

## TransRouter: a decision support tool for translation managers.

The TransRouter Consortium<sup>1</sup>:  
Contact author: Margaret King  
TIM/ETI (ex-ISSCO)  
University of Geneva  
54 rte des Acacias  
CH-1127 Carouge, Geneva  
Switzerland

email: Margaret.King@issco.unige.ch

### Abstract.

Translation managers often have to decide on the most appropriate way to deal with a translation project. Possible options may include human translation, translation using a specific terminology resource, translation in interaction with a translation memory system, and machine translation. The decision making involved is complex, and it is not always easy to decide by inspection whether a specific text lends itself to certain kinds of treatment. TransRouter supports the decision making by offering a suite of computer based tools which can be used to analyse the text to be translated. Some tools, such as the word counter, the repetition detector, the sentence length estimator and the sentence simplicity checker look at characteristics of the text itself. A version comparison tool compares the new text to previously translated texts. Other tools, such as the unknown terms detector and the translation memory coverage estimator, estimate overlap between the text and a set of known resources. The information gained, combined with further information provided by the user, is input to a decision kernel which calculates possible routes towards achieving the translation together with their cost and consequences on translation quality. The user may influence the kernel by, for example, specifying particular resources or refining routes under investigation. The final decision on how to

treat the project rests with the translation manager.

### 1. Introduction.

Introducing modern technology into traditional work has far reaching consequences on the way that work is done. We can already see how fundamental are the changes that can be brought about by thinking about how the introduction of text processing has revolutionised the work of document production, and about how the widespread use of PCs has further changed the nature of secretarial work. The secretary who once spent most of his working life laboriously producing clean typed documents from hand-written scribble or from a dictated tape may now only rarely if ever be called upon to carry out the same task, even using a text processor. The chances are that the original author will work directly onto his own computer, producing his own clean and printed copy.

When we consider the introduction of modern technology into the more complex task of producing a translation, the scale of change increases and becomes more varied. The way the individual translator works will change dramatically if he starts to work with a translation memory system, for example and within the overall work flow of preparing and producing a translation new tasks will appear in the form of a need to create and maintain the resources required for translation technology systems to be useful. It is not even clear who should carry out some of these new tasks; new types of expertise may cause new professions to emerge.

Such changes affect all translators, including both freelancers and those employed in translation services or by translation vendors. Other factors operate more

---

<sup>1</sup> TransRouter is project LE4-8345 in the Telematics Applications of Common Interest programme of the Fourth Framework Programme, supported by the Commission of the European Communities and by the Swiss Federal Office for Education and Science. The members of the Consortium are Berlitz, Dublin (Charles Hughes, John Micks), CST, Copenhagen (Bart Jongejan, Nancy Underwood), University of Edinburgh (Joe Calder), TIM/ETI, University of Geneva (Margaret King, Sandra Manzi), Lernout and Hauspie, Munich (Johannes Ritzke), LRC, Dublin (Keith Brazil, Conor McDonagh, Reinhard Schäler), University of Regensburg (Rainer Hammwöhner, Jürgen Reischer).

particularly within larger organisations. Translation services tend to experience ever growing pressure as document production increases and as the need for translation grows. The pressure may lead to solutions other than conventional translation of a whole text being used. For example, a summary in the desired target language(s) of a long text may replace a full translation, or even a quick verbal summary over the telephone. Depending on the use for which the translation is intended, machine translation may be proposed, with or without post-editing. Pressure here leads to innovative solutions. Globalisation of the market place has produced a different kind of pressure on providers of localisation services, where time to market, cost and quality issues dominate a cut-throat world and force the use of translation technology in the interests of maximising efficiency.

This paper concerns just one aspect of the changing face of translation work, that of deciding on the most appropriate way to achieve a translation of the quality required in the time allowed at an acceptable cost. Such decisions are made on a daily basis by translation managers and planning officers, but they are mostly made on the basis of experience and intuition rather than on the basis of hard information. The aim of TransRouter is to support the decision making process by supplying a suite of automated tools integrated into a decision support tool which will facilitate the task of the translation manager by providing him with information about the various routes open to him. where the possible routes include human translation, either by an individual translator or by a group of translators, human translation supported by the use of particular resources (specified glossaries or term banks, for example), human translation supported by the use of a translation memory system, combined perhaps with the use of a specified terminological resource, and use of a machine translation system - or any combination of the above: it would be perfectly plausible, for example, for some project to involve a highly sensitive covering letter which had to be dealt with by human translation, some background material which was a new version of a document already stored in a translation memory and some extensive information purposes only material which could be dealt with satisfactorily by machine translation. TransRouter will also estimate the cost of following any given route in terms both of monetary cost and of projected translation quality.

It is important to emphasise that our aims are relatively modest: we are not aiming to produce a full-blown expert system capable of making the appropriate decision totally automatically. Rather we propose a tool which will, for a given translation project, determine what possible routes exist and what the price in terms of quality and cost would be of following each possible route. The results are then presented to the translation manager who is responsible for the final decision. The translation manager may also intervene during the decision making process in order, for example, to block certain routes or to specify the use of some particular resource. Thus, although we

are attempting to isolate key factors which influence a translation manager's decision and to provide tools which will examine those factors for a particular translation project, we are *not* trying to produce a cognitive model of the translation manager's decision making process.

In the rest of this paper, we shall first examine what the key factors might be, looking also at the tools which analyse those factors within the TransRouter support tool. It should be noted that some of the tools may also be useful as stand alone tools, working independently of the TransRouter environment. Then we shall look at how the tools are put together with various profiles to form the TransRouter prototype. A final section will briefly outline some outstanding areas of difficulty.

## **2. Key factors in determining an appropriate translation route.**

Although machine translation systems have been around for some time, translation memory systems, and with them the use of local terminology management systems and bi- and multi-lingual concordancing systems, have only become widely available within the last five or six years. There is thus only limited practical experience to build on in deciding what the key factors in determining an appropriate route might be, and although some factors may be so obvious as to be banal, others are less self-evident. The list given below is based on experience within the localisation industry, on a study carried out in a previous project of the Translation Services of the European Commission and on the direct experience of the members of the consortium in consultancy roles. It makes no claim either to be error free or exhaustive. We hope to be able to refine the list as a result of feedback from third parties.

The list is divided into a number of sub-lists.

### **2.1 Factors to do with the translation project itself.**

The most obvious of these is clearly the languages concerned. This factor may also effectively block certain routes: there is no point in even considering machine translation if no system is available with the appropriate language pair and direction, for example, or conversely in suggesting human translation if no human translator with the appropriate language combination is available. The user is asked to supply information on the language pair(s) involved in the project.

A second rather obvious factor is the deadline by which a translation must be reproduced. To take a rather caricature example, if 5'000 pages have to be translated in the space of twenty four hours, the only route by which this might be achieved is machine translation without post-editing. In the vast majority of cases, this factor will interact with other factors, such as the length of the text and the availability of appropriate resources (glossaries, translation memories). Again, the user is asked to supply information on deadlines.

A third factor is the maximum cost. In general, this factor will not play a major role in TransRouter. since our base assumption is that the translation manager would like to know about all the possible routes for his project, and to be presented with comparative costs, but it may be useful in some cases to allow the user to stipulate a maximum cost and thus reduce TransRouter processing time.

## **2.2 Factors to do with the nature of the text to be translated.**

The first of these, and again the most obvious, is its length. A very short document may warrant no investment at all in using any technology other than dictation software or a word processor: it might take longer, for example, to set up a translation project for a translation memory system and to work interactively with the memory than just to dictate the translation. At the other extreme, as we have already noticed, a lengthy document combined with a too short deadline may force translation by a team of translators rather than by an individual, with possibly deleterious effects on final quality which may be palliated by the use of shared resources or by certain translation management tools. Document length is measured by a word counter.

Sometimes, of course, a document may have to be divided across a number of different translators not because of its length but because it requires different expertise from different translators. Although we recognise the existence of this factor, the current version of TransRouter does not offer any help in determining when or where this factor is present.

Formal characteristics of the text may affect what routes are appropriate. One such factor is average sentence length: it is well known that as a general rule, machine translation systems will produce more satisfactory results for shorter sentences than for longer ones. Similarly, sentence complexity will have an effect on the results of machine translation. Sentence simplicity, although intuitively related to sentence length is not a direct function of sentence length. For example, a sentence like "Destroying of bridges weakens native hostility" whilst not particularly long is rather complex. TransRouter provides tools which estimate both average sentence length and sentence simplicity.

If a document exhibits a high degree of internal repetition, it may be worth considering working with a translation memory system interactively, creating a memory as the translation progresses. The analysis functions provided with most translation memory systems provide statistics on the exact matches at phrase level contained in a text, but give no estimate of the degree of repetition at less than the phrase level, or of fuzzy repetition. (There are ways with some systems to hack round and get some sort of estimate, but the hack is complicated and does not give very satisfactory results). Furthermore, not all repetitions are equally interesting. Knowing that "to the" appears fifty times in a text is not very useful. Knowing that "unmitigated scoundrel" appears fifty times is.

TransRouter therefore provides a repetition detector which estimates the degree of repetition, expressing it numerically, and also constructs an ordered list of repeated sequences, putting the most interesting at the beginning of the list.

## **2.3 Factors to do with whether the text has been translated before.**

If a document has been translated previously, only the changes between the new text and the old text need to be translated. A version comparison tool identifies the changes.

## **2.4 Factors to do with the existence of linguistic resources.**

The unknown terms detector compares a text to be translated with an existing lexical resource and provides a list of all those sequences in the text which are likely candidates as terms but which do not appear in the lexical resource. The results serve as an indicator of the utility of using that lexical resource in translating the text. When the lexical resource is part of a machine translation system, the results also provide an indicator of the probable utility of the machine translation system as a whole.

Most translation memory systems provide a facility for comparing a text to be translated with one or more existing translation memories in order to determine the degree of overlap between the new text and what is already stored in the memory. TransRouter too makes use of this information in order to determine whether a route including use of an existing translation memory is likely to be fruitful.

## **3 Putting it all together: the TransRouter Prototype.**

The TransRouter system is intended to support users in choosing the best route by which to carry out a translation project. It will present the user with a description of one or more viable routes which could be taken, along with the associated cost, time required, output quality and any advantages or disadvantages for each route. The calculation of the viability of the different potential routes is carried out by the TransRouter decision kernel. The information on which the kernel's calculations are based is found in a number of different profiles. In this section we briefly describe the different types of profiles and the relationship between the component tools and the profiles and the decision kernel in the expected final version of TransRouter.

### **3.1 Profiles in TransRouter.**

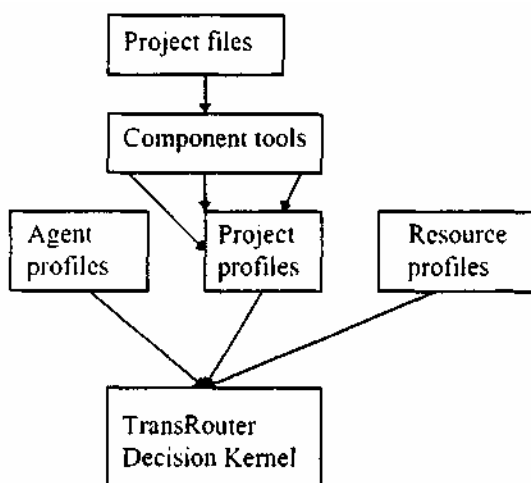
There are three different types of profile to be found in TransRouter: project profiles, agent profiles and resource profiles.

Project profiles: a project profile contains information concerning a translation project. It contains all information relevant to the project, including both

information on the requirements of the commissioner of the translation (such as the source and target language(s), the quality of translation required, and the deadline for the translation) and the inherent properties of the text to be translated which can affect the routes to be taken (see discussion in the preceding section). Whilst some of the information must be entered by hand, a number of features pertaining directly to the properties of the input text will be computed automatically by the component tools, as also discussed above.

Agent profiles describe “translation agents” which could be used in translating a project. The term agent covers not only agents directly carrying out translation such as translation memory systems, machine translation systems and human translators, but also applications supporting the translation enterprise such as electronic lexica, terminology and alignment tools.

Resource profiles describe the linguistic resources available within the different translation agents, such as translation memories (from previously translated texts), machine translation lexica, termbanks and stand alone lexica. Resource profiles are periodically up-dated off-line as resources are built up or acquired. The relationship between the profiles, component tools and decision kernel can be summarised by saying that component tools supply information to the project profiles, which then are used, along with the agent profiles and resource profiles by the decision kernel to calculate possible routes. This can be pictured as in the diagram below, where it should be noted that only the flow of information into the kernel is shown, nothing being said here about output from the kernel.



In the diagram it can be seen that the kernel takes information from the different profiles. Whilst agent and resource profiles are manually compiled off-line, the project profile is partly compiled by hand but also receives some information regarding the properties of the translation project files from the output of the component tools.

### 3.2 The architecture of the decision kernel

TransRouter is an object oriented decision support system. There are classes for all relevant profiles within the system – projects, agents and resources. Each class provides an associated viewer class allowing a form based interaction with an object of that class. Thus, the user may browse profiles and update them interactively. The decision process – allocation of resources and computation of routes – is associated to a specific object class which implements only a very generic decision function drawing heavily on class specific functions for projects, agents and resources.

The decision process comprises several main steps (minor details are omitted):

1. Selection of agents and resources according to project features:
  - a. Class oriented selection rules allow the selection (or rejection) of entire groups of agents/resources (For example, machine translation should be rejected if high quality is required).
  - b. Object oriented selection rules affect individual objects (Is the appropriate language pair available?).
2. Combination of agents and resources: Resources are allocated to agents, agents are grouped into teams. Class oriented rules and features define which objects may be combined.
3. Resource assessment: Resources like translation memories or termbanks are related to project features (What is the coverage, how many terms are unknown with respect to a dictionary etc?). This information is provided by specific component analysis tools connected to the decision kernel.
4. Route computation: The system will next set up a set of feasible routes. The route model of TransRouter distinguishes between several route types represented by object classes. Each route type consists of three steps only (pre- and postprocessing and a main translation step) but defines individually how the resources at hand are used (by defining the main translation agent, the role of translation memories etc).
5. Route assessment: The routes are validated with respect to time, cost and quality aspects. Each route type has its own time, cost and quality estimation routines which rely on the features of projects, agents and resources.
6. Refinement: The refinement step is performed by the translation manager. He may edit the route suggestions of the system and then run the assessment step again.

The object oriented architecture of TransRouter as outlined above has several advantages. New object types (some newly developed translation aid or a newly available translator, for example) may be integrated easily. Neither the overall interface nor the decision

process in general need be affected. Any individual software component is comparatively simple and easy to understand. The expressive power of the decision model mainly stems from the adequate combination of pertinent object types.

#### 4 Progress to date.

First versions of the component tools mentioned above have been developed. In the first phase of the project a conscious decision was made to make use of existing technology wherever possible in order to be able to concentrate on the design and development of the decision kernel. Thus the word counter implements access to the word counter built into Microsoft Word, and the translation memory coverage estimator makes use of the analysis functions provided with the Trados Translators Workbench software. This latter makes the first version of the TransRouter platform dependent on Trados. Subsequent versions will remove this dependency, by including other translation memory systems, although the translation memory coverage estimator will continue to be intimately connected to the translation memory system for which the memory was created, at least until the TMX exchange format for translation memories comes fully into use.

The project includes participants from very different backgrounds. In order to facilitate common understanding of the project's goals, it was thought important to produce a first prototype as soon as possible. This prototype was produced using rapid prototyping methods, rather than following the object-oriented architecture described above which will be used for subsequent prototypes. However, as can be seen from the brief description which follows, the main ideas of the core architecture have been included in the first prototype, which has thus served to lay the framework for the TransRouter system, by providing a reduced set of functionality to demonstrate the principles of its operation. Profiles have been defined for projects, agents (including translation memory systems), and for the resources to be associated with agents. Vendor price quotes are also captured through the agent profiles. A database storage and retrieval architecture was defined to allow entry, lookup, and modification to profile data relating to a project, agent or resources. As we have seen, the decision kernel examines, combines and derives results using the profile data, user input and analysis tool input. In this prototype the analysis performed is divided into three phases as follows.

1) The "Cutoff analysis" phase compares the components of the project against the each of translation routes. A translation route ('route') comprises the combination of human translators and agent systems and is seen as the method of processing a project. This analysis phase performs filtering to determine what can be processed by which route according to the defined compatibility issues.

- 2) The "Critical analysis" phase uses the functionality of the component tools of the TransRouter system to calculate leverage statistics of the project components in connection with each route. The raw output of this phase provides qualitative and quantitative measures of advantages and disadvantages for each of the defined routes.
- 3) The "Cost Analysis" phase combines this raw output with price quote data from the profiles to produce measures of cost for the translation process via each of the routes.

The results of this analysis are provided to the user to be used in assessing the viability of processing a project in a particular way. In subsequent prototypes it is planned that analysis results will be archived for potential re-use. It is also planned to carry the analysis a step further in taking user-supplied preferences as a basis for ranking results and providing recommendations. However it is not seen as the role of the TransRouter system to make the decisions.

The information processed in this first prototype is very simple and does not accurately describe the multitude of factors governing the translation process. One main aim for further development is to refine the definitions of routes and the data in profiles to approach more closely to a representation of the real process. However there is a trade off between accurate modelling and providing a simple, easy to use system. It is hoped to arrive at an acceptable compromise where useful results are quickly obtained and easily understood.

#### 5. Evaluation and feedback.

A videocam demonstration of the first version of TransRouter has been made and will be shown to selected groups of users who have indicated their interest in the project in order to elicit feedback, especially on the choice of factors which has been implemented and on the design of the user interface.

#### 6 Research themes and open issues.

One theme of some importance has emerged during work on the project. The consortium is made up of people with a wide variety of competencies, including translation managers, localisation experts, translation technology experts and computer scientists. Those who have the knowledge and experience to define plausible routes for a translation project do not have the programming competence nor the resources to translate their knowledge directly into computer code. Those who have the programming competence do not have the translation or localisation expertise. We were thus faced with a communication gap, which has been bridged by the creation of a language for route specification in the form of diagrams. At the time of writing, proposals for this language have been put forward, but we do not as yet have much practical experience with its use as a communication tool. If it should prove effective, it may

allow us also to communicate clearly and effectively with the members of the user groups mentioned above, offering them a way to make suggestions about how routes should be refined or modified, or about new routes to be added.

Another major research theme concerns the thorny question of translation quality. The decision kernel works out routes and reports on them giving for each route an associated cost and time frame as well as an estimate of the translation quality thus achieved. The first phase in this research has been to identify and represent the factors which affect translation quality.

Information necessary for estimating translation quality is gleaned from a number of sources in TransRouter. In the project profile the user defines the translation quality required, whilst agent and resource profiles can contain a number of attributes directly or indirectly pertaining to the quality which can be achieved using a particular agent and its associated resources. Thus for example, the profile of a machine translation system will contain attributes which indicate the general level of quality of "raw" translations (possibly divided into sub-attributes for different text-types and/or subject areas if relevant). Profiles of the lexical or terminological resources associated with the specific machine translation system will also have attributes indicating whether the resource has been validated, or used in a previous translation project and so on.

The properties of the source text which are discovered via the component tools also play a role in determining translation quality. For example the syntactic complexity of a text as identified by the sentence simplicity checker will affect the quality of a raw translation output by a machine translation system. Thus the output of the component tools can be seen as qualifying the quality-related attributes in the agent and resource profiles. The second phase of the research will be to investigate how to best model quality in the decision kernel itself. Quality related attributes in the profiles are given qualitative values. How these are utilised in modelling quality and whether they are converted into numerical values in the decision kernel, are as yet unresolved questions.

Since different potential routes will most often involve more than one step or agent, the decision kernel must also calculate the effects of combining the different steps or agents on the final translation quality. For example, including a human post-editing step into a machine translation route would be expected to have a positive effect on the quality of the final translation. Although quality modelling is still a current research topic a number of points are already clear. Translation quality is closely linked to the amount of time available to carry out the translation. A deal of flexibility in representing translation quality attributes will be necessary to account for different users' working practices and enable them to get meaningful results. For example, if a user already has a certain way of validating and categorising terminological or translation memory

resources then it should be possible for them to use the same validation categories in profiling these resources in TransRouter.

### **Bibliography.**

- Falkedal, K. (1994): Proceedings of the Evaluators' Forum, Les Rasses, ISSCO, Geneva.
- King, M. (1996): SdT: A Case Study. Appendix to the 1996 report of the EAGLES Evaluation Working Group, to be found at <http://issco-www.unige.ch/projects/ewg96/>
- King, M. (1998): Translation technology: integration in the workflow environment. EAMT Workshop Proceedings, WHO, Geneva.
- Steenkamp, J.B.E.M. (1989): Product Quality: an investigation into the concept and how it is perceived by consumers. Van Gorcum, Assen/Maastricht.
- Steiner, G. (1975): After Babel: Aspects of Language and Translation. Oxford University Press.
- TMX: the OSCAR special interest group within LISA is responsible for this initiative:  
<http://www.lisa.unige.ch/tmx/>
- Underwood, N.L. and Jongejan, B. (1999). Profiling Translation Projects: An Essential Part of Routing Translations. To appear in the Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation, TMI-99, University College, Chester, England.
- Van Slype (1979): Critical Study of the Methods for Evaluating the Quality of Machine Translation. European Commission, DG XIII, Report BR 19142.