Terminology

Monday 20 October 2003 (1.30 – 5.00 p.m.)

Session Agenda & Introductory Report

"Innovative translation techniques in the road field"

22.T1.E

SESSION AGENDA

1. Introduction

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

2. Dictionaries

Ms. Sylvie PROESCHEL (Member of Committee on Terminology/FRANCE)

3. Terminology CD-ROM

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

4. Continuous Updating

Mr. Daniel VERFAILLIE (Member of Committee on Terminology/BELGIUM)

5. Adding a Language

Mr. Theo MICHELS (Member of Committee on Terminology/THE NETHERLANDS)

6. Nordic Countries

Mr. Tore HOVEN (Member of Committee on Terminology/NORWAY)

7. Babylon

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

8. Discussion

Mr. Doug M. COLWILL (Secretary of Committee on Terminology/UK)

9. Machine Translation

Dr. James C. WAMBOLD (Member of Committee on Terminology/USA)

10. Demonstration

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

11. Limitations

Ms. Sylvie PROESCHEL (Member of Committee on Terminology/FRANCE)

12. Translation memory databases

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

13. Conclusion

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

14. Closure

Mr. Patrice RETOUR (Chairperson of Committee on Terminology/FRANCE)

CONTENTS

Dear Colleague and Delegate,

Translation concerns us all: to understand and be easily understood.

Although PIARC has been active in the field of terminology since the 1930s, there has never been a session at a World Congress. This anomaly will be corrected at the Durban Congress.

The session that we have prepared will offer you the opportunity to take stock of the tools of assisted translation and the dictionaries and program packages available; you will be able to see demonstrations on line and you will be able to take part in debates. Also demonstrations will take place at the PIARC's stand during the course of the Congress.

English and French, the official languages of PIARC, are the principal languages with which we are concerned, but the list of the languages in which PIARC already offered services is impressive; namely:

Arabic	*German*	*Polish*
Chinese	*Greek*	*Portuguese*
Czech	*Hungarian*	*Romanian*
Danish	*Italian*	*Russian*
Dutch	*Japanese*	*Slovak*
English	*Khmer*	*Spanish*
Finnish	*Lithuanian*	*Swedish*
French	*Norwegian*	*Vietnamese*

Come and see how it is –and will be– increasingly easy, fast and cheap to translate from one language to another!

On behalf of the PIARC Technical Committee on Terminology,
Patrice Retour
President

The following persons have been involved in drafting this report:

Doug COLWILL (United Kingdom)
Theo MICHELS (the Netherlands)
Sylvie PROESCHEL (France)
Patrice RETOUR (France)
Daniel VERFAILLIE (Belgium)

EXECUTIVE SUMMARY

The developments in computer science and the Internet can now be used to achieve a more rapid sharing of information. In recent years, software programs for translating into several languages have appeared on the market and they are already facilitating the exchange and pooling of knowledge.

The PIARC Technical Committee on Terminology has the role of facilitating the communication between the members of the road community and, more particularly, between those who use different languages. This is why the Committee is resolutely engaged in the assessment of modern tools (dictionaries and computer-aided translation) that facilitate the time consuming task of translating into different languages.

In the session, the state of the art regarding electronic dictionaries and lexicons specializing in the road field will be presented. It will be accompanied by demonstrations of translation assisted by computer, which will show the results and the limitations of these packages. The extensive terminology documents of PIARC, in electronic form, will also be presented.

Finally, the debate during this session will allow a first evaluation of the prospects opened up by the progress in computer-aided translation in the road field.

1. ELECTRONIC DICTIONARIES AND TRANSLATION SOFTWARE: INVENTORY OF TOOLS, APPLICATIONS WITHIN THE FIELD OF CIVIL ENGINEERING

1.1 The end of a myth…

The notion of a translation machine capable of overcoming in an instant all language communication problems between the world's populations is a dream as old as time along the same lines as the dream of flying. The Cold War period brought with it an extensive effort on the part of Americans to decipher documents written in Russian and led to considerable American investment in developing an automatic translation system. This initial idealism eventually paid off by intensive linguistic research. Recent discoveries in artificial intelligence and new information technologies, coupled with the recognition that language complexity lies beyond (at least for the time being) any full-scale automation, have debunked the myth of a translation machine. The processes inherent in current automatic translation programs no longer claim to rival the human brain in reproducing all of the nuances behind the thought expressed through the use of language. Translation software manufacturers and distributors cannot be faulted: they no longer promise 100% exactness in the translation output by an automatic process, which has not, however, prevented them from branding their products with miracle-sounding names!

1.2 The release of new tools full of promise…

The potential offered by Internet has served to catalyse and place 'centre stage' the entire range of far-reaching developments in the field of language sciences, providing Web users with access to all sorts of linguistic tools and information on the products being marketed. The objective is not to set forth an inventory of all the existing products or to describe recent trends in terms of morphological, lexical, syntactic or semantic analyses, but rather to present the various families of useful products.

1.2.1 Electronic dictionaries far and wide

The tremendous success generated by electronic dictionaries is owed to the browsing and search-enhancing functions used for consultation. Many dictionaries, whether monolingual or multilingual, may be accessed free of charge on the Internet, while others are marketed by specialised distributors. The latest dictionaries feature the translation of words and expressions in relation to their context.

Other products simply provide an electronic support that allows the creation of more personalized dictionaries.

1.2.2   Multilingual terminology databases:

The European Community, a major consumer of translation services, has developed the "Eurodicautom[1]" database to supply, at no cost, translations and definitions of useful terms in the field of civil engineering. Unfortunately, the level of congestion currently caused by the extremely heavy volume of database queries has seriously limited access.

Quebec, a pioneer in the domain of terminology development as a result of the province's longstanding institutionalization of bilingualism, has produced the Quebec Terminological Databank along with a number of resources, both free and paid, which may be accessed on the Web[2].

Moreover, a number of terminology management software products make it possible to constitute terminological databases specific to each organization and adapted to a particular set of needs. The availability of illustrations or photographs, despite their value in helping to comprehend the terms used to describe machine parts or structural elements, remains quite limited. Without necessarily trying to anticipate the future, considerable advances in integrating illustrations can be reasonably expected thanks to the enormous progress being made in hard disk and memory capacity as well in the more efficient use of computing tools. The full range of audiovisual media will be used within the foreseeable future.

1.2.3   So-called "translation" software

A distinction must be made here between software applications that provide translation-related assistance to the translator and applications that output "raw" translated text.

Translation memory bases *or archives* (built on the "alignment[3]" principle) assist the translator by making use of the repetitive nature of technical documents and by automatically displaying on the screen previously translated and validated sentences, text passages and phrases. This type of tool is without a doubt the most worthwhile and *de facto* the most reliable; when combined with a terminology management tool, it becomes extremely valuable and ensures homogeneity not only of specific terms but in phrasing as well. However, such a tool requires centralized organizational co-ordination in order to be truly operational and profitable (in the light of the collection, processing and validation steps involved). Furthermore, these software products are not intended to replace the professional translator or writer, who retains the authorship of the final output, as opposed to translation software whereby a complete translated document is generated in an autonomous manner.

---

[1]   http://europa.eu.int/eurodicautom/login.jsp
[2]   http://www.olf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/Internet/Index/
[3]   To align a translated text, is to create a two-column table in which each sentence of the source file is put in correspondence with the corresponding sentence(s) of the file of the target language.

> The PIARC CD-ROM **"CD-ROUTE"** (published in 1995, 1997 and 2000) constitutes an exceptional translation memory base in the field of road and road traffic.
>
> Its latest version contains some 16,000 pages of reports and technical articles published both in French and English. In the great majority of cases, the translations were validated by PIARC.
>
> All the texts were published according to the rule laid down by PIARC in 1991 and complied with ever since: the French text is on the left, the English text on the right and the two texts are aligned, paragraph by paragraph.
>
> All the files are PDF files (Acrobat of Adobe®); not only has an index been established for each report, but a general index has been established for the 12,000 pages as well.
>
> Consequently, when searching for a word or a phrase in a language, all the occurrences of this word or phrase are found almost instantly in the whole of PIARC's publications; and its translation in the other official language of the Association is accessed at once.
>
> As a **translation memory base**, this CD-ROM is a "gold mine" for translators of technical texts in the road field.

Automatic translation software programs have become increasingly widespread. The most well-known and undoubtedly one that still tops the list in terms of performance, despite its lengthy presence on the market is Systran®, a direct descendant of the early American work discussed above, but extensively developed for the European Commission. Systran® also happens to be the software selected by PIARC. Its principle is entirely different from that of a translation memory base: sentences are dissected word by word according to a set of linguistic rules, to such an extent that the first draft of the translation is produced without involving the translator. This type of tool implies the allocation of fewer human resources than a translation memory base as the centralization and coordination of terminological resources is not mandatory. The output, however, does require considerable individual, qualified effort in the way to edit (and in many instances revise) the text.

1.3   The PIARC product line…

It is now commonly recognized that terminology is the pillar not only of any translation aid, but also of any high-quality communication system. PIARC has for a long time acknowledged the major role inherent in both concept and terminology and, as far back as 1930, began compiling the first version of the "PIARC Dictionary" (already then featuring several pavement structure diagrams). More recently, in collaboration with both LCPC[4] and SETRA[5], PIARC studied the feasibility of developing an automatic translation tool especially devoted to the field of roads and civil engineering by upgrading the capabilities of Systran® by adding the PIARC terminology database. Over time, PIARC has thus refined its own array of terminology-related services and now offers several types of translation aids.

---

[4]   LCPC: the French Public Works Research Laboratory
[5]   SETRA: the Department of the French Road Directorate in charge of general studies

1.3.1 Conventional lexicographic instruments available either in hardcopy or electronically

- a French-English Dictionary[6] that provides, whenever the concepts inherent in the target terms vary from one language to the next, <u>reliable definitions</u> of such terms and concepts in both languages;

- a Lexicon[7] that remains a <u>simple toolbox</u> offering equivalences between the French and English technical terms without including any definitions;

- a number of dictionaries and lexicons for cross-reference with other languages.

1.3.2 More sophisticated electronic translation aids

- a tool for creating and managing specialized terminology data bases in various languages: the CD-ROM "PIARC Terminology";

- a well-adapted version of the French-English-French Systran® software enhanced by a set of specific terms stemming from the PIARC database;

- an adapted version of the French-English-French dictionary-building software "Babylon®" enhanced by a set of specific terms stemming from the PIARC database. (see section 4)

1.4 Some basic common sense principles…

An organization interested in acquiring translation aids should establish a set of minimum specifications that incorporate the following criteria:

- profile of the particular group expressing translation needs and their qualifications, especially linguistically;

- current needs and the types of texts to be translated, level of quality required depending on the target audience, distinction between translation into one's native tongue and translation into a foreign language (as one can rarely master a foreign language to the same extent as the mother tongue);

- analysis of needs expressed and assimilation of these needs with the various existing tools: lexicons, dictionaries, terminology databases and translation software programs, together with the equally very useful spelling and grammar correction tools;

- potential for effective combination of the various tools with other software, such as desktop publishing and graphics/multimedia packages, depending on the target applications;

---

[6] A dictionary is a "collection of words sorted into a specific order that provides a definition or information on the set of signs and symbols".

[7] A lexicon is a "succinct collection of scientific or technical terms associated with a specialized field".

- choice of relevant data fields, in particular the role of graphics, illustrations and/or definitions, according to the type of tools envisaged;

- limits imposed by financial and human constraints in managing the tools already acquired.

1.5 Example of approach adopted within a civil engineering research organization…

Researchers at the *Laboratoire Central des Ponts et Chaussées* (LCPC) experience major needs in the area of translation (or writing), especially into English, whether to communicate with their foreign partners, draft working documents, publish articles in scientific and technical journals, or write papers to be delivered at conferences. These needs, however, are quite distinct in nature and depend on the author's linguistic competence, specific features of the fields of interest and target use.

Following a joint effort by PIARC and SETRA, coupled with a needs-based survey, the LCPC has provided its researchers with several types of terminological tools in order to satisfy a very diverse set of needs stemming from potential users. These include dictionaries (language dictionaries, dictionaries containing French and English definitions of technical terms, English language encyclopedias, etc.), together with terminology management tools (Babylon®) that allow the creation of "customized" dictionaries and an actual translation software application (Systran®) whose fields of specialization will be described below. Unfortunately, the option of developing translation memory bases was rejected at a time when the required "alignment" techniques still entailed overly complex computer manipulations as well as major management and coordination functions. However, this option is being investigated.

2. CONTINUOUS UPDATING OF PIARC'S DICTIONARY AND LEXICON

The availability of the PIARC Dictionary and Lexicon in an electronic form is certainly a major stride forward towards a common terminology that is understood by professionals all over the world. Moreover, it is instrumental in promoting technology transfer under PIARC's Strategic Theme No. 5, which is to achieve appropriate levels of road and road transport development – especially in developing countries, countries in transition and rural and remote areas.

However, these useful communication and technology transfer aids would soon lose much of their value if they were not regularly updated to reflect the continuing developments in road and road-related technologies.

Committee T1 has developed a procedure for the continuous updating of its two main tools: the Technical Dictionary of Road Terms and the Lexicon of Road and Traffic Engineering. The presentation of this procedure will be restricted to the basic versions of the Dictionary and the Lexicon, that is, the versions in English and French (the two official languages of the Association).

2.1. The cyclical "hard copy" updating era

In the past, the preparation of a new "hard copy" edition of the Dictionary, with contributions from a number of PIARC Technical Committees, was long and tedious work for the four members of a Commission on Terminology chaired by the late Mr Coquand, who was known as "the father of the Dictionary". The 6th edition of the Dictionary was finally released in 1990, one year before the World Congress in Marrakesh.

In 1991, PIARC published the first edition of a Lexicon of Road and Civil Engineering, with over 12,000 terms in English and French in common use by road engineers and including – unlike the Dictionary – terms belonging to related fields.

Hence, a major challenge was facing the Commission on Terminology that was newly formed with some ten members after the Marrakesh Congress: in addition to its traditional task of updating the Dictionary, the Commission had to check the translations of over 12,000 terms of the Lexicon, many of which belonged to fields with which its members were unfamiliar.

In August 1992, a letter signed by the PIARC Secretary General (Mr Fauveau) was sent to all the Technical Committees of the Association sharing out the various chapters of both the Dictionary and the Lexicon, while ensuring a coverage of each chapter by at least two committees, and asking for their suggestions for additions and amendments. Even with two two-day meetings a year, in which only the "problem terms" were discussed, it took the Commission several years to process the replies from the Committees. The 7th edition of the Dictionary was released in 1997 and the 2nd edition of the Lexicon early in 2000, just after the World Congress in Kuala Lumpur.

The production of new "hard copy" editions of both the Dictionary and the Lexicon had taken a full cycle of eight years, that is, two inter-Congress periods, and there was every reason to believe that, when the same procedure was followed, the next cycle would be just as long.

A major disadvantage of this eight-year cyclical approach was the impossibility of keeping up with the rapid developments in terminology associated with the introduction and dissemination of new technologies and the extension of PIARC's activities to other fields of interest, such as intermodality and multimodality. By the time the Commission released a new edition of one of its reference works, it was sometimes out of date!

2.2   The continuous "soft copy" updating era

2.2.1   The big change

The big change came in 1999 when PIARC ordered a French software company to develop a computer program to make both the Dictionary and the Lexicon available in electronic form. This was the start of the "PIARC Terminology" project, which brought the PIARC Technical Committee on Terminology – as it was renamed after the World Congress in Kuala Lumpur – into the computer era.

The project resulted in the release of a CD-ROM of the same name in November 2000. In addition to the 7th edition of the Dictionary (in the basic French-English version and in a few other languages) and the 2nd edition of the Lexicon, this CD-ROM contained a number of specialized glossaries: one on sustainable transport, one on bridges, and one on weigh-in-motion techniques. Through the Internet, each of these terminology documents can be extended with translations into other languages, and new documents can be added. In this way, an electronic terminology database is built that can be updated at any time.

2.2.2   A new updating procedure: the "PIARC-Terminology" CD-ROM

To understand the approach, it is useful to bear in mind that the CD-ROM is intended for three levels of use:

- **consulting**: after installing the program and the terminology data files on the hard disk of his/her computer, a _user_ can consult a terminology document, or several at the same time, for a given term; the result of the search can be displayed in three languages at a time. Moreover, updates can be freely downloaded from PIARC's Web site as they become available (such as the Italian and German translations of the PIARC Dictionary and a Dutch translation of the 1st edition of the PIARC Lexicon);

- **editing**: using a password, an _editor_ has the additional option to make proposals for additions and amendments: new terms, new definitions, a new language for a given terminology document, etc.;

- **managing**: finally, the _administrator,_ representing PIARC as the owner of the copyright of the terminology database, makes the validated additions and amendments available on the Web site of the Association.

Each member of the Committee on Terminology has the status of an "editor". Some of them maintain specific contacts with a number of Technical Committees of PIARC – or, more particularly, with the *terminology correspondent* appointed within each Committee.

The terminology correspondents within the Technical Committees have the "editor" status as well. This means that they can enter the proposals from their Committee into the terminology database, export these proposals to a file in an appropriate format, and send this file to their respective contacts in the Committee on Terminology.

In this respect, a good way for terminology correspondents to collect proposals from their Committees is to add a glossary to each specialised report. Other valuable sources of information are the reports for the PIARC World Road Congresses and the articles published in the PIARC's magazine "*Routes/Roads*".

Once the contact persons have received the proposals from their terminology correspondents, they forward them –with comments – to a general *coordinator* in the Committee on Terminology.

The coordinator processes all the proposals received into one file per language (French and English) and sends the two files to all the members of Committee T1 for editorial comments and for their opinion of each proposed term, as to whether it should be included in the Dictionary and the Lexicon or in the Lexicon only, or whether it should be rejected. Using their replies, the coordinator allocates the proposals to the Dictionary and the Lexicon and prepares a document with any remaining "problem terms" for discussion in a plenary session of the Terminology Committee. After circulating the result of this work and discussion to all the Technical Committees for final comment and approval, the coordinator sends the "validated" files to the administrator.

The administrator then makes the update available on the Web site of PIARC.

The full procedure takes about one year. This means that updated basic versions of the Dictionary and the Lexicon become available every year – which is almost continuously when compared with the former eight-year cycle.

3. ADAPTATION TO ANOTHER LANGUAGE: CASE OF THE DUTCH LANGUAGE

The Netherlands has a long-standing tradition of contributing to multilingual dictionaries. Nevertheless, owing to the time-consuming nature of hard copy updating, the latest Dutch edition of the PIARC Technical Dictionary dates back to 1987 (as it was based on the original PIARC fifth edition of 1982).

After releasing the Lexicon of Road and Traffic Engineering in 1991, PIARC invited the Dutch National Committee to add a Dutch translation to it. CROW (the Dutch Information and Technology Centre for Transport and Infrastructure) was asked to co-ordinate this action with funding from the Dutch national roads agency ("*Rijkswaterstaat*"), and to seek the cooperation of the Belgian Road Research Centre (BRRC). The translation work was done by means of digital files containing some 12,000 terms, and was finalized in 1999. This Dutch version of the first edition (French-English-Dutch) was made available to users of the Terminology CD from the PIARC Web site, where it can still be found.

The second edition of the Lexicon (French-English – 16,500 terms, 4,500 of which are new and some 3,000 have been revised) is now available and it is periodically updated. Hence, there is a challenge to update the Dutch edition. Concurrently, the seventh edition of the PIARC Technical Dictionary of Road Terms was developed and released in 1997, and in 2001 a new Dutch nomenclature for roads and traffic (another co-production of CROW and BRRC) was released with some 2,400 definitions.

The availability of these new reference documents has led to a clear need to synchronize the updating of both the Lexicon and the Dictionary. In other countries there is a growing demand for facilitating access to Dutch language documents. The Dutch and Belgian National Committees of PIARC have, therefore, initiated this updating and arranged for funding. Work started in 2002 and is expected to be ready by the end of 2003 or early in 2004.

In the session at the Congress, experiences and problems encountered during the translation process are discussed, as well as the costs and benefits. Other countries considering adding their language are encouraged to join in this discussion.

4. THE BABYLON-PRO® PROGRAM PACKAGE

There are a large number of computer packages that make it possible to create, manage and consult electronic dictionaries. Among them, BabylonPro® is particularly interesting. This package is operational "on-line" (Internet connection) or "off-line".

It allows, by a simple click on a word (or an expression) displayed on screen (using Windows®), the immediate display of its translation, possible synonyms, an illustration, an Internet link, etc.

The package recognizes not only the word, but also, and in priority, the expression in which this word is used. It displays the results of the search in each of the electronic dictionaries selected.

4.1   Functions of the computer package

On-line or off-line access

On line, the package provides information from the dictionaries or from a comprehensive on-line index (without downloading). The search can then be extended to other on-line dictionaries. To work off line, it is sufficient to have downloaded the relevant dictionaries.

Translations

There are dictionaries in more than 55 languages, which provide bilateral translations. Among the languages available are: English, German, Spanish, Italian, French, Portuguese, Dutch, Hebrew, Swedish, Japanese and Chinese. These dictionaries contain more than 120,000 general, encyclopaedic or slang expressions.

Specialised Dictionaries

Babylon® offers more than 1,500 specialized dictionaries, from lists of short definitions up to encyclopaedic references. These dictionaries are of commercial or private origin. They cover a vast range of subjects and can be consulted on-line or when downloaded.

Selection of dictionaries

Users can determine the order in which the electronic dictionaries are consulted. For example, they can choose the dictionaries in the "French into English" direction in the following order: 1/ personal dictionary 2/ specialised dictionaries of the relevant subject area and 3/ general dictionary.

Conversions

On-line, the software allows for converting currencies, measurements and time zones.
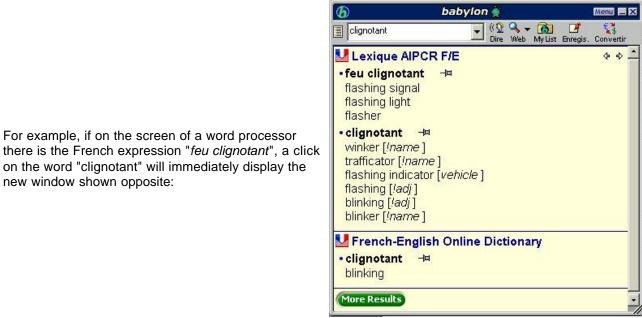
English pronunciation

The package includes the sound of the correct pronunciation of an English word.

Creation and importation of specialized dictionaries

The principal interest in this software ("Babylon-Builder®" module) lies in the possibility of creating, importing and updating specialized dictionaries for each subject area.

4.2   Example of use

The French Public Works Research Laboratory (LCPC) has produced an adaptation of the PIARC Lexicon –French-English and vice versa– and several other specialized dictionaries (weigh-in-motion, aggregates, etc.). These dictionaries are available on the internal Web site for use by all the staff of the Laboratory.

For example, if on the screen of a word processor there is the French expression "*feu clignotant*", a click on the word "clignotant" will immediately display the new window shown opposite:

Screen dump when clicking on
« *feu clignotant* »

Each user can create his/her personal dictionary. The software allows for the comparison of all the personal dictionaries in order to remove double entries, operate regroupings and, finally, enrich the common terminological base.

LCPC has acquired more than one hundred licences for this software for a total staff of approximately 600 people.

5. ADAPTATION OF THE PIARC LEXICON FOR SYSTRAN®

The more its specialized dictionaries are adapted to the field of the text to be translated, the more the use of machine translation software will be effective.

When a translation is undertaken by Systran®, the dictionaries are used in the following order:

1. *Personal dictionary* (none or only one);
2. *Specialized dictionaries*, provided by Systran® (up to four specialized dictionaries can be used simultaneously);
3. A *general dictionary*, also provided by Systran®.

In the field of civil engineering, the following *specialized dictionaries* provided by Systran® may be relevant: car, chemistry, electronic, computer science, earth sciences, mathematics, mechanics, engineering, physics. However, there is little information available on these specialized dictionaries and the size and the origin is not known.

This is why it seemed useful to build a "*personal*" dictionary starting from the PIARC Lexicon. Systran® is accompanied by a compiler that makes it possible to build *personal dictionaries* without the practical constraints of size.

The adaptation of the PIARC Lexicon for its use as a "personal" dictionary with Systran® imposed a very significant workload that took over a year, and was completed in the last quarter of 2001.

The richness of the PIARC Lexicon is to be found in its collection of expressions and synonyms. Therefore, it was important to have these expressions available in a personal dictionary compatible with the translation engine of Systran®. This adaptation of the terminology database of PIARC was entrusted to the University of Nantes[8] (France). In time and complexity, it required more effort from English into French than in the other direction, because an expression comprising a noun and an adjective in English is often translated by two nouns in French and because inflecting[9] adjectives is asymmetric in the two languages. For each pair of languages, all the expressions of the Lexicon were labelled[10] according to the recommendations of Systran®. Labelling is a long and tedious operation, but could be partly automated.

---

[8]   Specialists in natural language and computer science.
[9]   Inflecting: operation consisting in reaching the agreements (gender, plurals) between the nouns and adjectives.
[10]   Labelling: consist in allotting specific Systran® beacons to each word of each expression (source and target) in a dictionary. The beacons specify grammatical nature, gender, and degree of importance of each word in one expression. Thus in the expression "power station" it is the word "station" (substantive, neuter) which is most significant and which the Systran® software will inflect in the event of plural ("power stationS").

Systran®, with its general dictionary as well as all its specialized dictionaries, often proposes only one word translation for a word (except for several idiomatic expressions: "*Il pleut des cordes*" = "Its raining cats and dogs"). Thus the noun "*chaussée*" is always translated (general dictionary) by "roadway".

However, during the adaptation of the Lexicon for Systran® the decision was taken to offer a choice between several translations. Thus, when the PIARC Lexicon is used as personal dictionary, the word "*chaussée*" is translated by: {traveled way | roadway [USA] | pavement [*structure*] | carriageway [road component]}. The brackets delimit the options suggested and a vertical line separates them; context and origin can be specified between square brackets[11].

The use of Systran® with the PIARC Lexicon as personal dictionary is now possible. However, with use, the great richness of the Lexicon turned out to be a handicap because each expression with several synonyms was translated several times. The text then had to be "cleaned" by selecting only one of the translations suggested ("*autoroute*" resulted in "motorway", but could also result in "freeway [USA] " or "superhighway [USA]").

A reduced version of the personal dictionary based on the PIARC Lexicon was developed. It does not include any Americanism and reduced the number of synonyms to one, wherever possible. This version gives satisfactory translations from "British" English into French and vice versa. A French version "from French" into "Anglo-American" could be carried out in the same manner.

---

[11] In the PIARCLexicon, basic English is the one used in the United Kingdom. Americanisms have been included as synonyms, while specifying their origin [USA].

6 LIMITATIONS AND PROSPECTS OF AUTOMATED TRANSLATION

6.1 As for the disadvantages…

6.1.1 Some loss of linguistic richness…

The use of automatic translation software generates a cultural disadvantage because of the *de facto* tendency to remove richness from the languages being processed.

A translation software produces a higher-quality result when the source document is written in a simpler style. For instance, distinguishing between present participles and substantives is not handled particularly well and most of the time those pronouns and personal adjectives not easily associated with a gender or number will require correction. Moreover, there is no use loading a large number of synonyms onto these applications dictionaries because the appropriate term can only be found by conducting a highly advanced semantic and contextual analysis, a feature only very crudely programmed into the latest generation of tools.

Translation systems that rely on memory bases are also, in part, affected by this disadvantage insofar as they focus on repetitive use of terminology and thus contribute to compressing stylistic innovation.

On the other hand, simple language serves to ensure translation reliability; for that reason, the software is recommended for use with more simply worded and simply written source texts, as a means of enhancing accuracy in the translation output.

6.1.2 The rereading step: a potential hurdle!

All translations produced automatically must undergo "editing". It should be recognized however that correcting the prose of another author (whether human or machine) often proves a more difficult task than completely rewriting the text in one's own style. Editing a translation generated by an automated process may turn out to be a very laborious and thankless task!

6.2 …and now for the advantages!

Interestingly, one of the advantages highlighted is the availability of a software-produced translation in electronic form, which bypasses the manual keyboard entry of the translation! Some engineers prefer being provided with a pre-entered and pre-formatted text (tables, illustrations, etc.) in the destination language and then to proceed with the modifications deemed necessary directly onto the resulting computer file.

The on-line automated translation of Web pages already constitutes a substantial service. Similarly, reliance upon translation software provides the user with a cursory glance at the unedited content of a foreign-language text received at the last minute.

Fellow members of an international working group too will appreciate the immediate availability of translated draft versions of reports still in the production stage.

The greatest impact made by translation software in terms of time saving is found when texts feature recurrent content and a consistent form of presentation. These tools become competitive for documents such as assembly plans, equipment specification sheets or weather forecasts. Within this segment, simple sentence syntax, coupled with appropriate lexicological resources loaded onto the software, can lead to a guaranteed level of reliability in the final output.

6.3   Prospects and recommendations

Ongoing advances in the areas of language processing and linguistic and information sciences, together with considerable efforts undertaken by multidisciplinary research teams, will produce major progress in automated translation systems. The activity of translation, however, extends beyond the simple task of cross-coding from one linguistic structure to another (a task far from being accomplished with precision by current applications) and entails social and cultural understanding plus the capacity to interpret a message within a given context, all typically human faculties not programmable into software (regardless how sophisticated it may be).

Common sense would dictate that using translation software judiciously, and confining its application to those contexts where it can prove highly effective, without attempting to ascribe qualities that, to date, remain exclusively human. Furthermore, for documents with more exacting language requirements, the time and effort involved in the editing task should not be underestimated.

Nonetheless, some systems perform a given type of task excellently: the key is knowing when and how to use them!

CONCLUSIONS

The following draft conclusions will be submitted for approval to the delegates participating in the session at Durban:

General – In the last four years, the PIARC Dictionary and Lexicon have been produced in electronic versions rather than the conventional hard copy; this change has enabled the Committee to update the documents on a continuing annual cycle rather than the eight-year cycle previously possible with the hard copy. The existence of these electronic versions of the terminology documents has led to the possibility of more accurate computer-aided translation in the topics covered by PIARC.

For Decision-Makers - Electronic terminology databases and computer-aided translation software programs will provide more rapid, accurate and cheaper translation of documents thereby generating a better understanding between people in different countries and accelerating technology transfer. Consequently, decision-makers should support the development of terminology tools for activities in PIARC's field.

For Technical Experts – The more technical experts contribute to terminology in their speciality field, especially by improving the PIARC terminology data bases, the more accurate will be the computer-aided translation.

For PIARC - The development of electronic versions of the PIARC Dictionary and Lexicon, together with the computer-aided translation systems being developed, are facilitating the transfer of technology in the road field. In order to increase this transfer, it is recommended that:

- PIARC Technical Committees continue to feed the Technical Committee on Terminology with new terms in order to make the PIARC terminology database as up-to-date as possible;

- National Committees of PIARC should be encouraged to make translations of the PIARC terminology database into their respective language;

- before the next World Road Congress (Paris, 2007) PIARC should make the PIARC Dictionary and Lexicon available, at no cost for non-commercial purposes, on the Internet.