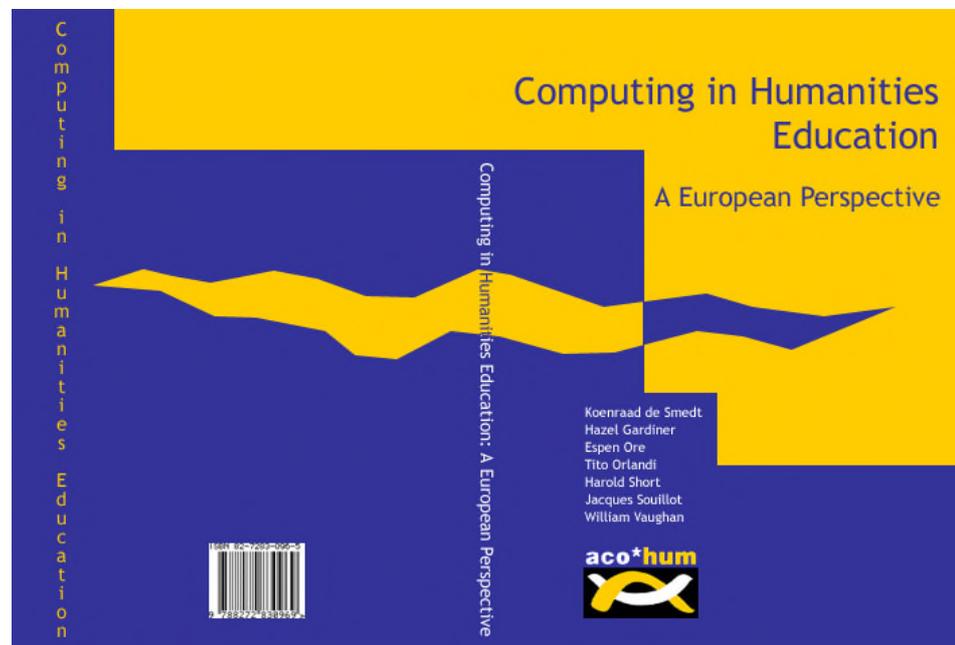


## DIGITAL PUBLICATION

Ore, Espen S, Harold Short, Donald Broady, Anthea Ballam, Lou Burnard, Elisabeth Burr, Stuart Lee, Lisa Lena Opas & Thomas Rommel (1999), "Chapter 3. European studies on textual scholarship and humanities computing," pp. 63–88 in: Koenraad de Smedt, et al., *Computing in Humanities Education. A European Perspective. SOCRATES/ERASMUS thematic network project on Advanced Computing in the Humanities*. Bergen: University of Bergen

Downloaded 6 Oct. 2015 from <http://www.hd.uib.no/AcoHum/book/>





## Chapter 3

# European studies on textual scholarship and humanities computing

**Espen S. Ore**

University of Bergen

**Harold Short**

King's College, London

**Anthea Ballam**

King's College, London

**Donald Broady**

Royal Institute of Technology/Nada, Stockholm

**Lou Burnard**

Oxford University

**Elisabeth Burr**

University of Duisburg

**Stuart Lee**

Oxford University

**Lisa Lena Opas**

University of Joensuu

**Thomas Rommel**

University of Tübingen

---

## 3.1 Introduction

### 3.1.1 History and background

The origins of computing techniques applied to textual scholarship lie with Father Roberto Busa SJ, who set out in the late 1940s to explore the new possibilities offered by computational machines, and who in due course created concordances and large lexical and morphological databases of the works of St. Thomas Aquinas.

Although his pioneering work aroused some interest during the following decade, it was in the later 1960s and early 1970s that a recognizably European dimension to the application of computing in the textual disciplines began to emerge, with significant work being done in a number of universities across Europe, notably in Italy (e.g. Pisa and Rome), Germany (e.g. Tübingen), France, the Scandinavian countries and the UK (Cambridge, Oxford and King's College London).

This new momentum and the developing sense of a community of scholars interested in exploiting the new techniques was marked by the founding of the [Association for Literary and Linguistic Computing](#) in

1973. However, it would be true to say that in this period the main emphasis in the application of computing was on research rather than teaching. Starting in the 1970s, Europe saw the gradual concentration of this research in institutional and national research centres, e.g. by the establishment of a national Norwegian Centre for Computing in the Humanities at Bergen in 1972.

During the 1980s and the early 1990s, the possibilities of the new technologies became more widely recognised, particularly when the significance of the personal computer revolution started to become apparent. It was in this period that the potential for computing to transform teaching and learning as well as research in the text-based disciplines began to receive attention, and courses and programmes formally involving the application of computing were initiated in a number of European institutions (as discussed in more detail in Section 3.3 below). In some countries this was accompanied by national initiatives, such as the [Computers in Teaching Initiative](#) in the UK. Although this initiative was broadly based, it involved the setting up of two centres relevant to the text-based disciplines - the CTI Centre for Textual Studies at Oxford, and the CTI Centre for Modern Languages at Hull.

The role of computing in learning and teaching is now widely understood, not only in terms of producing new generations of researchers and university teachers trained in the application of formal methods and able to apply the new technologies to their own work, but also in ensuring a substantial majority of European citizens and workers who understand computing, who are technically adept as they enter the work force, and who are equipped to contribute to political and cultural life of their society within a wider international context that is increasingly dominated by technology.

### 3.1.2 Computational techniques and tools

From the beginning, one of the most obvious applications of computing in textual disciplines was the creation of concordances. This technique remains important of course, but is relatively straightforward, and different types of concordance can be generated by a number of readily available software tools. However, the work of Father Busa and others ensured that from the outset more challenging work was undertaken in lexicography and semantics, even though the early tools for such work would now be considered primitive.

One natural development from the early work has been the increasingly sophisticated application of statistical methods. These are the cornerstone of stylometric studies in general, and of authorship attribution research in particular. A number of statistical packages are used, all developed for more overtly 'statistical' areas of activity. Another example of techniques and software developed for other disciplines playing a role in literary studies is the use of cladistic analysis, from the biological sciences, in tracing the 'genealogies' of manuscripts.

Text mark-up is another major area of activity which grew out of the early work. Mark-up can be used for something as basic as recording the structure and appearance of a source text, but is now widely used for more sophisticated analytical approaches, not only by encoding parts of speech to assist study of an author's use of language, but by making semantic or other encodings to enable more overtly interpretative analyses. The work of the [Text Encoding Initiative \(TEI\)](#) - funded in part by the European Commission - has played a significant role in the development of this work, ensuring that efforts that go into mark-up can be re-used by other scholars and preserved for the future. By adopting SGML as the basis for its recommendations, the TEI ensured that scholars would be able to take advantage of software tools developed for the commercial world.

With the development of XML as a new text encoding standard, the significance of the TEI for textual scholars is likely to increase further. It is also true that the analytical needs of scholars are by no means identical to the commercial objectives of publishers and others, so that there is likely always to be a need for the development of software that is specific to research and teaching. At the same time, there is likely

in turn to be potential for the wider application of such software and analytical methods in the commercial world in due course.

The dramatic reductions in digital storage costs has made it possible for new approaches to be followed in scholarly editing. It is now feasible to include transcriptions of all known sources for a particular work, along with images of the source manuscripts. Using mark-up, the editors can highlight the variations between sources.

Low storage costs have also made possible the increasing production of large digital resources. Dictionaries, mono-lingual as well as multi-lingual, were among the first of this type. However, increasingly large *corpora* of works are being produced, so that it is now realistic to expect to find 'all of Shakespeare', or 'all of 19th century French fiction', or 'all of Classical Greek literature', and so on. This in turn is enabling new kinds of scholarship.

This phenomenon also makes it possible for large multi-media resources to be created, so that image and sound material of all kinds can be stored alongside texts. This is serving to break down some of the barriers between disciplines, and between types of discipline - e.g. between the literary scholar and the historian or the art historian, so that tools developed for either of the latter groups becomes grist to the mill of the former.

The proliferation of the Internet and the associated increase in bandwidth has made digital resources of all kinds, whether small or large, more and more widely available. The internet explosion seems set to continue as it becomes driven increasingly by the commercial sector, but it seems clear that XML will play a significant role in this commercial development, which should enable the text-based disciplines to take continuing advantage.

### 3.1.3 The impact of computing on textual scholarship

All of these developments are having a dramatic effect on the way scholars view their work, and on the potential modes of research and teaching.

Increasingly, the essential reference works needed by researchers and students are available in electronic form (and often on-line). However, when these works are owned by commercial publishers, the electronic versions often come at a considerably higher price than printed books.

The possibilities for advanced study are enormous. Large corpora make it possible to carry out new kinds of analyses that take into account all the works of an author or all the works of a particular location, period or genre. Moreover, the integration of different types of materials - e.g. texts and images - enlarge the scope of interest of the textual scholar and encourage broader disciplinary views; this also encourages more inter-disciplinary approaches.

As an example of a project using computer methods in edition philology, we mention [The Wittgenstein Archives](#) at the University of Bergen. The aim of this project has been the transcription of Wittgenstein's complete literary estate (*Nachlaß*) into machine-readable form, the development of software for the presentation and analysis of the texts, the provision of access to the machine-readable transcriptions for visitors and scholars at the University of Bergen, and the publication of an electronic facsimile and machine-readable transcriptions of Wittgenstein's *Nachlaß* on CD-ROM. The work has been completed in 1999 and the CD-ROM volumes are being published in cooperation with Oxford University Press.

Wittgenstein's *Nachlaß* presented numerous problems for publication. Since Wittgenstein himself never prepared more than a negligible fraction of his writings for publication, most of his manuscripts and typescripts are full of various annotations, deletions, insertions, marginal remarks, critical instructions and

cross-references, alternative formulations for particular phrases, and even writings in secret code (*Geheimschrift*). Neither is it always clear which of such alternative formulations he finally decided upon. In order to reproduce the texts as completely as possible, The Wittgenstein Archives have developed their own text coding system MECS (Multi-Element Code System), which provides the basis for specially designed software that offers wide flexibility in the presentation and analysis of the texts.

The increasing sophistication of the computing techniques and tools makes essential new kinds of collaboration between textual scholars and technical experts, including computer scientists and engineers, and encompassing both the theoretical and applied domains. Also, the boundaries between 'research' and 'teaching', and especially between the materials used in each, are narrowing. Increasingly it is possible for students to work with source materials in digitized forms, and to use the same reference works as researchers.

As a consequence, the role of academic support staff and the nature of their relationship with researchers, students and teachers is changing. The developments in the services offered by 'the digital library' or 'the hybrid library' are now as significant for textual scholars as for any other group. Also the relationship between scholars and publishers is changing dramatically, affecting for example the selection and preparation of teaching materials.

The relationship between teachers and students and the wider 'cultural heritage' sector is also undergoing a profound change. The availability for study of cultural objects, wherever they reside in their original form, is changing learning and teaching in universities, and in return the work of students and researchers is serving to enrich the understanding and wider access to those objects.

### **3.1.4 Implications and questions for teaching and learning**

The increasing importance of information technology in the text-based disciplines, no less than in the society at large, raises a number of questions and has some important consequences for teaching and learning in Europe's higher education institutions.

First, teachers need themselves to understand the technologies and how they can be applied in their disciplines. This in turn means they must receive the necessary training. If they have not themselves had the opportunity for this during their own undergraduate studies then basic training is needed, but in the context of rapid changes, which are likely to continue for the foreseeable future, opportunities are needed for regular up-date information and training. This poses the question of how this training is to be provided.

At the same time, teachers need support from technical specialists, so they can concentrate on the discipline aspects rather than on the technologies. This poses the questions of what technical depth is appropriate for the textual scholars, and what should be provided by means of institutional or other support.

There are a number of different and successful models for how textual scholars can best be provided with specialist support. The key questions here are how the institutions who already have structures in place can best develop them, and how others can best put support structures in place.

Students need to acquire basic computing proficiency. This should be gained before or outside their textual studies. This raises the question of how institutions should provide it, at least until they can expect that every student will arrive from secondary school with the necessary skills. There are a number of models for the inclusion of computing components in textual studies curricula, with two important types of model, broadly speaking. In the first of these, the interdisciplinary aspects of the appropriate formal methods and techniques form the basis of components that are taught to students in a number of

disciplines; in the second, those computing techniques and tools that are held to be important for a specific discipline are included in courses offered in that discipline. Both approaches have their particular advantages, and in many cases they are related to institutional history and culture. However, in both cases some important questions arise:

- Who should teach the computing components?
- To what extent should the students be taught the *application* of the appropriate techniques, and to what depth should they be expected to understand the *underlying* computing or other technical *principles* involved?
- To what extent should the acquisition of *transferable skills* determine the scope and scale of such courses and programmes?

Furthermore, the Internet explosion offers particular opportunities for students who are used to dealing with multimedia resources, mixing text and images (and sounds), and who are proficient in the creation and manipulation of such resources. This is an area of activity that raises the 'transferable skills' issue in a particularly acute form.

A key aspect of computing in textual studies (as in other disciplines) is the need to develop students' analytical skills, which in turn make them particularly attractive in the wider labour market. This raises issues of how to ensure that prospective employers understand that the students have more to offer than merely the ability to use, for example, a database or a spreadsheet package.

The availability of user-friendly statistical packages make it possible to consider the inclusion of statistical methods in all humanities education, including the textual disciplines. One question raised by this idea is whether the general benefit to society of having truly numerate citizens is outweighed by the difficulties of applying measurements to humanities data in an appropriate way and the possible resistance to the idea on the part of humanities students.

### 3.1.5 The current situation

The remainder of this chapter attempts to provide a general analysis of the current situation in at least some European countries and attempts to arrive at some conclusions and recommendations. The discussions which provided the ground work for this chapter took mainly place in the context of ACO\*HUM, the SOCRATES thematic network project on [Advanced Computing in the Humanities](#), through working group meetings and participation at conferences.

The current opportunities for teachers interested in the subject include information, on paper and on websites, from those institutions where computing is included in some way in the textual disciplines, as well as gateway sites, such as HUMBUL, based at the University of Oxford. The two main professional associations in this area - the Association for Literary and Linguistic Computing (ALLC) and the Association for Computers and the Humanities (ACH) - also provide support. The ALLC, for example, has a programme of workshops and seminars aimed at training teachers and researchers. It also has a formal involvement in ACO\*HUM.

Information exchange is facilitated by journals, including those of the two associations and a number of more specialised publications. There are also a number of conferences at which teaching and curriculum issues are covered, including the joint annual conferences of ALLC/ACH and the annual Digital Resources for the Humanities conferences. There are a number of other conferences that are relevant, including those of ELRA, the European Language Resources Association.

However, it will be clear that the overview below is limited, and that more systematic action is needed towards a comprehensive mechanism for collection and dissemination of information, if European

countries are to exploit the full potential of curriculum development in these areas.

## 3.2 The computing environment for text-oriented disciplines

### 3.2.1 Introduction

The growth and development of facilities for the support and teaching of humanities computing in Europe have been evolutionary and the facilities have been established and organised on an ad hoc basis. The lack of uniformity of departments and centres would indicate that the subject is still undergoing development, being viewed (according to the institution involved) as anything from an important function to an academic sideline. In this section we will first take a look at the functions of textual scholarship and humanities computing facilities and the services they provide, and then an overview will be provided of the organisational position of such facilities at selected institutions. In this section we will not define the term *textual scholarship* either by itself or in the context of Humanities computing. Instead this will be discussed in section 3.3.

The facilities described in this section represent a small sampling of existing institution and the examples are not selected on a purely random basis for statistical use. Instead they are selected from among institutions known to the Working Group members. But as it was pointed out earlier the institutions described in this section show widely different approaches to the selection of services made available for Textual Scholarship and Humanities Computing (TS & HC).

### 3.2.2 Functions and services expected or needed in TS & HC

Research and teaching in TS & HC need services and facilities at different levels. Some of these services and facilities are not explicitly linked with TS & HC but rather with the use of information technology in general. At some level, however, the needs become specific for TS & HC although it is impossible to define exactly at which level this is. For all institutions/universities we expect that there is some basic information technology infrastructure such as networking and possibly network servers and printers.

The use of computer tools clearly requires training. There is no reason for specialists in TS & HC to train novices in basic computer literacy although this may in fact be done at certain institutions. This may reflect local practices but is not related to TS & HC. On the other hand there is a gradual scale from basic computer literacy to training in special and/or advanced software and hardware to training in computational methods in specific TS & HC related disciplines. Some of this training falls under the heading of the next section, "Computing in textual Scholarship Courses" while some of this training will take place on an ad hoc basis where it is needed and not as organised courses.

Apart from consultancy and advice on a variety of low to intermediate levels specialist support for users with a high level of competence is needed. This helps to focus and channel resources and create an awareness of existing tools, methods and techniques. Re-inventing the wheel should be avoided, and specialist support as provided by TS & HC is indispensable in most areas of interdisciplinary research.

Another central aspect of work in TS & HC is the synergetic effect of sharing knowledge and resources. Interdisciplinary in nature, TS & HC brings together experts from different fields that share fundamental methodologies or technologies.

One of the fundamental requirements of TS & HC is the electronic text. In most cases, support facilities for TS & HC function as repositories of data; they either hold electronic (textual) data themselves, or provide access to reliable/verified resources. By storing and providing data these centres play a vital role in the dissemination process of raw data and electronic tools for processing. Repositories often function

as centres for the teaching of advanced methodologies, providing cutting-edge tools, methods and techniques.

In addition, collecting, maintaining, and using textual data provides one of the areas that allow academia and industry (eg academic repositories and commercial publishers) to merge know-how and resources. This should prove mutually beneficial.

### 3.2.3 Organisational models

Support facilities for TS & HC can be (and are) organised in many different ways. One possible categorisation is:

- Institutions which provide no dedicated support for TS & HC
- Institutions providing dedicated support for TS & HC at a general computing centre
- Institutions providing dedicated support for TS & HC at a general Humanities computing centre
- Institutions with dedicated departments or offices for TS & HC

The institutions and departments/offices which have been studied so far do not show a meaningful separation between the last two categories, so instead the following classification in three groups is offered:

1. ***Institutions which provide no dedicated support for TS & HC***
  - Royal Institute of Technology (RIT), Stockholm, Sweden
  - University of Duisburg, Germany
  - University of Joensuu, Finland
2. ***Institutions providing dedicated support for TS & HC at a general computing centre***
  - Sheffield University: CICS, UK/England
  - University of Oslo: USIT, Norway
3. ***Institutions providing dedicated support for TS & HC at a general Humanities computing centre or dedicated departments or offices for TS & HC***
  - Austrian Academy of Sciences, Vienna, Austria
  - Exeter University: Pallas, UK/England
  - Glasgow University: HATII, UK/Scotland
  - King's College, London: CCH, UK/England
  - Leeds University: ACOM/LETC, UK/England
  - De Montfort: CTA, UK/England
  - University of Bergen, Norway (Humanities Information Technologies programme and Humanistic Informatics department)
  - University of Oxford: CHC, UK/England
  - University of Tübingen: LDDV/ZDV, Germany

According to this classification, most universities have a centre for Humanities computing and/or Textual scholarship. But this is rather the result of skewed sampling. If we look at only the institutions with members in the ACO\*HUM working group, we find that among those represented in the table three of the institutions belong to column 1 and four in column 3. Since the group members are active, also on an international level, in TS & HC, we presume that the situation in Europe in general is worse when it comes to institutional support.

### 3.2.4 Conclusions

As a preliminary conclusion, we first remark that a broader European study of the situation at other institutions will be needed to give us more hard data and to let us know more surely what the situation is. Countries which have not very actively been involved in the ACO\*HUM network, as well as the Eastern

European countries which so far have been excluded, should be part of any wider study. Based on such a study, a set of case studies could then be made available on the net in a digested form.

In any case (and we refer also to the general recommendations below), we suggest that institutions should provide dedicated support for TS & HC in more structured ways than is generally the case at present.

## 3.3 Computing in textual scholarship courses

### 3.3.1 Introduction

With the emphasis of computers moving from being computational devices to manipulators of text, the case for developing and indeed investing in humanities computing teaching courses has become inarguable. Despite this, it is not always easy in European universities to gain the resources needed to teach computer skills to humanities students, not to mention special courses in textual scholarship and humanities computing (TS & HC). However, here are few humanities faculties that do not need the staff, equipment and facilities either to teach (humanities) computing, or to teach textual scholarship techniques using computers (or vice versa).

Textual scholarship in a wide meaning of the term is in one way or another involved in almost all imaginable disciplines. The following is a non-exclusive list of elements which form a part of TS & HC or which use TS & HC:

- Text corpora and text coding
- Language learning
- Library science
- Documentation science
- Information retrieval, including Web searching
- Databases and analysis
- Literary and textual criticism
- Authorship and stylistic studies

During the the activities of the working group on TS & HC, we found that there is a division between courses being taught where the IT is the main core of the course and courses in TS related disciplines which may use IT without this being necessarily mentioned in the course description. It is difficult to make a European inventory of courses of the first kind and impossible for courses of the second kind. On the other hand the courses of the second kind show that IT has become established in TS teaching and research.

### 3.3.2 Content elements of TS & HC courses

TS courses cover a large area of possible applications of HC. Therefore a clear definition of the different target groups is very important. Course materials as well as course structures and setup depend on the various levels of sophistication of students. The lowest common denominator should always be computer literacy (as provided by computing centres), and the varying degrees of competence would reflect a need for different skills for different purposes.

General vs. task-bound skills require different training, and this depends to large degree on the nature of TS and/or the HC course. There are general or degree courses, courses embedded into a wider framework of interdisciplinary research, or courses designed specifically to link up with existing courses in the humanities (and humanities credits systems). There is, however, a certain amount of agreement between the different institutions and countries when it comes to course content.

All of the course elements outlined here have computing at the core of the tools and methods taught in academic disciplines:

- Methods and theory
  - Theories and methods in humanistic informatics
  - Methods and computers in the humanities
  - Computer applications in the Humanities
  - Artificial intelligence
  - Humanities computing in the future
  - A theoretical analysis of cultural aspects of IT
  - Digital culture
  - IT as a research tool
  - IT in TS related disciplines
- Metastructures
  - Text encoding and markup
  - Representation of data
  - Data handling
- Text analysis
  - Content analysis
  - Stylistic analysis
  - Linguistic analysis
  - Concordances, thesauri, indexes
- Corpora
  - Databases
  - Corpus linguistics
  - Corpus analysis
  - Lexicography
  - Data analysis/quantitative methods
- Web
  - Hypertext
  - Hypermedia
  - Text retrieval
- Multimedia
  - IT in teaching and dissemination
  - Computer aided learning;
  - Machine translation
  - Text and image capture
  - IT as a language tool

### 3.3.3 Courseware available online

There is a growing collection of courseware available online. This is valid both for courses in Computing in Textual Scholarship and in Humanities Computing but also for traditional Textual Scholarship courses. The courseware available may be general data collections or material provided for specific courses. Among the types of data available are:

- Text corpora, either for downloading or for searching and processing online.
- General text archives
- General and language specific bibliographies
- Image collections
- On-line articles on textual scholarship (theory, research) and humanities computing in general

- Self-learning components in TS & HC

The data available online would, as the conclusions in section 5 mention, be of a greater European usefulness if there were a well structured roofing Web page or a portal which draws together humanities computing/textual scholarship activities and on-line resources explaining what can be found behind each link, to be used as a resource to find information and to link to when teaching individual courses.

### 3.3.4 Conclusions

It can be concluded that the integration of computing in Textual Scholarship courses is a manifold activity. It would seem profitable to draw the various elements together and create an autonomous subject to be studied together with text based subjects. The subject itself is, however, only just gaining shape, thus more study of its outline in the years to come seems necessary. Although on the one hand, it can be expected that the subject will explore interdisciplinary links, including history, computational linguistics etc., it could on the other hand strengthen ties with traditional text-based subjects like literature. In any case the requirements for collaboration between different perspectives will present a challenge to university environments.

## 3.4 Computing in textual studies: Relations with the wider world

### 3.4.1 Introduction

In the present section of this chapter, a number of recommendations will be made with a view to developing the academic environments in the text-based disciplines in such a way that inclusion of computing components would be viewed as normal. There are a number of implications and potential benefits that would arise from the wide and co-ordinated inclusion of computing components in textual studies curricula in European institutions of higher education, and these are highlighted in point form in this section.

**Note:** For economy of expression, students who successfully complete courses in textual studies and computing of the kind proposed are referred to as *TSC graduates* in the remainder of this chapter.

### 3.4.2 Career prospects and transferable skills

When summarizing the desired profile of TSC graduates, they would first of all have some specific proficiency in the software tools used in their course. However, beyond this they would have a set of *applications* skills that would enable them to tackle new software packages with confidence, and - even more important - a range of *analytical* skills that would have been acquired in dealing, typically, with non-uniform data. The ability to identify or create information structures from complex data sets, and to design solutions for complex problems, represent skills that are highly marketable and potentially of significant value to employers.

Text capture and manipulation techniques, including mark-up, along with an understanding of metadata and version control, are important parts of *document management*, and are relevant in an increasing range of commercial and other spheres of activity.

Digital image capture and manipulation are likely to be part of a standard curriculum, and are increasingly important across all sectors.

Electronic publication is becoming the norm for all types of organisation. The techniques identified above are of course important for this activity, and in addition the TSC graduates would understand XML

(and HTML) and would be proficient in the tools that are needed to create and maintain websites.

Multimedia digital resources of all kinds are already playing an important role in a number of areas, not least in the media and communications, entertainment, and cultural heritage sectors. TSC graduates who understand the basics of the creation, management and preservation of such resources, as well as issues of resource description and discovery, and information retrieval, will be well placed to find employment in these sectors.

### 3.4.2 The commercial world

As indicated above, there are certain areas of the commercial sector, particularly those that involve publishing and information management, where TSC graduates could have an important role.

However, there is a general point to be made, namely that graduates with good analytical skills, allied to the training of the imagination that comes from textual studies, have a significant contribution to make to the development of commercial products and the management of commercial operations.

Computing and information technology is having a profound effect on commercial and administrative relationships, e.g. between *information providers* (including academics) and *publishers* (who increasingly include broadcasters). TSC graduates would have the necessary basic understanding to play a role in the negotiation of the new relationships that are being developed.

There is strong encouragement for new partnerships to be developed between higher education and the wider world, including the commercial sector. TSC graduates would be well placed to play a role in imagining and creating these new ways of collaborating.

Those TSC graduates who develop their technical skills to a high degree may be able to make an important contribution to the development of new software tools, not only in terms of imagining and creating new types of tool, but also in ensuring that new tools take an understanding of the people who will use the tools into proper account.

### 3.4.3 Cultural heritage

TSC graduates will be well placed to contribute to the development of closer links between higher education and the cultural heritage organisations, such as museums and galleries, in whichever of the two sectors they work.

Those who remain in HE will have a better understanding of the resources and responsibilities of the museums and galleries, and can help to ensure that the resources are more widely and more appropriately used in HE courses.

Those who work in the cultural heritage sector can make an important contribution to the development of high quality digital resources.

With their electronic publications skills, these graduates will also be able to assist in the dissemination of cultural heritage materials, for example within primary and secondary education, and to the wider public.

### 3.4.4 Multilinguality

Multilinguality is central to the textual studies disciplines, and TSC graduates will understand the technical aspects of representing alphabets and writing systems, both for contemporary and dead languages (we refer also to Chapter 5 on specific issues in this respect).

TSC graduates, particularly those with language interests or specialisations, will be familiar with multilingual corpora. TSC graduates are likely to have experience of using multilingual thesauri, particularly in relation to specialised terminology sets.

### **3.4.5 Mobility of learners and workers**

Systematic work to develop common curriculum elements in applied computing in textual studies could increase learner mobility in Europe to a significant extent, ensuring that students from one country are not at a disadvantage with respect to the analytical and practical computing skills required for successful study in other countries.

In institutions where TSC programmes and expertise are not available, collaborative arrangements could be made to take advantage of distance learning tools and frameworks, so that, for example, students could acquire at a distance the background and skills they need to pursue more advanced components.

The range of analytical and practical computing skills acquired by TSC graduates would help to ensure maximum mobility in seeking employment opportunities.

### **3.4.6 Preparedness for citizenship**

In the information-rich societies of the 21st century, it is imperative that the countries of Europe have citizens who have confidence and the skills in using, managing and creating information in electronic form. TSC graduates will be well prepared for this role.

The European commitment to the preservation and promotion of all its languages and cultures depends increasingly on the appropriate application of computing techniques, including digitization and multimedia electronic publication. TSC graduates will be able to make an important contribution in this area.

European commercial success in a harshly competitive world depends on a workforce that is highly literate and skillful in the management and manipulation of electronic information. TSC graduates will form an important component of this workforce.

## **3.5 Conclusion and recommendations**

### **3.5.1 Preliminary remarks**

From the above, it is clear that TS & HC is a field with good research perspectives and a high potential to link with many partners in society. However, many institutions of higher education have limited structural support for computing in the text-based disciplines, and the inclusion of computing components in textual studies courses is patchy and highly variable. Text-based disciplines cover a broad area of the humanities, but do not have a clear home base, apart from literature perhaps, but most literature departments have not shown sufficient openness toward interdisciplinary collaboration and have been reluctant to explore the potential offered by computational methods.

The recommendations made in this section arise from consultation and discussion over the course of the ACO\*HUM project, and are drawn from experience in institutions where computing is well established in textual studies as well as concerns expressed in institutions where this is not the case.

Some of the recommendations are general in nature, and concern infrastructural issues that are likely to be addressed by institutions as part of wider strategy covering provision for all students and/or teachers.

However, it is important the text-based disciplines should be fully supported in such provision; the time has passed when these disciplines can be regarded as having lesser requirements for computing facilities and tools.

A number of the recommendations concern specific action related to establishing a framework for more systematic gathering and maintenance of information.

**Note:** Following the practice in the previous section, the abbreviation *TS* means *textual studies*. Hence, *TS students* is used to designate the generalia of students in the text-based disciplines, while *TSC students* designates students in these disciplines whose courses include applied computing components.

### 3.5.2 General recommendations

- All students should have equal access to adequate computing facilities. These should include multimedia computers with high resolution monitors and high-bandwidth connections to the Internet.
- TSC students should have access to appropriate software, including tools for concordance creation, text mark-up and analysis, statistical analysis, digitization, image manipulation, and the creation and publication of information in XML and HTML format.
- All students should have access to introductory or remedial training in basic computing skills - such as those covered by the European Computer Driving Licence (ECDL).
- All institutions should provide support for TS teachers, however this support might be organised. TS teachers need both general IT support and also some degree of specialist support in the application of computing in the TS disciplines.

### 3.5.3 Recommendations on collection and dissemination of information

- A central repository should be established or designated to hold information on curriculum components relevant to the study of computing in TS disciplines. One difficulty faced by administrators and teachers alike is finding comprehensive information about current practice in Europe. The HUMBUL gateway at the University of Oxford is a possible candidate for such a role.
- A standard framework for the systematic collection of metadata about curriculum components would make such a repository much more useful than it would otherwise be. The proposed IMS standards should be investigated with this in mind.
- In addition to information about curriculum components, the central repository should also contain information about and, where appropriate, links to teaching or learning materials used in TSC courses.
- The information collection and dissemination would best be done in formal collaboration with a professional association. In Europe the most appropriate one is the ALLC; a possible link should be investigated.
- Funding should be sought to produce a set of 'case studies' of institutions where successful TSC courses are in operation. Such a publication would provide models and would give guidance on good practice. Such an approach would enable the experience of Europe's existing centres of excellence in this area to be disseminated more widely.

### 3.5.4 Recommendations on course development

- Funding should be sought to enable collaborative development of common curriculum components. The outline of a set of core components is fairly clear, but further systematic work is needed to establish how these components can best be adapted for each of the European countries, so as to ensure maximum compatibility across Europe to provide maximum mobility of learners. A

preliminary set of core components is given as an appendix to this chapter.

- A system for the accreditation of TSC courses should be investigated. In practice such a system might best be established within the framework of a professional association, such as the ALLC. Such a system would need to incorporate an appropriate peer review process. The purpose of the accreditation would be to provide assistance in the creation of courses, to ensure that the computing methodology is appropriately rigorous, and to provide a basis for confidence within and between institutions.
- In investigating such an accreditation system, the validation procedures adopted by bodies such as ELRA should be considered carefully. The range of institutional and national validation procedures in operation would also have to be taken into account.

### 3.5.5 European masters

- The development of a European masters degree should be investigated. It is possible that there should be a separate programme for the text-based disciplines. However, it is likely that there would be scope for some degree of collaboration with other humanities discipline areas, such as history, history of art and computational linguistics. It is possible that a general programme along the lines of Applied Computing in the Humanities might be developed, with optional modules constructed in such a way that particular emphasis could be given to particular discipline groupings, as appropriate.
- The emphasis in such a programme would be on methodology, including analysis and design, with the text-based disciplines providing some key components that would be of relevance across all the humanities disciplines.
- It is also possible that, once developed, some core elements could be incorporated into existing masters programmes to provide them with a computing component.
- In terms of learner mobility, it is possible that the development of a common curriculum could take place more rapidly in the context of a European masters than in undergraduate programmes.

### 3.5.6 Recommendations on institutional collaboration

- The scope for collaboration between institutions should be explored and exploited.
- The basic opportunity for this lies in agreement on common curriculum components; however, the existence of the European academic network makes it feasible to collaborate also on the development of course materials.
- Similarly there are opportunities for collaboration in the teaching of courses and assessment of students. If key course materials are available on-line, a number of technologies are now available to make it possible to teach courses in which the learners are widely dispersed.
- Collaboration of this kind could for instance take the following forms:
  - Specialised expertise could be shared between institutions.
  - Students in institutions with no TSC programme could participate in courses at other institutions
  - 'Mentor' schemes could be established, in which existing centres of excellence could assist other institutions to establish new TSC programmes. This might be of particular benefit for some institutions in parts of Europe where development programmes are in operation, e.g. certain of the Eastern European countries.

## 3.6 References

### Books, articles and CD-roms

Bergenholtz, Henning & Schaefer, Burkhard (eds.) (1979): *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Korpora* (Monographien Linguistik und Kommunikationswissenschaft 39). Königstein/Ts.: Scriptor.

Biber, Douglas, Conrad, Susan & Reppen, Randi (1998): *Corpus linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

ECI/MCI (1994): *European Corpus Initiative Multilingual Corpus 1*. CD-ROM.

Habert, Benoît, Nazarenko, Adeline & Salem, André (1997): *Les linguistiques de corpus*. Paris: Arman Colin/Masson.

Kennedy, G. (1998): *An introduction to corpus linguistics*. Addison Wesley Longman Higher Education.

Lancashire, Ian (1991): *The Humanities Computing Yearbook 1989-90. A Comprehensive Guide to Software and other Resources*. Oxford: Clarendon.

Marcos-Marín & Francisco A. (1994): *Informática y Humanidades*. Madrid: Gredos.

McEnery, Tony & Wilson, Andrew (1996/1998): *Corpus Linguistics*. Edinburgh University Press.

Sinclair, John (1991): *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sperberg-McQueen, C. M. & Burnard, Lou (eds.) (1990): *Guidelines for the encoding and interchange of machine-readable texts* (TEI P1). Chicago/Oxford: ACH-ALLC-ACL Text Encoding Initiative.

Sperberg-McQueen, C. Michael & Burnard, Lou (eds.) (1994): *Guidelines for Electronic Text Encoding and Interchange* (Electronic Book Library Nr. 2). Providence: Electronic Book Technologies.

## Journals

*Computers and the Humanities*. Kluwer Academic Publishers.

*ICAME Journal*. University of Bergen.

*International journal of corpus linguistics*. John Benjamins Publishing Company.

*Literary and Linguistic Computing*. Oxford University Press.

*Research in Humanities Computing*. Oxford University Press.

## URLs of organizations and projects

[Advanced Computing in the Humanities \(ACO\\*HUM\): http://www.uib.no/acohum.](http://www.uib.no/acohum)

[Association for Literary and Linguistic Computing \(ALLC\): http://www.allc.org/.](http://www.allc.org/)

[Computers in Teaching Initiative \(CTI\): http://www.cti.ac.uk.](http://www.cti.ac.uk)

[European Language Resources Association \(ELRA\): http://www.icp.grenet.fr/ELRA/home.html.](http://www.icp.grenet.fr/ELRA/home.html)

[Text Encoding Initiative \(TEI\): http://www.tei-c.org/.](http://www.tei-c.org/)

[Wittgenstein Archives \(at Bergen\): http://www.hit.uib.no/wab/.](http://www.hit.uib.no/wab/)

## 3.7 Appendix: Core curriculum components

The following are an initial set of core curriculum components that are recommended for inclusion, in some way, in the courses followed by students in the textual disciplines. The components may be provided together as part of a specialised course in 'textual studies computing' or 'humanities computing', or may be included, as appropriate, separately or in groups in a variety of courses offered in specific text-based disciplines.

Underlying all these components should be the basic principle that each component represents the introduction and use of formal methods, and the analytical aspects in each case are of more fundamental importance than the design and implementation skills.

It should also highly desirable that the design and implementation skills are taught in as general a way as possible, so that maximum transferability is enabled, and attachment to the specific software tools used in the course is discouraged.

Initial set of core curriculum components:

- Information analysis, modelling and system design
- Information retrieval and filtering, including the use of large digital resources, such as bibliographies, reference works and corpora
- Representation of character sets and writing systems
- Text mark-up (including the TEI)
- Automated text analysis
- Electronic (non-critical) editing and publication (including XML and HTML)
- Digitisation, and image manipulation
- Tabular data, including spreadsheets and simple databases
- The creation and management of digital resources, including metadata and resource discovery and preservation
- The role of computing in society and its implications for cultural life and the transmission of cultural heritage

Other candidates for inclusion in the core set are likely to include:

- Statistical methods
- Design of complex relational database systems