LETRAC Curriculum Modules

Deliverable D3
April 1999

Authors: Toni Badia, Karl-Heinz Freigang, Johann Haller, Christoph Horschmann, Dieter Huber, Belinda Maia, Ursula Reuther, Paul Schmidt
## MODULE A
### INTRODUCTION TO COMPUTER SCIENCE

**UNIT 1: INTRODUCTION TO COMPUTER SCIENCE**

Sub-Unit 1.1: Methods and Devices
Sub-Unit 1.2: Data Input – Hardware
Sub-Unit 1.3: Data Input – Software
Sub-Unit 1.4: Data Storage – Hardware
Sub-Unit 1.5: Data Storage – Software
Sub-Unit 1.6: Data Manipulation – Hardware
Sub-Unit 1.7: Data Manipulation – Software
Sub-Unit 1.8: Data Output – Hardware
Sub-Unit 1.9: Data Output – Software

---

## MODULE B
### IT AND DTP FOR TRANSLATORS

**UNIT 1: DTP**

Sub-Unit 1.1: General overview
Sub-Unit 1.2: Hardware Components (scanner, printer, computer, cf. Module A)
Sub-Unit 1.3: Core Concepts of DTP

**UNIT 2: TELECOMMUNICATIONS/INTERNET**

Sub-Unit 2.1: Telecommunications, Technical Basis
Sub-Unit 2.2: E-mail
Sub-Unit 2.3: Dialogue Systems
Sub-Unit 2.4: Internet - Technicalities
Sub-Unit 2.5: Internet – Contents
Sub-Unit 2.6: Internet – Resources for Translators

**UNIT 3: ADVANCED SKILLS**

---

## MODULE C
### LANGUAGE ENGINEERING

**UNIT 1: GENERAL ASPECTS AND TOOLS**
Sub-Unit 1.1: Word processing and checkers
  1.1.1 Multilingual word processing
  1.1.2 Checkers (from spelling to style)
  1.1.3 Interaction of word processors with dictionaries

Sub-Unit 1.2: Controlled language

Sub-Unit 1.3: Project and document management (basics)

Sub-Unit 1.4: Terminology systems

UNIT 2: TRANSLATION-SPECIFIC ASPECTS AND TOOLS

Sub-Unit 2.1: Translation memories

Sub-Unit 2.2: Machine translation systems: working with TM and MT

UNIT 3: LANGUAGE FORMALISATION

Sub-Unit 3.1: Structural and functional approaches

Sub-Unit 3.2: Grammars and parsers

Sub-Unit 3.3: Unification-based approaches to grammar

UNIT 4: MACHINE TRANSLATION

Sub-Unit 4.1: Historical overview

Sub-Unit 4.2: Basic concepts

Sub-Unit 4.3: Evaluation

Sub-Unit 4.4: Unification-based MT systems

Sub-Unit 4.5: EU-MT and NLP projects

UNIT 5: CORPUS LINGUISTICS AND LOW-LEVEL ANALYSIS

Sub-Unit 5.1: Corpus design and organisation

Sub-Unit 5.2: Corpus annotation

Sub-Unit 5.3: Low-level parsing

Sub-Unit 5.4: Information extraction from corpora

UNIT 6: CASE STUDIES

Sub-Unit 6.1: Project, document and terminology management

Sub-Unit 6.2: Controlled language in industrial environments

Sub-Unit 6.3: Complete localisation project (e.g. one source language, several target languages)
Introduction

Background
The present Curriculum Modules have been developed within the LETRAC project and represent one of the core elements of the project results. They are meant to be integrated into translation curricula in order to prepare the students in an adequate and up-to-date way for their professional life with respect to language technology related issues.

The basis for the present Curriculum Modules was an in-depth analysis of the user requirements formulated by the language and translation industry, by individual translators and by students and teachers of translation studies (available as separate documents).

Structure
The LETRAC Curriculum Modules, covering a wide range of language technology related aspects, consist of three more or less self-contained modules which are however interrelated in some respect:

- **Module A**  
  Introduction to Computer Science

- **Module B**  
  IT and DTP for Translators

- **Module C**  
  Language Engineering

In all there are **10 units** (module A consisting of 1 unit, module B consisting of 3 units, and module C consisting of 6 units) some of them obligatory and other optional.

The **compulsory units** (5) are considered to include basic knowledge in the respective domain which should be taught to all students of translation. They represent the minimum knowledge a translator should have working nowadays in this rapidly increasing service sector. Of course, some of this knowledge, like DTP, is not specific to the translation sector, but respective course elements are included with a special focus on translation related issues.

- The **optional units** (5) are intended either to deepen knowledge already acquired in a compulsory core unit (e.g. in module C, unit 6.2 builds on and goes beyond unit 1.2 on Controlled Language)

  or

- to give students the opportunity to specialise in a domain which does not belong to the basic core units, e.g. Language formalisation (unit C 3) for future language engineers or software localisation (within unit C 6) for future localisers.

How it fits together
As for interrelations within and among the three Modules, there are three types of relations, i.e.

1. overlapping parts
2. some parts being a prerequisite for others
3. complementary parts

**Note 1:** This is the case where the same topic is introduced twice in different modules, a fact which cannot be avoided since omitting it in one module or the other would result in an incomplete and incoherent description of the module in question (e.g. hardware components like scanner and printer are referred to in both Module A and Module B).
**Note 2:** In some cases basic units are defined as a prerequisite for other units (e.g. Module A plus unit B1 or B2 are prerequisite for B3).

**Note 3:** Some units are defined as being complementary to each other, either because of the division into theoretical and practical sessions or because of the content (e.g. Module A ‘Data Manipulation’ is complementary to unit C2 ‘Translation specific aspects and tools’ and to C3 ‘Language Formalisation’).

All cases of interrelationship are marked as such and are partly represented as hyperlinks within the document.

The overall structure and the relationships of the modules can be read off the following figure:
Size
The overall size of the LETRAC Curriculum Modules is 230 hours spread over the whole curriculum, which comes down to an average of approximately 2 weekly term hours. By assuming this size it was intended, as a rule of thumb, not to exceed 10% of the total amount of weekly term hours of an average translation curriculum. However, we are well aware that this percentage cannot always correspond to the real conditions, given the various degrees and university systems in Europe.

Nevertheless, in order to allow for European compatibility, it was decided to adopt the European Credit Transfer System (ECTS); details on the distribution of credit points are still to be settled.

How to implement the Modules
Some of the contents covered by the three Modules, ten units in all, may be part of already existing Master's or undergraduate translation studies.

Therefore, if a translation department decides to implement the LETRAC Curriculum Modules, it is necessary to first check which parts are covered already, either by an identical course or, as is perfectly possible, 'in the guise' of another course; for instance, a 'traditional' course on general linguistics is very likely to include elements of unit C3 ‘Language Formalisation’, or, basics of terminology are taught in LSP classes, etc.

In this case, the modular design of the curriculum elements also allows partial integration of units into an existing translators’ curriculum.

If a translation course or a complete translation curriculum is going to be set up anyway, a complete implementation of the proposed modules ensures completeness and consistency of the Language Technology elements.
Module A

Introduction to Computer Science

1 Unit (compulsory)
Total: 30 hours
ECTS credit points: 3

Contents:
The purpose of this module is to provide students in interpreting and translation with some basic concepts of computer science which are relevant for their future careers as professional interpreters and translators. Rather than presenting the usual general purpose introduction to the techniques and technologies of computers and data processing by machine given for first-year students of computer science, the course elements within this module are focused on the specific needs to be met by any computer system applied to the solution of natural language problems, both in written language and speech and both within one particular language or across language barriers. With this aim in mind, the structure of this course follows a general progression from data input via data storage and data manipulation to data output. Thus, after a brief presentation of the main components and devices of computers, each of these four aspects is presented separately both from a hardware- and a software-oriented point of view, and presents a stratified introduction to the fundamentals of computers and their use in multilingual data processing.

Method:
The teaching method should be based on a very practically oriented approach, when hardware components of a computer are presented. A computer could be taken apart in class and components like hard disk drive and motherboard can be shown around. This helps a lot to decrease the respect of technology for students who aren’t acquainted with it yet. A practical approach should as well be followed for issues like file organisation and data storage. Other topics like character sets, programming languages etc. will of course be dealt with in a more theoretical manner. Furthermore, an indication about current hard and software standards should be given in this class, in order to enable students to buy an appropriate computer.

Unit 1: Introduction to Computer Science

30 hours (compulsory)
Focus on theory and practice

Professional requirements addressed by this unit:
The use of computers in the translation workflow of today goes without saying. Therefore, every student has to acquire basic skills in this domain, which are meant to lay the foundation for further study modules.

Overview:
In this unit general concepts of computer science are presented. The contents already put a certain focus on issues relevant for translators and interpreters.
Prerequisites:
None

Relation to other LETRAC Modules and Units:
Many of the subjects which are introduced in this module only on a rather theoretical level will be treated more extensively and in depth in the other LETRAC Modules. These relations are marked by links within this document.

Relation to the rest of the curriculum:
Besides interrelations with certain LSP translations, which deal with texts about information science and technology, it is hard to see an immediate relation to the rest of the curriculum proper. Instead, there is undoubtedly a relation to the rest of life outside the academic, where an information society increasingly grows and flourishes, which the translator in training will be soon a part of. At the latest after he leaves the university and starts to exercise his profession in practice, he will be expected to work in an environment characterised by a high integration of communication and information technology and he will need any of the knowledge acquired in this unit.

Sub-Unit 1.1: Methods and Devices
4 hours

Contents:
An introduction to
- Data and information processing
- Input, Output and Central Processing Unit (CPU)
- Operating systems and Systems Software
- Arithmetical and logical operations
- Bits and Bytes
- Algorithms and Programs
- Programming Languages
- Compilation and Interpretation
- Application software

Sub-Unit 1.2: Data Input – Hardware
4 hours

Contents:
Description and demonstration of hardware devices used for language data entry, both on-line and off-line, for interactive as well as batch processing, are described and demonstrated.
- Terminal keyboard (and screen)
- Scanner
- Optical Character Reader (OCR)
- Optical Mark Reader (OMR)
- Networked data entry (LAN, WAN, Internet, WWW, etc.)
- CD-ROM and DVD
- Tape
- Diskette
- Microphone
- Audiotape
- Camera
- Videotape

**Sub-Unit 1.3: Data Input – Software**

*4 hours*

**Contents:**
Transforming hand- and machine-written text, speech and video-data into the binary code of sequences of 0 (zeros) and 1 (ones), which is the only "language" that computers can handle, is called digitisation. Different digitisation methods have been developed to represent alphabetic, numeric, acoustic and visual data in computer systems.

UNICODE and ASCII/ANSI
- Representing alphabetic and special characters
- Representing numeric values
- Optical Character Recognition OCR (both handwriting and print)

Digitisation of Audio and Video
- Data Sampling
- Anti-Aliasing
- Quantisation
- Speech Recognition

**Sub-Unit 1.4: Data Storage – Hardware**

*4 hours*

**Contents:**
Once converted into binary code, any kind of data can be stored in different sorts of computer memory designed for a variety of purposes, both within and outside the central processing unit (CPU). Depending on memory space and data density, accessibility, addressability, the physical organisation and format of arrangement, retrieval speed and the actual material used as the carrier of the memory function, various schemes of classification for computer storage devices are used.

Main Computer memory
- RAM
- ROM
- PROM
- EPROM
Auxiliary Storage Devices
- Diskette/Floppy disk
- Hard disk
- Winchester disk
- Disk cartridge
- ZIP disks
- JAZ-disks
- Magnetic tape
- Tape cartridge
- Bubble and flash memory
- Writable CDs

Magnetic core memory
- Semiconductor memory
- Optical disk systems
- Transistors

**Sub-Unit 1.5: Data Storage – Software**

*4 hours*

**Contents:**
Software for data storage basically performs one of the two basic tasks: sorting (including file organisation and data compression) and searching (data retrieval). Relevant topics to be covered are the following:

**Sorting:**
- Coding, compression and en/decryption
- En/decryption
- Sequential file organisation
- Indexed file organisation
- Indexed text bases
- Direct or relative file organisation
- Data bases (hierarchical, networked, relational)
- Data base management

**Searching:**
- Sequential retrieval
- Random retrieval
- Hashing
Sub-Unit 1.6: Data Manipulation – Hardware
2 hours

Contents:
The details of this section will be linked to the respective description of Module C, mainly Language Formalisation and Translation specific aspects and tools.
Data manipulation comprises all kinds of linguistic data and information processing including such diverse tasks as e.g. lemmatisation, morphological analysis, syntactic parsing, multilingual text processing and text generation, and automatic translation and interpreting.

Sub-Unit 1.7: Data Manipulation - Software
3 hours

Contents:
The details of this section will be linked to the respective description of Module C, mainly Language Formalisation and Translation specific aspects and tools.
Application software for a variety of linguistic tasks has been developed within the academic fields of computational linguistics and natural language processing.

Sub-Unit 1.8: Data Output – Hardware
4 hours

Contents:
- Terminal screen
- Printer
- Plotter
- Networked data output (LAN, WAN, Internet, WWW, etc.)
- CD-ROM and DVD
- Tape
- Diskette
- Loudspeaker (natural or synthesised speech)
- Audiotape
- Videotape
- Braille board

Sub-Unit 1.9: Data Output – Software
3 hours

Contents:
Data output, to a large extent, implies the reverse procedures to the digitisation processes performed in data input, i.e. the binary code of computer internal data processing an
communication has to be transformed back into the analogue data perceptible to and understandable by human computer users.

- D/A Conversion
- Text generation
- Text-to-speech synthesis
Module B
IT and DTP for Translators

3 Units (compulsory and optional ones)
Total: 60 hours
ECTS credit points: 8

Contents:
This module is called IT (information technology) and DTP (desk top publishing) for lack of a term that covers various topics such as technical and information retrieval aspects of the WWW, Desktop publishing, programming and knowledge about operating systems.
According to the information that was collected during the user needs analysis, we assume three modules, twenty hours each, two of them compulsory covering desktop publishing (DTP) and telecommunications, and one optional module covering advanced computer skills such as different operating systems and basic programming.

Method:
The fact that translators are mainly supposed to use these tools efficiently – and not to design them - should always be borne in mind. Thus, education should prefer a practical approach in conveying the respective skills instead of putting too much stress on theory, underlying concepts and architecture. Still, students have to know about these things, but there is no reason why they shouldn’t face these subjects after a hands-on experience with soft- and hardware. In other words, a bottom-up approach should be followed.

Unit 1: DTP
20 hours (compulsory)
Mainly practically oriented

Professional requirements addressed by this unit:
The market survey of the professional environment gave strong evidence to the observation that today's translator is frequently required to work within DTP-processed documents. He must be able to handle different applications designed for professional layout purposes. Even though he is not meant to design material of professional layout from scratch, he should nevertheless be able to modify text and graphics received in DTP-processed form.

Overview:
This unit is meant to enable students to handle different DTP-Tools like Framemaker, Pagemaker, QuarkXpress and the like, because this is a common requirement in the modern translation environment of a translator.

Contents:
The overall purpose of the unit is to acquire the ability to produce professional high quality printed material on the basis of modern DTP systems. This serves the most prominent requirement from the questionnaires and other investigations of the user needs analysis. A
didactic principle that holds for this course as well as for others is that it is not just the ability to make mechanical use of a technology should be taught. Students should be enabled to make 'creative' use of the techniques. An other important goal is that people should be enabled to explore the DTP facilities themselves, as the 20-hour class described below cannot guarantee that they will become absolutely professional in DTP. This unit can be split into different sub-units.

**Prerequisites:**

- **Background knowledge:** There are two prerequisites: One is basic knowledge about computing, i.e. Module A. Another is basic text processing. The DTP course should not include basic text processing.

- **Software:** DTP packages like Framemaker, Pagemaker, QuarkXpress, Ventura Publisher and the like. Graphical programs like Corel Draw, Photoshop, Paint, Paintbrush etc. Presentation software like Powerpoint and/or Visio etc.

**Relation to other LETRAC Modules and Units:**

Even though this unit is not meant to teach students Word Processing, some advanced features of text treatment might overlap with Unit 3 of this module. Practice has shown that students are often not familiar with general aspects of file handling, which can be important in the DTP context, as far as e.g. the embedding of graphics is concerned. These skills are of course to some extent included in [Module A](#), but a realistic approach should focus on the practical application of these principles in terms of "learning by doing".

**Relation to the rest of the curriculum:**

If in certain LSP-translations DTP files are processed and translated, this unit can lay the basis for students' ability to handle these files and formats.

**Sub-Unit 1.1: General overview**

4 hours

**Contents:**

What is DTP? Which systems exist? What are the basic differences, major operations and facilities of DTP systems? Where can further information about DTP systems be found? (Cf. e.g. [http://www.fask.uni-mainz.de/caf/kurse/dtp/dtp.html](http://www.fask.uni-mainz.de/caf/kurse/dtp/dtp.html))

This overview should give an introduction accompanied by practical exercises. Rather than presenting one system, several systems should be available. Very few selected functions should be presented. The orientation is introductory.

**Sub-Unit 1.2: Hardware Components**

*(scanner, printer, computer, cf. [Module A](#))*

2 hours

**Contents:**

There should be an introduction to the role of hardware components in DTP. The function of a scanner, its practical use and the use of OCR software should be explained. Exercises should allow everyone to use a scanner and OCR software.
As far as printers are concerned, the principles of inkjet printing and laser printing should be explained. Printing software should be introduced. It should be explained what Postscript is and what a printer driver is.

**Sub-Unit 1.3: Core Concepts of DTP**

*14 hours*

**Contents:**
This sub-unit introduces the core concepts of DTP and offers extensive exercises with DTP systems.

This relates to three dimensions: textual, graphical and layout.

- **The textual dimension** relates to the creation of text with the advanced facilities of a DTP system.
- **The graphical dimension** relates to the creation and the manipulation of graphics, vector graphics, bitmaps. Common software for manipulation and creation of graphics has to be introduced and practised intensively.
- **The layout dimension** relates to the combination of text and graphics and to the overall creation of professional printed material.

The core of this sub-unit is an extensive introduction to one of the DTP programmes, e.g. Pagemaker, Framemaker, QuarkXpress, Ventura Publisher. The exercises must lead to the creation and manipulation of professional material.

One sub-area of DTP is the creation of slide presentations with software like Powerpoint and/or Visio. Since these techniques are just about to become a standard in professional presentation, a translator should be able to make use of these tools. This should be included in the core unit of DTP.

**Unit 2: Telecommunications/Internet**

*20 hours (compulsory)*

*Mainly practically orientated*

**Professional requirements addressed by this unit:**
Information and terminology retrieval via Internet is a common part not only of the general translation workflow but in the age of technology as a whole, just like e-mail has become a most common means of communication. A modern translator just cannot do without.

**Overview:**
This unit is meant to introduce the translators to all aspects of telecommunications that are relevant for them.

**Contents:**
This unit refers to technical aspects such as proper handling of tools like ftp, telnet, and proper handling of all technicalities that are related to e-mail, and use of the Internet. There are two sides to using the Internet, the first being technical dealing with the use of search engines and browsers, or the creation of HTML documents. The second should demonstrate
how to make optimal use of resources available on the Web which are relevant for the
translator, i.e. how to handle the Internet properly from an information science point of view.
(The following link provides an example for a presentation of all the mentioned aspects of
telecommunications: http://www.ask.uni-karlsruhe.de/books/inetbuch/all.html)
The technical and content aspects of telecommunications are both subjects of the courses that
are described in the following sub-units.

Prerequisites:
- **Background knowledge**: Basic knowledge about computing, i.e. Module A and basic
text processing.
- **Software**: Internet browser like Netscape or Microsoft Explorer. HTML editors like
AOL-Press, FrontPage Express or Netscape Composer.

Relation to other LETRAC Modules and Units:
As far as other LETRAC-Modules are concerned, the unit Telecommunications/Internet
comprises some aspects which are also dealt with in Module C (e.g. MT-Systems,
terminology or text corpora on the web). Since the treatment of these aspects in this module is
meant to focus on a general overview and not on architecture and underlying concepts of
these systems, there are no real overlaps. On the contrary, students who have already been
presented these systems in the Internet context, have already a certain familiarity with this
subject when it comes to the theoretical aspects of MT, Terminology etc..

Relation to the rest of the curriculum:
The same argument is valid for the introduction of on-line dictionaries and data bases, which
will also be introduced and used in computer-based LSP-translations. In the latter classes the
focus will be put on the practical use of language specific dictionaries and databases, while in
Module B the existence of these resources can be only pointed at, and one or maybe several of
them can be shortly demonstrated.

In addition, it must be stated that the contents of this module are virtually a general
prerequisite for any professional profile in modern life. In a technologically advancing world,
the use of the inter- and intranet, e-mail and all kind of computer applications has become a
standard requirement which anyone simply cannot do without.

Sub-Unit 2.1: Telecommunications, Technical Basis
2 hours

Contents:
This sub-unit provides the principles and the technical basis of telecommunications. It is
purely technically oriented and is meant to be an introduction given by the teacher. No
exercises are included. It comprises the following topics:
- Electronic mail
- Dialogue systems (ftp, telnet)
- Networks (principles and protocols)
- Client-Server Principle
Sub-Unit 2.2: E-mail
4 hours

Contents:
This is an intensive course on all aspects that are relevant for e-mail. E-mail involves three activities:

Sending mail, receiving mail and managing mail.
- **Sending mail:** The major problem with sending e-mail is sending files via e-mail. It includes techniques like attachment, packing, formats, encrypting, distribution lists etc..
- **Receiving mail:** Receiving mail again creates problems for handling sent files: (unpacking, formats, problems with respect to decoding attachments).
- **Managing mail:** (address books, storing mail, distribution lists etc.)

See: [http://www.fask.uni-mainz.de/cafl/kurse/komm/komm-12.html](http://www.fask.uni-mainz.de/cafl/kurse/komm/komm-12.html)

Sub-Unit 2.3: Dialogue Systems
2 hours

Contents:
This sub-unit is about practising telnet (remote access to computer terminals), ftp (remote access to files) and other facilities, such as chat.

See: [http://www.fask.uni-mainz.de/cafl/kurse/komm/komm-13.html](http://www.fask.uni-mainz.de/cafl/kurse/komm/komm-13.html)

Sub-Unit 2.4: Internet - Technicalities
4 hours

Contents:
This module is about the technicalities of the Internet: The use of browsers (introduction to functionalities) handling of HTML and SGML (details might be part of simple programming), the creation of Web Pages, the use of HTML editors.

See: [http://www.fask.uni-mainz.de/soli/html/selfhtml.htm](http://www.fask.uni-mainz.de/soli/html/selfhtml.htm)

Students should get an overview of different browsers and not only see one of them. This leads to a deeper understanding about HTML and the way it is handled by the browser. The same can be said of HTML editors. It may be doubted whether is makes sense to teach HTML scripting to translators, since they will not see the need for it as long as they are able to use the editors. Instead, the structure of a HTML document can be presented via the view page-function of the browser. HTML scripting itself could be part of the optional course in simple programming.
**Sub-Unit 2.5: Internet – Contents**

*4 hours*

**Contents:**
The general structure of the Internet as a distributed information system with its different available kinds of information must be presented. Students must know about handling search engines and develop strategies for acquiring the intended information.

**Structure of the Internet:**
There should be an introduction into the structure of the Internet: URLs, Web sites, WWW, gopher, etc.

**Search strategies:**
An important point is to develop proper search strategies: The use of search engines, search options, search strategies, information harvesting in general, bookmarks (saving, editing, etc.).

‘Creativity’ in information acquisition, specific information for translators should be a topic in this course as should validation and verification of information found on the Internet.

**Sub-Unit 2.6: Internet – Resources for Translators**

*4 hours*

**Contents:**
This sub-unit should be a central one. It should introduce the Web sites that are most important for translators. WWW is a source of information about virtually anything, online dictionaries, glossaries and databases like EURODICAUTOM, Euterpe (accessible via Trados), Celex, Machine Translation Systems: Systran, T1 and others, electronic reference text material, aligned corpora, articles and journals, free software and many other kinds of resources can be consulted. These should be visited and explanations should be given on the benefit offered by these sites and how translators can make use of the resources in their daily work. See: [http://www.fask.uni-mainz.de/inet/translation.html](http://www.fask.uni-mainz.de/inet/translation.html)

Another aspect relevant for translators are mailing lists and newsgroups.

**Unit 3: Advanced Skills**

*20 hours (optional)*

**Focus on theory and practice**

**Professional requirements addresses by this unit:**
Not only in the localisation industry but in any domain of translation, professionals are frequently expected to work autonomously in a highly computational environment. The skills conveyed in this unit will enable future translators to meet these requirements to a even higher extent than the compulsory units of this curriculum.
**Overview:**
Deepening of knowledge and further understanding of the functioning of devices of modern information and communication technologies.

**Contents:**
This unit is an advanced course that goes beyond the basic skills described in the previous two modules. It should be optional. It may contain elements from the above units such as advanced HTML, advanced Web design, advanced DTP, a more in-depth introduction to computer science, programming, e.g. PROLOG for NLP, C, C++, AWK, Perl.
It should also have an introduction to operating systems: Advanced DOS/Windows, MacOS. LINUX/UNIX. (http://www.xnet.com/~blatura/linapps.shtml and (http://gc.bcm.tmc.edu:8088/gc_help/unix.html)

**Prerequisites:**
- Module A
- Units 1 and 2 of this Module
- A general interest in concepts and techniques in information science.

**Relation to other LETRAC Modules and Units:**
This optional unit is designed to be an object of specialisation. Therefore it is closely related to any other module of the model curriculum, since it will require skills acquired in other units or modules, just like a general interest in a deeper understanding of information technology.
On the other hand, contents conveyed in this unit will lead to a better understanding of any of the subjects treated in the Modules B and C, because it makes the underlying processes more transparent.

**Relation to the rest of the curriculum:**
Where ever a computer is in use, someone is needed who is able to handle it. The more someone is specialised in information technology, the more he is able to understand the processes and mechanisms any application is based on. Somebody with specialised skills in IT/Telecommunications will get a better insight to the functioning of any computer application, and will be able to handle all tools and aspects of language technology relevant for translation more easily.
Module C

Language Engineering

6 Units (compulsory and optional ones)
Total: 140 hours
ECTS credit points: 19

Contents:
This is the ‘core module’ of the entire curriculum – in the sense that special skills are taught
which have to do with the processing of linguistic data by the computer.

It contains the following elements:
- 2 compulsory units (one with general tools and one with translation specific tools)
- 4 optional units (two of them covering areas of language formalisation, one on MT
  and one for advanced users of LT)

For each unit, we propose a number of hours, divided into ‘Theory’ and ‘Practical Exercises’.

Method:
The focus in this Module is rather put on the practical side for the 2 compulsory units, and on
the theoretical units for the optional ones. Totally 97 hours are devoted to Theory and 43 to
Practical Exercises.

Unit 1: General aspects and tools

30 hours (compulsory)
Theory: 20 hours
Practical Exercise: 10 hours

Professional requirements addressed by this unit: not specific to the translation environment
but relating to every profession which aims to create and manipulate texts.

Overview:
This module serves as a general introduction to tools and their principal functions which are
absolutely necessary for a future translator (and for every profession which wants to create
and manipulate texts in the global world of technical and scientific communication).

Contents:
Word processing and checkers
Controlled language
Project and document management
Terminology systems
The contents of the Unit are further developed at sub-unit level.
Prerequisites:
General knowledge about word processing is a general prerequisite for this unit. For sub-units 1.2 and 1.4 basic knowledge in linguistics and terminology is needed.

Relations to other LETRAC Modules and Units:
The contents of this unit presupposes the one of Module A.

Relation to the rest of the curriculum:
Practice in word processing could obviously be acquainted in other courses and specially in translation courses. The sub-units devoted to controlled language and terminology systems should be related to contents included in linguistics and general terminology courses.

Sub-Unit 1.1: Word processing and checkers
10 hours
Theory: 5 hours
Practical Exercise: 5 hours

Contents:
This Sub-Unit should concentrate on the special features of word processors which are needed for handling texts in various languages.

1.1.1 Multilingual word processing
The students should learn how to install and use different character sets and keyboard drivers for different languages, especially in the case of non-Latin character sets. They should also have some notion of the principles of character encoding (ASCII, Unicode) and how to handle this in their word processor. They should know how to switch between different languages within their word processor and see the implications of selecting a language e.g. for some paragraphs in their documents.

1.1.2 Checkers (from spelling to style)
Some word processors already offer spell and grammar checkers for different languages. Students should be able to install and use such proofing tools, which exist in abundance for the English language:

- Word Grammar Checker  
  http://www.reed.edu/webdoc/Macintosh/Word/grammar.html
- WinProof Grammar Checker  
  http://www.leconsulting.com/arthurhu/95/01/winproof.htm
- Proof Positive  
  http://tag.sgml.com/01202003.htm
- Grammarian  

but also for French:
and for other languages.

1.1.3 Interaction of word processors with dictionaries

Machine-readable dictionaries (general as well as LSP) often offer interfaces to word processors, allowing lookup of terms from within the word processor without explicitly toggling between different windows. Students should learn to use these interfaces to highlight terms in their text, to look them up in the dictionary and to paste target language equivalents into their document (e.g. PC-Lingua, Langenscheidt Series, http://www.langenscheidt.de/).

Sub-Unit 1.2: Controlled language

Theory: 5 hours

Contents:
This module should give a first insight about the principles of ‘Controlling Language’.

Very didactic definitions are given in the ‘Controlled Language Homepage’
http://www-uilots.let.ruu.nl/www/Controlled-languages/faq.html

A controlled language (CL) is a precisely defined subset of a natural language, on the one hand constrained in its lexicon, grammar and style, and on the other hand possibly extended by domain-specific terminology and grammatical constructions.

The motivation for the use of controlled language is that it makes all aspects of text manipulation (both human and computational) easier. By eliminating sources of ambiguity and prescribing style rules, controlled languages aim for improved readability, understandability, maintainability and easier computational processing, e.g. machine translation. Most internationally operating organisations have to produce both their internal documentation and their product's user manuals in a number of languages. Good technical documentation is an important factor in the overall quality of the organisation's products.

Use of controlled languages can improve the quality of the documentation and make it available earlier. As such, the use of controlled languages can improve the economic competitiveness of the organisation.

The student should gain a general insight into these concepts and some examples of restricted vocabulary and syntax, for example from the AECMA Simplified English.

Sub-Unit 1.3: Project and document management (basics)

Theory: 5 hours

Contents:
This module should give a basic overview of the tasks involved in project management. Students should learn how to develop workflow plans for different working environments and they should be able to design a database structure reflecting all information categories important in managing and supervising translation and interpretation projects. The course should also give a basic introduction into quality management and quality assurance procedures for translation projects. The student should learn the basic principles of designing a project management database using a universal relational database system (e.g. MS Access) and they should learn to use spreadsheet software for project calculation. As a prerequisite for
designing translation projects (especially such projects where Translation Memory Systems and term banks are used in an integrated environment), they should be able to organise reference material, background information and documents on their hard disk or other media, also using archive systems, file compression systems (ZIP, Lharc etc.) and systems for document indexation and retrieval (see also Module A, Data Storage and Manipulation).

**Sub-Unit 1.4: Terminology systems**

**Theory:** 5 hours  
**Practical Exercise:** 5 hours

**Contents:**  
This module should give a structured overview of terminology management systems including those that are available on the market. Students should learn the different approaches like "fixed entry structure" (e.g. TermStar) vs. "freely definable entry structures" (e.g. Multiterm). The course should cover the basic methodological principles of computer-aided terminology work, including discussion of different entry models and data categories as defined in ISO standards. The students should learn to create and manage their own terminological database, which they should use in connection with regular LSP translation courses. Practical PC exercises should introduce students to using the interface between a word processor and a terminology management system. The module should also contain an introduction to terminological resources available online on CD-ROM or in the Internet; the students should become acquainted with strategies for terminology search on the Internet and with the most important Internet addresses in the field of terminology (e.g. Terminology ForumVaasa, [http://www.uwasa.fi/comm/termino](http://www.uwasa.fi/comm/termino); EURODICAUTOM, [http://www2.echo.lu/edic](http://www2.echo.lu/edic)).

**Unit 2: Translation-specific aspects and tools**

30 hours (compulsory)  
**Theory:** 13 hours  
**Practical Exercise:** 17 hours

**Professional requirements addressed by this unit:**  
This unit aims at enabling students to handle translation relevant software tools and specially translation memories.

**Overview:**  
Translation memories are becoming more and more important for both freelancers and employed translators. Furthermore several big institutions use integrated systems where translators have at their disposal several tools. The aim of this unit is to make students familiar with the basic functionalities of such integrated systems.

**Contents:**  
Translation memories  
Machine translation systems  
The contents of the Unit are further developed at sub-unit level.
Prerequisites:
Students should have fluency in word processing and text analysis, and should have basic translation skills.

Relations to other LETRAC Modules and Units:
The contents of this unit will be linked to the respective description of Module A, mainly Data Manipulation.

Relation to the rest of the curriculum:
The contents in this unit are specially related to LSP-translation and terminology courses.

Sub-Unit 2.1: Translation memories

Theory: 5 hours
Practical Exercise: 5 hours

Contents:
Translation memories are one of the most important tools today in the commercial field of translation. Especially where texts have a high degree of repetition (change of model, machine series etc.), they can save a substantial amount of time and effort for individual translators and big companies. It is important to understand their basic functions, the implications for the organisation of the document chain and their influence on translation strategies and techniques. This course should also provide a basic knowledge of text analysis procedures for evaluating the possible efficiency of translation memories for the translation of a given text. Students should also get a basic insight into problems, methods (criteria) and tools for aligning existing source language texts and their target language translations. A very important feature of translation memory systems is the "fuzzy-matching"-algorithm for the recognition of similar sentences in texts. Advanced students can also become involved in discussions about improving retrieval performance of these systems by creating a knowledge base of linguistic information for segmenting and identifying translation units.

Translation memories are not always simple to handle; a certain amount of practical training has to be done (project creation, data organisation ...) before they can be used with the effects mentioned. Once the student has acquired the basic skills, a considerable part of this practical work with TM can be done in regular LSP translation classes. Teachers in these classes also have to pay attention to the fact that using translation memory systems tends to emphasise a string-by-string or sentence-by-sentence translation strategy. This, of course, also means that traditional translation teachers themselves have to become aware of this problem.

Sub-Unit 2.2: Machine translation systems: working with TM and MT

Theory: 8 hours
Practical Exercise: 12 hours

Contents:
It is common knowledge today that the use of ‘real’ machine translation systems is meaningful only under certain conditions. Several big institutions and companies today use integrated systems where the translator is able to use several tools (as do other employees who need only a quick translation for information purposes). This module contains examples
where an organisational effort has led to a ‘physical integration’ of different systems (terminology, memories, machine translation): the biggest effort in terms of time and money has probably been invested in the EURAMIS system of the European Commission where most employees have access to the Commission’s own SYSTRAN system, the TRADOS translation memory and terminology handling system and several other tools and databases in parallel. This integration is likely to grow with the introduction of the EPTAS tools (also called UNITOOLS, some details of which can be found at http://www.epp.de/), where all tools are controlled and accessed by a single screen.

Another success story of MT is the SAP case where the translation department started early to buy, adapt and use commercial MT software.

All students looking for employment with international firms or organisations should have studied these cases.

In this unit, there is also an introduction to the use of integrated systems which are marketed either by MT or TM vendors. The translator can adjust parameters (when MT or TM is to be used), and store post-edited MT translations in the memory.

If possible, practical demonstrations should be offered showing, on the one hand, memory systems like Transit or Trados Workbench integrated with LOGOS machine translation, or on the other hand, machine-translation systems like T1 or PersonalTranslator plus integrated translation memory. Practicability and user-friendliness of such a system integration should be discussed.

In most cases translation memories come with a terminology management system, and it is important to know how the interaction between the two components is organised (colours, cut and paste, automatic replacements). A basic function of such integrated systems is the automatic terminology recognition feature, which is able to identify inflected word forms using special (sometimes linguistically based) fuzzy matching algorithms. Practical work with such integrated systems can - after an introductory stage - be included in regular LSP translation classes.

A special translation task where these systems play a prominent role is software localisation, i.e. the translation of user interfaces, help screens and other textual material which comes together with software. On the one hand, there is a high degree of repetition; on the other, consistent terminology is of crucial importance in this text type. An overview of basic principles and methods of software localisation should also emphasise the fact that different text types are involved in software localisation: printed manuals, on-line help, menus, dialogue boxes, error messages, etc. This can also be used as a starting point for discussing problems of different functions of texts with the same contents and for a discussion of translation strategies.

Special tools have been created for the translation of software material which extract messages from programs or executable files. In some cases, they even offer the possibility of displaying and adapting dialogue boxes to the different sizes of a message in different languages. As software localisation is a fast-growing business, the student should at least have basic insights into the possibilities of these tools and how to use them: e.g.: XL8, CATALYST, PASSOLO (http://catalyst.corel.ie or http://www.pass-engineering.com/).
Unit 3: Language formalisation

20 hours (optional)
Theory: 12 hours
Practical Exercise: 8 hours

Professional requirements addressed by this unit:
The unit is addressed to those students that are interested in the processing of natural language, either because they are willing to work on NLP/MT projects, or because they are interested in understanding the specificities of the processing of natural language.

Overview:
The use of computational tools using natural language is increasing not only in the field of translation. This unit is intended to provide the students with the basic concepts underlying the computational treatment of language.

Contents:
Structural and functional approaches
Grammars and parsers
Unification-based approaches to grammar

The contents of the Unit are further developed at sub-unit level.

Prerequisites:
Basic linguistic courses.

Relations to other LETRAC Modules and Units:
The contents of this Unit will be linked to the respective description of Module A, mainly Data Manipulation.

Relation to the rest of the curriculum:
The contents of this unit are specially related to linguistics.

Sub-Unit 3.1: Structural and functional approaches
Theory: 3 hours

Contents:
A small module is devoted to the structural and functional approaches to syntax (i.e. sentence structure) and morphology (i.e. word structure), which are the two major approaches in current formal and computational linguistics.
Sub-Unit 3.2: Grammars and parsers
Theory: 3 hours

Contents:
Students are introduced to the distinction between the linguistic knowledge represented in grammars and the processing strategies being adopted. The different types of grammars and the basic parsing techniques are presented and discussed.

Sub-Unit 3.3: Unification-based approaches to grammar
Theory: 6 hours
Practical Exercise: 8 hours

Contents:
The core of this module consists in presenting the unification-based approaches to grammar, which are the state-of-the-art approaches to grammar in computational linguistics. The basic features of such approaches are presented: lexicalism and surface-orientedness. For clarification, different linguistic phenomena (both monolingual and bi-/multi-lingual) are presented within such approaches. Students acquire not only the linguistic knowledge explicitly described but also the formal framework in which these formalisms are shaped (features, types, inheritance...), which is a common core in many formal and computational applications using natural language. Finally some implementations of unification-based formalisms are presented (DCG, PATR, Alep, CUF, ALE...); at least one of them is examined in detail.

Unit 4: Machine Translation
Theory: 20 hours (optional)

Professional requirements addressed by this unit:
To be able to face the necessity of evaluating and choosing an appropriate translation system in a given specific user situation.

Overview:
In order to analyse MT critically and realistically, the student should acquire basic notions about history and basic concepts of machine translation. The same holds for the principles of translation evaluation where different approaches have been presented over the years.

Contents:
Historical overview
Basic concepts
Evaluation
Unification-based MT systems
EU-MT and NLP projects
The contents of the Unit are further developed at sub-unit level.
Prerequisites:
Knowledge on language formalisation as well as fluency to face currently translation divergences and techniques to solve them.

Relations to other LETRAC Modules and Units:
The contents of this Unit will be linked to the respective description of Module A, mainly Data Manipulation and Data Output.

Relation to the rest of the curriculum:
Some contents of this Unit will be linked to general translation courses as well as to linguistic courses.

Sub-Unit 4.1: Historical overview
Theory: 4 hours

Contents:
Students should have an overview of the development of MT ideas since its beginning, the reasons for failure and partial success of some systems, and the political and commercial reasons behind this history. They should know about the following periods: Origin and first ideas of MT (Weaver, etc.), 5 decades of ups and downs (Georgetown, euphoria, ALPAC, basic research, revival (EU/Japan), realism and commercialisation).

There are many books on this subject. One of the most didactic chapters comes from Doug Arnold in his book MACHINE TRANSLATION: An Introductory Guide (http://clwww.essex.ac.uk/~doug/book/book.html).

Sub-Unit 4.2: Basic concepts
Theory: 4 hours

Contents:
The student should have a knowledge of the most important MT concepts and should be able to classify systems according to these concepts (direct/indirect methods, analysis, synthesis, interlingua, transfer, dictionaries, programming, procedural/declarative systems, modularity, formalisms, statistical approaches). He should understand the names and functions of the major commercially available systems and research systems.

The same book as in the preceding course is recommended.

A short text concerning the basic concepts can be found at http://www.lsadc.org/Kay.html.

Sub-Unit 4.3: Evaluation
Theory: 6 hours

Contents:
With the increasing number of systems on the market, decision-makers are confronted with the problem of choosing the system (or even a package of system components, i.e. MT & TM) which is appropriate for their specific user situation. Standards against which MT systems could be measured are not yet available or are reduced to mere recommendations.
The well-known best-buy overviews apply to abstract user types, i.e. the prototypical MT stakeholders like translators, information consumers, managers, researchers, etc., rather than to specific user requirements. Last but not least, MT evaluation is difficult since there is no one ‘right’ translation. Therefore, conventional benchmarking is quite useless. Approaches implemented by the EAGLES initiative (http://www.cst.ku.dk/projects/eagles2.html) have for these reasons excluded MT evaluation up to now, although there are plans to develop a quality model for MT. Work on extending and adapting the ISO 9126 standard for software quality is in progress. The central idea of this extension is that user requirements monitor the evaluation of any software in use.

Additionally, the definition of quality metrics and the evaluation process itself are now subsumed under different ISO headings, the latter being part of the 14000 series, which is also related to product quality evaluation.

An introduction to MT evaluation as part of the MT sub-unit should cover the general framework of EAGLES since it is the European focal point of NLP evaluation standards. However, more practical work on MT evaluation has been carried out under the umbrella of the DARPA (Defence Advanced Research Projects Agency) MT initiative. At present, work concentrates on a so-called translation proficiency scale (TPS). Here, MT user requirements are related to the concrete purpose of MT in terms of text-handling tasks, ranging from publishing to mere information filtering, where MT systems are then measured according to their ability to meet these needs. The application of the TPS could serve as a basis for evaluation exercises where students are trained within a true (perhaps their future) user environment, simulating a number of scenarios for MT ‘in use’.

**Sub-Unit 4.4: Unification-based MT systems**

*Theory: 4 hours*

**Contents:**

A short special section is devoted to unification-based MT systems which are considered to be the most modern approach in computational linguistics but which have not (yet) reached a commercial exploitation stage. Starting with the EUROTRA programme of the European Commission, a series of research projects have been analysing the pros and cons of this approach. During the different phases of EC promotion, these projects have changed from a basic research orientation to a more application-oriented perspective. Augmenting computer power gives hope that such systems can come nearer to the performance of their older MT brothers (EUROTRA, LFG, CAT2, ALEP, - one of them in detail).

**Sub-Unit 4.5: EU-MT and NLP projects**

*Theory: 2 hours*

**Contents:**

The Research Framework Programmes of the EU contain small sections called ‘Language technology’ within Action Lines like ‘Multimedia Contents’ (5th Framework Programme). It is worthwhile considering a selection of EU projects in this field, especially because these activities are likely to go on in the future and can serve as additional help in international and multilingual environments which tend to grow in the context of globalisation. These projects belong(ed) to the research programmes called subsequently ‘Linguistic Research and Engineering’, ‘Language Engineering’, ‘Human Language Technology’.
The students should have an overview on the working programmes, the most important and successful projects (http://www2.echo.lu/langeng/en/) and the general information about the functioning of such projects. In their future professional work, they may face the opportunity or the necessity to lead or organise the participation of their employer in such a project.

**Unit 5: Corpus linguistics and low-level analysis**

*20 hours (optional)*

**Theory:** 12 hours  
**Practical Exercise:** 8 hours

*Professional requirements addressed by this unit:*

The ability to use any kind of software for information extraction.

**Overview:**

Within the field of applied linguistics, corpora have been an essential tool in lexicography work (see COBUILD Dictionary), but they are becoming an indispensable tool for translators as well. A translator may use a corpus to have real-life examples of the use of words in order to choose the adequate translation, or to obtain a statistical analysis of the most frequently used word forms or sentence structures in certain types of texts. Furthermore, multilingual corpora with strictly parallel texts (originals and translations) can be easily converted into translation memories.

Translators should be able to extract linguistic information from corpora (e.g. terminology extraction) but also eventually manipulate/create ad hoc corpora for their own purposes. Thus, this module should offer the student an overview of corpus exploitation, as well as corpus development.

**Contents:**

Possible topics are (taken from http://www.ccl.umist.ac.uk/teaching/modules/3000/3009.htm):

- Why corpora are necessary; user needs
- Review of relevant statistical techniques
- Corpus design; criteria for composition; text typology
- Representation, annotation and analysis of corpora
- Tools and techniques for corpus processing
- Results of corpus processing; their exploitation
- Corpus-based NLP and NLP system design methodologies
- Probabilistic and statistically based NLP
- Sublanguage-based NLP

A very complete collection of concepts and their definition is given by the Systematic Dictionary of Corpus Linguistics: http://donelaitis.vdu.lt/publikacijos/SDoCL1.htm#CL1.

The contents of the Unit are further developed at sub-unit level.
Prerequisites:
Basic linguistic knowledge

Relations to other LETRAC Modules and Units:
The contents of this unit will be linked to the description of Module A, mainly Data Storage.

Relation to the rest of the curriculum:
Practice in information extraction could be useful in much more other courses as translation courses as well as in general

Sub-Unit 5.1: Corpus design and organisation
Theory: 2 hours

Contents:
Corpus design is obviously constrained by the aim it is created for (learners, terminology extraction in a certain field, contrastive linguistic studies, lexicography...). An overview of some of the most important available general corpora for different languages but also of some multilingual corpora should inform this module in order to make students acquainted with different design principles.
Sampler of corpus projects: ACL/DCI, TEI, NERC, BNC, Treebanks, EAGLES, ELRA.

Sub-Unit 5.2: Corpus annotation
Theory: 4 hours

Contents:
It should be shown that the more richly tagged corpora are, the more fine-grained enquiries can be made. Structural annotation for sentences, paragraphs, titles, etc., as well as morphology tagging and lemmatisation constitute a minimal annotation for corpora. Furthermore, corpora are being annotated with syntactic and semantic tags.

Sub-Unit 5.3: Low-level parsing
Theory: 4 hours
Practical Exercise: 4 hours

Contents:
In order to handle corpora and their annotations, students should understand low-level parsing techniques (partial parsing, regular expressions, constraint grammars...). Corpus-based NLP and NLP system design methodologies should be presented so that the students become acquainted with probabilistic and statistically based NLP and sublanguage-based NLP.
Sub-Unit 5.4: Information extraction from corpora

Theory: 2 hours  
Practical Exercise: 4 hours

Contents:
Students should become familiar with KWIC programs and other software for information extraction. Exercises on extraction of linguistic information could be done in general- and LSP translation courses.

Unit 6: Case studies

20 hours (optional)

Overview:
Some students will specialise in the advanced facilities offered by the new technologies. This is especially the case if they expect employment not only as translators but also as technical writers who produce technical documentation from the source. For management tasks, this module is of high priority.

Professional requirements addressed by this unit:
The ability to produce multilingual technical documentation and to face any kind of management tasks.

Contents:
Project, document and terminology management  
Controlled language in industrial environments  
Complete Localisation project
The contents of the Unit are further developed at sub-unit level.

Prerequisites:
Knowledge on information extraction software, terminology data-bases, translation memories and any kind of translation specific tools.

Relations to other LETRAC Modules and Units:
The contents of this Unit will be linked to the content of Module B, mainly Advanced Skills.

Relation to the rest of the curriculum:
This Unit relates mainly to translation and general terminology courses.
Sub-Unit 6.1: Project, document and terminology management

Theory: 8 hours

Contents:
This specialisation can be provided, e.g. in project seminars, where students can start with defining their own project management databases for different workflow environments. They will have to design workflow plans for in-house translation in team work, outsourcing projects and tele-working. Among the different project tasks are maintenance of terminological and translation memory databases during a project, including regular updating for all translators involved in the project.

Such project seminars can also include the integration of various software tools in a single working environment (for advanced students the development of interfaces between different applications (API) might even be possible).

Teamwork in such projects also requires tasks like terminology management, including terminology interchange. Interchange formats like MARTIF or SGML have to be used for data export and import.

In projects dealing with systematic terminology work in special fields, the application of software for terminology extraction from texts (e.g. SystemQuirk) can be considered.

A complete translation project, including the processing of graphics, pictures (e.g. screen shots), also contains practical exercises in maintaining layout information in MT and TM systems as well as the treatment of various DTP or word processor formats in translation memory systems.

Sub-Unit 6.2: Controlled language in industrial environments

Theory: 6 hours

Contents:
In this sub-unit, the work of the basic sub-unit 1.2. ‘Controlled language’ is continued. The student should perform several case studies where controlled language is used or seriously considered, where tools for automatic checking of the conformity are used and where first experiences are reported (AECMA http://www.aecma.org/ietplex.htm, CASL ...). Students should understand the interdependence of SGML, DTDs and terminology. They should also be acquainted with automatic checking tools (see again http://www-uilots.let.ruu.nl/www/Controlled-languages/faq.html). For LANT see http://www.lant.be/CL.htm, which also contains a detailed list of examples from Simplified English and their possible correspondents in ‘normal’ English. Other possible topics are ALETH-CL (from ERLI) and the recent MULTIDOC project of the European automotive industry (http://www.iai.uni-sb.de/multidoc/home.html and http://www.iai.uni-sb.de/en/multidocs.html). One of the best summaries of approaches and problems can be found at http://www.herc.ed.ac.uk/AnnualReport96/Text/5.4-www.html.
Sub-Unit 6.3: Complete localisation project (e.g. one source language, several target languages)
Theory: 6 hours

Contents:
A complete software localisation project can depart from software in one source language and create localised versions in several target languages, thus also providing the possibility for cooperation beyond language departments. In such a project, different text types involved in software localisation should be treated: menus, dialogue boxes, online help, error messages, printed documentation (also online documentation using hypertext). Different translation strategies should be discussed for the different text types. In order to maintain consistency in terminology and style, terminological glossaries and translation memories should be used, both those belonging to translation memory systems like Trados, Workbench/Transit and those coming with specialised software localisation tools. Possibilities of data interchange between the systems can be analysed (interchange formats like MARTIF or TMX, link to: http://www.lisa.unige.ch), so that students are able to reuse the translations in a different machine assisted translation system.