in TransRouter terminology) for the use of which he charges specific rates and schedules specific timelines.

This information is stored in the TransRouter database. It should not have to be maintained by the TransRouter user but by a sales or vendor management team. In a networked environment, it should also be possible for service providers themselves to update relevant information online.

All they would have to do is to enter information in forms about the languages they can support, the cost and time for translating text into each language, the types of agents they can work with, and the impact those agents have on the cost and time factors specified by them.

7.4 Stand-alone and full analysis

TransRouter component tools can be used to generate reports on characteristic features of individual projects *(project profiling).*

Component tools can be launched stand-alone from the TransRouter platform.

In the case of a full analysis, all information available to TransRouter in its knowledge base (stored and organized in the profiles and defined in its kernel) is evaluated.

---

[6] In a networked environment, especially involving the Internet agent developers could be invited to maintain the profiles of the tools developed by them online.

The result of the analysis is a full report on all possible translation scenarios.

7.5 TransRouter report and ranker

The report generated by TransRouter is presented by target language and ranked using default settings. These default settings can be modified easily by the user with TransRouter's ranking tool.

As different users have different requirements for different projects, TransRouter allows users to easily modify the presentation of the report generated by it, according to their preferences in relation to factors such as time, quality, cost, vendor and routing.

8 Conclusions

When the TransRouter project finished in April 2000, the project had achieved some major innovative results. The project:

- produced a template for and prototype implementation of the electronic profiling of essential translation technology applications and of different types of translation projects;

- created a model for the simple but effective approach to routing based on user requirements (cost, time, quality);

- implemented a prototype of this model making use of electronically available information on projects, translation technology applications and linguistic resources in an open architecture.

TransRouter puts up-to-date information on translation technology applications and their respective linguistic resources at your fingertips.

It enables users to automatically analyse the implications on time, cost and quality of each of a number of possible translation routes involving different vendors - in-house and third party - using different approaches ('routes') to the translation of a project.

It provides an added level of confidence in the decision to choose a particular translation *route* and will hopefully lead to a wider use of translation technology applications.

For users of *Translation Portals* TransRouter has the potential to help them identify the most suitable vendor for specific projects.

TransRouter has already attracted considerable interest from commercial consultancy companies, translation tools developers and translation service providers.

These contacts will be further developed. It is hoped that one of the parties who have shown an interest in the TransRouter project will be in a position to develop this prototype into a commercial application.

9 References

Eagles home page: http://www.ilc.pi.cnr.it/EAGLES/home.html (consulted 03.10.00).

Reinhard Schäler, *TransRouter - a decision support tool for translation managers,* The ELRA Newsletter, Vol.4, n.4, October-December 99.

TransRouter Consortium (contact author: Margaret King), *TransRouter: a decision support tool for translation managers,* Proceedings of the MT Summit, Singapore 13-17 September 1999.

TransRouter home page: http://kc.csis.ul.ie/TransRouter/ (consulted 03.10.00).

TransRouter page on the European Commission's web site for Human Language Technologies: http://www.hltcentral > projects > TransRouter (consulted 03.10.00).

Nancy Underwood and Bart Jongejan, *Profiling Translation Projects,* Proceedings of the International Conference on Theoretical and Methodological Issues in MT (TMI'99), Chester, England, August 1999.