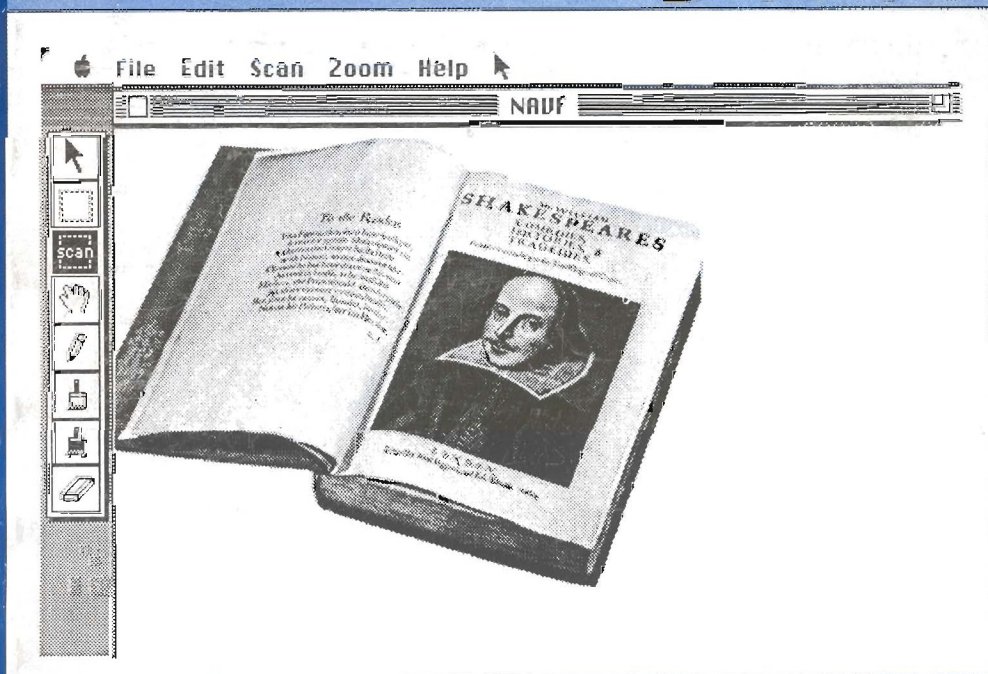


h

UMANISTISKE

DATA



NAVF'S EDEB-SENTER
FOR HUMANISTISK FORSKNING
NORWEGIAN COMPUTING CENTRE
FOR THE HUMANITIES

2-87

SENTERETS RAPPORTSERIE

RAPPORTER UTGITT F.O.M. 1980

RAPPORT nr. 17. *Svein Lie: Automatisk syntaktisk analyse*. Del 1. Grammatikken. Desember 1980. ISBN 82-7283-014-0 Pris kr. 30.

RAPPORT nr. 18. *Datateknologi og humanistisk forskning*. Bidrag til en NAVF-utredning. Desember 1980. ISBN 82-7283-015-9 Pris kr. 30.

RAPPORT nr. 19. *Statistiske metoder på arkeologisk materiale*. Rapport fra et seminar på Bryggens museum, Bergen 24.-26. november 1980. Mars 1981. ISBN 82-7283-017-5 Pris kr. 35.

RAPPORT nr. 20. *EDB-prosjekter i humanistiske fag 1980*. Juni 1981. 2. opptrykk oktober 1981. ISBN 82-7283-018-3 Pris kr. 45.

RAPPORT nr. 21. *Rune Johansen: Bruk av EDB i teatervitenskapelig forskning*. Mai 1981. ISBN 82-7283-019-1 Pris kr. 35.

RAPPORT nr. 22. *Årsmelding 1980*. NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-020-5 Gratis.

RAPPORT nr. 23. *Stig Welinder: A program package for archaeological use*. 1981. ISBN 82-7283-021-3 Pris kr. 45.

RAPPORT nr. 24. *Rapport fra seminar om bruk av edb innen teater og teatervitenskap*. Januar 1982. ISBN 82-7283-026-4 Pris kr. 50.

RAPPORT nr. 25. *Ole Lauvskar: Diskriminantanalyse i SPSS*. Desember 1982. ISBN 82-7283-028-0 Pris kr. 55.

RAPPORT nr. 26. *Stig Welinder: Paleodemography*. Oslo 1982. ISBN 82-7283-030-2 Pris kr. 55.

RAPPORT nr. 27. *Årsmelding 1981*. NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-029-9 Gratis.

Forts. 3. omslagsside.

HUMANISTISKE DATA

2-87

***NAVF's EDB-SENTER
FOR HUMANISTISK FORSKNING***

***The Norwegian Computing Centre
for the Humanities***

HUMANISTISKE DATA

utgis av NAVFs edb-senter for humanistisk forskning.

REDAKSJON

Jostein H. Hauge, Kristin Natvig (red.).

ADRESSE

Harald Hårfagesgt. 31, Boks 53 — Universitetet, 5027 Bergen.
Tlf. 05-212954/55/56.

ABONNEMENT

Gratis for enkeltpersoner, kr. 70,- for institusjoner (3 nr. pr. år).

Artikler, rapporter og meldinger mottas gjerne.

Medarbeidere fra Senteret i dette nummer:

Jostein H. Hauge, Knut Hofland, Kristin Natvig, Espen S. Ore, Øystein Reigem, Per Vestbøstad.

Redaksjonen avsluttet 24. september.

Forsidebildet er scannet inn fra en AGFA 200S til en Apple Macintosh og bearbeidet ved hjelp av Apples program for desktop publishing PageMaker. Scanningen og bearbeidingen er utført hos Kontorutvikling AS - avd. Bergen.

NAVFs EDB-SENTER FOR HUMANISTISK FORSKNING

ble opprettet av Norges allmennvitenskapelige forskningsråd i 1972. Senteret skal arbeide på nasjonal basis for bruk av edb i forskningsarbeidet i de humanistiske fagene.

Sentrale oppgaver er edb-tjenester (veiledning og betalte oppdrag) og program- og metodeutvikling. Senteret holder kurs, seminar og konferanser om bruk av edb i humanistiske fag. Foruten utgivelse av Humanistiske Data omfatter informasjonstjenestene en rapportserie, årsmelding og elektronisk informasjonsformidling.

Senteret er sekretariat og operativt edb-organ for International Computer Archive of Modern English (ICAME), og utgir tidsskriftet ICAME Journal. Senteret er ansvarlig for administrasjon og drift av Norsk tekstarkiv og har det administrative ansvar for NAVFs sentral for informasjon om forskningsprosjekter (SIF) og Fagtjenesten for informasjon om humanistiske forskningsprosjekter (SIF-H).

Humanistiske Data is published by The Norwegian Computing Centre for the Humanities. Editorial group: Einar-Arne Drivenes, Jostein H. Hauge, Kristin Natvig (ed.).

The journal can be ordered free of charge from the address above. Contributions are welcome. On request the Centre can supply the addresses of contributors to the journal.

GRAFISK FORMGIVNING: Svart på Hvitt.

SATS: Gerhard Datasats A.S.

TRYKK: Bergen Trykk A.S.

INNHOOLD

ARTIKLER

The Hull Domesday Database Project. <i>J.J.N. Palmer</i>	4
Programming in SPITBOL for Historians. <i>Daniel Greenstein</i>	23
Tradisjon og teknologi. <i>Kristin Natvig</i>	34
Desktop Publishing. <i>Per H. Jacobsen</i>	40
Tekstbehandling og tegnssett. <i>Espen S. Ore</i>	60
Ei frekvensordbok for nynorsk. <i>Per Vestbøstad</i>	68

RAPPORTER

Fem undervisningsprogrammer fra Datasekretariatet. <i>Espen S. Ore</i> ..	74
The MLA <i>Bibliography Generator</i> . <i>Ralph Jewell</i>	80
Politiske holdninger. <i>Einar-Arne Drivenes</i>	83
CALICO '87. <i>Jostein H. Hauge</i>	86
Computers and Teaching in the Humanities. <i>Kristin Natvig</i>	90
Optica '87. <i>Øystein Reigem</i>	99
«The Use of Computers in the Teaching of Language and Languages». <i>Per Vestbøstad</i>	104
ICAME 8th. <i>Knut Hofland</i>	107
XIV ALLC Conference. <i>Espen S. Ore</i>	111
Utviklingsseminaret 1987. <i>Kristin Natvig, Øystein Reigem og Per Vestbøstad</i>	112

MELDINGER	116
------------------	-----

SUMMARY	126
----------------	-----

THE HULL DOMESDAY DATABASE PROJECT⁽¹⁾

J.J.N. Palmer

Introduction

The Hull project to computerise Domesday Book is an unusual Arts enterprise in a number of ways, the most important of these being that we are aiming to produce a package in which the data is completely integrated with the software needed to process it. The software itself will offer an unusual combination of facilities, which will include database creation, a database editor, text analysis, concordancing, graphics and mapping, all of which we have found to be essential to the proper exploration of this fundamentally important source.

The project has evolved through several stages. The early work of data input was done on an ICL 1904S machine, with a character set which did not include lower case letters, which therefore had to be indicated by escape characters. With the combined English and Latin texts amounting to some two and a half million words, or 15 million characters, editing the data in this form would have been a nightmare! Fortunately, we were able to transfer our files to a machine with a full character set before much revision was required (a Harris 800). All our data is currently lodged on an ICL 3980 ('Estriel') machine. In addition to these three mainframes, microcomputers have been used for various tasks, originally BBC's and latterly IBM PC compatible machines. At present, all software development is concentrated on the PC and we hope to make the entire package available on this machine, perhaps in CD-ROM, if present plans come to fruition.

Our experience with a variety of hardware and operating systems has made us very conscious of the problem of software compatibility, and every effort is therefore being made to make the bulk of the software transportable, though there will always be problems with mapping and graphics. The release of the new range of downward compatible IBM machines does, however, seem to guarantee a reasonably long life for software on PC compatibles.

Over the past three years, we have also used a number of software packages in our development work: the Oxford Concordance

Program, SPSSx, GIMMS, FAMULUS, REFLEX, and a number of line and context editors, word processors, and operating systems. The problems of moving data in and out of a variety of packages which require it to be in many different formats can be extremely frustrating and has made us acutely aware of the advantages of bundling the data with as much of the essential software as possible, into one integrated unit. This is not a strategy which would be feasible for many purposes. But for the single most important source of social and economic data for several centuries of English History, it seems an effort worth making. Domesday Book is, arguably, such a source.

The source

Domesday Book is unique. It is the oldest and most famous English public record. There is nothing directly comparable from any other part of the world, and nothing even vaguely comparable for several centuries after 1086 in England. Much of English History effectively begins with Domesday. It is the first national survey in English History; and has claims to be the first census, the first survey of land usage, and the first survey of landownership. Of the 13,500 towns and villages described in its folios, about 90% are recorded for the first time in their history, and more than 95% are described there for the first time. The history of the overwhelming majority of English towns and villages therefore begins with Domesday. So, too, does the continuous history of English families and of English estates. Domesday Book is the most important single document in English social history.

The importance of Domesday Book as a source is reason enough to computerise the record. But in addition Domesday is a very substantial document and therefore not easily analysed by 'manual' methods. It also has a complex structure, unsuitable for most of the purposes for which historians, geographers, linguists and place- and personal-name students would like to utilise it. The reasons for computerising the records are therefore all but overwhelming.

Overview

Unlike a modern statistical document, the statistics in Domesday are embedded in a text written by a scribe who has been described by one authority as 'revelling in the use of synonym and paraphrase'.⁽²⁾ The same piece of information may therefore be phrased in a variety of ways; and divergent information may also be presented as though it were essentially similar. This makes the task of extracting even the

most basic statistics a complex, error-prone, and potentially contentious exercise. As the greatest of Domesday historians has observed, 'two men not unskilled in Domesday might add up the number of hides (tax units) in a county and arrive at very different results, because they hold different opinions as to the meaning of certain formulas which are not uncommon'.⁽³⁾

Faced with this situation, there are two possible strategies for computerising the record. One is to discard the purely textual material, extract the statistics, and create a database which is a matrix of numbers, which may then be 'crunched' by such commercial or academic packages as SPSS, SAS, Minitab etc. This is the strategy which has been adopted by the largest of the Domesday research teams, that at Santa Barbara, California, funded by a federal grant rumoured to be over 3 million dollars.⁽⁴⁾ The second possible strategy, and the one adopted at Hull, is to create a text database — or a *textbase*, as databases of this type are sometimes called — and to write the software which will enable us to analyse the statistics in their context. This alternative avoids the main danger of the first approach, that just one of many possible interpretations of the statistics in the text are 'frozen' into the database. A further advantage is that the text itself is available for analysis. But it cannot be denied that this strategy adds very considerably to the complexity of the task of computerising Domesday. It does so in three major respects: the software must be customised, the text structured, and all individuals uniquely identified.

(1) SOFTWARE DEVELOPMENT

The Databases: The first complication of this strategy is that it requires us to write our own software. Our initial specification for the software was that it should be portable, designed for interactive use, and ultimately available on micros. Since the Latin and English text of Domesday are each well over a million words, this specification entailed that the databases should be made randomly accessible; and since we knew that many additions would be made to each text in the course of our work, it was essential that they could easily be recreated from time to time. Both the structuring of the texts, and the identifications of individuals, would entail a large volume of textual insertions which would be all but impossible to make with reasonable speed and accuracy using standard line and context editors. A customised database editor was thus another essential ingredient. The main relationships between the database software and the editor are shown in fig. 1.

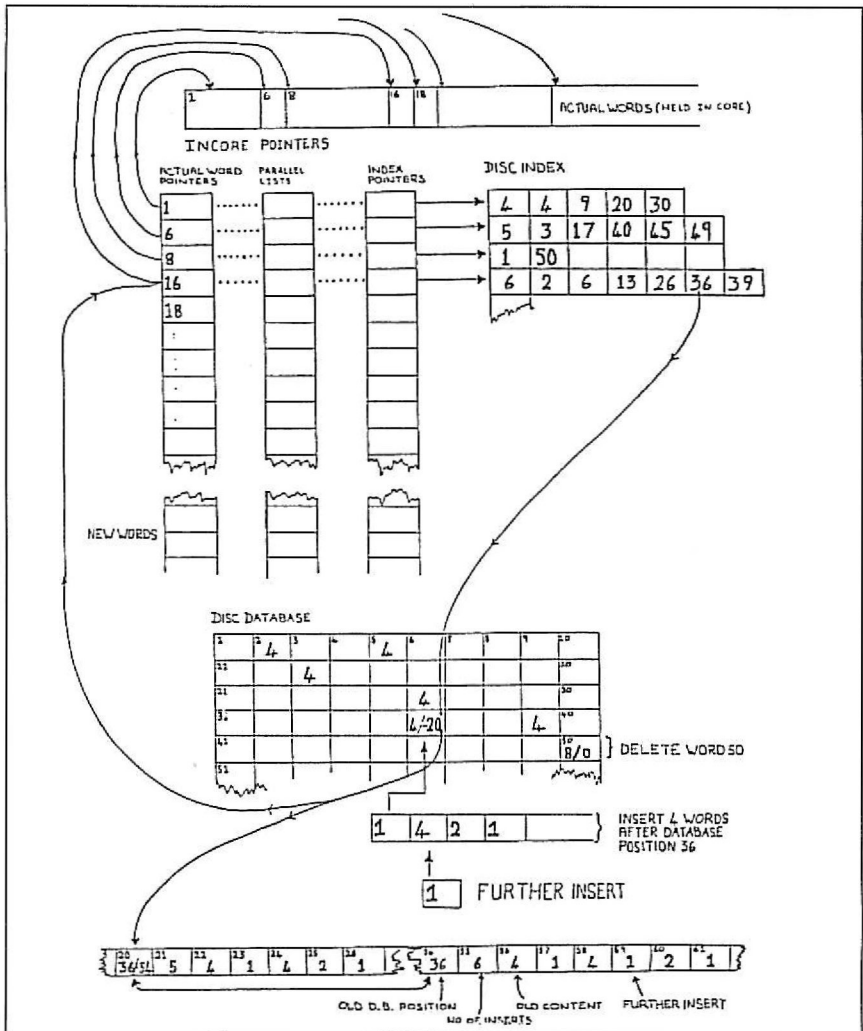


Figure 1. The Database. The diagram illustrates in a simplified manner on a small segment of data the relationships between the incore index, pointers, disc database, and editor. The fourth word (alphabetically speaking) is 'an', occupying character positions 16-17 in the memory array. It has 6 entries in the disc index, each of which points to an absolute position in the disc database, where the number 4 (for 4th word) points back to the index via the pointers. The lower half of the diagram shows how inserts are made by the editor without disrupting the database.

Related Files: Since each Domesday database is well in excess of a million words, efficiency has been a vital consideration, given that the database is to run on a micro of limited capacity. One of the steps we have taken to maximise efficiency is to confine the database itself solely to the text of Domesday Book, which has been stripped of all punctuation and embedded codes. Punctuation, the marking codes described below, and various invisible characters (such as line feeds) are stored in a related file which, like the database itself, is randomly accessible to the software. This increases efficiency considerably. Searches for phrases which cross line boundaries, or are separated by coding or punctuation, require search algorithms of considerable complexity which are therefore markedly slower in operation than a straightforward search for adjacent words, or words separated by a defined number of other words.

A further advantage of this approach is that the related database files may contain information about the context of words which, because it is assembled in one place, can be accessed very speedily. One instance should make this clear. 'Pigs' in Domesday normally occur as a measure of woodland, not as a description of livestock. As a measure of woodland, 'pig' will appear within a phrase or sentence marked out by codes which delimit text describing the appurtenances of majors. To detect 'pigs' as livestock, it is necessary to locate occurrences of the word which occur outside such contexts. Normally, it would be necessary to examine every occurrence of the word 'pig' in the text, then to scan backwards and forwards in search of delimiting codes, a very slow process. The related file allows this information about context to be extracted with a single read.

The system also allows considerable flexibility. At a future date, we shall probably lemmatise the entire text, so that it is possible, for instance, to distinguish homographs (lead, as in leader, from lead, as in lead balloon). This will involve no more than the addition of a code to the related file, which may then be extracted in one read.

Search software: Text may be retrieved by searching for words or phrases in more or less complex ways. The first step in any search is to define a context in which the results are to be displayed. The user may choose to define the context as a span of words, a line of text, a sentence, an entry (the default), a folio, a fief or a quire. If either of the first two are selected, the results may be displayed a page at a time, in KWIC-format if required. This provides in effect an inter-active concordance facility.

Within this context, words or phrases may be located using standard boolean operators, with or without various wild card options or

case conscious facilities. Phrases or formulae may be sought within a specified span. This **span** may be defined in terms of the number of words on either side of a search word within which a second word is to be located, or in terms of specified words or phrases between which the search word is sought. This facility allows very considerable flexibility in search patterns. At its simplest, a two word phrase may be sought by defining the search words and a span of one word. But the facility also allows searches for formulae in which the key words are separated from each other by text which may vary. For instance, the Ploughland statistics in Domesday are almost invariably introduced by the formula 'Land for x ploughs', the 'x' being a variable word or number. It is therefore possible to extract all such formulae from the text by defining a span of 3 words and searching for: **Land AND plough***. For a heavily formulaic text such as Domesday, this will prove an invaluable search tool.

In effect, the span allows the user to define any portion of the text as a record (in normal database parlance). Applied to a textbase, this is an extremely powerful facility. By defining the Domesday entry as the span, for instance, it is possible to locate every entry in Domesday which contains (or does not contain) some particular feature. The single command: **NOT slave*** would scan the million word text and select from some 22,000 entries only those Domesday manors with no slave population.

Restructuring: Finally, the software allows the database to be restructured. As is well-known, the structure of Domesday Book is just about the most unhelpful which the Conqueror could have devised for later generations. The most obvious example of the unsuitability of the existing structure of Domesday Book is the way in which the villages of Domesday England have been dismembered by the Domesday scribes and their fragments scattered around the text of each county, making it necessary to undertake a substantial amount of work in 're-constituting the vills' before an examination of the characteristics of villages can even be begun. Similarly, the possessions of Anglo-Saxon landowners are randomly distributed throughout the text, making any examination of the social structure of Anglo-Saxon England out of the question without the considerable preliminary labour of bringing together all the manors owned by each Anglo-Saxon lord or tenant.

The database will allow the user to reconstruct the text of Domesday in a variety of ways. Since Domesday Book is overwhelmingly concerned with landowners and their manors, most of the possible structures relate to one or other of these categories. The command **Structures** will produce a menu of structures from which the user may

select, the default being the order of the Domesday text. Each of the structures and its individual components may be directly listed. The command **List Vill**s, for instance, will produce an alphabetical listing of villas, and **List Norman Lords** a list of Norman tenants in chief. Once a structure has been defined, all subsequent searches are carried out within that framework, the results appearing in that sorted order.

Finally, all information extracted by any searches may be mapped or graphed. The software for doing this is effectively in place, though much work remains to be done in assembling the data required for the mapping.

(2) STRUCTURING THE TEXT

In order to allow the software to interpret it in an intelligent manner, the text itself must first be structured. This is the second complication of the strategy we have adopted. 'Structuring' is effected by inserting codes (invisible to the user) which highlight special characteristics of any part of the text. Such codes may mark the beginning and end of entries, fiefs, counties or folios; or may mark particular categories of landowner, type of property, or group of statistics. Some indication of the scale and complexity of this operation may be gleaned from the fact that well over 20,000 different codes, totalling more than 250,000 insertions, will be required to complete the task. The sophistication of the codebook, and the accuracy with which the coding is done, is therefore critical to the success of this strategy of computerisation, as it would be for any electronic text database. A recent report on the progress of the computerisation of the *Oxford English Dictionary* makes this point very clearly. The Oxford text database is a substantial one by any standards, requiring some 500 million keystrokes for data input (10 times the size of Domesday), employing 120 skilled typists and 70 proofreaders, and drawing on the resources of OUP, IBM and the University of Washington. The report noted that 'the explicit realisation' of the structure of the database was 'the central problem around which nearly everything revolved', the 'explicit realisation' of the structure taking the form of the insertion of 'structural tags', or coding, into the database.⁽⁵⁾

A full description of our coding scheme cannot be given here since our codebook is as long as this article. The general principles of coding texts are, however, well-known and were referred to in the previous issue of this journal. I shall therefore confine myself to a brief outline of the more important groups of codes and how they are used in practice. It should first be recalled that Domesday consists

of 7 circuits, which are in turn subdivided into counties, fiefs, hundreds, and 'manors'. The 'manor', or entry, is the basic unit of the survey; and it is at this level that most coding of the text takes place.

The codes may be divided into two main groups, dealing respectively with textual and subject matter. The textual codes mark such features as the physical make-up of Domesday (volumes, quires, folios and columns); its conceptual arrangement (circuits, counties, fiefs, hundreds, and 'manors'); and such textual features as marginalia, interlineation, deletions, blanks, editorial insertions etc. Circuits, counties, volumes and quires may be reconstituted by simple algorithms and so do not require explicit coding. Folios (1-413, 1-475) and columns (a-d) are indicated by angular brackets; fiefs are separated from each other by three blank lines, entries by two blank lines, and sub-entries by a single blank line. Textual features are marked by a combination of brackets and characters which do not appear in the text (e.g. ?, \$, ^, *, %).

The more important markers for the subject matter may be subdivided into four groups of codes for: (1) landowners, (2) places, (3) administrative divisions, and (4) statistics. The text to which the codes for landowners, places and administrative divisions refers is indicated by enclosing it within round brackets. Text and codes containing the various categories of statistics are delimited by hashes. The brief entry for Wraysbury (Bucks) which follows gives specific examples of some of the codes which are used in the English and Latin texts:

```

? .XX.7 TERRA (T1 ROBERTI GERNON). IN (H1 STOCHE) HDO.

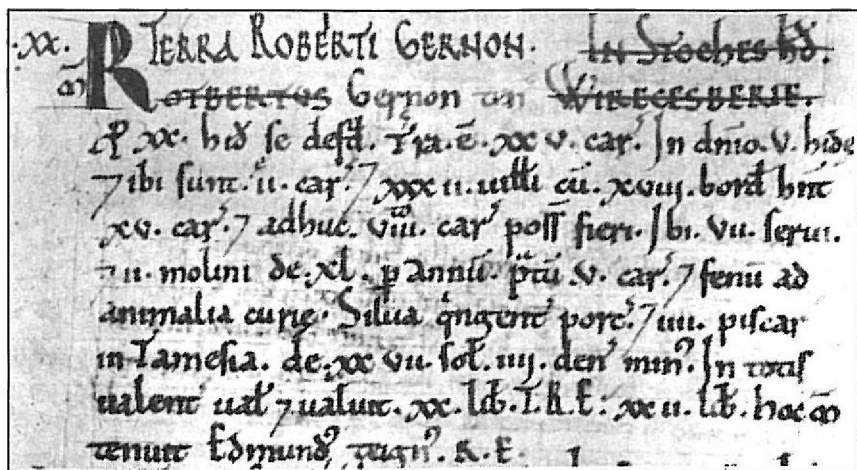
7M27 #A (TO ROBERTUS Gernon=213) ten1 (p1 WIRECESBERIE
=BKM1709). pB XL. hido fo defdO. A# #T Tira .eO .XXV.
car1. In dn0lo .v. hidae. 7 ibi fuit .II^ae^ car1. 7 #M
XXXII. uill0i cuO .XVIII. bordO hndt XV. car1. 7 adhuc
.VIII^Lo^ car1 pofaO fieri. T# ibi .VII. ferui. 7 M# #W
II. molini de .XI. p2 AnnoO. p^a^tuO .v. car1. 7 fenuO
ad animalia curiao. Silva q^i^ngent1 porci. 7 IIII.
pifcar1 in Tamofia. de .XXVII. fo1O .IIII. den1 min9 W# #V. In
totia valent1 valO 7 valuit .XX. libO. T.R.E: XXII. libO V#.
Hoc M2 tenuit (T9 Edmund=5226) teign9. (TB R. E=10000).

7207 LAND OF (T1 ROBERT GERNON=213)

In (H1 STOKE) Hundred
7M7 #A (TO Robert Gernon=213) holds (p1 WRAYSBURY=BKM1709).
It answers for 20 hides. A# #T Land for 25 ploughs; in lordship 5
hides; 2 ploughs there.
#M 32 villagers with 18 smallholders have 15 ploughs; a further 8
ploughs possible. T# 7 slaves; M#
#W 2 mills at 40 [s] a year; meadow for 5 ploughs;
hay for the cattle of the court; woodland 500 pigs;
4 fisheries in the Thames at 27s less 4d. W#
#V The total value is and was £20; before 1066 £22. V#
(T9 Edmund=5226), a thane of (TB King Edward's=10000)

```

Fig. 2: Domesday entry for Wraysbury (Bucks).⁽⁶⁾



Facsimile of the Domesday entry for Wraysbury.

Landowners: Codes are assigned to all landowners, individuals or institutions appearing in the text, and may therefore be attached to phrases ('said claimant'), pronouns, and common nouns ('Earl'), as well as to personal names. The main purpose of this coding is to indicate landholding structures, so persons appearing other than in a landholding capacity are coded as unclassified (T0). There are two major types of code: those classifying landownership, and those uniquely identifying the individual or institution. The coding system associated with identifications is described below, in the section dealing with this subject. The main categories of landownership are:

T0	: Unclassified personal name
T1	: Tenant in chief 1086
T2	: Tenant 1086
T3	: Subtenant 1086
T4	: Tenant of part of entry 1086
T5	: Landholder between 1066 & 1086
T6	: Claimant
T7	: Claimed against
T8	: Lord 1066
T9	: Tenant 1066
T10	: Subtenant 1066
T11	: Possessor of 'jurisdiction' 1066
T12	: Tenant of part of entry 1066
T13	: Royal Official

The above markers may be combined with the following ('n' indicating any of the code numbers 1-13):

Tn.a	:	Crown
Tn.b	:	Bishops (of English sees)
Tn.c	:	English monastic houses
Tn.f	:	Foreign monastic houses
Tn.d	:	Other clergy
Tn.q	:	Lapsed fiefs
Tn.br	:	Breton
Tn.s	:	Anglo-Saxon landholder in 1086
Tn.fl	:	Flemish
Tn.nf	:	Northern French
Tn.sf	:	Southern French
Tn.w	:	Women

Uncertain classifications are indicated by inserting 'u' after the stop: **T2.usf** therefore refers to a tenant who was probably from southern France. Codes may be combined by repeating the element after the stop: **T1.w.fl** indicates a female Flemish tenant-in-chief.

Places: The main object in coding places is to permit the reconstruction of the manorial structure of Domesday England. Place names which occur outside this structure are coded as unclassified (P0); and place names which have a structural significance but which are not the subject of an entry, are indicated by an upper-case marker (P). The codes p1-p7 and p10 may take either upper or lower case markers. Special provision is made for places which are not referred to by name ('the same village') or which the scribe has accidentally omitted from the record.

As with the codes for landholders, there are two types of codes for places: those indicating the manorial characteristics of the place, and those uniquely identifying the place (mainly for the mapping software). The skeletal outline of structural codes is as follows:

p0	:	Unclassified placename
p1	:	Manor/Manorial centre 1086
p3	:	Berewick or outlier
p4	:	Dependency of a manor
p5	:	Sokeland or jurisdiction
p6	:	Thegnland
p7	:	Uses formula 'x holds ... in y'
p8	:	Manorial centre 1066

p10 : Entry not a landholding
 p11 : Manor, called such in text
 p17 : Formula 'x holds ... in y'; called manor

These codes may also be combined with the following:

pn @1 : doubtful placename
 pn @2 : duplicate entry
 pn @3 : cancelled entry

Places are uniquely identified by appending a code to the place-name which uniquely identifies its main characteristics to the mapping software. This code has three elements: a three letter county code, a two figure hundred code, and a two figure parish code. In the example text given at the head of this section, the code: =BKM1709 identifies Wraysbury to the system software as the ninth parish in the seventeenth hundred in Buckinghamshire. This coding system allows the software to select areas on the basis of the county, hundred or parish to which they belong, while uniquely identifying them. It should be emphasised that the code identifies a mapping unit, to which several Domesday vills may belong. The administrative units of Hundred, Lathe, Rape, Riding, Wapentake etc., are coded in a similar manner, the codes being preceded by the letter 'H'.

Statistical codes: Finally, the following pairs of coded markers are used to delimit portions of text according to the statistical information which they enclose:

#A #A : Geld assessment
 #T #T : Ploughlands and Ploughteams
 #M #M : Population
 #W #W : Manorial Appurtenances
 #V #V : Values

(3) IDENTIFICATIONS

The third complication introduced by the strategy we have adopted is that it makes it essential to identify different forms of the same name which occur in the text as belonging to a single individual. Persons referred to by pronouns or descriptive phrases, or identified by the context; or names which are spelt or punctuated slightly differently, or abbreviated, all present major problems for the system software

which are best tackled by coding the text to make it more explicit. But the problem of making the text intelligible to the software is only a minor part of the wider problem of identifying individuals in Domesday Book. Computerising the whole text of Domesday has, however, forced us to face this larger problem which we might otherwise have felt tempted (as others have been) to evade. But the problem of identifications is too serious to be evaded by anyone wishing to produce an enduringly useful computerisation of Domesday, because the majority of landowners in Domesday are named only by their christian names and are thus very insecurely identified. There are, for instance, 1232 occurrences of the name 'Hugh', 1155 of 'Ralph', 1446 of 'Robert', 1262 of 'Roger', and no less than 2494 of 'William'. Some of these christian names occur in conjunction with surnames; but even so these five names may represent anything between a few score and more than 5,000 individuals. Margins of error so huge totally preclude any useful social analysis of the landowning classes before or after the Conquest. Since the ownership of land at those two dates is mainly what Domesday Book is about, this situation very severely restricts the usefulness of the document as a whole.

Excursus: The problem of identifications is so fundamental to the exploration of Domesday that a brief excursus on the centrality of the subject may be excused. The history of the English landed classes begins with Domesday Book. Less than half a dozen families can trace their ancestry beyond 1066, whereas the continuous history of hundreds of families and thousands of estates, can be traced from Domesday Book onwards. For this reason alone, the identification of Domesday owners and tenants has been a major concern of generations of Domesday scholars and local historians. Since the majority of individuals, and the overwhelming majority of the tenants in Domesday, are named only by their christian names, there is vast scope for such research. Virtually all of the work of identification undertaken to date, however, has been antiquarian in conception, its objective being the identification of particular individuals, of the owners of particular manors, of the tenants of a particular lord, or of the members of a particular family. No previous attempt has been made to identify all the individuals named in Domesday over a wide area in order to analyse the social groups to which they belonged. To attempt this for the entire country is, in a sense, absurdly ambitious. But a sensible methodology permits a surprisingly large proportion of the possible identifications to be achieved within a reasonable timescale. By profiting from the work of previous scholars, and by using the basic original printed sources and the most accessible secondary materials, we have

tried to achieve maximum results within a modest period, leaving manuscript sources and the more recondite original and secondary sources for later investigation. On the basis of work done on circuit three, where all the sources have been explored for some counties, we estimate that 90-95% of all possible identifications will be achieved by the strategy we have adopted, leaving the remaining 5-10% for later consideration.

Methodology: Most serious Domesday scholars have attempted some identifications, yet none have explicitly described their method in any detail. The following description may therefore prove helpful. Two preliminary points need to be made. First, given the many thousands of tenants named only by their christian names and the very large numbers of the more popular christian names, it is impractical to search for all potential linkages. Any attempt to establish linkages between all possible combinations of the 1446 occurrences of 'Robert', for instance, would by itself have consumed all the time devoted to researching identifications to date! Hence the emphasis in what follows on looking for initial clues as to likely connections. Secondly, this description deals only with the more important methods of identification. There are a variety of lesser techniques which have been pressed into service. Identifications may be suggested, for instance, by plotting the distribution of properties held by men with the same christian name on a map; and connections between Domesday manors and their dependencies often afford valuable clues to further identifications.

Preliminaries: Within each county, a separate index card is prepared for each name, as all names at this stage are assumed to represent different individuals unless the text makes it clear otherwise. Thus, if two successive entries name a Roger as tenant, separate cards are written for each unless the second entry uses the phrase 'the same Roger', or its equivalent. The following information is recorded on the cards: Name, County, Hundred, Saxon Predecessor, Tenant in Chief, these items being crucial to the identification process. More miscellaneous information which may prove useful, or which may be required for other purposes, is written on the back of the card. When all cards for a county have been compiled, they are sorted into alphabetical order of tenants.

Identifications which have been made by other historians are first examined, to eliminate the easy cases. The notes to the Phillimore edition, the Domesday chapters in the volumes of the *Victoria County History*, and the detailed manorial descents given in the same work (where they have been published), are the most natural starting

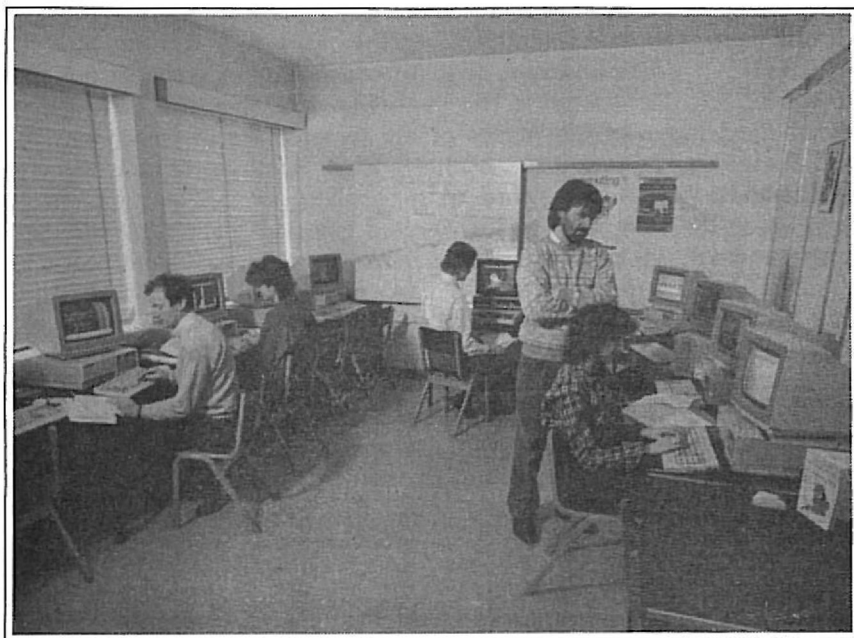
points, followed by inspection of the works of Clay, Farrer, Salzmänn, Round, the best of the county historians, and the articles scattered for the most part among the county record society and antiquarian publications (to be used with caution). Finally, original materials with a strong local bias are best investigated at this stage. Among these are satellite surveys, or Domesday-like surveys of a slightly later date, collections of local charters, and the charters of monastic houses in the locality. Identifications which are felt to be well established by any of these sources are noted, and the relevant cards are stapled together to indicate this fact. (More tentative identifications are indicated by joining the cards with a paper clip).

Using the Database: At this point, the Domesday database can assist the process of identification if the county under scrutiny has had its marker codes inserted. In such cases, the Domesday software can be useful in a variety of ways. Domesday Book itself is sorted by tenant in chief, all the properties of each tenant in chief within each county appearing consecutively. This is the most unhelpful arrangement which could have been devised from the point of view of anyone attempting to make identifications. It is normally a simple matter to identify the tenant in chief, since the great majority have surnames. Having their properties grouped together therefore confers little or no benefit. It does, however, tend to conceal other clues to possible identifications. Restructuring the entries so that they are arranged in alphabetical order of the names of the tenants, for instance, may reveal that some tenants in chief have large numbers of tenants with the same christian name (which is suggestive), or that two tenants in chief share the same (or similar) limited group of names among their tenants, which is even more suggestive of potential identifications. Similarly, sorting the entries by Saxon predecessor will sometimes disclose that a particular Norman tenant has apparently stepped into the shoes of the same Saxon lord each time, a strong clue to identification. Finally, and most usefully, if the entries are restructured by alphabetical order of the villages in which they lie, a crop of identifications are often indicated. Such an order may well reveal, for instance, that all or most of the properties in the vill are held by tenants with the same christian name, even though held from a variety of tenants in chief. In such cases, it frequently transpires that all the tenants are in fact the same individual. Further, it is not uncommon to find several tenants in the same vill with the same christian name, which is also the christian name of a tenant in chief holding property in the vill. This is a strong indication that the tenant in chief is the same man as the tenants bearing his christian name, he having tried to obtain total

control of the village by acquiring tenancies in the vill from other tenants in chief.

It must be stressed that none of these indications may be taken as proof of identification at this stage. At most, they may lead to different index cards being joined by paper clip. But they do point strongly to possible links which should be investigated. Knowing what to look for saves very considerable amounts of research time.

Finally, the basic sources for manorial history in the twelfth and thirteenth centuries, such as *Feudal Aids*, the *Book of Fees*, the *Red Book of the Exchequer*, and the *Calendar of Documents: France*, are searched to see whether manors held by men with common christian names in Domesday were held by the same family at a later date. If they were, this is taken as proof of identification. The list of Domesday vills and their tenants generated from the Domesday database are invaluable at this stage, since most of these later sources (unlike Domesday) are arranged geographically, by county and vill. Having the Domesday information sorted in the same order as that of these later sources allows comparisons between the two to be made much



The Computing Laboratory, Domesday Unit. (University of Hull Photographic & Copy Service, Brynmor Jones Library).

more efficiently, resulting in considerable economies of research time and effort.

Once all tenants within a particular county have been investigated in this manner, their index cards are conflated with those of other counties within the same circuit, and sorted together. This process will throw up further identifications across county borders. Ultimately, the cards for all seven circuits will be combined into a single index.

Updating the Database: As the identifications for each circuit are completed, all occurrences of names within the circuit have a number appended to them which is unique to each identified individual (not to each name). This serves to confirm the identifications and to link occurrences of the same name which belong to one individual, distinguishing them from other occurrences of that name which belong to other individuals. This is useful to the user but absolutely indispensable for the software, which would otherwise identify all occurrences of a particular christian name as references to a single individual, thereby reducing the number of landowners in Domesday to a few score individuals! The unique reference number also identifies for user and software the different name forms borne by one individual. Only actual names are uniquely tagged in this manner. Anonymous individuals or groups — ‘a cripple’, or ‘two men at arms’ — are not assigned these numbers.

The numbers have been assigned circuit by circuit as the optimum means of avoiding duplication. To diminish this possibility further, the following precautions have been taken. First, all tenants in chief were assigned their numbers before the process of dealing with tenants on a circuit by circuit basis began. As each new circuit was dealt with, individuals who were tenants in chief anywhere in the country but who held as tenants within that circuit, were first eliminated. A search was then made among tenants to establish whether any had previously been assigned a number in another circuit (a process facilitated by the fact that the circuit indexes are held in alphabetical sequence of tenants). Finally, tenants without such numbers were assigned them sequentially, in alphabetical order. The numbers were written in the top right hand corner of the index cards. At a later stage, these numbers were inserted into the database by appending them to all occurrences of the name belonging to each individual.

Database related files: When the project is complete, users will have access within the Domesday system to the assembled proofs for the identifications we have made. The unique reference numbers assigned to each individual will serve as a pointer, which links the Domesday text database to an external file containing electronic ‘foot-notes’ to this and other pertinent material. The software for this is not

yet in place, and as an interim measure we are using the bibliographic package 'Famulus' to store this information. The transfer of data from Famulus to our customised software will present no problems. Within the Famulus file, one record is created for each identified individual, with the following fields (not all of which need be used for each record):

UNUM	Unique personal number
DBNM	Form of name in Domesday
SRNM	Surname (or additional name) if known from other sources
FAMN	Later Family name, or name of family to whom bulk of the lands descendes
COUN	Counties in which this individual held land
BIBL	References to information directly relating to this individual, primary sources being referred to in the COMM field
RELT	Relations who appear in Domesday
XREF	Other tenants with whom this individual might possibly be identified
COMM	Reasons for identifications supported by primary sources

CONCLUSIONS

This description of the project does not cover all aspects of the work we are doing. Although it has attracted a lot of attention, I have not described the mapping, partly because we have fundamentally rethought our approach in the past few months and are now rewriting the software, but mainly because this aspect of the project has recently been described elsewhere.⁽⁷⁾

A question we have frequently been asked is when, and how, the package will be made available to others. The 'how' has been implicitly answered in this paper. All the software is being re-written for IBM PC compatible machines, on which we hope it will ultimately be available. We are producing a prototype version on the PC for Oxford University Press which will be tested by them later this month. On the basis of those trials, a decision will be made about a possible electronic publication by O.U.P.

Even if the decision is positive, however, it is almost impossible to predict at the moment when it will become available. The main constraint will be the research costs of completing the coding of the data and the research on identifications. The work to date has been funded by the Economic and Social Research Council and various

lesser grants. But our major funding has all but expired and it is difficult to persuade funding bodies (not many of whom make large grants in the Arts) to support a partially completed project. In different circumstances, younger members of staff might have become involved in a project of this kind. But younger members of staff are very thin on the ground in English universities these days. The youngest member of this department is over forty, the last appointment at the assistant lecturer or lecturer level having been made as far back as 1974 (despite a 40% loss of staff in the interval). The continued funding of two research assistants is therefore crucial to the completion of the project. This is a very frustrating situation. It is a truism that there is a dearth of good software for teaching and research in the Arts. We have developed software which could make a major contribution to both teaching and research applications on a body of material of interest to a wide scholarly community.

But it would be misleading to end on a downbeat note, for the project has a number of achievements to its credit. In the past twelve months, the three members of the Domesday Unit have delivered two dozen lectures (including 9 conference papers), given 30 software demonstrations, mounted 6 exhibitions, written 10 commissioned articles, and travelled some 20,000 miles in the process. Despite its unfinished state, the project has also begun to make valuable scholarly contributions to the discipline.⁽⁸⁾ Hopefully, this is the beginning of a steady stream of such contributions.⁽⁹⁾

NOTES

1. This article is based upon research funded by the Economic & Social Research Council (ESRC) reference number B00232095.
2. J.H. Round, *Feudal England* (London, 1985), p. 26.
3. F.W. Maitland, *Domesday Book and Beyond* (Cambridge, 1897), p. 407.
4. The Santa Barbara database is described in: Robin Fleming, 'A Report on the Domesday Database Project', *Medieval Prosopography* vol. 7 (1986), pp. 55-61.
5. E.S.C. Weiner, 'The New OED: problems in the computerisation of a dictionary', *University Computing* vol. 7 (1985), pp. 67-71; J.C. Gray, 'Creating the Electronic New Oxford English Dictionary', *Computing and the Humanities* vol. 20 (1986), pp. 25-49.
6. The coding added to the text is shown in bold type. The arabic numerals in the Latin text are codes used in data input to signify the curious characters in the original which do not appear in any normal character set. These characters have to be graphically defined, but it is a fairly simple task to do so on the BBC and IBM PC machines. They can be seen in the facsimile.
7. J.J.N. Palmer, 'Computerising Domesday Book', *Transactions of the Institute of British Geographers* new series vol. 11 (1986), pp. 279-289.

8. See previous note, plus J.J.N. Palmer, 'The Domesday Manor', in: *Royal Historical Society Domesday Essays*, ed. J.C. Holt (London, 1987), pp. 139-153; Andrew Ayton & Virginia Davis, 'Ecclesiastical Wealth in England in 1086', in: *The Church & Wealth* (Studies in Church History, 24), ed. W.J. Sheils & D. Wood, London 1987. We are also producing two substantial volumes of indices to the 40 volume edition of Domesday Book published by Phillimore & Co. between 1975 and 1986.
9. The two research assistants who contributed to these activities are Dr. Virginia Davis and Mr Andrew Ayton. All the software for the project has been written by Mr George Slater of the Computer Centre, the one absolutely indispensable member of the team. We count ourselves extraordinarily fortunate to enjoy his enthusiastic contribution.

Dr. J.N.N. Palmer is Head of the Domesday Unit and Senior Lecturer in the Department of History, University of Hull.

PROGRAMMING IN SPITBOL FOR HISTORIANS

A REPORT ON WORK IN PROGRESS ON THE HISTORY OF THE UNIVERSITY OF OXFORD

Daniel Greenstein

In 1968 a small research group was established at Oxford University; its brief was to research and write the official history of the University and to present this work in at least six volumes, two volumes for each period of the University's history. Since its inception, the dimensions of the project have changed to include the publication of eight volumes chronologically arranged: two on the middle ages; three on the early modern period to 1800; two on the nineteenth century; and a final volume on the twentieth century.⁽¹⁾ An integral part of the research has been an analysis of the University's role in the process of social selection and differentiation in Britain. In general this has entailed the construction of collective biographies about Oxford students, their socio-economic or class origins, their University achievements, and their occupational and social destinations after leaving the University. Anyone who has ever conducted genealogical research or compiled a collective biography of historical figures can appreciate the immense amount of work involved in producing a representative sample of names, and in looking through a variety of disparate professional, business, and city directories and elite biographical registers for information about the lives of individuals in the sample. In January 1986 when work began on the final volume of the History of the University of Oxford, it immediately became clear that the origins and destinations study of Oxford's twentieth-century students could be facilitated with published resources the likes of which were simply not available to researchers working on earlier periods. In fact, it was established that we could use an optical scanner to computerize some basic information about the names and academic achievements at Oxford of the 150,000 students who had registered at the University

since 1900. If that information could then be entered into a database, it would be possible to generate a genuinely representative sample of Oxford's twentieth-century students. Moreover, the task of collecting biographical information about the people in our sample would be made less onerous by the fact that as many as two-thirds of the sample population would still be living and could therefore be surveyed. The use of an optical scanner to record information and then to enter it into a database, however, involved a number of computing problems. Those problems and their solution using the text-oriented language called SPITBOL, form the topic of this paper.⁽²⁾

Since 1565, Oxford University has been required by statute to keep a written record of each of its matriculants.⁽³⁾ That is, a record of the name of every student who was admitted to the University, their father's occupation, and the amount and type of schooling they had received prior to coming to Oxford. Since at least 1870, a portion of the matriculation register has been published a few times a year in the weekly *Oxford University Gazette*.⁽⁴⁾ These published matriculation lists contain the name of each matriculant, and the name of the Oxford college or society that they attended at the date of their matriculation. This latter piece of information is particularly important since Oxford is a collegiate University in which differences between colleges may be an important variable in an analysis of the University's role in the process of social selection. Since 1870 the University has also published class lists in some of the issues of its *Gazette*.⁽⁵⁾ These are lists of the names of those individuals who succeed in passing their final examinations which include for each individual on the list, the name of their college at the time of their examination, the subject in which they took their final examination, and a record of their performance on the examination. With the use of an optical scanner, (in our case a Kurzweil Data Entry Machine, or a KDEM) the information that is available on matriculation and class lists can be entered into a computer where it is recorded in exactly the same format as it appears on the printed page.

The immediate problem, of course, is that a computerized replica of the published class and matriculation lists is not conducive to typical database operations (like sorting and selecting) which are necessary to generate a sample of individuals for whom more detailed biographical information can then be collected. With all of the material on the class and matriculation lists entered into the computer, we have two rather different sets of information; two files full of names if you prefer. The first file contains information about students who matriculated. The second file contains information about students who success-

fully completed their courses (graduate and undergraduate) at Oxford. The first objective, then, is to combine for each individual in the sample, the information about that individual that is contained in the two separate files, thus creating a third comprehensive file. But even before negotiating the problem of 'record-linkage', it is necessary to clean up the material in the two original files by separating information about each individual into its component parts.

The cleaning-up operation is demonstrated in tables 1 to 4 below. Table 1 shows ten typical entries as they might appear in the file of matriculants created by reading the published matriculation lists with the optical scanner.⁽⁶⁾ Table 2 shows how we want that information to look after processing. The format of Table 2 should be more familiar to anyone with some experience of standard databases. Each record is broken up into its component parts, last name, first names, college, and date, and the different component parts are entered into separate columns.

Table 1.

7 May 1941

TRINITY COLLEGE

Arthur, Thomas John Hampden

Barbel, Anthony Richard Mark

Butling, Joseph Goold

Carlin, William James

Clan, Philip Rhodes

Cooker, Basil Kieth

Corn, John Court

Colt, Neill Dudley

MAGDALEN COLLEGE

Curson, Peter Thomas

Dandy, Arthur James

Table 2.

LAST NAME	FIRST NAME	OTHER NAME 1	OTHER NAME 2	COLLEGE	DATE
Arthur	Thomas	John	Hampden	TRINITY	7 May 1941
Barbel	Anthony	Richard	Mark	TRINITY	7 May 1941
Butling	Joseph	Goold		TRINITY	7 May 1941
Carlin	William	James		TRINITY	7 May 1941
Clan	Philip	Rhodes		TRINITY	7 May 1941
Cooker	Basil	Kieth		TRINITY	7 May 1941
Corn	John	Court		TRINITY	7 May 1941
Colt	Joseph	Dudley		TRINITY	7 May 1941
Curson	Peter	Thomas		MAGDALEN	7 May 1941
Dandy	Arthur	James		MAGDALEN	7 May 1941

The information contained in the file containing the class lists is then transformed in a similar fashion. Table 3 shows a sample of names drawn from the file created by reading the published class lists with the optical scanner. The table provides data about students who passed their final examinations. It shows, for example, that in the summer of 1933, Hamish, Prince, and Adams, obtained a first class distinction in their final examination in Physics. It also shows that in the same year, Altock, Collson, Dithers, and others, obtained a second class degree in Physics. The table also gives an abbreviated form of the Latin name of the college that each individual attended at the time he or she took the final examination in Physics. Hamish was at New College, Adams at St. Hugh's, Hoodwink at St. John's College. As with the information on the matriculation lists, the material in the class lists has to be broken down into its component parts and entered into a table each row of which contains information on one person, and each column of which contains a discrete kind of information. Table 4 shows how the information in Table 3 will look after it has been transformed.

Table 3.

11 July 1934

IN PHYSICA

CLASSIS I

Hamish Michaelmas A. e. Coll. Nov.

Prince Franciscus V. e Coll. Tri.

Adams Mitch G. Coll. e S. Hug.

CLASSIS II

Altock Ericus ex Aul. Priv. S. Petri

Collson Bernardus A. e Coll. Reg.

Dithers Galfridus F. e Coll. Nov.

Hoodwink Thomas W. e Coll. D. Jo. Bapt.

James Willelmus P. e Coll. Reg.

Lewis Edwinus T. e Coll. Aen. Nas.

Morris Robertus e Coll. Nov.

Table 4.

LAST NAME	FIRST NAME	OTHER NAME	COLLEGE	SUBJECT	CLASS	DATE
Hamish	Michaelmas	A.	Coll. Nov.	PHYSICA	I	11 July 1933
Prince	Franciscus	V.	Coll. Tri.	PHYSICA	I	11 July 1933
Adams	Mitch	G.	S. Hug.	PHYSICA	I	11 July 1933
Altock	Eric		S. Petri.	PHYSICA	II	11 July 1933
Collson	Bernardus	A.	Coll. Reg.	PHYSICA	II	11 July 1933
Dithers	Galfridus	F.	Coll. Nov.	PHYSICA	II	11 July 1933
Hoodwink	Thomas	W.	Coll. D. Jo. Bapt.	PHYSICA	II	11 July 1933
James	Willelmus	P.	Coll. Reg.	PHYSICA	II	11 July 1933
Lewis	Edwinus	T.	Coll. Aen. Nas.	PHYSICA	II	11 July 1933
Morris	Robertus		Coll. Nov.	PHYSICA	II	11 July 1933

Much as the transformations may seem dramatic, they rely on straightforward computer programming in SPITBOL, a language which is designed to recognize and manipulate text. The information in Table 1, for example, is cleaned up with a programme which is based on a number of rules which provide for the correct interpretation of the material contained in the matriculation file.

Rule 1. Each record (each line of text) is either a date, the name of a college, or the name of a person.

Rule 2. Dates are identified when a record contains an uninterrupted sequence of four digits (the year). Every time a date is identified it should be assigned to the variable DATE. DATE should then be associated with each subsequent record containing an individual's name until a record is reached which contains another date. When this happens, the second date is then assigned to the variable DATE.

Rule 3. College names are identified when a record contains the character string 'COLLEGE'. When a record contains a college name, the name (minus the character string 'COLLEGE')

should be assigned to the variable COLL. The character string contained in COLL should then be associated with every record containing a person's name until the next occurrence of a record which contains the word 'COLLEGE'. When a second college name is identified it should replace the character string contained in the variable COLL.

Rule 4. A record is identified as containing a name when it contains a comma. The name is then broken down into its various component parts. Letters preceding the comma are assigned to a variable LAST, a sequence of letters following the comma and preceding either the end of the line or the next blank space, are assigned to the variable FIRST. Subsequent uninterrupted sequences of letters are assigned to variables SECOND, THIRD, and so on until the end of the line is reached.

Having divided the name into its component parts the individual's name, college, and date of matriculation can then be output in a fixed-field format using a SPITBOL utility which allows variables (in this case, LAST, FIRST, SECOND, THIRD, COLLEGE, and DATE) to be padded to either the left or the right with blank spaces (or any other characters for that matter). After the variables have been padded with the appropriate number of blank spaces they can be output as a row in a table where each column is left-hand justified.

OUTPUT = LAST FIRST SECOND THIRD COLL DATE

Before reading the next record, all the variables containing parts of the name e.g. LAST, FIRST, SECOND, THIRD are set equal to null.

A similar programme is used to transform the information contained in the file containing the class lists to produce a data table like the one shown in Table 4. Only the programme to clean up the class list file has to use slightly different rules to ensure that individuals' names are associated with the appropriate examination dates, subjects, and results. Records containing dates are identified and assigned to DATE when they contain an uninterrupted sequence of four integers. Subjects and examination results are identified and assigned to the variables SUBJ and RES respectively when a record consists entirely of upper-case letters. Exam results are distinguished from subjects by the occurrence of the character string 'CLASSIS'. Finally, since the class lists file does not use commas between last and first names, names must be identified by the occurrence in a record of lower-case

letters. When a record containing a name is found, the name is broken into its component parts distinguished by the occurrence of blank spaces between uninterrupted sequences of characters up to the occurrence of the character string 'e' or 'ex'. All characters after the string 'e' or 'ex' are known to contain the name of the individual's college and those characters are placed in the variable COLL. Following these simple rules, a programme is written to reorganize the data shown in Table 3 so that it appears in the form shown in Table 4.

Once both files have been tidied up, it is possible to produce a third file in which the information about each individual that appears in the class and matriculation files is combined. It is necessary to work in this direction, from the class file to the matriculation file since we can be sure that everyone who passed a final examination matriculated in the University while we cannot be certain that everyone who matriculated in the University eventually passed a final examination. An abbreviated sample record in this third comprehensive file would look something like this:

LAST NAME	FIRST NAME	SECOND NAME	COLLEGE AT MAT.	DATE OF MAT.	COLLEGE AT EXAM	DATE OF EXAM	SUBJ.	CLASS
Owen	John	James	Tri.	11 Nov. 1934	Tri.	26 May 1937	Math	II

Record-linkage is achieved with a programme developed which reads records one at a time from the class file. After each record is read from the class file, the component parts of the name are identified and placed in separate variables, LAST, FIRST, SECOND. The identification of the component parts of a record is now much easier since the data appears in a fixed-field format. That is, we can be sure that the last name, for example, is contained in the first x number of characters in a record where x is the width of the last-name field. The programme then instructs the computer to search through the matriculation list for an occurrence of a record in which the component parts of the name are identical to those names contained in the variables LAST, FIRST, and SECOND. An identical record is considered to be found only after one or several previously defined conditions or rules have been satisfied. A sample of the rules that are required in this programme include the following, ranked in order of their priority:

Rule 1. At Oxford there is a set number of years within which an individual must complete his or her course. The number of

years depends on the course. An undergraduate degree in Chemistry, for example, must be completed within seven years while a Diploma course in Anthropology must be completed in three years. Therefore it is possible to establish the rule that a person who is known to have taken an examination in year x must have matriculated between the years $x - 1$ and $x - y$ where y is the maximum amount of time allowed for the completion of a given course. A number of ancillary rules are included to ensure that an appropriate value is assigned to y each time a record is read from the class list. Having read the first record in Table 4, for example, PHYSICA is identified as the subject and 7 is assigned as the value of y based on a rule which specifies that the Physics course must be completed within 7 years of matriculation. This rule and its subsidiaries are very important as they cut down on computer search time by instructing the computer to search for the name read from the class file in those records of the matriculation file which contain a year within the specified range. In effect, this will limit the search for any one name to a maximum of 12,000 records in the matriculation file which contains a total of approximately 150,000 records.

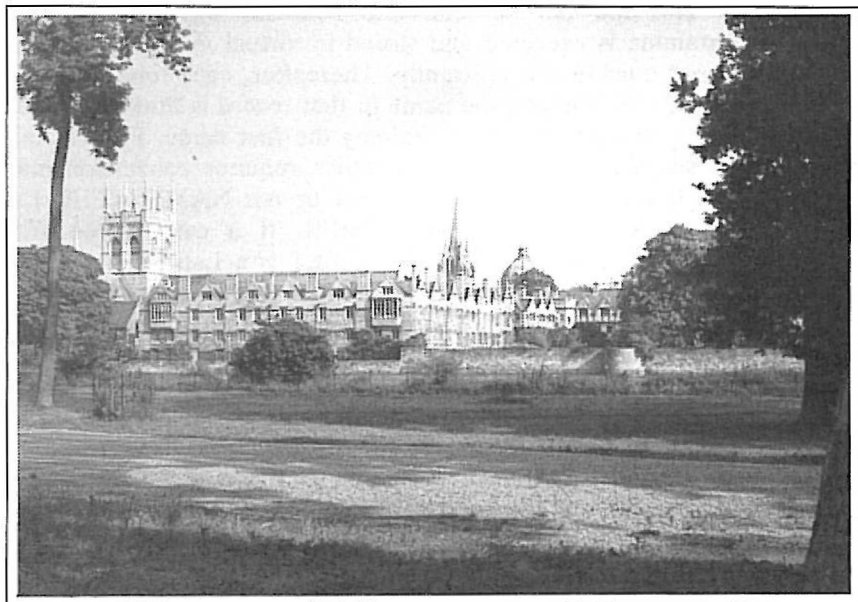
Rule 2. Having established where to look in the matriculation list, it is then necessary to determine what to look for. Ideally the computer is searching the relevant portion of the matriculation file for the occurrence of a name which is identical in all respects to the one contained in the record that has been read from the class file. Therefore, we establish a rule which says that when an exact match of LAST, FIRST, and SECOND names occurs within the range of years specified by Rule 1, combine the information from the record in the matriculation list with the information of the record in the class lists and put it into our new file.

Rule 3. Notice in Table 3 that the class file only provides the first initial of the second name. This is not true of the entire class file, as some of the records contain the name in its entirety. We do have to allow, however, for the possibility that one or other of the files will contain records which only contain a last name followed by initials. A third rule is therefore designed to ensure that if Rule 2 is not satisfied, then a match

will be said to have occurred if the last name and all subsequent initials of a record read from the class list are identical to a last name and subsequent initials of a record read from the relevant range of the matriculation list, the relevant range having been determined by Rule 1.

Of course, there is one other small problem that we have not yet accounted for. That is that the class file uses the Latin equivalent of English first names. Therefore, before we can link a record in the class file with the relevant record in the matriculation file, it is necessary to convert the Latin first names back into their English equivalents. Fortunately, each Latin name only has one English equivalent. Names such as Mitch which have no Latin equivalent are printed in English.

Given the consistency with which Latin names were substituted for English names, it is possible to convert any one Latin name back into its original English form. In SPITBOL this procedure is very straightforward with the use of a TABLE. A TABLE can be conceptualized as a two-dimensional array with two columns. Unlike a normal array, however, the first column of a TABLE does not contain a variable.



*Change without the appearance of change?
Merton College, Oxford.*

Instead, it contains a unique label which identifies the information contained in the second column of the same row. Because the information in the first column of a row identifies the row, no two labels in the first column can be identical. Table 5 shows data which can be used to construct a simple TABLE called NAMES.

Table 5.

Ricardus	Richard
Michaelmas	Michael
Galfridus	Geoffrey

In a programme which uses NAMES as a TABLE, NAMES<Ricardus> refers to the second column of NAMES in the row labelled Ricardus and contain the character string 'Richard'. Similarly, NAMES<Michaelmas> contain the character string 'Michael'. As importantly if a variable called VAR contains the character string 'Ricardus' then NAMES<VAR> will contain the character string 'Richard'.

Since there is a finite number of Latin first names, a file can be prepared which contains a table of Latin first names and their English equivalents. This file can be read into NAMES when the record-linkage programme is executed and stored in virtual memory where it can be accessed quickly and efficiently. Thereafter, each time a record is read from the class file and the name in that record is broken up into its component parts the variable containing the first name, FIRST, can be tested to see if it is a Latin name which requires conversion into English. This is done by checking whether or not NAMES<FIRST> exists in the look-up table called NAMES. If a call to NAMES <FIRST> succeeds then we know that FIRST is a Latin name which is listed in NAMES with an English equivalent. In this case the computer is instructed to search the matriculation file for a name which matches the name from the class file represented as LAST NAMES <FIRST> SECOND. If the call to NAMES<FIRST> fails, then we know that FIRST is not a Latin name. This being the case, the computer is instructed to search the matriculation file for an occurrence of a name represented as LAST FIRST SECOND. Other TABLES tables can be constructed and incorporated in the record-linkage programme so that Latin college names and subject names are also converted into their English equivalents before any combined records are output into our new file.

In actual fact, the data structures are more complicated than the ones shown in the tables presented above. Hence many more rules

required in the programmes that I have described. Nevertheless, we have found that SPITBOL is perfectly suited to these problems of data management and record-linkage which are so often encountered in historical investigation of this sort. In fact, the facilities which a text-oriented language provide have made the entire project possible. Without them, the information contained in the published class and matriculation lists could only be entered by hand at great expense, and accurate record-linkage would have proven well nigh impossible. More importantly, by using the routines described above, we are able to produce a database which contains basic information about each of Oxford's 150,000 twentieth-century students. The database can then be used to generate a smaller sample of Oxford students for whom biographical information is collected. With the information contained in the database the sample can be made representative of Oxford's student body as a whole as we can ensure that each college, subject, and year of matriculation is represented in the sample in the same proportion that it is represented in the entire population. Without the database, less representative samples would inevitably have been constructed and the validity and accuracy of our analysis of the social origins and destinations of Oxford's twentieth-century students would have been severely limited.

NOTES

- 1) Under the late general editor, T.H. Aston, the following volumes have already been published: J.I. Catto, ed., *The History of Oxford University. Volume I, The Early Oxford Schools* (Oxford, 1984); J. McConica, ed., *The History of the University of Oxford. Volume III, The Collegiate University* (Oxford, 1986); Dame Lucy Sutherland, *The History of the University of Oxford. Volume V, The Eighteenth Century* (Oxford, 1986).
- 2) MACRO SPITBOL is an implementation of the SNOBOL4 computer language initially designed and implemented by Robert B.K. Dewar of the Courant Institute of Mathematical Sciences, and Prof. A.P. McCann of the Department of Computing at Leeds University. Essential references include Griswold, Poage and Polansky, *The SNOBOL4 Programming Language* (Addison, Wesley, 1971).
- 3) See J. McConica, *op.cit.*, pp. 49-50.
- 4) The largest single annual list of matriculants appears in an autumn issue of the *Gazette* since most students entered the University in October.
- 5) Class lists tend to be found in the May and June issues of the *Gazette* since this is when most students took their final examinations.
- 6) Due to the fact that our licensed status under the Data Protection Act is still pending, all of the names in the Tables have been altered.

Daniel Greenstein is conducting the research on the social origins and destinations of Oxford's twentieth-century students for Volume 8 of The History of the University of Oxford.

TRADISJON OG TEKNOLOGI

EDB OG HUMANIORA VED UNIVERSITY OF OXFORD

Kristin Natvig

I april besøkte undertegnede University of Oxford for å rapportere om den senere tids utvikling innen edb og humaniora ved dette tradisjonsrike universitet. Oxford University Computing Service (OUCS) har i en årrekke stått for sentrale tjenester for humanister — både nasjonalt og internasjonalt — som optisk lesing, fotosetting og sist men ikke minst, Oxford Text Archive.

Alle disse tjenestene, i tillegg til Oxford Concordance Program, ble beskrevet i HD 1-86. Her skal bare «siste nytt» omtales, og et par edb-prosjekter vil bli presentert. Takket være *Lou Burnard* og *Catherine Griffin* ved OUCS fikk undertegende anledning til å intervju flere humanistiske forskere ved universitetet som arbeider med edb-baserte prosjekter. En av disse forskerne, *Daniel Greenstein*, beskriver sitt prosjekt i en egen artikkel i dette nummeret av HD.

OXFORD UNIVERSITY COMPUTING SERVICE

Edb-tjenesten for humanister ved OUCS ledes av *Susan Hockey* og har følgende medarbeidere: *Lou Burnard*, som er ansvarlig for både Oxford Text Archive og databaseapplikasjoner (sammen med *Paul Salotti*), *Catherine Griffin*, som driver fotosettingstjenesten og fire personer som arbeider med optisk lesing. En stilling som ekstern rådgiver for edb i humaniora er p.t. ledig, men skal besettes i slutten av september.

Fotosetting

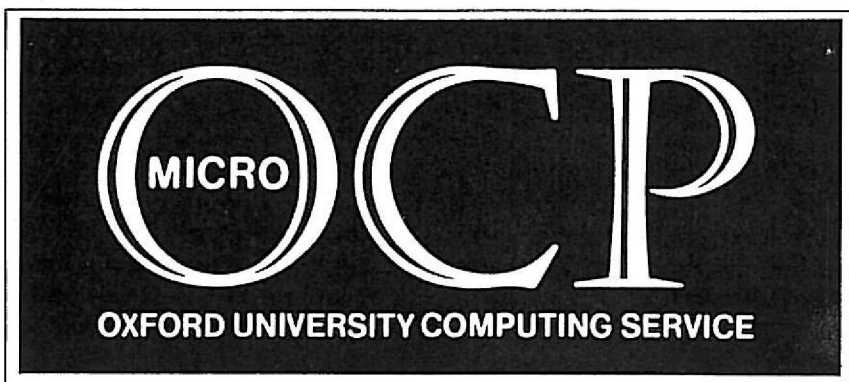
Catherine Griffin arbeider særlig med fotosetting av ikke-europeiske alfabeter, og har nylig skrevet et program for fotosetting av hieroglyfer som et oppdrag for en professor i egyptologi. Etterspørselen etter fotosettingstjenester er stor, men mange forskere har ikke bruk for

trykksaker av høy kvalitet. Det er derfor nylig kjøpt inn en laserskriver som er koblet til OUCS' nye VAX Cluster. Laserskriveren har PostScript og mulighet for et stort antall fonter, både kommersielle og egenutviklede.

Nytt om Oxford Concordance Program

Lou Burnard kunne fortelle at en ny stormaskinversjon av Oxford Concordance Program (OCP) skal lanseres senere i år. Grensesnittet er det samme som for versjon 1. Sortering foretas mye hurtigere enn tidligere, og datalagringen er mer økonomisk.

En mikromaskinversjon av OCP er kommet på markedet nå, i regi av Oxford University Press. Denne versjonen kan kjøres på IBM-kompatible maskiner med harddisk og minne på minst 512 kb. Den har et enkelt skjermorientert grensesnitt slik at brukeren kan definere kommandoer selv, men er ikke fullstendig interaktiv. Micro-OCP er for øvrig identisk med versjon 2 av stormaskinutgaven. Den selges i en pakke med to programdisketter og en diskett med et utvalg tekster for demonstrasjons- og øvingsformål. Pakken koster \$275 for enkeltpersoner, \$395 for institusjoner og kan bestilles fra: *Oxford Electronic Publishing, Oxford University Press, Walton Street, Oxford OX2 6DP.*



Burnard ble spurt om forskjellen mellom OCP og WordCruncher (et lignende program utviklet ved Brigham Young University). Fordelelen med det sistnevnte programmet er at det er mulig å konvertere en tekst til en leksikonfil før videre behandling, slik at sorteringen foretas én gang for alle. OCP sorterer kun de ordene som brukeren ber om. WordCruncher har imidlertid et vesentlig mer komplisert grensesnitt, og fungerer ikke så effektivt på ikke-europeiske alfabeter som OCP. For øvrig har de to programmene svært like fasiliteter.

Programvare for CAFS

CAFS (Content Adressable File Store) er tilleggsutstyr til ICL stormaskiner spesielt utviklet for sorteringsformål. I stedet for bruk av programvare er det selve *maskinvaren* som gjør fritekstsøking mulig i store mengder ustrukturerte tekstdata. Innholdet i en tekst blir direkte identifisert, slik at indekseringsprogrammer er overflødige.

Lou Burnard har skrevet et generelt gjenfinningsprogram som utnytter CAFS' potensiale slik at Oxford Text Archives' omfattende tekstkorpora kan utnyttes bedre. Programmet er hittil blitt anvendt ved Oxford på bl.a. en database over Shakespeares verk, tekstene til Thomas Hardys romaner, i eksperimentelt arbeid på tekster i korpuset Thesaurus Linguae Graecae og katalogen til the Bodleian Library fra før 1920.

Et nytt prosjekt

Arbeid har nettopp startet på et to-årig prosjekt finansiert av the Computers in Teaching Initiative. Formålet med prosjektet er å gjøre pensumlitteraturen i språklige fag og litteraturstudier tilgjengelig for on-line søking og produksjon av konkordanser og indekser via en hvilken som helst IBM-kompatibel arbeidsstasjon ved universitetet. Arbeidsstasjonene er koblet til den nye VAX Cluster.

Prosjektmedarbeiderne skal utvikle «idiotsikre» omgivelser for standard programvare for tekstanalyse. Dette systemet skal også være tilgjengelig for hovedfagsstudenter og det vitenskapelige personale.

Prosjektet utføres som et samarbeid mellom edb-tjenesten for humanister ved OUCS, The Computing Teaching Centre, og instituttene for engelsk, moderne språk, orientalske språk og klassiske fag. Det skal gis undervisning i bruk av systemet i gammelengelsk, italiensk renessanselyrikk, Lucretius og Catullus og klassisk arabisk.

To medarbeidere er engasjert til prosjektarbeidet. Det første prosjektåret skal gå med til skriving av programvare og tilrettelegging av tekster, og i det andre året skal systemet evalueres som undervisningsredskap.

Framtidsplaner

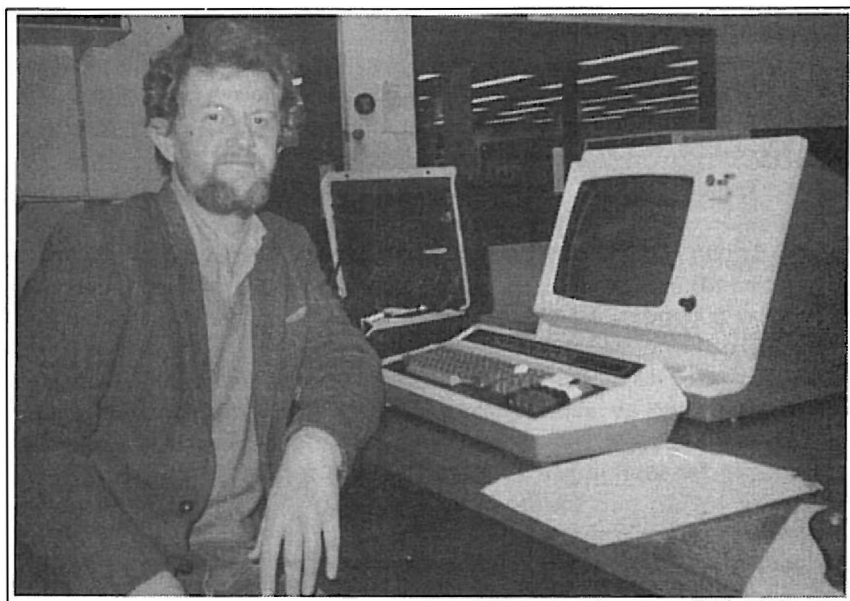
OUCS har nylig fått innvilget et forskningsstipend av the British Library fra 1.1.88. Stipendiaten skal være ansvarlig for en ett-årig undersøkelse av pågående og potensielle faglige anvendelser av maskinleselige tekster.

Lou Burnard har søkt om finansiering til produksjonen av en CD-ROM som skal inneholde et stort korpus med engelske renessanse-dramaer. På ønskelisten står også utvikling av enklere metoder enn de nåværende for koding og integrering av tekster, og utgivelse av en veiledning i koding av litterært materiale.

FOTOSSETTING AV ARABISK

John Cooper, Institute for Oriental Studies, St. John's College har arbeidet med fotosetting av arabisk i fire år. På OUCS' Lasercomp-fotosetter bruker Cooper et Monotype-program for moderne arabisk. Dette programmet kan håndtere vokaltegn, men ikke de antikverte grafiske representasjonene av «vokalene» som fins i de eldste utgavene av Koranen. Cooper har derfor utviklet sitt eget mapping-system som brukes sammen med programmet for å løse dette problemet. I systemet er de grafiske lydrepresentasjonene kodet og lagt inn i en tabell. Monotype-programmet velger så ut riktig bokstavform ut fra kodingen og tabellen slik at lydene gjengis riktig på fotosats.

Cooper tar på seg fotosettingsoppdrag for andre og bruker mapping-systemet i sitt eget arbeid med arabiske manuskripter. Han har



John Cooper.

produsert all fotosatsen til sin arabisk/engelske utgave av *Tabari. Commentary on the Qu'ran*, vol. 1, Oxford University Press 1987.

LEXICON OF GREEK PERSONAL NAMES

Elaine Matthews, som har arbeidsplass ved University of Oxford, er en av redaktørene av *Lexicon of Greek Personal Names*. Matthews fortalte at dette kjempeprosjektet ble startet i 1972 og er finansiert av The British Academy. Formålet er å lage en database/leksikon over alle greske egennavn fra den tidligste historiske tid til den islamske erobring av Det nære østen i midten av det 7. århundre. Materialet vil kunne brukes av forskere i en rekke humanistiske disipliner i studier av bl.a. det antikke greske samfunn, spredningen av religioner, filologiske studier og i arbeid med å restaurere navn på papyrus.

Egennavnene har vært ekstrahert fra gresk (og i en viss utstrekning latinsk) litteratur, papyri, mynter, gjenstander og framfor alt, inskripsjoner. Geografiske og mytologiske navn er ikke inkludert. Det er samlet inn ca. en 1/2 million navn i alt, fordelt på ca. 13.000 ulike navn. Innsamling og bearbeiding av data har vært utført av forskere over hele verden, som ikke har støttet seg til eksisterende indekser.

I begynnelsen av prosjektperioden foregikk registrering av dataene på maskinskrevne arkivkort. I 1975 ble edb-registrering tatt i bruk, og programmet FAMULUS ble anvendt for sorteringsformål fram til 1982. Mellom 1983 og 1985 ble tekstfilene, som på det tidspunktet bare inneholdt fire felt for hver post, konvertert til en atskillig utviklet IDMS nettverksdatabase spesielt utviklet for ICL-maskinen ved OUCS. Individdata fra samme region er kodet og lagt inn i regionsfiler som deretter er blitt samsortert.

The British Library skal avgjøre hvorvidt databasen skal bli tilgjengelig online. Materialet skal iallfall utgis av Oxford University Press i 6 bind à ca. 500 sider. Første bind skal utgis i løpet av året. Hvert bind skal dekke bestemte regioner og inneholder ca. 68.000 innførsler. Materialet skal først bli presentert i alfabetisk rekkefølge, deretter skal forekomster av samme navn gis i henhold til region, og hver region skal deles inn i en kronologisk rekkefølge. Det er OUCS som er ansvarlig for utviklingen av programvaren for fotosetting av materialet til den trykte publikasjonen (se prøven).

Mer detaljert informasjon om prosjektets tekniske aspekter er beskrevet av Elaine Matthews og Sebastian Rahtz i «Designing and Using a Lexicon of Greek Personal Names» i *Proceedings of the ALLC Conference, Nice 1985*.

Εύφωλος	'Εχειράτης	
Εύφωλος ΡΗΘΙΟΣ: —ΚΑΜΗΡΟΣ: (1) s. v. inc <i>TCam</i> 160 (s. <i>Εύφωλος</i> , f. <i>Πρωτόδος</i>)	Εύφωλος ΚΥΡΕΝΟΣ: —ΙΘΑΛΙΟΣ: (1) iii-iv ad <i>HCC</i> p. 553 no. 1046 (<i>Ερωτήσανός?</i>)	'Εχειράτης ΚΥΡΕΝΑΙΚΑ: —ΚΥΡΕΝΕ: (1) f. 1 ad <i>SEG</i> ix (82, 7, xvii 799, 12 (f. <i>Τυμάρτα</i> ; date J.M.R.)) —ΤΙΒΙΑ: (2) inc 1 ad <i>IG</i> xii (3) Suppl. 1301
Εύφωλιμος ΚΥΡΕΝΑΙΚΑ: —ΜΑΡΜΑΡΙΚΑ: (1) s. ii ad <i>PMarm</i> x, 42	Εύφωλιμος ΚΥΡΕΝΕ: —ΤΑΒΡΙΑ: (1) imp. <i>IC</i> 2 p. 208 no. 10 (f. <i>Τυμάρτης</i>)	'Εχειράτης ΚΥΡΕΝΕ: —ΛΑΤΟ: (1) hell.-imp. <i>IC</i> 1 p. 138 no. 25, 2; p. 142 no. 31, 2 ('Εχειράτος (n. gent.))
Εύφωλιος ΚΥΡΕΝΟΣ: —ΚΙΤΙΟΣ: (1) imp. <i>HCC</i> p. 553 no. 1945 ('Εχειράτης—ed. name T.B.M.)	Εύφωλιος ΚΥΡΕΝΟΣ: —ΙΕΡΑΦΥΤΑ: (1) e. 110-67 no. <i>BCH</i> 103 (1970) p. 81; cf. <i>MH. Cain</i> i p. 322 nos. 5-6; pp. 323 ff.; <i>Le Rider, Monnaies crétoises</i> p. 207 (date); (2) s. inc <i>IC</i> 3 p. 68 no. 37 (s. <i>Κριάλλος</i>)	'Εχειράτος ΚΥΡΕΝΕ: —ΛΑΡΡΑ: (1) f. ii no. <i>BCH</i> 45 (1921) pp. 4 ff. III, 115 (f. <i>Άρχων</i>)
Εύφωλα ΚΥΡΕΝΕ: —ΑΙΔΟΣ: (1) f. ad <i>IC</i> 2 p. 77 no. 41 (s. <i>Βρωτόλος</i>)	Εύφωλα ΚΥΡΕΝΕ: —ΛΥΤΤΟΣ: (1) imp. <i>IC</i> 1 p. 214 no. 110 B	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΚΥΡΕΝΟΣ: —ΑΜΑΤΙΘΙΟΣ: (1) imp. <i>Excav. Cyprus</i> p. 96	Εύφωρη ΚΥΡΕΝΕ: —ΚΑΑΤΕΙΣ (Per. Syme): (2) imp. <i>AE</i> 1915, p. 132 no. 5	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΚΥΡΕΝΕ: —ΛΕΡΤΙΑ:	Εύφωρη ΚΥΡΕΝΕ: —ΚΑΑΤΕΙΣ (Per. Syme): (2) imp. <i>AE</i> 1915, p. 132 no. 5	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
—ΖΑΥΚ: (2) iv/iii bc <i>IG</i> xii (9) 245 A, 64 (s. <i>Εύφωρη</i>) ? = (3), (3) ~ ib. l. 142 (f. <i>Εύφωρη</i>) ? = (2)	Εύφωρη ΚΥΡΕΝΕ: —ΑΝΑΦΗ: (1) imp. <i>IG</i> xii (3) Suppl. 1287 (f. <i>Μαύστη</i> 'Ε')	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
ΡΗΘΙΟΣ: (4) 53 ad <i>IG</i> xii (1) 2, v (f. <i>Εύφωρη</i> II); (5) ~ ib. (II s. <i>Εύφωρη</i> I)	Εύφωρη ΚΥΡΕΝΕ: —ΚΑΜΗΡΟΣ: (1) iv/iii bc <i>TCam</i> i III, 14 (<i>Εύφωρη</i> ; <i>Εύφωρη</i> ; <i>Εύφωρη</i> ; <i>Εύφωρη</i>)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρίδας ΚΥΡΕΝΕ: (1) 193/2 no. <i>IG</i> v (2) 203, 20 (s. <i>Σωμ</i>)	Εύφωρη ΚΥΡΕΝΕ: —ΑΒΑΔΕΣ: (1) vii/vi bc Hoffmann, <i>Early Cretan Artifacts</i> p. 12 no. M6 (s. <i>Ερραμείος</i>)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρίνος ΡΗΘΙΟΣ: (1) iii. inc bc <i>IG</i> xii (1) 112; <i>Suppl. Rod.</i> 51 (s. <i>Αριστόβροτος</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΚΑΡΠΑΘΟΣ: —ΑΡΚΑΣΙΑ: (1) iv/v ad <i>Cl. Rh.</i> 6/7 p. 568	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
ΚΩΣ: (2) i bc <i>PH</i> 160 (f. <i>Εύφωρη</i> II); (3) ~ ib. (II s. <i>Εύφωρη</i> I)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
ΡΗΘΙΟΣ —ΚΑΜΗΡΟΣ: (4) f. ii bc <i>TCam</i> Suppl. p. 220 no. 157 h 11, 11 (<i>Εύφωρη</i> I)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
ΤΕΝΕΔΙΟΣ: (5) imp. <i>IG</i> xii (8) 140 (f. <i>Φαλαμάρης</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρίνος ΤΙΒΙΑ: (1) imp. <i>IG</i> xii (3) 384	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΚΩΣ: —ΚΑΡΤΗΙΑ: (1) s. iii bc <i>IG</i> xii (5) 610, 52; p. 334 (inc.)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρίδας ΤΕΛΟΣ: —ΤΕΛΙΟΣ: (1) f. ii bc <i>IG</i> xii (3) 43, 1-2 (s. <i>Καλλιστόδος</i> , f. <i>Γερραλός</i> , <i>Αριστομένης</i> , <i>Κλεισσύνης</i> , <i>Καλλιστόδος</i>); (2) ~ ib. l. 3 (f. <i>Καλλιστόδος</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΚΥΡΕΝΕ: —ΟΡΥΤΥΣ: (1) iii bc <i>IC</i> 4 p. 312 no. 260 (s. <i>Σαρχων</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρων ΡΗΘΙΟΣ: —ΙΑΛΥΣΟΣ:	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
—ΚΥΡΑΝΕΙΑ (Per.): (1) ii-i bc <i>IG</i> xii (1) 285 (f. <i>Ελευγιά</i>) ? = (2), (2) i bc <i>Cl. Rh.</i> 2 p. 177 no. 6, 56 (f. <i>Επιμαρόντος</i>) ? = (1)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΕΛΒΙΔΑ: —ΧΑΛΚΙΣ: (1) ii/vi bc <i>IG</i> xii (9) 952 II, 9; <i>BCH</i> 103 (1979) pp. 184 ff. (date) (f. <i>Μακαριώτης</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΕΛΒΙΔΑ: —ΕΡΕΤΙΑ:	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
—ΧΥΤΙ: (1) iv/iii bc <i>IG</i> xii (9) 246 A, 273 (f. <i>Σάρος</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωρη ΕΛΒΙΔΑ: —ΧΑΛΚΙΣ: (1) iii-iv bc <i>IG</i> xii (9) 1026 (s. <i>Διοσέντος</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)
Εύφωλης ΜΕΛΟΣ: (1) s. v bc <i>IG</i> xii (3) 1156 (f. <i>Μιλισσός</i>)	Εύφωρη ΚΥΡΕΝΑΙΚΑ: —ΚΑΛΙΣΤΟΣ: (2) imp. <i>SEG</i> ix 289; [290] (<i>Αντ. Εύφωρη</i> ?)	'Εχειράτος ΣΙΚΙΝΟΣ: (1) imp. <i>IG</i> xi Suppl. p. 102 no. 188 (s. <i>Διοσέντος</i>)

DESKTOP PUBLISHING

FRA HOBBY TIL BEDRIFTSTJENESTE OG VITENSKAPELIG DOKUMENTASJON

Per H. Jacobsen

INNLEDNING

Denne artikkelen skal ta for seg et tema som virkelig fyller spaltene i mange fagblader for tiden. Begrepet *Desktop Publishing* er temmelig nytt, og henger sammen med utviklingen av program- og maskinvare for mikromaskiner. *Publishing* har funnet sted helt fra Gutenbergs tid, håndverksmessig, mekanisert og i vår tid elektronisk. I løpet av de siste par årene har også denne teknikken blitt tilgjengelig på våre skrivebord, som *Desktop Publishing*.

Samspillende, og i markedsmessig og økonomisk forstand *tilgjengelig* maskin- og programvare danner et konsept som selges under merkelappen *Desktop Publishing (DTP)*. Temaet er vidløftig og det finnes mye brukbart utstyr og programvarer på markedet. DTP kan betraktes og drøftes fra mange synsvinkler. I denne framstillingen skal vi begrense oss til å se på:

- *Macintosh-konseptet*, som spesielt har vært med på å sette fart i utviklingen av DTP og som vel var det første maskinsystemet som brakte dette til torgs,
- Med dette eksempelet som utgangspunkt drøfter vi noen egenskaper ved DTP-systemer i relasjon til *typografi* og tar opp enkelte *estetiske* forhold knyttet til bruk av DTP
- Vi skal også komme litt inn på behandling av fremmedspråklig tekst.

Framstillingen vil være uteknisk og allment lagt opp som en presentasjon av hva DTP er, og hvordan det foregår på en Macintosh. Vi skal se at det er mange programmer og maskinløsninger (konfigurasjoner) å velge mellom, men vi gir ingen markedsveiledning. Målet med innlegget er å formidle ideen om DTP, og inspirere til at man tar dette

meget nyttige *kommunikasjons-, dokumentasjons- og informasjonsverktøyet* i bruk.

HVA ER DESKTOP PUBLISHING?

DTP er egenproduksjon av dokumenter av ulike slag på en utrustning som består av en kombinasjon av et mikromaskinsystem, programmer for generering av dokumentets innhold, et layoutprogram eller sidedesignprogram («page composition program») og en laserskriver. Nøkkelordet er *egen*-produksjon, dvs. at vi selv ved hjelp av vår mikromaskin og våre programmer går gjennom alle faser i produksjonen av et dokument til vi sitter igjen med en trykkeklar original av bra kvalitet.

DTP passer for en lang rekke typer dokumenter, f.eks. nyhetsblader, prislister, skjemaer, foldere, reklame, tidsskrifter, kataloger, artikler, avhandlinger osv., kort sagt alt fra hobbydokumenter til bedriftsdokumenter og vitenskapelige publikasjoner.

Et DTP-produsert dokument nyter godt av redskaper utover de vi finner i vanlige tekstbehandlere (editorer og formaterere). Layoutprogrammet *etterbehandler* tekst fra tradisjonelle tekstbehandlere, og blander teksten sammen med grafikk og bilder. Noen layoutprogrammer inneholder selv visse muligheter for tekstbehandling også. Vi skal kunne utnytte fullt ut ulike *fonter* og *typografiske hjelpemidler*, *grafikk*, og *innskannede bilder*:

- *Fonter* eller ulike skrifttyper som kan opptre i forskjellige størrelser og i ulike former
- *Typografiske hjelpemidler* som f.eks. «Letraset»-lignende virkemidler (border, bakgrunnsmonstre, logoer, streker, piler m.m.)
- *Grafikk*, dvs. informasjon i grafisk form f.eks. *presentasjonsgrafikk* i form av kurver, diagrammer, regneark eller *frihåndstegning* eller *strukturtegnning*
- *Bilder*, dvs. «Avis»-bilder i form av portretter e.l.

Vi skal kunne blande tekst og bilde, skyve elementene maskinelt rundt på arket til vi er fornøyd, og skrive ut resultatet på en skriver som gir bra kvalitet. Det endelige utseende vil avhenge av kvaliteten på tilgjengelige fonter og (laser)skriverens nøyaktighet (oppløsning), og dessuten ikke minst av det grafiske, layoutmessige og typografiske *craftsmanship* vi besitter.

Hva er det som er så nytt med dette? Inntil nylig var forskjellige typer programmer atskilt fra hverandre. Vi brukte tekstbehandleren

Noen tekstfonter for LaserWriter Plus
gjengitt i 12 punkters størrelse og i
forskjellige skriftstiler

Avant Garde
Avant Grade Oblique
Avant Garde Demi
Avant garde Demi Oblique

Bookman Light
Bookman Light Italic
Bookman Demi
Bookman Demi Italic

Courier
Courier Oblique
Courier Bold
Courier Bold Oblique

Helvetica
Helvetica Oblique
Helvetica Bold
Helvetica Bold Italic

Helvetica Narrow
Helvetica Narrow Oblique
Helvetica Narrow Bold
Helvetica Narrow Bold Oblique

New Century Schoolbook Roman
New Century Schoolbook Italic
New Century Schoolbook Bold
New Century Schoolbook Bold Italic

Палатино Роман
Палатино Италици
Палатино Болд
Палатино Болд Италици

Times Roman
Times Italic
Times Bold
Times Bold Italic

Zapf Chancery Medium Italic

9 punkter Times
10 punkter Times
12 punkter Times
14 punkter Times
18 punkter Times
24 punkter Times

Gangen i det hele: forarbeid i ulike
programmer og sammenstillingen i
layoutprogrammet

Tekstbehandling
på andre maskiner

Blablabla

Blablabla

Blablabla

Blablabla

Blablabla

Blablabla

Tekstbehandling
på Macintosh

- MacWrite
- MS Word
- WriteNow

Grafikk

Fritegning

- Macpaint
- FullPaint
- SuperPaint

Strukturtegning

- MacDraw
- Macdraft
- Cricket Draw

Presentasjon

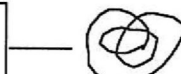
- MS Chart
- Cricket Chart

Clip Art

"Letraset"
Egne logoer etc.

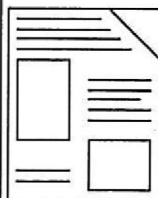
Bilder

Scanner



Page Composition

PageMaker
Ready,Set,Go!
Xpress
og andre



LaserWriter

Tegninnholdet i Times (14 punkter)

	Shift
<1234567890+´	>!"#\$%&/'()=?
qwertyuiopå¨	QWERTYUIOPÅ^
asdfghjkløæ	ASDFGHJKLØÆ
zxcvbnm,.-'	ZXCVBNM;:_*
Option	Shift-Option
≤ ¡ « £ ¢ [\] ≠ ± º	≥ Â » Ê Æ È Ë Ì Í Î Ï ð
• Σ ™ ® ¨ † ‡ μ ¶ § Ω π °	◊ , ^ % ¸ , º Ú Ô Ï Ò Ì
à ß ð Ñ © ^ Δ œ − ø æ	Á Ÿ ¢ > 2 Û Ç Æ Æ
§ ≈ ¢ √ ∫ ~ ¸ ∞ ... − @	fi fl Ç † : 1 " ' ¸ — ÷

til innskriving og redigering av tekst, grafikkprogrammet til å lage stolpediagrammer osv., men hadde små muligheter til å føre grafikken inn i teksten, og enda mindre muligheter til å designe dokumentet med flere spalter, overskrifter og ingresser, bruke bokser, border og rastre. DTP-konseptet, og spesielt layoutprogrammet, lar oss hente de ulike dokument-elementene fra de forskjellige applikasjonsprogrammene og blande dem sammen i dokumentet.

MAC-KONSEPTET: INNHOLD

Vi skal ikke begi oss inn på tekniske detaljer angående maskinutstyr her, bare sette opp hvilke byggeklosser som inngår i Macintosh DTP-konseptet. Vanligvis inngår følgende elementer, men ikke alle er nødvendige:

Maskinelementer:

- Mikromaskin:* Apple Macintosh Plus, Macintosh SE (eller senere, Macintosh II)
- Storskjerm:* Det fins mange i ulike kvaliteter og prisklasser
- Laserskriver:* Apple LaserWriter Plus
- Pinneskriver:* ImageWriter II
- Scanner:* ThunderScan, Abaton Scanner
- Nett-løsning:* Appletalk
- Kommunikasjon:* Modem

Storskjerm vil være en god investering hvis vi bruker Macintosh mye til DTP, idet den innebygde skjermen er nokså liten, og ettersom vi i praksis ønsker å designe hele A4-sider, og derfor helst vil se dem i sin helhet i ekte format på skjermen.

Pinneskriver er en vurderingssak — trenger vi det i tillegg til laser-skriveren?

Nettløsning er bare nødvendig hvis vi er flere personer med hver sin maskin som ønsker tilgang til felles ressurser som en LaserWriter, eller til felles filer.

Modem trenger vi hvis vi ønsker å overføre data (f.eks. tekst) til og fra vår mikromaskin, eller hvis vi ønsker å bruke vår mikromaskin som terminal mot en stormaskin.

Programelementer (eksempler):

<i>Tekst:</i>	MacWrite, Microsoft Word, WriteNow
<i>Grafikk:</i>	Frihånd: MacPaint, FullPaint, SuperPaint Struktur: SuperPaint, MacDraw, MacDraft, Cricket Draw Presentasjon: Cricket Graph, Microsoft Chart
<i>Bilde:</i>	Thunderscan, Abaton Scanner
<i>Annet:</i>	Regneark: Excel Idebehandler: MORE - og sist men ikke minst:
<i>Layout:</i>	PageMaker, Ready,Set,Go!, Quark Xpress

Fonter og annet:

Videre trenger vi et passende utvalg *fonter* og kanskje noen «Letra-set»-hjelpemidler (Clip Art). Et utvalg bra fonter følger med LaserWriter Plus, endel færre følger med LaserWriter. Mer om fonter senere.

Switcher:

Det finnes dessuten et nyttig hjelpemiddel som heter Switcher. Dette er et system som gjør at inntil fire programmer kan dele hukommelsen i maskinen, slik at alle fire programmene ligger i hver sin del av hukommelsen samtidig. Dermed kan vi veksle mellom programmene uten å lagre det vi holder på med i det ene programmet, stenge og gå ut av det, åpne neste program, hente inn det vi skal arbeide med i dette programmet osv. I et arbeidsmiljø hvor vi nettopp ønsker en fleksibel veksling mellom programtyper er derfor Switcher meget nyttig og tidsbesparende, idet vi med et muse-klikk kan hoppe fra det ene

til det andre programmet. Vi kan f.eks. montere vår yndlingstekst-behandler, et fritegningsprogram, et program for presentasjonsgrafikk og kanskje layoutprogrammet sammen. Det blir plass til flere store programmer ved siden av hverandre, dess større hukommelse maskinen har.

PostScript:

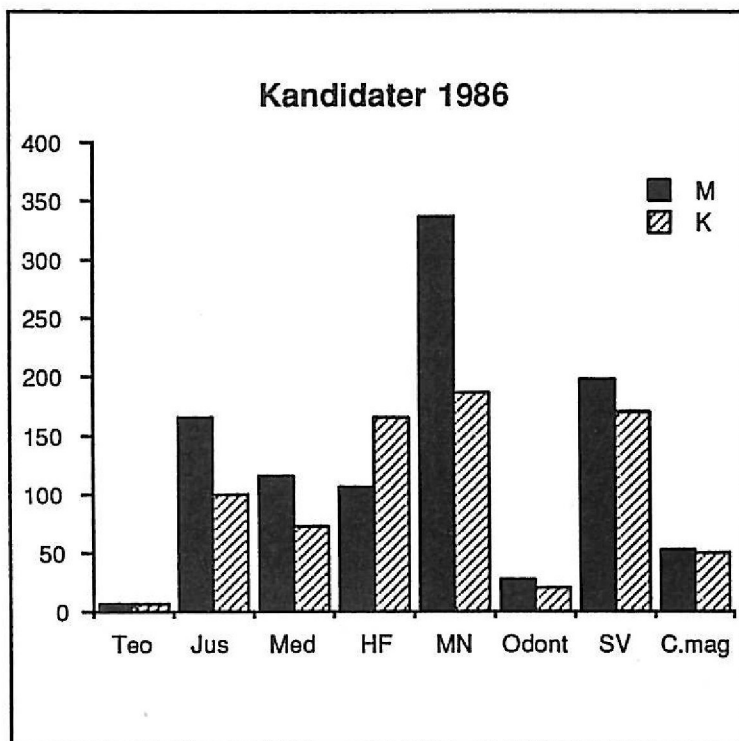
Dette er et interessant språk for sideoppbygging. PostScript tillater en øvet bruker å «programmere» laserskriveren selv for å oppnå spesielle effekter på sidene. Mange layoutprogrammer lar oss legge inn PostScript-direktiver rett inn i dokumentet. Disse fanges opp av laserskriveren, som fortolker og utfører våre kommandoer. Hverdagsbrukere behøver ikke bekymre seg om PostScript, idet layoutprogrammet og laserskriveren samarbeider bak kulissene slik at den design på dokumentet som vi bygger opp på skjermen, oversettes til PostScript-direktiver som er det laserskriveren skjønner.

Vi nevner PostScript her fordi det ser ut til å vinne stadig større markedsandeler som en de facto «standard» for denne typen «sideutleggsspråk». Grovt sagt er det slik at vi kan ta ut et DTP-dokument fra Macintosh på en hvilken som helst utskriftsmaskin (laserskriver eller fotosetter) som forstår PostScript. LaserWriter er én slik PostScript-drevet maskin, og mange andre kommer etter på markedet, bl.a. fotosettere som kan gi en helt annen kvalitet enn den rimelig brukbare vi får ut av LaserWriter.

MAC-KONSEPTET: EGENSKAPER

Vi kan ikke gå nøye gjennom Macintosh sine egenskaper her, men vi skal trekke fram et par forhold av betydning for DTP-bruk. For det første er Macintosh i utgangspunktet en meget *grafisk* orientert maskin. Grafikkmuligheten er innebygd i systemet, og utnyttes av programvarene. Macintoshens bruksmåte, med mus, menyer og vinduer, gir oss dessuten en behagelig arbeidssituasjon hvor vi slipper å lære oss en mengde kommandoer og tastetrykk for å få brukt programmene. En like viktig egenskap er at de aller fleste programmene *benytter et standard brukersnitt*. Dermed blir det lettere å veksle mellom forskjellige programmer. Vi gjenfinner mange av de samme menyvalgene i alle programmene, og valgene betyr det samme over alt. Dermed kan vi si at alt forarbeidet inntil vi har alle elementene i et dokument som skal designes, dvs. tekstbehandlingen, generering av

Presentasjonsgrafikk laget i Cricket Graph



grafikk osv., skjer i programmer med likelydende brukersnitt. Vi kan gi følgende karakteristikkk:

- generell lettbrukthet
- enkel integrasjon mellom programvarer («klipp & lim»)
- bruksmessig likhet programmene imellom
- mangfoldighet (i programtilbud, fonter, Clip Art etc.)
- og komponentene spiller godt sammen ytelsesmessig og kvalitetsmessig til en overkommelig pris. Men som vi skal se senere, også Macintosh-konseptet har sine begrensninger.

Sammenblandingen av tekst- og grafikkelementer, og hvor lett, elegant eller kronglete dette gjøres, sier noe om DTP-systemets anvendelighet. Integrasjonen er også en av de ting som skiller DTP fra tradisjonell dokumentbehandling. En annen viktig forskjell ligger i

*Innskannet bilde (av Fartein Valen) i
skjermbildekvalitet, ikke laserkvalitet*



forfatterens nære kontakt, eller *interaksjon* med sitt dokument helt fram til trykkingen. Hele dokumentframstillingsprosessen endrer karakter med de konsekvenser dette har, f.eks. for utviklingen av den typografiske fagtradisjon og den typografiske kvalitet på framstilte dokumenter.

FORARBEID

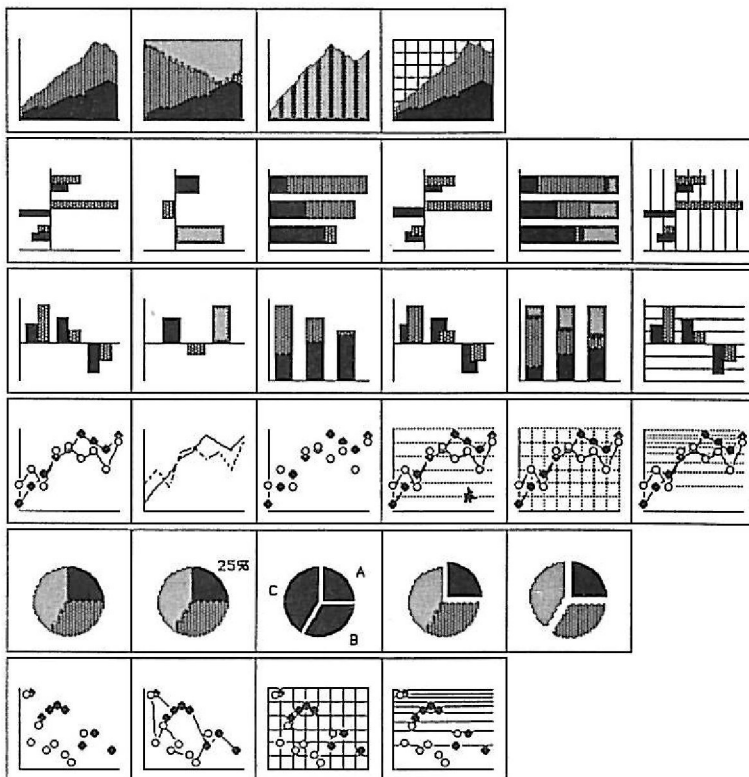
Vi skal se litt nærmere på arbeidsprosessen *forut* for selve sidedesignen, og deretter si noen ord om layoutprogrammet.

Forarbeidet består i produksjon av tekst og grafikk som skal med i dokumentet. Vi bruker passende tekstbehandlere, grafikkprogrammer osv. Vi må passe på at vi bruker programmer som lagrer tekst eller grafikk på en slik måte at layoutprogrammet senere kan hente dette inn. Ikke alle programmer passer sammen, men de beste layoutprogrammene kan bruke data som er laget i de mest populære applikasjonsprogrammene på markedet.

Vi skal ikke gå gjennom spesielle programmer her, vi viser til noen av illustrasjonene som gir en pekepinn om hva slags dokumentelementer vi behandler i de ulike programtypene.

Imidlertid, det som er viktig å merke seg er at størsteparten av

Noen typer presentasjonsgrafikk i Excel



arbeidstiden ligger i dette forarbeidet. Det er arbeidssparende at de elementene som skal inn i dokumentet gjøres så *ferdige* som mulig før vi setter igang med layoutprogrammet. Dette krever igjen at vi *planlegger* vår publikasjon på forhånd, at vi tenker gjennom hvordan våre illustrasjoner skal se ut etc. Vi bør kanskje lage en skisse på papir av hvordan vi vil ha det — dette er også nyttig etterpå, under designen av dokumentet i layoutprogrammet. Kan hende mange av oss trenger *faghjelp* til dette — vi er ikke grafikere, typografer eller designere.

FERDIGSTILLING

Når vi har samlet sammen alt stoffet som skal inn i dokumentet, bruker vi layoutprogrammet til å samle dette sammen og sette det opp på sidene slik vi vil ha det. Det fører for langt å gå gjennom alle mulighetene i slikt program, f.eks. PageMaker, vi skal bare gi noen generelle kommentarer og tips.

I layoutprogrammet bestemmer vi hvordan dokumentet skal se ut. Hvor skal de forskjellige bitene plasseres? Hvor mange tekstspalter vil jeg ha? Hvordan skal de se ut — like brede, av ulik bredde? Hvordan skal sidene nummereres? Skal jeg ha et løpende sidehode/sidebunn? Etter at bitene er noenlunde på plass, gir programmet oss mulighet til en siste (viktig) finish med pynt, justeringer, utskifting av fonter i overskrifter, gjennomgang av teksten for uheldige ordde-linger (teksten fylles inn i spalter av annen bredde enn i tekstbehand-leren), bruk av streker og bokser, i det hele tatt, den endelige «touch-up» som gjør dokumentet attraktivt og vellaget.

Det kan være lurt å dele opp en lang tekst på flere filer, da får vi større frihet når den skal fylles inn i spaltene. Dess mer endelig fer-dig teksten er på forhånd, jo bedre. Merk at layoutprogrammet f.eks. kan la teksten flyte rundt figurer på flere måter. Merk også at teksten kan hentes inn i layout-programmet med *attributter* hvis den er lagret slik med tekstbehandleren. Dette kan være tab-posisjoner, font-informasjon etc. Likevel vil vi ofte oppdage at de valg vi gjorde i tekstbehandleren ikke alltid passer så bra i det endelige sideutlegg, så det blir litt omprøving, også i forarbeidsprogrammene, før vi er fornøyd.

Det er åpenbart en fordel med storskjerm, slik at vi kan se hele A4-arket på én gang under redigeringen. Vi kan dessuten se to tilstø-tende sider samtidig, slik at vi kan vurdere hvor godt de står til hverandre.

Det er mange muligheter i et slikt layoutprogram, men mye av

nyttan henger sammen med våre evner til å tenke visuelt, bruke grafiske og typografiske virkemidler og ha estetisk sans.

Denne typen programmer er dessuten i rivende utvikling ettersom DTP-markedet har eksplodert. Nye, bedre versjoner av eksisterende programmer lanseres, og nye gode programmer dukker opp med stadig mer avanserte muligheter. Layout-programmene er nokså maskintunge, dvs. det foregår intense beregninger hele tiden mens vi skyfler bitene våre rundt og prøver oss fram. Likevel virker det som om Macintoshen som regel klarer dette rimelig bra — det er sjelden spesielt lang ventetid mens programmet setter opp en side på nytt og justerer påfølgende sider osv., men det kan av og til oppstå «regnepauser». (Dette henger bl.a. sammen med hvor mange frihetsgrader vi gir programmet, f.eks. hvor lange tekstbiter vi opererer med. En endring på side 1 som påvirker side 2,3,4,5,6 osv., kan ta tid.)

Ved utskrift på laserskriveren, må vi se til at vi hovedsakelig har benyttet laserfonter, slik at resultatet blir best mulig, og vi må sørge for at de fontene vi bruker finnes i systemet på den Macintoshen vi bruker *ved utskrift* (det kan jo tenkes at vi veksler mellom flere maskiner). Ved utskrift kan vi velge om LaserWriter skal bytte ut ukjente fonter med kjente eller ikke. Her kan vi få «morsomme» overraskelser. Kort sagt, vi må ha orden på sakene våre: oversikt over fonter, orden i mengden av småfiler med biter av innhold til det endelige dokumentet osv.

TEKST & FONTER

I de fleste tilfeller er det et *tekstlig* budskap vi vil ha fram. Det er derfor verdt å dvele litt ved framstilling og oppsett av den tekst som skal inngå i publikasjonen. Kvaliteten på den ferdige teksten avhenger av mange forhold:

- tekstbehandlerens formateringsegenskaper
- kvaliteten på fontene, f.eks. utformingen av de individuelle symbolene, sammenstillingen av dem og utlegg av paragrafene
- vår evne til layout, design og typografi

Tekstbehandlingen

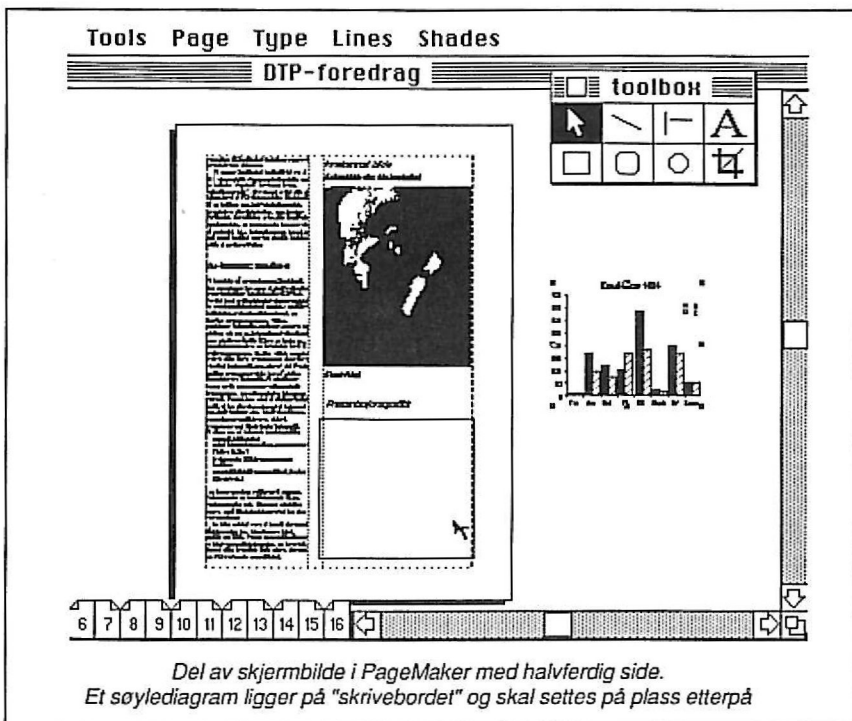
Tekstbehandlerne er stort sett slik vi kan vente oss, med de vanlige formateringsegenskaper. Enkelte typografiske egenskaper mangler imid-

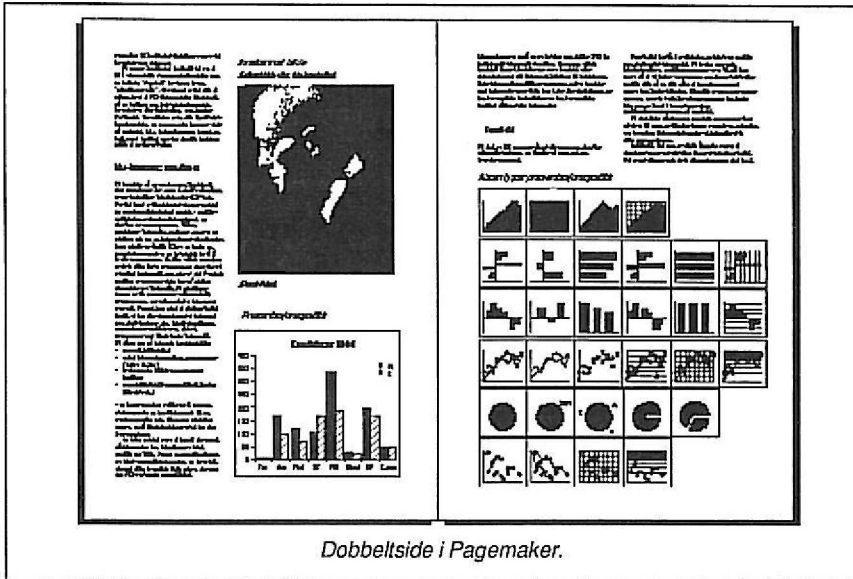
lertid ofte, f.eks. *kjerning*, bruk av *ligaturer*, brukerkontrollert ordmellomrom, bruk av typografiske enheter som *punkter*, *cicero*, *pica* m.fl. og orddeling.

Det varierer også hvor gode typografiske egenskaper selve layout-programmet har, men her finner vi gjerne noen flere muligheter, slik at vi i noen grad kan etterarbeide teksten typografisk.

Fontene

En *skrifttype* er bokstaver, tall og symboler med et felles utseende som skiller dem fra andre skrifttyper. Times er én skrifttype, Helvetica en annen. En *skriftfamilie* oppstår når en skrifttype opptrer i forskjellige *skriftstiler*, f.eks. Helvetica roman, *italic*, **bold**, **bold italic**, condensed og expanded, og i ulike størrelser, kanskje fra 5 til 72 punkters størrelser. En *font* er et komplett sett bokstaver, tall og symboler tilrettelagt for bruk på en bestemt utskriftsmaskin. På LaserWriter Plus



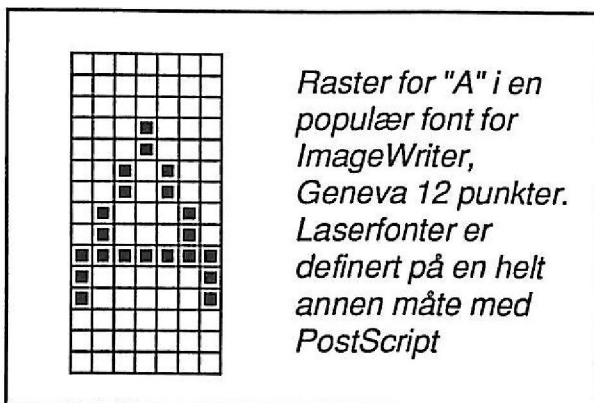


kan vi f.eks. bruke fonten 12 punktets Helvetica Bold Roman, og får da skrifttypen Helvetica. Andre skrivere kan også tenkes å ha skrifttypen Helvetica, der representert i en annen font.

På Macintosh finnes det to hovedgrupper av fonter, *laserfonter* og fonter for pinneskriver. Disse to typene er definert på helt forskjellig måte. Pinneskriverfontene er *bit-mapped*, dvs. definert som rastermønstre (se illustrasjon), mens laserfontene er definert i PostScript som instruksjoner som forteller hvordan kurvene og linjene i symbolene skal se ut. Dette medfører bl.a. at en laserfont kan skaleres opp og ned og likevel beholde kvaliteten.

Rasterfontene på Macintosh-skjermen opptrer i 72 dpi (dots per inch) oppløsning, og vil se temmelig likedan ut på skjermen og på pinneskriveren (selv om denne har endel bedre oppløsning). Merk at det må lages rastermønstre for alle tegnene og for alle størrelser av rasterfonter. På LaserWriter vil rasterfonter se bedre ut enn på pinneskriver, men de utnytter ikke laserskriverens muligheter, og gir ikke et typografisk godt resultat.

Laserfonter kan være *residente*, dvs. lagret i laserskriverens ROM, eller *down-loadable*, dvs. de kan overføres til laserskriveren for bruk. Laserfontene opptrer i to forkledninger, for skjermen og for LaserWriter. Laserfontene ser ikke spesielt bra ut på skjermen, men de er lagt opp til å skulle etterlikne utseendet på laserskriveren så bra som



mulig. Grunnen er selvsagt forskjellen i oppløsning mellom skjerm og laserskriver, den siste har 300x300 dpi oppløsning, dette er mye mer enn skjermen har.

Vi trenger et fornuftig sammensatt utvalg av fonter, og sammensetningen avhenger av bruksområdet. Vi kan ha fonter med forskjellig bruksområde:

- for vanlig løpende tekst
- for overskrifter og blikkfang
- for spesielle effekter, eller (pynte)symboler
- for spesialsymboler, f.eks. for matematikk
- for ikke-latinske språk, f.eks. kinesisk eller arabisk

Og sist men ikke minst, vi trenger å kunne sette sammen fonter i et bestemt dokument slik at utseendet blir best mulig. Fontene må stå godt til hverandre, opptre i riktige størrelser og brukes med fornuft. Det er lett å overlesse! Vi bestemmer selv hvilken portefølje av fonter vi vil bruke, og vi kan lett sette sammen vårt fontutvalg ved å plukke ut fonter fra vårt fontbibliotek.

Det er et meget rikt utvalg fonter på markedet, alt fra ganske bra utgaver for LaserWriter til elendige hjemmesnekrete ting. Noen få eksempler er tatt med som illustrasjoner, blant annet gjengis det fontsettet som følger med LaserWriter Plus i noen utvalgte variasjoner og størrelser.

Det er dessuten såpass lettbrukte programmer for *fontredigering* på Macintosh at det kan innby til fusk i faget. Det må tilføyes at utvikling av fonter er lettere for pinneskrivere enn for laserskrivere. Fontdesign

er imidlertid en alvorlig sak, og absolutt fagarbeid. Husk at de skrifttyper vi bruker til daglig har utviklet seg over lang, lang tid.

Installasjon av fonter

Nye fonter installeres lett i systemet vårt, slik at det ikke byr på vanskeligheter å sette sammen et passende utvalg, eller å endre dette av og til.

FREMME SPRÅKLIG TEKST

Innenfor Macintosh-konseptet finnes det som nevnt en mengde fonter av ulike slag, også fonter for fremmede språk. Mange av oss har behov for å behandle tekst i andre vestlige eller ikke-europeiske språk. Vi kan skille mellom to typer fremmede språk:

- *Latinske språk*, i.e. språk som anvender de 26 latinske bokstaver eventuelt pluss enkelte spesialbokstaver og aksentuerte bokstaver.
- *Ikke-latinske språk* som bruker andre tegnssett, f.eks. det kyrilliske, det greske alfabet, arabisk osv.

Latinske språk

Når det gjelder andre latinske språk er problemet hovedsaklig knyttet til forekomst av nødvendige ekstra nasjonale bokstaver og aksenter i tillegg til de vanlige 26 latinske bokstavene i de fontene vi har tilgjengelig (se illustrasjon). Det finnes fonter som er spesialsydd for et bestemt språk, men disse må skaffe/kjøpes. Orddeling kan også være et problem.

Selv om vi kommer godt på vei med de aksenter og spesialtegn de vanlige fontene inneholder, kan det være nødvendig for oss å skaffe spesialfonter. Vi bør da bl.a. undersøke fontenes:

- *kompletthet*: er alle tegn vi trenger til å skrive det språket vi skal bruke tilgjengelig?
- *effektivitet*: er tegnene fordelt på tastaturet slik at teksten blir behagelig å skrive, eller slik at eventuelle transkripsjons-standarder følges?
- *kompatibilitet*: kan fontene brukes på den laserskriveren vi har?
- *størrelser*: finnes de størrelser vi trenger?

Språk & aksenter	
<p><i>Dette er en oversikt over aksenter i språk som bruker det latinske alfabet. Listen er ikke komplett</i></p>	
Aksenter	
acute	· á
grave	· à
diæresis/umlaut	¨ ä
circumflex	ˆ â
tilde	~ ã
cedilla	ç
macron	strekk over
breve	bue over
caron	omv. ˆ
dobbel acute	¨
ring	° á
dot above	
ogonek	krøll nede
I Helvetica finnes	
<p>áàãäå ç èéêë ìíî ï ñ òóôõ öüÛ ŷ æøå œ ç ß</p> <p>ÁÀÃÄÅ Ç ÈÉÊË ÌÍÎ Ï Ñ ÒÓÔÕ ÖÜÛ Ü ŷ ÆØÅ Œ Ç</p>	
<p>Noen Vest-Europeiske språk:</p> <p>Dansk: á æ ø Nederlandsk: Ingen Finsk: ä ö Fransk: â ã ç é è ë î ï ó œ ú û ü Tysk: ä ö ü ß Italiensk: à è ì ò ù Latin: a e i o og u-macron og breve Norsk: æ ø å (samt å é ó) Portugisisk: á â ã ä ç é ê ë ì í î ï ó õ ö ü ú û Spansk: á é í ñ ó ú ü Svensk: å ä ö</p> <p>Noen Øst-Europeiske språk:</p> <p>Tsjekisk: á d' é í ó l' u-ring ú y-acute c e n r s og z-caron Ungarsk: á é í ó ö ú ü, o og u-dobbel acute Polsk: a-ogonek c-acute e-ogonek polsk-l n-acute ó s-acute z-acute z-dot Tyrkisk: â ç g-breve g-caron, i, ö s-cedilla ü ü, l-dot</p> <p>Noen spesielle bokstaver:</p> <p>eth, thorn, yogh, polsk-l, D med strek over, 3, tysk skarp ß, fransk ligatur œ Ç, færøysk ð m.fl.</p>	

Ikke-latinske språk

For bruk av ikke-latinske språk er det ofte ikke nok med en font for tegnsettet i språket. Mange språk er så særegne i forhold til latinske språk, at vi egentlig trenger en *tekstbehandler* for dette språket. Noen språk skrives fra høyre mot venstre, noen består av tusenvis av symboler, noen har ligaturer og tegn som endrer utseende med deres plassering i ordet. Og kanskje vi ønsker å *blande* et fremmed språk og et latinsk — f.eks. skrive hebraisk og engelsk om hverandre i samme dokument.

Det finnes flere brukbare fonter og flere tekstbehandlere for en lang rekke «eksotiske» språk på markedet for Macintosh, se liste.

Eksempler på eksotiske fonter

Diverse: Gresk (og gammel-gresk), Kinesisk, Japansk (fire skrivesystemer), Koreansk, Hebraisk, Arabisk

Øst-Europa: Armensk, Kyrillisk, Devonisk, Georgisk, Serbisk, Makodonsk, Russisk

Tibet: Mongolsk, Tibetansk

India og Sørøst-Asia: Bengalsk, Burmesisk, Devanagarisk, Gujaratisk, Kannadisk, Khmersk, Laotisk, Malayalamsk, Sanskrit, Telegusk, Thai

WYSIWYG

Det er begrenset hvor avansert formatering av tekst og symboler som kan foregå på et typisk WYSIWYG-system (What You See Is What You Get). For å få til riktig komplisert formatering, må vi ty til *mark-up* systemer. Formateringen kan bli komplisert på flere plan, selve den løpende teksten kan bli mettet av formler og symboler som er vanskelig å få satt opp riktig, eller dokumentet skal følge en fagstandard for vitenskapelig dokumentasjon med mange løpende spesialiteter som fotnoter, referanser, seksjoner og avsnitt som skal nummereres automatisk selv om vi redigerer på teksten og bytter rekkefølge på elementer, stikkordregister, abstracts etc. Her kommer WYSIWYG-systemer gjerne til kort.

På Macintosh kan vi alltid se på skjermen hvordan det vi lager blir seende ut når det skrives ut, eller rettere sagt, vi kan se en god tilnærming til hvordan det blir. (De fleste WYSIWYG-systemer viser seg å være mer eller mindre WYSIMOLWYG, i.e. *What You See Is More Or Less What You Get*).

PROS & CONS

Macintosh-konseptet for *Desktop Publishing* er meget fleksibelt og lett å bruke, krever lite forkunnskaper i databehandling og består av programmer som har lav læringsterskel og felles brukersnitt. Konseptet innbyr til lekenhet, utprøving, kreativitet og kanskje også til overlesing, mylder og rot. Likevel mener vi at det etter dagens standard er

et meget nyttig hjelpemiddel. I forhold til det øvrige marked for DTP-systemer oppfatter vi Macintosh som det mest *naturlige egnete* maskinsystemet for slik bruk, kanskje spesielt p.g.a. maskinens grafiske egenskaper.

Macintosh-konseptet er kanskje særlig velegnet til litt mer «uformell» produksjon, eller produksjon som nyttiggjør seg de gode mulighetene for blanding av tekst, grafikk og bilde. Dette må ikke misforstås dit hen at Macintosh ikke er velegnet til framstilling av «tørr» tekst som rapporter, avhandlinger, dokumentasjoner og håndbøker.

Hvis vi stiller strenge typografiske krav støter vi på konseptets begrensninger, f.eks. font-kvaliteter, tekstformateringsgenskaper, laser-

Hva koster moroa?

Vi tar med en liten liste over priser for endel Macintosh-relatert utstyr. I prisene er tatt med moms og gjeldende rabatter for Universitetet i Oslo pr. dags dato. Prisene er omtrentlige. Merk at *den første arbeidsplassen* blir mye dyrere enn de påfølgende, fordi endel utstyr, f.eks. LaserWriter Plus deles mellom flere brukere hvis maskinene koples sammen i et nettverk. For pengene mener vi at man får et ganske profesjonelt utstyr med stort potensiale

Macintosh SE	27000
Storskjerm	25000 *
LaserWriter Plus	70000 †
Abaton Scanner	30000 *
Nett-utstyr	9000 *
Pinneskriver	4000 *
Programvarer	20000 ‡

Totalt	185000
--------	--------

- * Hvis nødvendig
- † Pris uten rabatt
- ‡ Et bra utvalg

skriverens oppløsning (pr. idag). Her skal vi dog være klar over at PostScript raskt vinner popularitet, og at mange tyngre laserskrivere og fotosettere allerede leveres med PostScript.

KONKLUSJON

Denne anvendelsen av mikromaskiner er dessuten i eksplosiv vekst, og mange mener denne utviklingen helt vil forandre framstilling og distribusjon av trykt informasjon, og på lengre sikt også muligens påskynde en utvikling bort fra papir og over til elektronisk kommunikasjon. Macintosh har vært med på å dra igang denne utviklingen, og er fremdeles et av de aller beste konseptene for DTP på mikromaskiner.

LITTERATUR

Desktop Publishing

Det kommer for tiden en god del bøker om DTP i varierende kvalitet både når det gjelder innhold og hva utseende og design angår. Langt fra alle er gode eksempler på det område de omhandler! Vi har under arbeidet med denne framstillingen vært innom følgende bøker:

- James Cavouto: *LaserWrite It! A Desktop Publishing Guide To: Reports Resumes Newsletters Directories Business Forms And More*. Addison-Wesley. 1986.
ISBN 0-201-11327-9.
- Kelvin Rardin: *Desktop Publishing On The Mac. A Step-By-Step Guide To The New Technology*. Plume/Waite, New American Library. 1986.
ISBN 0-452-25902-9.
- Tery Ulick: *Personal Publishing with the Macintosh*. Hayden Books. 1986.
ISBN 0-8104-6572-8.
- Tony Bove, Cheryl Rhodes, Wes Thomas: *The Art Of Desktop Publishing. Using Personal Computers To Publish It Yourself*. Bantam Books. 1986. ISBN 0-553-34307-6.
- Daniel J. Makuta, William F. Lawrence: *The Complete Desktop Publisher*. Compute! Publications Inc. 1986. ISBN 0-87455-065-3.
- Bill Grout, Irene Athanasopoulos, Rebecca Kutlin: *Desktop Publishing from A to Z*. Osborne McGraw-Hill. 1986. ISBN 0-07-881212-7.
- John Seybold, Fritz Dressler: *Publishing from the Desktop*. Bantam Books. 1987.
ISBN 0-553-34401-3.

Videre har vi brukt de to Macintosh-bladene *MacUser* og *MacWorld*. Her har det det siste året vært mye stoff om DTP og relaterte emner. Vi har også brukt håndbøkene for *Pagemaker* og *Ready, Set, Go!* — samt for andre programmer vi har brukt til illustrasjoner m.v.

Dessuten bør nevnes:
PostScript Language: *Reference Manual*. Addison-Wesley, 1986. ISBN 0-201-10174-2.
PostScript Language: *Tutorial and Cookbook*. Addison-Wesley, 1986 ISBN 0-201-10179-3.

Generelt

Videre er følgende bøker brukt:

J. Nievergelt et al.: Document Preparation Systems. North-Holland, 1982. ISBN 0-444-86493-8.

Hsieh S. Hou: Digital Document Processing. John Wiley & Sons, 1983. ISBN 0-471-86247-9.

The Chicago Manual of Style. The University of Chicago Press, 1982. ISBN 0-226-10390-0.

Per H. Jacobsen er avdelingsleder ved Universitetets Sentrale EDB-tjeneste (USE), Brukerstøtteavdelingen, Universitetet i Oslo.

TEKSTBEHANDLING OG TEGNSETT

Espen S. Ore

Tema for denne artikkelen er problemer som oppstår ved databehandling av tekster som er skrevet på et annet språk enn engelsk. Disse problemene kan oppstå når vi ønsker å skrive inn en tekst: det kan hende at tastaturet ikke passer og/eller at vi ikke kan gjengi de tegn vi ønsker på skjermen. Vi møter dem også når vi skal skrive ut teksten: ikke alle skrivere er like gode til å takle sanskrit. Og har vi arbeidet med løsninger på skriver/skjerm/tastatur-problemer, sitter vi igjen med kanskje det viktigste: hvis vi vil prosessere teksten, f.eks. ved å lage ordlister, er det slett ikke sikkert at vi har noen programmer som kan takle annet enn engelsk.

En variant av disse problemene har nesten alle i Norge som bruker mikromaskiner til tekstbehandling, møtt: ÆØÅ-problemene! Lett forenklet er årsaken til disse problemene at datamaskiner opprinnelig er utviklet i engelskspråklige land av personer med en matematisk-naturvitenskapelig bakgrunn. Deretter ble næringsliv og samfunnsvitenskaper store brukere, mens filologer lenge har stått i siste rekke. Det er viktig her å se de forskjellige behov for fremmedspråklig tekstbehandling. Innen næringslivet vil det ofte være tilstrekkelig med en lokal ad hoc løsning som gjør det mulig å skrive ut brev med et brukbart tysk tegnssett (med f.eks. ä og ß). For en filolog er det derimot viktig at data skal kunne utveksles, at tekster skal kunne behandles med mer eller mindre generelle konkordansprogrammer osv. Dessverre har filologer ofte fått inntrykk av at det er umulig å dekke deres behov dersom de har henvendt seg til tradisjonelle edb-miljøer. Derfor har de ofte måttet finne seg i nettopp slike ad hoc løsninger som gjør at data og programmer ikke kan utveksles effektivt mellom forskere som arbeider med samme språk.

TEGNKODING OG -GJENGIVELSE

Bakgrunn

Når vi ser en bokstav på en dataskjerm, er den som de fleste vet,

satt sammen av punkter. (Se fig. 1) På en Olivetti M24 er hver bokstav tegnet innenfor en firkant som består av 8x16 punkter (pixler). For å lagre en slik bokstav med full informasjon om dens utseende trenges det 16 byte. Et skjermbilde (i tekstmodus) på en vanlig PC har 25 linjer à 80 tegn. Et skjermbilde med 16 bytes bokstaver krever altså 25x80x16 byte = 32.000 byte. Det sier seg selv at en vanlig diskett med plass til ca. 360.000 byte fort blir sprengt dersom man skriver dokumenter med tegn lagret på denne måten.

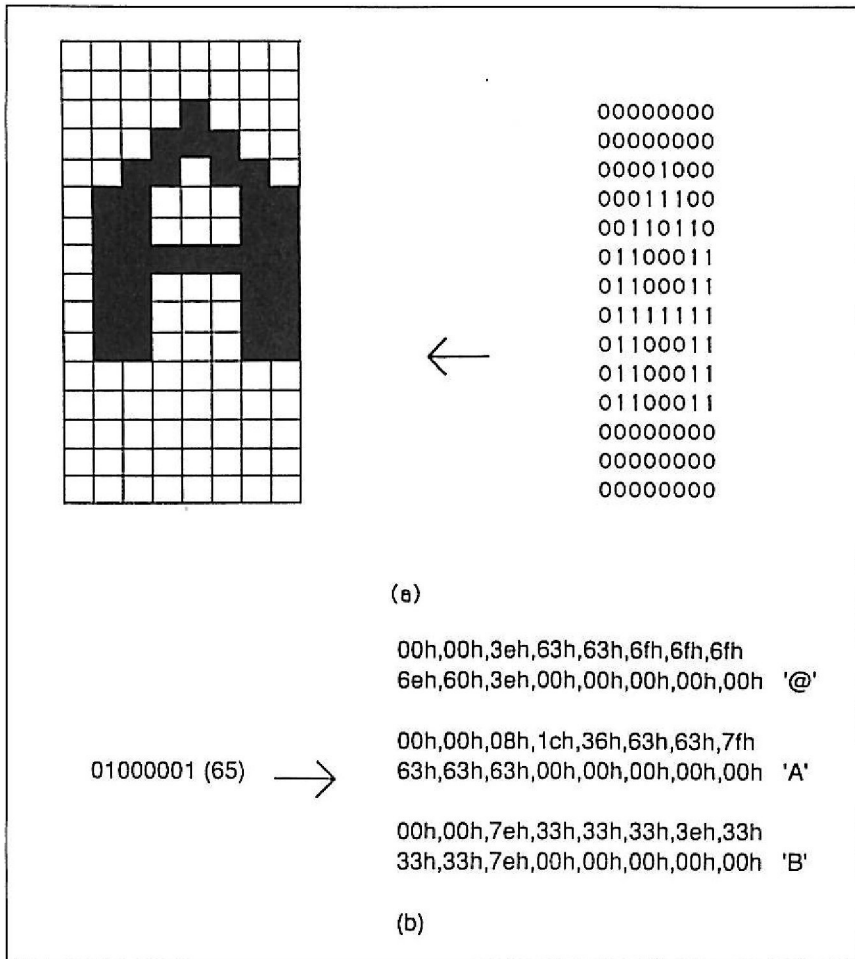


Fig. 1. Bokstaven 'A' vises på skjermen (a) ved at tegnkoden brukes til å slå opp i tabellen som beskriver punktmønsteret i bokstaven (b).

8x16 punkter gir oss $2^{(128)}$ forskjellige kombinasjoner. Imidlertid bruker vi aldri tegnsett med så mange forskjellige tegn, derfor vil det bli lagret mye unødvendig informasjon dersom tegnmønstrene skal spares komplett for hvert tegn. De samme forholdene som gjelder for lagring, gjelder også for datakommunikasjon. Jo flere punkter vi må sende pr. tegn, jo langsommere går det. Løsningen er å kode bokstaver i et system der koden er lang nok (men ikke lengre) til at vi kan skjelner mellom tilstrekkelig mange tegn. I en datamaskin kan vi måle kodelengden i *bit* (binærsiffer). 2 bit gir 4 forskjellige koder, 8 bit gir 256. Når et tegn skal vises på en skjerm eller skrives ut, kan koden så brukes i en tabell som angir formene.

Slik tegnkoding i bit er eldre enn datamaskinen. I fjernskrivere har 5-bit Baudot-koding vært brukt. (Oppkalt etter den samme Baudot som også har gitt navn til et mål for overføringshastighet: baud.) Denne koden ble fastsatt av CCITT (Comité Consultatif International Télégraphique et Téléphonique) i 1932 som deres alfabet nr. 2 (morsekoden var nr. 1). Med 5 bit kan man uttrykke 32 forskjellige tegn. Dette er i minste laget hvis man både skal ha bokstaver, tall og spesialtegn som f.eks. !, (og). I 5-bit Baudot-kode blir problemet løst ved at et spesielt tegn angir at tegn som følger etter, skal tas fra en annen tegntabell (omtrent som når vi på et tastatur bruker en ekstra tast for å angi at vi vil ha stor bokstav).

I datamaskinenes yngre dager ble det brukt forskjellige lengder på tegnkoder. Dessuten brukte produsentene av datamaskiner sine egne kodesystemer, slik at en kode med 6-bit lengde, f.eks. 110011 stod for 'A' fra leverandør X, mens den stod for 'B' fra leverandør Y. Imidlertid fikk man etterhvert internasjonale standarder. En av de vanligste er basert på en lengde på 7 bit. Den er fastsatt av ISO og er strengt tatt en familie av kodesystemer med spesialtilpasninger for forskjellige språk. I datasammenheng er denne kodingen best kjent som ASCII, som egentlig er navnet på den versjonen som er godkjent av den amerikanske standardiseringsorganisasjonen ANSI. I den norske versjonen av denne standarden er Æ, Ø og Å plassert på plassene etter Z. Dette gir oss bl.a. greie muligheter for automatisk alfabetisering.

Tegnstandarder og mikromaskiner

ISOs 7-bit tegnstandarder er ikke de eneste som finnes. IBM benytter sitt eget system, EBCDIC, på stormaskiner, og flere andre leverandører har fulgt dem. IBMs markedsdominans har også satt sitt preg på mikromaskinene. Da de kom med sin IBM PC i 1981 hadde den et utvidet tegnsett med 8-bit tegnkoder. Tegnkodene f.o.m. 32 t.o.m. 126

følger ASCII. Kodene f.o.m. 128 t.o.m. 255 brukes til 128 ekstra tegn. Blant disse ekstrategnene finner vi enkelte som brukes i ikke-engelske skriftspråk basert på det latinske alfabet, f.eks. 'Æ'. En av tankene bak et slikt kodesystem var å gjøre maskinene internasjonale. Tidligere måtte man bytte en leselagerenhet (ROM) som inneholdt tegntabeller avhengig av hvilket land maskinen skulle brukes i. Med det nye systemet skulle det være mulig å la programmene velge blant ekstrategnene i stedet. Dermed kunne samme maskin brukes nær sagt overalt bare med små endringer i programmene.

Dessverre var det ikke så vel. Det har vi bl.a. erfart i Norge der uttrykket «ÆØÅ-problemer» snart kan tas inn i Norsk dataordbok som egen fagterm. Da IBM laget sitt utvidete tegnssett, ble 'Ø' uteglemt. Man må altså fortsatt lage egne ROM-brikker for Norge og Danmark. I disse tegnssettene blir 'Ø' og 'ø' plassert på de plassene ¥ og ¢ har. Det er imidlertid et annet, større, problem forbundet med IBMs måte å løse konflikter mellom forskjellige nasjonale tegnssett på: Sammensatte tegn (f.eks. tegn med aksentmarkeringer) finnes i det utvidete tegnsettet som ett tegn, dvs. é er lagret som et tegn som er forskjellig fra e. Det er altså på forhånd bestemt hvilke tegn som kan ta diakritika, og hvilke diakritika de enkelte tegn kan ta. Dette går rimelig bra så lenge man holder seg til fransk eller italiensk, men det skjærer seg grundig med én gang man skal bruke enkelte av de slaviske språk som benytter latinsk alfabet. En tsjekkisk ś er f.eks. umulig å fremstille med IBMs utvidete tegnssett.

Tekniske årsaker

IBMs PC-familie (og kopier/kloner) har et utvalg av forskjellige teknologier for å fremstille tegn på skjermen. Dette har vært et bevisst valg, det skal være lett å skifte videoprosessor i maskinen dersom man f.eks. ønsker seg grafikkmuligheter med høy oppløsning. Problemet er at tegnsettene er avhengig av videokortene som brukes. Det opprinnelige «Monochrome Graphics Adapter»-kortet har f.eks. en tabell over hvorledes tegn skal vises på skjermen der oppløsningen er større for det enkelte tegn enn det vi finner på «Color Graphics Adapter»-kortet. Til gjengjeld tillater CGA-kortet at et egendefinert tegnssett brukes når det arbeider i grafisk modus. Dette siste er også en del av PC-familiens egenskaper: videokortene skifter mellom skriftmodus og grafikk-modus. Når kortet er i skriftmodus, brukes en egen lagerenhet til **kodene** til de enkelte bokstavene som vises på skjermen. Det er altså ikke definisjonen av punktene som utgjør tegnet som lagres her. Siden det er et én til én forhold mellom lagerplass og plas-

sering på skjermen, og det bare er plass til koden for ett tegn på en gitt lagerplass, er det umulig å kombinere to tegn når kortet er i skrift-modus. Av samme grunn er det også umulig å gjengi proporsjonalskrift på skjermen i denne modus. Når kortet er i grafikk-modus brukes derimot denne lagerenheten til å holde opplysninger om de enkelte skjerpunktene (det er dette som på engelsk kalles bit mapped graphics).

Det finnes også andre videostyrekort, f.eks. Hercules Color Card og Enhanced Color Adapter. Disse gir andre muligheter for bruk av forskjellige tegnssett. Det store problemet for de som ønsker å bruke tegn utenfor standardtegnsettet, er at programvaren må ta hensyn til hvilket videokort som er installert i maskinen. Siden de forskjellige kortene tillater forskjellig kvalitet på erstatningstegnsettene og et varierende antall av dem, vil data som ser utmerket ut på maskin X kanskje ikke kunne vises overhodet på maskin Y. Det er dessverre også slik at programmer som kan bruke tegn utover det vanlige ikke har noen felles standard for hvorledes tegnenes utseende skal beskrives eller tegntabellene settes opp. Mikromaskiner som følger den standarden som ble satt med IBMs PC er altså fleksible i den forstand at vi kan sette forskjellige videokort inn i dem og dermed oppnå en større eller mindre mulighet for skjermmanipulering. På den annen side er vesentlige sider ved maskinens muligheter til å bruke alternative tegnssett knyttet til elektronikken i videokortene, og denne mangelen på standardisering kan skape problemer når data eller programmer skal overføres mellom systemer.

Apple's MacIntosh

I denne maskinen finner vi en annen strategi og arkitektur enn den som er beskrevet ovenfor. MacIntoshen er i utgangspunktet **lukket** og skal prinsipielt bare åpnes av godkjente Apple leverandører. (Det er her bare tale om den opprinnelige modellen opp t.o.m. MacIntosh Plus. Modellene SE og II har helt andre egenskaper enn de som kommer frem i denne artikkelen.) Skjermen er innebygget i systemenheten og det er ikke lagt inn noen mulighet til å skifte videokort (det finnes ingen utskiftbare kort). På tross av at denne konstruksjonen ved første øyekast ikke ser ut til å gi videre rom for variasjon, er maskinen meget fleksibel.

Mye av det som på maskiner av PC-typen er innebygget i elektronikken, er programstyrt på MacIntoshen. Det finnes heller ikke noe

skille mellom grafikk-modus og tekst-modus. Maskinen er alltid i grafikk-modus. På modellene før Mac Plus var alle tegnsett programstyrte og ble lastet sammen med operativsystemet. Derimot er en verktøykasse med kraftige grafikkrutiner innebygget i maskinens ROM. Disse tillater bl.a. at et tegn skrives ut på skjermen slik at det «blander» seg med det som står der fra før. Det er f.eks. mulig først å skrive s, så ^ for å få ś. I motsetning til skrift-modus hos PC-familien er det heller ikke noe i veien for å vise proposjonalskrift på skjermen.

Men alt er ikke bare rosenrødt på en MacIntosh. Det finnes også her forskjellige tekstbehandlingsprogrammer som bruker sine egne systemer for å lagre informasjon om skriftbildet. Disse kan skrive ut en tekst uten koder (tilsvarende «ren ASCII»), men da går informasjon tapt. Siden tegnsettene for mange tekstbehandlingsprogrammer blir lastet sammen med operativsystemet, er man avhengig av å ha et tilsvarende system på mottakermaskinen dersom data skal overføres på en fornuftig måte. Det eksisterer imidlertid programmer som kan skrive tekster ut i et overføringsformat der også kodene kommer med.

8-bit tegnkoder en restriksjon på tegnsett

Med 8 bit kan vi uttrykke 256 forskjellige tegn. I de fleste tilfelle blir bokstaver med diakritiske tegn lagret som egne, unike tegn (i praksis også på MacIntosh). Det er innlysende for de fleste at dette ikke gir et stort nok utvalg av tegn for kinesisk billedskrift, men vi har også andre, mindre ekstreme tilfeller der vi får for få mulige tegn.

Skriver vi en tekst i forskjellige språk, f.eks. engelsk med arabiske kommentarer, kan man bruke en ekstra kode som betyr at tegntabellen skal skiftes. På den måten kan bitgruppen 01000001 (=65) stå for et tegn (i ASCII 'A') så lenge det skrives på engelsk, men etter en kode som viser at vi har skiftet til arabisk tabell, kan den samme bitgruppen stå for noe helt annet. Denne løsningen er ikke så rett frem hvis vi holder oss innen ett og samme språk, og altså ikke egentlig ønsker å markere at vi skifter tegnsett. Hvis vi ser på klassisk gresk inklusive tekstkritiske markeringer, finner vi at antallet mulige tegn fullstendig sprenger grensen på 256 dersom hver sammensatte bokstav defineres som et eget tegn. Bare med vanlige aksenter, iota subscript og spiritus, finnes det tilsammen 96 forskjellige utgaver av de tre vokalene α , γ og Ω . En løsning som etter undertegnede mening har mye for seg, er å gå over til tegnkoder med en lengde på 16 bit, ihvertfall i utvekslingsformater. (På IBMs stormaskiner er det faktisk åpnet for 16 bit tegnkoding.)

Forholdene er, om mulig, enda mer kaotiske på skriverfronten enn de er når det gjelder videokort til mikromaskinene i PC-familien. Vi kan her først og fremst konsentrere oss om matriseskrivere siden de er vanligst, og siden de ofte kan programmeres til å vise forskjellige tegnsett. Typenhjulskrivere faller naturlig nok utenfor ettersom de er fysisk avhengige av hvilket hjul som er plassert i dem. Laserskrivere kan gi interessante muligheter, men de er fortsatt temmelig dyre (i det minste dersom de tillater noe særlig programmering).

En matriseskriver (eller nåleskriver) bruker omtrent det samme prinsippet som skjermen: bokstavene er satt sammen av punkter. Men, mens punktene på skjermen er fast plasserte, er det ikke noe i veien for at punkter fra en skriver kan skrives delvis overlappende. På denne måten blir skriftbildet penere enn om bokstavene var bygget opp av helt adskilte punkter. Det er denne utskriftsmetoden vi finner på matriseskrivere som påstås å ha brevkvalitet (letter quality) utskrift. Siden de fleste matriseskrivere kan styres på punktnivå, er det heller ikke noe i veien for at de kan tegne et hvilket som helst skriftbilde. Mange har også et lite lager som kan ta imot en tegnsesetabell med beskrivelse av de enkelte tegns punktmønster. Problemene her oppstår først og fremst fordi skriverne benytter forskjellige styringskoder. Derfor vil man finne at de fleste tekstbehandlingsprogrammer har ferdige rutiner for å styre et stort utvalg av skrivere, og at man gjerne går gjennom en installeringsprosedyre før man bruker et slikt program for nettopp å fastsette hvilken skriver som er aktuell (se fig. 2).

LOKALE VS. GENERELLE LØSNINGER

Som sagt i innledningen blir det dessverre ofte slik at man innen et institutt e.l. velger egne ad hoc løsninger på problemet med tegnsett og databehandling. Slike løsninger vil typisk være tilpasset den program- og maskinvare som er aktuell i øyeblikket. Det sier seg selv at data da ofte får en form som gjør at de ikke er allment tilgjengelige for andre forskere med en annen datamaskinell utrustning. Men slike løsninger kan være det som gir det beste resultatet for den lokale forsker. Det synes altså som om det kan være ønskelig med et felles utvekslingsformat som beholder all informasjon om tekstformatering, tegnsett osv., samtidig som det må være mulig å arbeide med de verktøy som lokalt gir best mulig resultat. Dette er mulig dersom alle som utvikler tekstbehandlingssystemer **samtidig** lager overføringsrutiner til et slikt fellesformat. Dette formatet kan kanskje bygges på SGML

(Standard Generalized Markup Language) — et system for tekstkodning som er utarbeidet av ISO.

```
WORDSTAR:

Printer is currently :           Brother 2024

***** STANDARD PRINTER TYPES *****

Select the letter of your printer from the list below.
This is menu #1 of 2; to view another menu press the
appropriate number.

A C. Itoh/TEC Starwriter/F10  G Epson MX80/100-no Graftrax
B Centronics 353              H Half line feed printer
C Centronics 739              I IBM Parallel printer
D Diablo 630                  J MPI 88G/99G
E Diablo/Xerox 1610/1620      K NEC 8023A matrix printer
F Diablo/Xerox 1640/1650      L NEC Spinwriter 3550

Enter the letter of your choice,
or enter the appropriate menu number,
```

```
ENABLE:

                                PROFILDEFINISJON
Ønsker du å forandre skriveropsjonene ?  Ja  Nei

Velg skriverkode fra listen nedenfor:  2

Skal skrivertilkoblingen verifiseres ?  Ja  Nei

Hvilken type skrivertilkobling ?  Parallell  Serie

Hvilken parallellport bruker du ?  1=LPT1:  2=LPT2:  3=LPT3:

0=Ikke i listen          7=NEC Spinwriter 7710      E=Epson LQ1500
1=FX80/100/105 - S115    8=Brother Twinwriter 5    F=Epson RX80
2=IBM Graphics Printer   9=HP Think Jet           G=HP Laser Jet
3=Diablo 630             A=Canon LBP 8            H=Qume Sprint
4=Epson MX80 / MX100     B=ORI 92/93              I=Epson LQ800/1000
5=NEC Spinwriter 2050    C=Generell 8-bit         J=
6=FACTI 4513/4514       D=Sperry Modell 5        K=
```

Fig. 2. Et eksempel på menyer for skrivertilpasning i to tekstbehandlingssystemer (WordStar og Enable).

EI FREKVENSBOK FOR NYNORSK

Per Vestbøstad

BAKGRUNN

Tanken om ei nynorsk frekvensordbok voks fram i prosjektperioden for Norsk tekstarkiv (1980-84). Formålet med denne typen ordbøker er kanskje ikkje heilt openbert for folk flest. Lat meg difor peika på at frekvenssorterte ordlister er nyttige når

- a) pedagogane skal freista å finna ut kva for gloser elevane bør læra først,
- b) språkforskarane vil samanlikna ordtilfanget i nynorsk med andre språk, eller dei vil studera kvantitative særtrekk ved språket.
- c) me vil laga maskinleselege ordlister til bruk ved automatisert rettskrivingskontroll.
- d) Språkrådet treng kvantitative opplysningar til å kasta ljøs over språknormeringsspørsmål.

Når dessutan både svensk, dansk og norsk bokmål har fått sine frekvensordbøker, er det klårt at tida er inne for å laga ei for nynorsk. Det kan i seg sjølv vera interessant å samanlikna to så nærskylde målformer som bokmål og nynorsk.

Ankepunktet mot tidlegare norske frekvensordlister har vore at dei berre gjer greie for ulike former av eit ord kvar for seg. Oftast vil ein vera oppteken av å få opplysningar om alle formene t.d. av verbet *vera* utan at ein må slå opp *er*, *var*, *vera* og *vore* og så rekna i hop delfrekvensane sjølv. For å bøta på dette er grunnlagstekstane for Nynorsk frekvensordbok homografseparerte og lemmatiserte. Det vil seia at ulike ord som har likt skrivne former, er talde opp kvar for seg. Dette ser såleis ut til å verta den første allmenne norske frekvensordboka som gjev frekvensar både for oppslagsformer (=lemma) og underformer.

TEKSTGRUNNLAGET

I arbeidet med prosjektet har det lukkast å leggja til rette heller store tekstmengder for datamaskinell bruk, m.a. i eit samarbeid med Norsk språkråd.

Tekstane er lagde inn på ulik vis: Dei fleste vart tasta inn frå terminal på tradisjonell måte, andre vart henta frå diskettar og magnetband, som aviser og trykkeri var generøse nok til å overlata oss, og ein mindre del vart lesen inn med optisk lesar.

Utvalet av tekstgrunnlaget for Nynorsk frekvensordbok vart fastsett i samarbeid med Fagleg råd for Norsk tekstarkiv, der det sat representantar for Språkrådet og dei nordiske institutta ved dei fire universiteta våre. Fordelinga mellom teksttypane er slik:

50% avistekstar frå tida 1980-83; halvta av dette er henta frå regionale og nasjonale dagsaviser, den andre halvta er tekstar frå lokalaviser.

10% er stoff frå *Dag og Tid* (1981).

20% er artiklar frå *Syn og Segn* (1978 og 1983). Her er berre nytta artiklar som er skrivne på nynorsk opphavleg — ikkje omsette artiklar.

20% romantekstar utjevne i tida 1977-82: Jamnstore bitar frå «Dalen Portland» (Kjartan Fløgstad), «Born av støv» (Johannes Heggland), «Carolus, klovn» (Jon Hellesnes), «Melding frå Petrograd» (Edvard Hoem), «Alltid fleire dagar» (Ragnar Hovland), «Sjå Jæren, Gamle Jæren» (Tor Obrestad) og «Hild Rogne» (Tormod Skagestad).

BEARBEIDING AV TEKSTANE

Arbeidet med tekstane har vore gjennom desse hovudfasane:

1. Innskriving og korrekturlesing
2. Køyring av frekvensliste
3. Homografseparering
4. Lemmatisering
5. Konsekvenskontroll

1. Innskriving/innlesing og korrektur. Denne fasen var naturleg nok langvarig og arbeidskrevjande. Me prøvde å bruka optisk innlesing både på avistekstar og romanar, men fann at det vart dyrare enn beinvegs inntasting. Førtesekr. *Doreen Grotdal* har største æra for

at dette gjekk svært så bra. Sivilarbeidarane *Audun Skjervøy* og *Odd Ketil Rangnes* gjorde også ein god innsats i denne fasen, som strekte seg over åra 1981-84. Under korrekturen vart m.a. alle bokmåls-sitata tekne bort.

2. Til frekvensteljingane brukte me Senteret sitt KVIKKIS-program. Lista over ord med frekvens 1 og 2 vil alltid avdekkja ein del feil, så første frekvenskøyringa vart eit supplement til korrekturlesinga. Grunnlagsmaterialet vart no talt opp til 1.029.000 ord i alt. Før den vidare bearbeidinga tok me ut alle reint utanlandske ord og alle namn utanom norske geografiske namn. Dei finst i eigne lister, om nokon skulle ha interesse av å studera dei.
3. Homografsepareringa gjekk føre seg ved at alle ordformer som *kunne* ha oppslagsform med ulike ordklassar, vart lista ut i kontekst, slik at me kunne sjå kva for oppslagsform dei skulle sorterast under. Berre homografar som har ulik ordklasse, genus eller følger ulike bøyingsmønster, vart skilde åt. *Ein fyr* og *eit fyr* vart såleis talde opp kvar for seg, medan det ikkje vart gjort noko skilje mellom ein fyr i meininga *mann* og fyr i meininga *flamme*, sidan desse har same genus og bøyingsmønster. Grunnen til at grensa vart sett her, var at det elles er svært vanskeleg å finna objektive kjenneteikn på kvar grensa mellom ulike lemma og interne tydingsvariantar går.
4. Lemmatiseringa vart gjort ved at me først sorterte alle ordformer etter oppslagsform. Eit spesialprogram laga av konsulent *Solveig Hovland*, talde så i hop frekvensane for alle underformer for kvart lemma. Programmeringsarbeidet her var gjort parallelt med fase 3.
5. Kontroll av ordklassemarkeringane og tilvisingsprinsippa: Ordklassemarkeringane var problematiske når det galdt å skilja adjektiv frå partisippformer og substantiv frå verbale -ing-former. Rettesnora var at stivna verbalformer skulle merkjast som adjektiv eller substantiv, så snart det ikkje var naturleg å tenkja dei inn i eit fullt bøyingsmønster for det opphavlege verbet. (Døme: oppskrytt, adj. sidan infinitiven å *oppskryta* ikkje finst.) Avgjerdene *må* her byggja mykje på skjøn.

Dei endelege listene er kontrollerte 3 gonger, men likevel reknar me med at feil og inkonseksensar kan finnast. Me trur likevel at kvaliteten no er god nok til at Nynorsk frekvensordbok er til å lita på.

MEDARBEIDERAR

Konsulent Hovland og eg har følgd arbeidet heile tida. Driftsass. *Jarle*

Helle sette på ordklassemerke, medan cand. mag. Odd Ketil Rangnes og cand. mag. Unni Utvik har vore faglege medarbeidarar i fase 3, 4 og 5. Rangnes hadde den daglege leiinga av homografsepareringa og har gjort mykje av kontrollarbeidet.

Prosjektansvaret har eg hatt høve til å dela med direktør Jostein Hauge (NAVFs edb-senter), og gode støtter i det faglege arbeidet har vore professor Egil Pettersen (Univ. i Bergen) og førsteamanuensis Gulbrand Alhaug (Univ. i Tromsø).

PUBLISERING

Utanom den *frekvenssorterte* lista med rangnr. og samla frekvens for kvart lemma, har Solveig Hovland laga program som gjev *baklengssortert* (*finalalfabetisert*) lemma-liste (utan underformer) og ei *alfabetisk liste* over alle lemma med både samla frekvens og frekvens for kvar underform (sjå prøvene bak denne artikkelen). Desse listene er det meininga å gje ut i bokform gjennom eit etablert forlag, men me har ikkje gjort nokon endeleg avtale enno.

Mykje av det språkmaterialet som er lagt til rette under arbeidet med Nynorsk frekvensordbok, kan ha interesse for andre. Me tek difor sikte på å kunna levera materiale i ulike bearbeidingar både i utskrift og på diskettar eller magnetband til interesserte. Slike tenester reknar me med å ta noko betaling for.

Rang nr.	Frekv.	Kum. frekv.%	Rang nr.	Frekv.	Kum. frekv.%
1 den, <i>pron</i>	35965	3.499	52 gjera, <i>v</i>	2684	45.908
2 i, <i>pp</i>	35130	6.916	53 der, <i>av</i>	2678	46.169
3 og, <i>k</i>	34415	10.264	54 slik, <i>pron</i>	2666	46.428
4 vera, <i>v</i>	33055	13.479	55 opp, <i>av</i>	2583	46.679
5 ein, <i>a</i>	19997	15.424	56 du, <i>pron</i>	2569	46.929
6 på, <i>pp</i>	19346	17.306	57 ved, <i>pp</i>	2466	47.169
7 til, <i>pp</i>	18236	19.080	58 ny, <i>ad</i>	2297	47.392
8 ha, <i>v</i>	17433	20.776	59 over, <i>pp</i>	2260	47.612
9 med, <i>pp</i>	13786	22.117	60 dag, <i>s</i>	2179	47.824
10 å, <i>inf.mrk</i>	13296	23.410	61 berre, <i>av</i>	2110	48.029
11 for, <i>pp</i>	12830	24.658	62 heil, <i>ad</i>	2069	48.231
12 dei, <i>pron</i>	12784	25.902	63 inn, <i>av</i>	2064	48.432
13 at, <i>k</i>	12603	27.128	64 når, <i>k</i>	2046	48.631
14 av, <i>pp</i>	11956	28.291	65 eller, <i>k</i>	2010	48.826
15 han, <i>pron</i>	10276	29.291	66 så, <i>k</i>	1939	49.015
16 ikkje, <i>av</i>	10203	30.283	67 her, <i>av</i>	1913	49.201
17 om, <i>k/pp</i>	7954	31.057	68 stå, <i>v</i>	1854	49.381
18 denne, <i>pron</i>	7678	31.804	69 fram, <i>av</i>	1814	49.558
19 eg, <i>pron</i>	7539	32.537	70 tid, <i>s</i>	1755	49.728
20 vi, <i>pron</i>	7538	33.270	71 som, <i>k/pron</i>	1741	49.898
21 men, <i>k</i>	6319	33.885	72 mot, <i>pp</i>	1719	50.065
22 kunna, <i>v</i>	6152	34.484	73 liten, <i>ad</i>	1677	50.228

Alfabetisk liste over ord med frekvens høyere enn 1

80	prinsipp, <i>s</i>	29	prinsipp	2	prissystem, <i>s</i>	1	prissystem
		13	prinsippa	1		1	prissystemet
5	prinsipp-program, <i>s</i>	38	prinsippet	2	pristilbud, <i>s</i>	1	pristilbud
		1	prinsipp-program			1	pristilboda
		3	prinsipp-programmet	2	pristilhøve, <i>s</i>		pristilhøve
		1	prinsippprogram	2	pristilskot, <i>s</i>		pristilskot
2	prinsippsspørsmål, <i>s</i>	1	prinsippsspørsmål	2	pristudeling, <i>s</i>		pristudelinga
		1	prinsippsspørsmål	2	prisverdig, <i>ad</i>		prisverdig
2	prinsippvedtak, <i>s</i>	1	prinsipp-vedtak	122	privat, <i>ad</i>	53	privat
		1	prinsippvedtak			1	privat-
74	prioritera, <i>v</i>	1	prioriter			68	private
		11	prioritera	3	privatbank, <i>s</i>		privatbank
		10	prioriterast	8	privatbil, <i>s</i>		privatbil
		17	prioritere			3	privatbilar
		11	prioriterer			3	privatbilen
		17	prioritert	4	privatleigd, <i>ad</i>	2	privatleigd
		7	prioriterte			2	privatleigde
23	prioritering, <i>s</i>	18	prioritering	2	privatfolk, <i>s</i>		privatfolk
		4	prioriteringa	3	privatthus, <i>s</i>		privatthus
		1	prioriteringar	3	privatisera, <i>v</i>		1 privatiserande
13	prioriteringsliste, <i>s</i>	5	prioriteringslista	6	privatisering, <i>s</i>	2	privatisert
		7	prioriteringsliste	8	privatliv, <i>s</i>	3	privatisering
		1	prioriteringsliste			3	privatliv
2	prioriteringsregel, <i>s</i>	1	prioriteringsreglane			5	privatlivet
		1	prioriteringsreglar	6	privatperson, <i>s</i>		privatpersonar
18	prioritet, <i>s</i>	17	prioritet	3	privatpraksis, <i>s</i>		privatpraksis
		1	prioritets	4	privatpraktiserande, <i>ad</i>		privatpraktiserande
3	prioritetsliste, <i>s</i>		prioritetslista	4	privatføre, <i>s</i>		privatføeren
241	pris, <i>s</i>	81	pris	2	privatøkonomisk, <i>ad</i>		1 privatøkonomisk
		4	pris-			1	privatøkonomiske
		29	prisane	4	privilegium, <i>s</i>		1 privilegia
		22	prisar			1	privilegie
		103	prisen			1	privilegiet
		2	priser			1	privilegium
4	prisa, <i>v</i>	3	prisa	5	priviligert, <i>ad</i>	1	priviligert
		1	prisar			2	priviligert
15	prisauke, <i>s</i>	4	prisauke			2	priviligerte
		11	prisauken	2	pro, <i>pp</i>		pro
			prisdirektoratet	2	proarabisk, <i>ad</i>		1 pro-arabisk
9	prisdirektorat, <i>s</i>		prisdirektorene			1	pro-arabisk
2	prisforskrift, <i>s</i>	1	prisforskriftene	466	problem, <i>s</i>	278	problem
		1	prisforskriftene			87	problema
2	prismessig, <i>ad</i>	1	prismessig			1	problemane
		3	prismemnd			5	problemer
4	prismemnd, <i>s</i>	1	prismemnda			94	problemet
		1	prismnivå	2	problemartikkel, <i>s</i>		1 problemartikkelen
5	prismnivå, <i>s</i>	4	prismnivået				1 problemartiklar
			prispolitikken	3	problematikk, <i>s</i>		1 problematikk
2	prispolitikk, <i>s</i>		prisreguleringane			2	problematikken
2	prisregulering, <i>s</i>	3	prisstigninga	3	problematisera, <i>v</i>		1 problematiseringa
6	prisstiging, <i>s</i>	2	prisstigning				1 problematiserer
		1	prisstigninga				

Baklengssortert liste over oppslagsord

indar s, 5	overliggar s, 1	attgangar s, 1	forsongar s, 2
haugesundar s, 1	tverrliggjar s, 1	hangar s, 2	korsongar s, 1
kristiansundar s, 1	steinhoggar s, 3	handdlangar s, 1	storsongar s, 1
formyndar s, 3	bitopphoggar s, 3	tilhengar s, 21	bluessongar s, 1
syndar s, 13	giesteinnbyggjar s, 1	nynorsktilhengar s, 2	kastratsongar s, 1
medsyndar s, 1	eigar s, 88	kyrkjeggengar s, 3	borgar s, 1b
farussyndar s, 1	nybil-eigar s, 1	kinogengar s, 1	medborgar s, 1
dekkodar s, 1	medeigar s, 7	forjengar s, 7	samfunnsborgar s, 1
tilskodar s, 16	jordeigar s, 2	dobbeljengar s, 2	statsborgar s, 6
elefantskodar s, 1	hundeigar s, 2	foigjengar s, 13	byborgar s, 2
åskodar s, 1	sanatorieigar s, 1	slengar s, 9	småborgar s, 2
leigemordar s, 1	kyrkjeigar s, 2	pengar s, 423	lugar s, 6
skrivebordsmordar s, 1	plantasjeigar s, 1	kassapengar s, 1	blodsugar s, 3
sjefsmordar s, 1	aksjeigar s, 2	bladpengar s, 1	støvsugar s, 1
lydar s, 12	drosjeeigar s, 1	brødpengar s, 2	flygar s, 12
forredar s, 2	eineigar s, 2	oljepengar s, 1	friflygar s, 1
landsforredar s, 1	tomteigar s, 2	vekepengar s, 1	einsamflygar s, 5
europear s, 2	hytteigar s, 4	sjukepengar s, 4	sågar av, 2
austeuropear s, 1	sauceigar s, 2	vekslepengar s, 2	kanadiar s, 1
storbyeuropear s, 1	elveigar s, 3	reklamepengar s, 1	plebeiar s, 1
trear s, 12	skogeigar s, 9	lommepengar s, 2	feiar s, 1
eritrear s, 16	fabrikkengar s, 4	reisepengar s, 1	leiar s, 184
farisear s, 1	leigar s, 3	skattepengar s, 5	geriljaleiar s, 2
far s.n, 3	bileigar s, 8	dagpengar s, 5	klubbleiar s, 4
far s.m, 375	lastebileigar s, 2	materialpengar s, 1	hovudleiar s, 1
oldefar s, 11	hotelleigar s, 2	bompengar s, 18	områdeleiar s, 1
familiefar s, 2	motelteigar s, 1	deltakarpengar s, 1	studieleiar s, 11
bekkefar s, 3	uteigar s, 2	fondspengar s, 1	landsstudieleiar s, 1
gamlefar s, 1	sameigar s, 1	kulturrådspengar s, 1	serieleiar s, 4
oksefar s, 3	reineigar s, 11	klesspengar s, 1	byggjeleiar s, 3
fotefar s, 1	tamreineigar s, 1	sundagspengar s, 1	musikkuleleiar s, 1
bestefar s, 24	maskineigar s, 2	inggangspengar s, 1	stjerneleiar s, 1
ættefar s, 3	grunneigar s, 80	årspengar s, 1	gruppeleiar s, 2
elvefar s, 2	bensinstasjonleigar s, 1	skjysspengar s, 1	venstreleiar s, 2
grovefar s, 2	godsleigar s, 11	statspengar s, 1	reiseleiar s, 3
skaffar s, 1	bruksleigar s, 2	idrettspengar s, 1	viseleiar s, 1
fulltreffar s, 1	huseigar s, 7	bruspengar s, 1	møteleiar s, 2
loffar s, 3	sjukehuseigar s, 1	dynamittpengar s, 2	lagleiar s, 1
omstrefar s, 3	sjøseigar s, 1	elevpengar s, 1	ringleiar s, 9
farfar s, 1	noteigar s, 1	småpengar s, 2	forsøksringleiar s, 1
verfar s, 1	fjordhesteigar s, 1	meslingar s, 2	songleiar s, 1
morfar s, 3	båteigar s, 6	springar s, 7	industrielleiar s, 2
trebarnsfar s, 1	småbåteigar s, 1	hallingspringar s, 1	partielleiar s, 3
husfar s, 4	sjolveigar s, 2	bøheringspringar s, 1	leikleiar s, 1
gar s, 1	krigar s, 3	vestlandsspringar s, 1	målleiar s, 1
jetjagar s, 1	gallarkrigar s, 1	tingar s, 15	programleiar s, 13
klagar s, 3	sigar s, 8	førehandstingar s, 1	kristentleiar s, 3
trommeslagar s, 3	rypefangar s, 1	songar s, 29	renntleiar s, 3
hovslagar s, 4	kvalfangar s, 2	operasjonar s, 1	nestrenntleiar s, 1
dragar s, 2	støytfangar s, 1	folkesongar s, 1	innleiar s, 10
oppdragar s, 2	gangar s, 4	visesongar s, 6	nynorskinnleiar s, 1
siloutleggar s, 1	gjengangar s, 1	stagersongar s, 1	industrivernleiar s, 1

RAPPORTER

FEM UNDERVISNINGSPROGRAMMER FRA DATASEKRETARIATET

Espen S. Ore

Fra Datasekretariatet mottar NAVFs edb-senter for humanistisk forskning programmer til utprøving. Siden Senteret ikke er en utdanningsinstitusjon, går vi lite inn på pedagogiske sider ved programmene når vi prøver dem. Istedet ser vi på programmeringsmessige forhold slik som sårbarhet for gale data fra brukeren (går f.eks. programmet i stå om vi oppgir navnet til en ikke-eksisterende fil?) og brukervennlighet.

Her vil jeg beskrive 5 programmer: *Artist*, *Ballong*, *Dataflora*, *Full fart med Newton* og *Vevplan*. Programmene er beregnet på forskjellige skoleslag og aldersgrupper. I bruksanvisningen til Newton står det at programmet kan brukes fra 4. klasse i grunnskolen til universitetsnivå. I den grad det er angitt noe for de andre programmene, faller det innenfor denne rammen. Alle programmene er prøvd på en eldre Tiki-100 med 2x188 KB diskettstasjoner. (Maskinen er såpass gammel at den hilser velkommen med «KON TIKI», og det er en ekte skolemodell uten lydgenerator.) Bruksanvisningen for noen av programmene er beregnet på Scandis-versjonen, men siden det her er tale om testversjoner, lar jeg ikke dette trekke ned i mine vurderinger.

Under prøvingen har jeg først forsøkt å starte programmene ut fra instruksjoner som er skrevet på disketten (f.eks. «Diskett i venstre sp. Trykk SHIFT+BRYT (Ikke skrivebeskytt.)»). Dette virket ikke alltid

like godt på min maskin, av og til ble grafikken/fargene uklare. Problemene løste seg imidlertid hvis jeg først startet maskinen med en av mine egne oppstartdisketter slik at skjermen ble satt i riktig tilstand. Deretter kunne jeg sette inn Datasekretariat-disketten og starte skoleprogrammet «manuelt» (altså ved skrivehake gi f.eks. kommandoen NEWTON). Dette var ikke så stort problem for meg, men slikt bør unngås i ferdige programmer som skal ut i skolen.

Selv om jeg alltid har prøvd å komme så langt som mulig med et program *uten* å lese bruksanvisningen, har jeg også sett på hvor egnet den virker. Siden alle programmene har vært prøveversjoner, kan jeg ikke si noe eksakt om bruksanvisningenes tilstand i dag, men det er verdt å merke seg at ingen hadde indeks. Artist og Ballong hadde heller ikke innholdsfortegnelse i sine bruksanvisninger. Alle bruksanvisningene var gjerrige på tekniske opplysninger som filformater o.l. Dette er muligens gjort bevisst for å sikre mot tyveri av ideer, men personlig oppfatter jeg dette som en svakhet ved dokumentasjonen.

ARTIST

For dette programmet er det ikke angitt hva slags undervisning det er beregnet på, og det er bare rimelig. Programmet er et generelt tegne-/maleprogram for Tiki. Bortsett fra at det styres fra tastaturet og ikke fra en mus, kan det betegnes som Tikis (farge)svar på MacPaint med ett viktig forbehold: mens god MacIntosh programvare stort sett er selvforklarende, var dette det mest uforståelige av de fem programmene jeg prøvde. Dette var den eneste gangen jeg var nødt til å lese i bruksanvisningen før jeg overhodet fikk noen resultater.

Bruksanvisningen til Artist er temmelig stor og ufremkommelig. Dette er synd, for programmet er i seg selv meget kraftig. Man kan tegne linjer med varierende tykkelse, fylle flater med farger, bruke egne kommandoer for å konstruere sirkler og rektangler osv. Bildene man har laget kan lagres på filer. I bruksanvisningen står det at filene tar 31 KByte. Dette må bety at skjermbildet lagres «rått» med 4 bit pr. skjermepunkt (pixel). Det foregår altså ingen datakomprimering. Det hadde her vært interessant å vite litt mer om filformatet med tanke på kobling mot egne programmer.

Alt i alt er Artist et kraftig program med mange gode muligheter (som man oppdager når man har lest bruksanvisningen). Det er vanskelig å se hva det er ved programmet som spesielt gjør det til et skoleprogram. Jeg synes det er bedre å kalle det et godt generelt tegneprogram for 8-bits TIKI som utnytter de mulighetene for farger og grafikk som denne maskinen gir.

BALLONG

I bruksanvisningen til dette programmet blir det heller ikke sagt noe om hvilket klassesertrinn programmet er beregnet for. Ut fra vanskelighetsgrad og språkbruk går jeg ut fra at det er ungdomsskolen. Programmet simulerer en tur med en varmluftsballong og anskueliggjør bl.a. hvorledes ballongens oppdrift er betinget av forskjellen mellom temperaturen utenfor og inne i ballongen.

Programmet var nærmest selvforklarende. Det eneste man ikke uten videre kunne få på skjermen, var en tabell over ballongens lasteevne basert på forholdet utetemperatur/temperatur i ballongen. Siden denne tabellen blir fremstilt som to rettlinjede kurver (én for en vinterdag med -12°C og én for en sommerdag med 15°C), skal det bare to forsøk for hver av utetemperaturene til før man kan lage sin egen tabell uten å ha åpnet bruksanvisningen. At både velkomstmenyen på skjermen og bruksanvisningens illustrasjon av den bruker uttrykket «metereologiske data» får tilskrives at det kun er tale om en foreløpig versjon fra mars 1987. Likevel synes jeg det er litt merkelig at man ikke innen skoleverket er spesielt opptatt av ortografi.




















I velkomstmenyen står det «Dette programmet er IKKE et spill, men et fysikkprogram.» Jeg er redd det likevel lett kan bli oppfattet som et spill, ikke minst fordi bruksanvisningen slutter med en oppfordring til brukeren om å slå rekorden for økonomisering med drivstoffet. Det gir også inntrykk av arkadespill at hele skjermbildet rister når ballongen krasjlander. Derimot må det være en pedagogisk frihet forfatterne har tatt seg, når de lar ballongføreren dø av surstoffmangel i en høyde av 3000 meter.

Som en slags oppsummering kan jeg si at programmet virker litt tynt. Grafikken og fargene på Tiki er nesten ikke utnyttet i det hele tatt, og det skulle vært unødvendig å fikser de meteorologiske data til to dager uten noen variasjon. Til gjengjeld var bruksanvisningen full av artige illustrasjoner.

DATAFLORA

Om dette programmet blir det sagt klart fra at det er beregnet på bruk i «naturfagsundervisningen (botanikk) i grunnskolens ungdomstrinn. (7.-9. klasse)». Videre bygger programforfatterne på at ungdomsskoleelever har vansker med å bruke de bestemmelsesnøkklene vi finner i tradisjonelt oppbygde floraer. (Det er kanskje ikke bare ungdomsskoleelever som kan ha problemer her ...)

Programmet var så lett å bruke at jeg tror man kan klare seg helt

ELEVARK.		KRYSS AV FOR PLANTENS KJENNETEGN				
BLOMST	<input type="checkbox"/> Stånd	 Enkeltblomst	 Kurv/Høpe	 Skjerm	<input type="checkbox"/> ANDRE	
	<input type="checkbox"/> Fange	 Kjølvet/ørre	 Kronblad	 Kronblad	 Blottersilene	<input type="checkbox"/> ANDRE
	<input type="checkbox"/> Farge	BLÅ	GUL	HUIT	RØD ROSA	FLER- FARGET
BLAD	<input type="checkbox"/> Stålling	 Krans	 Høtsatt	 Rosett	 Spredt	<input type="checkbox"/> ANDRE
	<input type="checkbox"/> Fønte	 Vanlig blag	 Voldgrasblat	 Blått/Koblet	 Falket	<input type="checkbox"/> ANDRE
	<input type="checkbox"/> Håndner	 Busknervet	 Blånervet	 Håndnervet	 Retnervet	<input type="checkbox"/> ANDRE

uten skriftlig bruksanvisning. Kort beskrevet inneholder systemet en planteliste der hver plante er registrert med felter som *Familie*, *Latinsk navn* etc. Man kan slå direkte opp i listen ved hjelp av norsk navn, eller man kan bruke ett eller flere av følgende kriterier: *Bloms-terstand*, *Krone*, *Farge*, *Bladstilling*, *Bladform* og *Bladnerver*. Ved å bruke den siste metoden får man en liste over ingen eller flere norske plantenavn. Man kan så gå gjennom denne listen og få komplett informasjon om de enkelte blomster for å avgjøre tvilstilfeller.

Programforfatterne skriver at de har lagt vekt på at en blomst skal kunne bestemmes ut fra enkle, visuelle kjennetegn. Dette har man som sagt forsøkt å gjøre slik at det skal bli lettere enn å bruke flora-nes systematiske nøkler. Men ved å lage et program som dette har man oppnådd noe i tillegg: *parallellitet*. De seks kriteriene som ble nevnt ovenfor er likeverdige! Dette står i skarp motsetning til den tradisjonelle nøkkelen som vi kan forestille oss som et tre der uheldig valg av forgrening på et tidlig tidspunkt gjør at alt som følger blir meningsløst.

De som har oppfattet meg dithen at dette programmet også kan brukes utenfor klassesituasjonen, har delvis rett. Dessverre er kriteriene få, og valgmulighetene innen de enkelte kriteriene likeså. Som et eksperiment prøvde jeg å bestemme Bergfrue (*Saxifraga Cotyledon*). Etter å ha oppgitt verdier for alle søkekriterier (unntatt norsk navn, selvfølgelig) satt jeg igjen med to blomster: Bergfrue og Jåblom. Disse er riktignok i familie, men må sies å skille seg klart fra hverandre med «enkle, visuelle kjennetegn». Det er mulig å redigere data i plantelisten, men man kan i praksis ikke endre verdiene for oppslagskriterier uten å forvrengte fakta. Det finnes heldigvis andre felter enn oppslagskriterier i plantelisten, og her kan noe gjøres ved f.eks. å legge inn en mer nøyaktig beskrivelse i feltene for «Opplysninger».

Tross innvendingene ovenfor likte jeg dette programmet. Det gjør ikke noe særlig bruk av TIKI-grafikk, men er ellers svært brukervennlig. Dersom det ble laget en versjon med flere (evt. mer tekniske) kriterier kunne dette programmet egne seg som et generelt bestemmelsesprogram for blomster.

NEWTON

I bruksanvisningen står det: «Programmet passer for elever fra 4. klasse og oppover, gjerne opp til universitetsnivå.» Som et bidrag til diskusjonen om nivåsenkning på norske universiteter er dette hard kost. Programsettets fulle navn er «FULL FART MED NEWTON». Litt forenklet virker programmet slik: På skjermen er det avbildet en

flate (sett ovenfra) der det befinner seg en puck (eller et eller annet annet som kan skyves bortover en flate). Denne pucken kan man så daske til ved å bruke piltastene på tastaturet, og har man da satt friksjonen lik 0,0 (unnskyld, 0.0 - selv om dette programmet skal være norsk har man her innført desimalpunktum i stedet for -komma), går pucken rundt på skjermen til man blir lei og slår av maskinen (eller gir en stopp-kommando). Er friksjonen ikke 0, stopper pucken før eller siden.

I **Den bakvendte familie**boken viser forfatterne (André Bjerke, Odd Eidem og Carl Keilhau) til elevforsøk i fysikk der eleven kan stikke et termometer inn i smeltende sne og se at temperaturen forblir 0°C. Dette programmet ligger i samme gate.

Etter så mye negativt vil jeg gjerne gjøre oppmerksom på følgende: Programmet virker teknisk sett bra, og det ga fornuftig respons da jeg prøvde å hente et skjermbilde fra en ikke-eksisterende fil.

VEVPLAN

Det er ikke angitt noe klassetrinn eller skoleslag for dette programmet, men det virker som om det kunne egne seg fra ungdomsskole og oppover. På samme måte som Artist (og til dels Dataflora) er dette et verktøy, ikke et kursprogram som skal gi en klart definert informasjon (som Ballong og Newton etter min mening er).

Programmet skal brukes der en før satt med ruteark og blyant og mye manuelt arbeid under konstruksjon av vevmønstre. Det gir ikke noen innføring i veving, og mine begrensede kunnskaper her ble tøyd til det ytterste og vel så det da jeg prøvde programmet. Likevel kom jeg ganske godt igang uten å lese bruksanvisningen, selv om det der anbefales i det minste å lese kapitlet «Første gang — for den utålmodige».

Programmet kan best sammenlignes med Artist. Det har mange muligheter, men krever litt arbeid i innlæringen. Forsøk på å oppgi en ikke-eksisterende datafil ble ikke taklet særlig elegant — jeg kom ikke ut av dette modus på annen måte enn ved å oppgi en fil jeg visste eksisterte og så fjerne de innleste data.

Jeg likte programmet, og dersom det kommer en versjon for brikkevev, vil jeg bestille den.

THE MLA BIBLIOGRAPHY GENERATOR:

HOW DOES IT COMPARE WITH THE BIBLIOGRAPHY FUNCTIONS OF GENERAL TEXT-PROCESSING SYSTEMS?

Ralph Jewell

Shapiro, Marvin, and Ted Salzman. **Bibliography Generator.**

A Derivative Work of The MLA Style Manual and The MLA Handbook for Writers of Research Papers (2nd Edition). Published by Educational Activities, Inc., under licence from the Modern Language Association of America.

The text-processing programme **Nota Bene** has been designed with the needs of academic writers especially in mind. All aspects of its design bear witness to this clear fact. Even in its first version, 1.0, it drew high praise from its users. Version 2.0 saw many additions and enhancements — too many to review here. Prominent among influential recommendations of **Nota Bene** is the one given to its members by the Modern Language Association of America (MLA).

With the needs of academic writing so obviously forming the focal design consideration of **Nota Bene**, it is not surprising that among its strong points the handling of bibliographical information is one of them. Yet in an information leaflet distributed among its members the MLA appended to its recommendation of **Nota Bene** a further recommendation of a separate utility programme, the **Bibliography Generator**, published by Educational Activities, Inc. under licence from the MLA.

Given **Nota Bene**'s own strengths in the field of handling bibliographical information the question might well occur to the active **Nota Bene** user: What purposes are served better by the free-standing **Bibliography Generator** than by **Nota Bene** alone? The question is not without interest to users of other text-processing systems capable of

handling bibliographies efficiently. How, then, does the **Bibliography Generator** work, and what are its special strengths, if any, over and above the bibliography capabilities of more comprehensive text-processing systems such as **Nota Bene**?

The **Bibliography Generator** is supplied in the form of a programme diskette and an eight-page user guide. Through a series of menus and prompts it asks specific questions about each work to be registered, processes the answers and then assembles a citation in conformity with the strict style and format rules concerning order of presentation, punctuation, spacing, and abbreviations, as set out in the *MLA Style Manual*.

After the user has opened an appropriate file the first set of choices the programme presents is the menu for **ADDING TO FILE**. This is a list of the main categories of likely types of works to be cited:

- 1 — Book or Pamphlet
- 2 — Work or Article in a Collection
- 3 — Article in a Reference Book
- 4 — Article, Editorial, Letter in a Periodical
- 5 — Review in a Periodical
- 6 — Introduction, Preface, Foreword, or Afterword
- 7 — Government Publication
- 8 — Dissertation
- 9 — Other, Miscellaneous

The next screen-menu seeks information about the names and functions of the persons associated with the work, whether author, translator, editor, or compiler. In accordance with the requirement of producing strictly standard formats the programme goes a long way towards suppressing entry errors, e.g. by capitalising names and initials and entering periods after initials even when incorrectly entered by the user.

The third screen-menu concerns the titles of the works to be cited, and requests responses to the following prompts:

- Name of Publisher:
- City of Publication:
- Edition (if not first):
- Year of Publication:
- Is this a republished book?
- Are there multiple publishers?

The final screen-menu provides the opportunity for entering additional information of a kind that had not been covered by the previous prompts.

After these stages the programme assembles the complete formatted citation and displays it on the screen. This may then be edited or, if without error, it may be saved.

Two further screen-menus cover printing options and other utilities. The most important of these further optional utilities is MAKE AN ASCII FILE which enables files generated by the **Bibliography Generator** to be incorporated into documents produced by other text-processing systems which, like **Nota Bene**, can accept ASCII data files.

The strength of the **Bibliography Generator** is that it does all the formatting automatically and strictly according to fixed convention. This strength involves a weakness. The **Bibliography Generator** does not receive into its files bibliographical entries already prepared with other text processing systems. The **Bibliography Generator** is water-tight against intrusion by text written elsewhere.

If we return to **Nota Bene** we may see a part of the explanation for the double recommendation that the MLA made to its members. Once the **Bibliography Generator** has assembled a citation or file of citations, and then transformed it to an ASCII data file, **Nota Bene** or any other ASCII-receptive text-processor may receive it and continue to use it exactly as is. Most text-processors allow the defining of blocks and for the subsequent assigning of a specific key-combination to them. A simple method for multiple entry of the imported citation would then be to assign the whole citation-text to a so-called phrase key by making it a defined block and then to assign it to the required key-combination — ALT and any alphanumeric key in the case of **Nota Bene**. Then, whenever a full citation is required in the text, exactly the same properly formatted citation is available. On making notes at the computer each note could be equipped with the full bibliographic citation of the source by using the two-key combination, for example. And since multiple citations to the same work would then have not only a standard format regarding style and punctuation but would be also identical, duplicates would be weeded out during the automatic compilation and alphabetic ordering of a document's final bibliography or bibliographies.

The **Bibliography Generator** is undoubtedly very useful in eliminating some of the uncertainties and pitfalls of producing bibliographies, even if its strength involves a limitation as I have indicated. There is an attractive simplicity and reliability in the way it draws responses

to specific requests and then proceeds to take steps to rectify, assemble and format just the very things that for many or even most authors of scientific papers are an unusually exacting challenge to typing accuracy — bibliographical citations. For those who take part in collaborative bibliographical enterprises with distant colleagues, and where complete standardization is essential, the **Bibliography Generator** would be a great and obvious boon.

Ralph Jewell is Senior Lecturer at the Department of Philosophy, University of Bergen.

POLITISKE HOLDNINGER

LÆREPAKKE FOR DEN VIDEREGÅENDE SKOLE

Einar-Arne Drivenes

Lærepakken som består av en lærebok/arbeidsbok og statistikkprogrammet NSD-STAT med tilhørende datasett, er laga ved Norsk samfunnsvitenskapelige datatjeneste (NSD) i Bergen. Lærepakken består av ei hendig arbeidsbok/instruksjonshefte og en diskett. Den er utarbeidd med sikte på bruk i den videregående skolen både i samfunnsfagsundervisninga og i eventuelle samarbeidsprosjekter mellom edb og samfunnskunnskap.

Når vi vet at bruken av edb-verktøyet i sterk grad har vært knytta til realfagene, må selv kun forsøk i den retning NSD har gjort, applauderes. Det er svært positivt at institusjoner som NSD også retter oppmerksomheten mot skoleverket. Og så langt jeg kan se er dette blitt en lærepakke som egner seg godt både som en innføring i samfunnsvitenskapelig analyse og de metoder/hjelpemidler (herunder edb) samfunnsviteren nytter i sine undersøkelser.

Datasettet som denne lærepakken bygger på, er henta fra Valgundersøkelsen 1981 som ble gjennomført av professor *Henry Valen* og Statistisk Sentralbyrå. Datasettet omfatter 1586 personer og inneholder opplysninger om ulike holdninger til aktuelle samfunnsspørsmål — fra abort til økonomisk vekst. I alt inneholder datasettet 25 variabler.

I læreboka/instruksjonsboka legges det stor vekt på å sette de samfunnsvitenskapelige problemstillingene i fokus. Med stor konsekvens greier en å fange leserens interesse om det viktigste — hvilke spørsmål er det en ønsker å få svar på og hvilke teoretiske og praktiske problemer må en tenke gjennom før en kan begynne å trykke på knappene for å få de nødvendige tabellene opp på dataskjermen.

Læreheftet gjennomgår formålet med meningsmålinger og gir ei grei innføring i hva ei meningsmåling er. Videre blir datamaterialets oppbygning og innhold gjennomgått. Et eget kapittel tar opp representativitetsproblemet når det gjelder datamaterialet i denne lærepakken.

De siste fire kapitlene går så inn på ulike problemstillinger som dette datamateriale kan bidra til å belyse; endringer i holdninger, holdningsdannelse, konfliktmønstre, holdninger og politikk. I gjennomgangen av dette gis en innføring i de ulike samfunnsvitenskapelige metoder/teknikker som en kan gjøre bruk av under analysen av materialet.

Læreheftet skal være et supplement til andre lærebøker og en kan selsagt ikke forlange at en på så få sider kan få fram alle metodiske nyanser og at alle formuleringer skal være like presise. Men i et par, tre tilfeller synes jeg nok forfatterne beveger seg på elefantføtter i så måte. Formuleringene om årsaksvariabler og virkningsvariabler og forholdet til bakgrunns-, holdnings- og personlighetsvariabler på side 50 er et eksempel på dette. Det samme gjelder på side 57 der en kan få inntrykk at grunnen til at vi innen samfunnsfag ikke kan operere med like entydige årsaksforklaringer som innenfor naturfagene, er av teknisk/metodisk karakter.

Prioriteringa av det samfunnsfaglige er prisverdig, men i noen tilfeller synes det som om dette har gått på bekostning av en skikkelig instruksjon i bruk av edb-verktøyet. Det forekommer bl.a. at en gir øvingsoppgaver uten i det hele tatt å gi henvisninger til hvor i heftet en kan finne informasjon om hvordan en skal gå fram for å få maskinen til å gjøre det den skal (se f.eks. s. 41). Jeg synes nok også at en skulle ha spandert noen sider til på vedlegg B: Hvordan bruke NSD-STAT? Den er til tider vel knapp for nybegynnere, f.eks. er velg-kommandoene dårlig forklart.

NSD-STAT fungerer stort sett greit og jeg har ikke ved en gjennom-

v12: ABORT: Syn på abortspørsmålet						
v21: BARNEHAGER: Bør forslaget om å bygge flere barnehager gjennomføres						
v12 -->	Abort al: Hvis Kv.: Sosiale : Selvbest:				SUM	
	dri tilla:	liv i fa:	kriterier:	ent abort::		
Meget godt forslag.	4	58	80	348 :	490	
Godt forslag	4	80	117	293 :	494	
Er lite opptatt	9	73	65	195 :	342	
Dårlig forslag	3	38	31	58 :	130	
Meget dårlig forsla	5	24	17	27 :	73	
SUM:	25	273	310	921 :	1529	
Inkludert:	1529	Utelatt:	67	Totalt:	1596	

v12: ABORT: Syn på abortspørsmålet						
v21: BARNEHAGER: Bør forslaget om å bygge flere barnehager gjennomføres						
v12 -->	Abort aldri: Hvis Kv. li: Sosiale kri: Selvbestemt:				SUM	
	tillate:	v i fare:	terier:	abort.::		
Meget godt	16.0	21.2	25.8	37.8 :	32.0	
Godt forsl	16.0	29.3	37.7	31.8 :	32.3	
Er lite op	36.0	26.7	21.0	21.2 :	22.4	
Dårlig for	12.0	13.9	10.0	6.3 :	8.5	
Meget dårl	20.0	8.8	5.5	2.9 :	4.8	
SUM:	100.0	100.0	100.0	100.0 :	100.0	
N=	25	273	310	921 :	1529	
Inkludert:	1529	Utelatt:	67	Totalt:	1596	

Eksempler på tabeller som kan skrives ut av programmet NSD-STAT. Tabellen øverst viser antall svar og den nederste tabellen viser svarene i prosent.

gang av øvingsoppgavene kommet over noen store mangler i programmet. Noen småting er irriterende. Avbruddkommandoene når det gjelder omkodning, fikk ikke jeg til å fungere. Når det gjelder omkodning, kan en også få ubehagelige overraskelser dersom en ikke legger inn komma etter de kodene (selv om det bare gjelder én kode) som skal inkluderes i den nye omkodede kategorien. Dette står det ingenting om i instruksjonene. En uvesentlighet til slutt, det er unødvendig og tungvint at programmet ikke går direkte tilbake til tabellversjonen når en har hatt oppe diagram-versjonen.

De positive sider ved denne lærepakken oppveier likevel definitivt de få skjønnsfeil jeg har funnet. Det er derfor all grunn til å anbefale lærepakken på det varmeste. De som har etterlyst nye fagområder når det gjelder anvendelsen av edb i skolesammenheng, har fått et handterlig program. Lærepakken er også et godt eksempel på at det går an å lage lære-«pakker» som formidler fagkunnskaper på en forsvar-

lig faglig — og samtidig pedagogisk fornuftig — måte. Det er ingen dårlig prestasjon.

En utfordring til NSD til slutt: Det ville være en klar pedagogisk gevinst om programmet også gav muligheter for å legge inn egne data. Er det mulig for den neste versjonen av NSD-STAT?

Einar-Arne Drivenes er førsteamanuensis ved Institutt for samfunnsvitenskap, Universitetet i Tromsø.

CALICO '87

***MONTEREY,
CALIFORNIA 6.-10. APRIL 1987***

Jostein H. Hauge

CALICO-konferansen arrangeres av Computer Assisted Language Learning and Instructional Consortium, som er den sentrale amerikanske organisasjonen for datastøttet språkopplæring. Organisasjonen utgir også bladet CALICO Journal (\$45 for private mottakere, \$75 for institusjoner), som kan anbefales til alle som arbeider med datamaskinbasert teknologi i språkundervisning.

CALICO-konferansen arrangeres årlig, og den neste vil bli holdt ved Brigham Young University i Utah (se egen kunngjøring i bladet). Organisasjonen har for øvrig sin basis ved BYU.

I tillegg til den årlige konferansen arrangerer også CALICO sommerundervisning (CALICO Summer Institute). I år ble det holdt sommerkurs i følgende emner:

- CALL (Computer Assisted Language Learning) Materials Development
- Design of Interactive Audial Video for CALL
- CALL Programming/Authoring
- The Application of Artificial Intelligence to Language Teaching

Nærmere opplysninger om slike arrangementer fås fra: *CALICO*, 3078 JKHB, Brigham Young University, Provo, Utah 84602.

Ved henvendelse til samme adresse kan en også få opplysninger om de spesielle interessegrupper som et medlemskap i CALICO gir adgang til (bl.a. for interaktiv video).

Den fjerde CALICO-konferansen i april i år ble holdt ved Monterey Defence Language Institute, et stort og moderne språkinstitutt som i sin egen undervisning legger vekt på å utnytte ny teknologi i språkopplæring og som også har en rekke FoU-prosjekter i gang knyttet til datastøttet læring.

På konferansen, hvor det deltok flere hundre deltakere hovedsakelig fra USA, ble det gitt ca. 70 foredrag med stor tematisk bredde: Rene teknologiske foredrag om utviklingen av CD-ROM, interaktiv video m.v., bruk av forfatterspråk, spesialiserte programsystemer for språkopplæring på pc og presentasjon av en lang rekke interaktive språkprosjekter i en serie nasjonalspråk (hebraisk, koreansk, japansk, engelsk, spansk, tysk, fransk m.fl.).

I en rekke av foredragene ble pedagogiske grunnlagsproblemer tatt opp. Man forsøkte bl.a. å analysere de muligheter som de nye elektroniske medier gir for en adekvat språkopplæring og hvilke deler av den mer tradisjonelle lærerbaserte/klassebaserte språkopplæringen som bør tas med inn i den interaktive læresituasjonen. Flere talere hevdet — og prøvde å vise — at de nye interaktive media som CD-ROM og interaktiv video gir helt nye muligheter for studentene til selv å styre språkopplæringen, legge inn øvingssekvenser etter eget behov og skaffe seg bakgrunnsinformasjon på de punkter hvor han/hun ønsker det. De nye mediene gir først og fremst mye bedre muligheter til å presentere autentiske språksituasjoner. Studentene ser og hører samtidig, og kan dessuten på sin pc se en tekstversjon av f.eks. en samtale, eventuelt på flere språk. Dessuten kan man repetere sekvenser, få langsommere uttale av ønskede deler og eventuelt en utskrift av en dialog til nærmere studium senere.

Men ny undervisningsteknologi koster. Et interaktivt videoprojekt som vil presentere bruk av språket gjennom en samtalesituasjon, krever opptak, redigering og produksjon av videoplate. Språkinholdet i



Carolyn J. Kuhn fra Facts On File Publications demonstrerer The Visual Dictionary på CD-ROM.

talt form må gjerne skrives ut, studeres i detalj for å finne frem til trekk som språkoppleringen skal fokusere på, og så må en ved hjelp av forfatterspråk bygge opp egnet læremateriell. En av foredragsholderne hadde erfart at det tok to uker å lage to minutter øving i et interaktivt språkoppleringsprogram. Dessuten tok det en uke å tilrettelegge læringssegmentet i pc-form.

I flere foredrag kom man inn på det enorme potensiale som ligger i en kobling av pc-utstyr, CD-ROM og interaktiv video. På en CD-ROM kan en eksempelvis lagre ca. 170.000 A-4 sider tekst som i pedagogisk sammenheng kan brukes til å gi visuelt instruksjonsmateriale og til å gi tekstlig bakgrunnsinformasjon for det språkstoffet som presenteres på videoplaten. Eventuelt kan hele språkoppleringen bygges opp på materiale lagret på CD-ROM.

Illustrerende i så henseende var demonstrasjonen av «The Visual Dictionary CD-ROM», hvor en med utgangspunkt i ordboken med samme navn har laget en CD-ROM som inneholder informasjon om en del av de objekter som blir beskrevet i papirutgaven. Det presenteres også ordboksdefinisjoner, tegninger i ulik forstørrelse og uttale av alle viktige termer på både fransk og engelsk. (Nærmere opplysninger om dette tiltaket kan gis av forfatteren).

Selv om teknologien gir enorme muligheter til å simulere reelle språksituasjoner, er det tydelig at mange av de miljøer som har kastet seg inn i utviklingsarbeidet, ennå er på begynnerstadiet. I flere av foredragene — som på forhånd så imponerende ut — kokte det hele ned til en presentasjon av videodelen av prosjektet, gjerne en videoplate sammensatt av tilrettelagte segmenter fra tv-utsendelser. Her kunne det spenne fra Fidel Castros taler til reklameinnslag og såpeoperaer. I en rekke av prosjektene som ble presentert, ble for øvrig materiale fra The BBC Foreign Language Video Disc Project benyttet.

På konferansen var det derfor mange som snakket om «Vapourware» — den programvare som mange sier de har eller vil utvikle, men som ingen har sett til nå.

I pedagogisk sammenheng er ofte et forfatterverktøy på pc grunnlaget for det pedagogiske utviklingsarbeidet. Flere foredrag tok derfor opp anvendelsen av slike forfatterspråk (f.eks. Pilot, Quest, Maestro) eller gav eksempler på hvordan man utifra spesielt utarbeidete evalueringsskjema kan skille mellom ulike typer og kvaliteter av forfatterspråk.

Et av de største språkopplæringsprosjektene i USA i dag er det såkalte ATHENA Language Learning Project ved MIT, Boston. Prosjektet startet i 1985, har store forskningsbevilgninger og har gjennom et samarbeid med DEC og IBM til disposisjon maskinvarer for mer enn \$50 millioner. Målsettingen er å bygge opp et fremtidsrettet språklaboratorium med MIT («The language lab of the future») i form av et senter som inneholder et arsenal av elektroniske audiovisuelle hjelpemidler og AI-baserte programvarer. Senteret skal kunne brukes av alle språkstudentene ved MIT ved hjelp av et nettverk som skal kunne koble på flere tusen kraftige personlige datamaskiner. Utviklingsarbeidet baseres på UNIX operativsystem. En basis for prosjektets satsing på avanserte arbeidsstasjoner til studentbruk er forventningen om at slike maskiner vil falle drastisk i pris. I sitt foredrag opplyste prosjektlederen *Janet Murray* at maskiner som i dag koster \$15.000 i 1990 vil bli solgt til studenter for \$1300 + \$500 i serviceavtale. ATHENA-prosjektet er et prestisjeprosjekt ved MIT og et prosjekt som ifølge konferansedeltakerne svært mange i USA ser på med stor spenning — og en viss skepsis. Prosjektet skal gå over tre år og oppgir pr. i dag å ha laget prototyper til sentrale deler av språkopplæringsystemet. Det var derfor mange som gikk skuffet fra foredraget etter i hovedsak å ha sett videoplater som er laget til bruk i prosjektet. Har vi her også foreløpig å gjøre med «Vapourware», var det mange som spurte seg.

CALICO '87 var mitt første møte med organisasjonens konferanser. Opplegget var meget profesjonelt og konferansen ble ledet med

militær (!) presisjon. En rekke utstillinger av programvarer og teknisk utstyr bidro også til utbyttet. Og ikke minst: Amerikanernes faglige engasjement er imponerende — og vedvarende — både i og utenfor konferanselokalene.

COMPUTERS AND TEACHING IN THE HUMANITIES

SOUTHAMPTON, 10.-11. APRIL 1987

Kristin Natvig

En konferanse om edb og undervisning i de humanistiske fag ble arrangert i Southampton i april av Office for Humanities Communication, University of Leicester og Department of Electronics and Computer Science ved University of Southampton. 100 personer, hovedsakelig fra institusjoner for høyere utdanning i Storbritannia og Europa deltok på konferansen.

Konferansen var en av de første av sitt slag i Storbritannia. Arrangørene ønsket å samle både de som allerede har introdusert edb i undervisningen og de som har planer om det. Arrangementet bestod av en blanding av plenumssesjoner om emner av generell interesse, spesielle sesjoner om enkelte felt og metoder, og spesialsesjoner for deltakere som ikke hadde forhåndskunnskaper i edb. Ettersom storparten av konferansen foregikk i form av 3-4 parallelle sesjoner, var det umulig å få med seg mer enn et lite utvalg av alle foredragene.

Dette forholdet var frustrerende både for undertegnede og andre deltakere. Et annet irriterende moment var at mange av foredragsholderne ikke holdt seg til oppgitt emne og/eller ga svært rotete framstillinger. I det følgende gis det derfor kun referat av de mest interessante foredragene, delvis basert på skriftlig materiale som ble utdelt.

INNLEDNING

Richard Ennals, Kingston College of Further Education holdt innledningsforedraget. Ennals konsentrerte seg om forholdet mellom edb og humaniora — han interesserer seg bl.a. for det han betegnet som «*humanity in computing*.» Særpreget til de humanistiske fagene kan være et nyttig bidrag til edb og videre til samfunnet. Etter hans oppfatning er det et stort gap mellom humaniora og edb som kan minskes via kunstig intelligens.

Edb-teknologi kan virke frigjørende på humaniora idet den kan gjøre videre utforskning av fagene mulig. Noen mener at bruk av edb strukturerer de humanistiske fagene i for høy grad og resulterer i innsnevring. Men ved å finne fram til regler for bruk av edb innenfor humaniora, er det mulig å studere mer inngående komplekse betydningsstrukturer. Dessverre konkretiserte ikke Ennals sine påstander, og de ble heller ikke relatert til pedagogiske spørsmål.

SPRÅK OG LITTERATUR

Bruk av konkordanser

Philip King, University of Birmingham, presenterte MICROCONC, en konkordanspakke utviklet for undervisning i teknisk engelsk. Formålet med pakken er å gjøre det mulig for studentene å studere betydningsforskjeller ved at de blir presentert med ord i ulike kontekster. Pakken inneholder bl.a. en øvelse i å fylle ut «*key word in context*». Programvaren omfatter fasiliteter for både satsvis og interaktiv kjøring for valg av eksempler. Det arbeides også med automatisk seleksjon. Det fins en opsjon for joker-notasjon og konteksten kan alfabetiseres på begge sider. Kontekstmengden er variabel — studentene kan oppgi så mange karakterer som de ønsker.

Korpuset pakken utgår fra, består av fire foredrag om transportteknologi og seks akademiske artikler om bl.a. plantegenetikk. Senere skal fagområdene utvides til å omfatte administrative emner og humaniora. Hver tekstfil er kodet for å angi bl.a. muntlig og skriftlig framstilling og fagområde.

Denne presentasjonen vakte en livlig debatt. King ble kraftig kritisert fordi tekstsamlingen er altfor liten til å være representativ for («god») engelsk språkbruk. Etter fleres mening burde han ha basert konkordanspakken på et av de store engelske tekstkorpora. Andre

mente at tekstgrunnlaget er stort nok til å utgjøre en eksempelsamling for teknisk engelsk.

Tyskfaget

Gordon Burgess, University of Aberdeen presenterte et edb-kurs for tysk-studenter som ble opprettet i 1981. Formålet med kurset er først og fremst å gjøre studentene oppmerksomme på mulighetene og begrensningene ved edb-bruk i litterær og lingvistisk analyse av fremmedspråklige tekster. Kurslærerne ønsker dessuten å provosere studentene til å stille spørsmålsteget til etablerte tilnæringsmåter til analysen av litterære tekster. I tillegg er lærerne seg bevisst det voksende behovet for edb-kunnskaper hos universitetskandidater, og selv om kurset ikke er spesifikt innrettet mot arbeidslivet, håper de at det vil være til hjelp for studentene i deres videre karrierer.

Kurset består av fire deler: 1. Generell introduksjon — databehandlingens historie (spesielt i forhold til humaniora), edb-kommunikasjon, grunnleggende filhåndtering og redigeringsteknikker. To ukers undervisning, obligatorisk frammøte. 2. Bruk av databaser, spesielt lagring og gjenfinning av bibliografiske data. 3. Programmeringsteori og -praksis med lingvistisk og litterær analyse som formål. Programmeringsspråket som brukes, er SNOBOL, som er spesielt utviklet for behandling av strenger. Etter sigende læres SNOBOL raskt p.g.a. en gjennomsliktig struktur og enkel behandling av variabler. Det forlanges ikke at studentene skriver sine egne programmer — de får presentert et stort antall eksempelprogrammer og rutiner og blir bedt om å revidere/legge til disse. Programmene blir deretter testet på tekstmateriale som språklig spenner fra middelhøytysk til moderne tysk. 4. Tekstredigering i teori og praksis, illustrert ved arbeid knyttet til generering og anvendelser av konkordanser (Oxford Concordance Program), inkludert statistisk og stilistisk analyse. Her lærer studentene også om behandling av egenavn, fotnoter osv. I løpet av hele kurset blir studentene oppfordret til å trekke slutninger av både pragmatisk og teoretisk art basert på egne erfaringer.

Undervisningen, som varer i to semestre, foregår i form av forelesninger og arbeid i små grupper. I stedet for en semesteroppgave må studentene utføre praktiske øvelser. Eksamen består av en skriftlig oppgave om de ulike aspektene av edb i humaniora.

Kurset har vist seg å være enormt populært og ifølge Burgess har studentene bare godt å si om undervisningsopplegget. Erfaringer viser at studentene blir stimulert til å reflektere over emner som lingvistisk

teori og kommunikasjonsproblemer som oppstår i spenningsfeltet naturlig språk/kunstig språk. Studentene blir seg bevisst behovet for og fordelene ved kommunikativ presisjon. Dessuten letter kurset tilegnelsen av lingvistisk teori.

Diverse

Andre konferansedag ble tre presentasjoner knyttet til språkopplæring klemt inn i en sesjon på 1 times varighet. *Graham Davies*, National Centre for Computer-Assisted Language Learning (NCCALL) i London snakket om fundamentale spørsmål med hensyn til den rollen edb-metoder bør ha i språkopplæring. Davies uttrykte bekymring over denne rollen i framtida — etter hans mening prøver mange språklærere å få datamaskiner til å utføre umulige og irrelevante oppgaver slik som talesyntese og -analyse. Andre lærere er blitt forført av mulighetene som kunstig intelligens innebærer, og bruker altfor mye tid og penger på å utvikle ekspertsystemer, som etter hans mening setter språkundervisningen 20 år tilbake. (Davies fikk dessverre ikke tid til å underbygge disse påstandene).

Etter Davies' mening blir mange forfattersystemer for høyere utdanning med rette kritisert for å være kjedelige. Programvare utviklet for undervisning ved videregående skoler er ofte av høyere kvalitet. Et eksempel er «Grandville» (stavemåten er usikker) — et avansert simuleringsprogram. Ved bruk av dette programmet «ankommer» elevene Frankrike, hvor de skal klare seg i fem dager på et lavt budsjett og bare ved å bruke fransk. Ifølge Davies bør dette programmet sette et eksempel for lignende programmer. Raffinerte simuleringsprogrammer av denne typen kan gjerne utvikles for undervisning ved institusjoner for høyere utdanning.

Monique L'Huillier, Brunel University, demonstrerte CONCORDE, et forfattersystem for trening av verbbygging basert på morfologisk analyse av bøyingsfeil. Programmet ble utviklet for å håndtere franske verb, men kan tilpasses andre språk. Det ble understreket at dette er et hjelpemiddel til utforming av *treningsprogram*, ikke *undervisningsprogram*. Programmet gir læreren maksimal assistanse i programutvikling, men på bekostning av fleksibilitet i øvelsesutforming og testmuligheter. CONCORDE tillater læreren å legge inn egne data, og søker å gi studentene maksimal interaksjon.

En demonstrasjon ble gitt av et program laget ved bruk av systemet. Dette programmet tilbyr automatisk analyse av feil når det gjelder stammer og endelser, eller hjelpeverb + perfektum partisipp. Det gis

også indikasjon på hvorvidt det uriktige svar er for langt eller kort, og når det gjelder sammensatte verb — hvorvidt ett av leddene mangler eller hvis mellomrommet mellom hjelpeverb og perfektumform mangler. Studentene kan teste sine kunnskaper enten i en bestemt rekkefølge eller ved fritt valg, hoppe over spørsmål, forlate programmet før øvelsen er slutt, og velge mellom kommentarer til feil på enten engelsk eller fransk. På sin side kan læreren legge inn kommentarer til svarene.

CONCORDE er rettet inn mot artium-nivå eller høyere, og er kun tenkt som et upretensiøst supplement til vanlig undervisning. Programmet er nok lett å bruke for både lærere og elever/studenter, men virket rett og slett ... kjedelig. Flere av deltakerne på denne sesjonen uttrykte også skepsis vedr. programmets pedagogiske fundament.

Til slutt i denne sesjonen ble det gitt en presentasjon av et program utviklet ved Polytechnic of North London som trener studenter i tekstkonstruksjon. Programmet heter «Text Jumbler» og fins i forskjellige versjoner for fransk, tysk og spansk. Demonstrasjonen ble gitt på fem minutter så det var vanskelig å få et skikkelig inntrykk av programmets formål og funksjoner.

Litteratur

Thomas N. Corns, University College of North Wales, og *Margaret E. Smith*, Further Education Unit, London innledet sesjonen om edb i litteraturundervisning med en gjennomgang av problemene som oppstår når edb-anvendelser skal tilpasses litteraturstudier. Vanskelighetene ligger bl.a. i forsøk på å få datamaskinen til å fungere på et semantisk nivå, modellering av dialogmodus og fagområdets kontroversielle natur. Det fins få vellykede edb-assisterte litteraturprosjekter i Storbritannia, ifølge foredragsholderne. Stilistisk analyse — den disiplinen som lettest kan foretas ved bruk av edb-metoder — er ikke «på mote» for tida. Ifølge Corns er det også et problem at tilhengerne av både marxistisk og feministisk teori — som begge har en dominerende stilling i Storbritannia — er i opposisjon til automatisering. (!)

Det fins imidlertid en del strakstiltak som kan gjennomføres. I stedet for semesteroppgaver o.l. kan studentene lage konkordanser basert på korte tekster med tilhørende egenproduserte innledninger. Statistiske programpakker kan tas i bruk, f.eks. for å lage statistikk om et teaters aktiviteter. Databaser kan med fordel produseres, f.eks. for oversikter over sekundærlitteratur. Tekstbehandling er et egnet redskap for å hjelpe studentene med strukturelle og andre typer forbedringer av skriftlige framstillinger.

Noen tanker om den videre utviklingen på feltet ble presentert til slutt. Den største hindringen i overskuelig framtid vil nok være manglende finansiering. Framover bør det satses på mer avanserte programmer enn de som er nevnt over, deriblant programmer for undervisning i gammelengelsk. Slike programmer vil gjøre det litt mer spennende for studentene å lære grammatikken — som jo er en forutsetning for å kunne studere pensumlitteratur på gammelengelsk.

HISTORIE

Innenfor historie-sesjonen ble det presentert et meget imponerende prosjekt ved University of Glasgow. «The DISH Project» (Design and Implementation of Software in History) ble startet i 1985, har 10 medarbeidere og ledes av *Richard Trainor*. DISH er et samarbeidsprosjekt mellom bl.a. universitetets tre historiske institutter, instituttet for informatikk, og universitetets arkiv og edb-tjeneste. Prosjektet har sitt eget undervisningskompleks med et tre-roms edb-«laboratorium» som tilbyr 16 mikromaskiner med oppkobling til universitetets stormaskin. Prosjektets viktigste formål hittil har vært å introdusere edb-basert undervisning til alle historiestudenter på cand.mag.-nivå. Så langt er edb brukt i undervisningen av tre ulike emner for viderekomne studenter. Flere edb-baserte kurs på dette nivået er under planlegging, likeledes for studenter i deres første og andre studieår.

Prosjektmedarbeiderne erkjenner at det er vanskelig å introdusere studentene til store historiske databaser p.g.a. materialets omfang og behovet for å mestre kompliserte operativsystemer. En av prosjektets målsettinger er å utvikle mer brukervennlig programvare som er basert på vinduer, ikoner (skjerm-symboler) og mus. Formålet er å gjøre kildemateriale med en kompleks struktur like lett tilgjengelig som det trykte ord.

Programvareutviklingen i regi av DISH er rettet mot dekning av behov som blir dårlig ivaretatt av eksisterende programvare. Når alle de planlagte programmene er utviklet, skal de integreres i en pakke som skal styres av et system for databasedrift. Programmene som hittil er utviklet, er beregnet for dataregistrering, filhåndtering/statistikk, søkbare litteratur/pensumlister, og enkel lenking/behandling av informasjon i ulike typer databaser.

Et parallelt formål med prosjektet er å gi historiestudentene tilstrekkelige kunnskaper og erfaring til å kunne bli aktivt med i tilrettelegging, behandling, administrasjon og gjønfinning av data. Studentene skal



UNIVERSITY OF GLASGOW

The DISH Project at Glasgow

Design and
Implementation of
Software in
History

kunne bruke programvare utviklet innenfor prosjektet til videre dyktiggjøring innenfor historiefaget, samtidig som de skaffer seg grunnleggende edb-kunnskaper.

Undervisningsopplegget består av fire faser: 1. enkle eksempeløvelser i fellesskap, 2. egne edb-opplegg som utføres med lærerassistanse,

3. integrasjon av output med egenprodusert tekst, 4. videregående anvendelser knyttet til eksamensoppgaver. I fasene 3 og 4 skal studentene ha anledning til å bruke små egen-genererte databaser.

DISH-prosjektet går også inn for å fremme historieforskningen generelt. Adgangen til databaser både ved University of Glasgow og andre universiteter er gjort lettere tilgjengelig, og innenfor prosjektet er det opprettet fire nye historiske databaser. Hovedfagsstudenter har nå fått adgang til edb-utstyr/programvare og får støtte av edb-medarbeidere slik at de kan opprette sine egne databaser og bruke disse effektivt.

En annen av DISHs oppgaver er å videreformidle undervisningskompetansen og programvaren som er utviklet innenfor prosjektet. Kontakt er derfor etablert med skoler hvor edb-aktiviteter foregår. Et sommerkurs for historielærere ble arrangert 6.-10. juli, hvor teoretiske og metodiske spørsmål ble diskutert og hvor deltakerne hadde anledning til å opprette prøvedatabaser over medbrakt kildemateriale.

Samarbeid pågår også med andre humanistiske institutter ved britiske universiteter hvor edb-basert forskning og undervisning fremmes. I DISHs regi undersøkes dessuten hvorvidt edb-kunnskapene som er hverves i løpet av historiestudiet gjør det lettere for kandidatene å få innpass på arbeidsmarkedet.

KLASSISKE STUDIER

Det kan nevnes kort at *Kevin O'Connell*, University of Exeter presenterte et enkelt program til hjelp i studiet av klassisk gresk litteratur. Programmet hjelper studentene til å holde rede på persongalleri, stedsnavn, årstall, begivenheter m.m. ved bruk av databaser. Formålet med programmet er å lette tilegnelsen av faktiske opplysninger i litterære verk slik at studentene kan konsentrere seg om spørsmål vedr. tema, dramatisk struktur osv. Ikke noe revolusjonerende program akkurat, men en original edb-anvendelse ifølge flere tilhørere.

AVSLUTNING

En meget celeber gjesteforeleser var invitert til å holde konferansens avslutningsforedrag — prof. *J.-C. Gardin*, Ecole des Hautes Etudes, Paris. Gardin skulle snakke om «The Future Influence of Computing in the Interplay between Research and Teaching», men kom aldri så langt. Starten var lovende, til tross for hans uttalelse om at konferansens aktiviteter bare hadde bestått av trivialiteter — en vurdering som naturlig nok virket svært fornærmende på tilhørerne.

Gardins utgangspunkt var det han påstod var en allmenn oppfatning — nemlig at kraftige edb-redskaper og økt lingvistisk ekspertise betyr at problemene ang. *representasjon* var løst. Gardin var ikke enig i dette synet. Han har selv arbeidet i 30 år med edb-metoder innen arkeologi, og finner det fremdeles vanskelig å beskrive gjenstander med tanke på registrering i en funksjonell database. Ifølge Gardin er mange edb-anvendelser i humaniora forsøk på å bruke matematikk for å skape orden av kaos.

Det er umulig å referere resten av foredraget. Gardin fortapte seg i vitenskapsteoretiske modeller med utstrakt bruk av tavla til oppstilling av matematiske formler som illustrasjon. Ikke bare undertegnede, men mange andre tilhørere mistet tråden fullstendig.

OPPSUMMERING

Selv om Gardins foredrag må karakteriseres som totalt mislykket i denne konferansens sammenheng, var en dose med teoretisk analyse for så vidt på sin plass. Teori var stort sett mangelvarer i de fleste foredragene undertegnede hørte — spesielt *pedagogisk* teori, et ankepunkt uttalt av flere deltakere. Mange av foredragsholderne bar preg av at de først og fremst var *forskere*, ikke pedagoger. En del var også tydelig preget av «edb-er-gøy»-holdningen.

Det var likevel en svært interessant konferanse — deltakerne fikk et godt inntrykk av mangfoldet innenfor edb-assistert undervisning i humaniora i Storbritannia, og ble selvsagt tilbudt anledning til å knytte nyttige faglige kontakter. Nivået innenfor feltet er kanskje ikke så høyt som en kanskje kunne tro i et geografisk område med såpass mange universiteter og institusjoner for høyere utdanning, men dette kan jo skyldes det faktum at humanister i Storbritannia er enda sterkere rammet av økonomiske nedskjæringer enn i Norge. Til tross for manglende penger og dårlige arbeidsforhold viste foreleserne imponerende pågangsmot og optimisme.

Det arrangementstekniske knirket litt på enkelte måter — mest irriterende var de trange og kokvarme grupperommene, og et altfor hektisk program. Demonstrasjonene gikk til dels i full fart, og det var ofte for liten tid til spørsmålsstilling og debatt. Konferansen kunne med fordel ha vart én dag til, slik at det ikke var nødvendig å holde så mange parallelle sesjoner samtidig.

Deltakerne fikk for øvrig utdelt ei utmerket bok med spesialskevne artikler i tilknytning til konferansens tema: Sebastian Rahtz (ed.): *Information Technology in the Humanities: tools, techniques and applications*. Ellis Horwood, Chichester 1987. ISBN 0-7458-0148-X. Pris: £25.

OPTICA '87

Øystein Reigem

Hittil har konferanser om optiske lagringsmedier vært mangelvare i Europa, men i år er det faktisk avholdt og planlagt avholdt en rekke slike arrangementer. Et av disse var konferansen Learned Information arrangerte i Amsterdam 14.-16. april 1987 — «Optica '87», med undertittelen «The International Meeting for Optical Publishing and Storage». Optica '87 var stort lagt opp, men gjennomførelsen var dessverre ikke så profesjonell som andre konferanser med samme emne undertegnede har deltatt på tidligere. Likevel må det sies at den var et kjærkomment tilskudd til konferansetilbudet i Europa, og vi ser den nok igjen neste år (men for å unngå forveksling med en etablert konferanse om optikk, vil senere konferanser ikke ha «Optica» i navnet).

Det følgende er ikke et strengt referat fra Optica '87, men en blanding av referat, nyheter annensteds fra og egne kommentarer. Kun CD-ROM er omtalt, og det skrevne kan ses som et supplement til CD-ROM-artikkelen i HD 2-86.

CD-ROM er i dag tatt i bruk i en rekke anvendelser, og det er en raskt voksende interesse for mediet til elektronisk publisering av databaser, leksikon og programvare, til lagring av kart, lyd og bilder, og til bruk i underholdning og undervisning. I USA regnes nå produksjon av CD-ROM som seriøs virksomhet, og i det siste er Europa kommet etter. I slutten av -86 var det mer enn 70 kommersielt tilgjengelige titler på CD-ROM. I tillegg er det produsert en mengde testplater og plater til intern bruk i bedrifter og institusjoner.

LEVENDE DIGITAL VIDEO, CD-I

På CD-ROM og andre digitale optiske plater kan man lagre bl.a. stillbilder, grafikk og bevegelig grafikk, men levende «naturlige» bilder, dvs. digital video, har lenge vært problematisk. Et digitalisert video stillbilde utgjør i seg selv en stor datamengde, og levende video inneholder 25 (Europa) eller 30 (USA) bilder («ruter» eller «frames») i sekundet. Selv når hver rute komprimeres med avanserte teknikker er digital video svært plasskrevende. Løsningen ser ut til å være å ta

i bruk teknikker som komprimerer ved å sammenligne påfølgende ruter. Mer om dette siden.

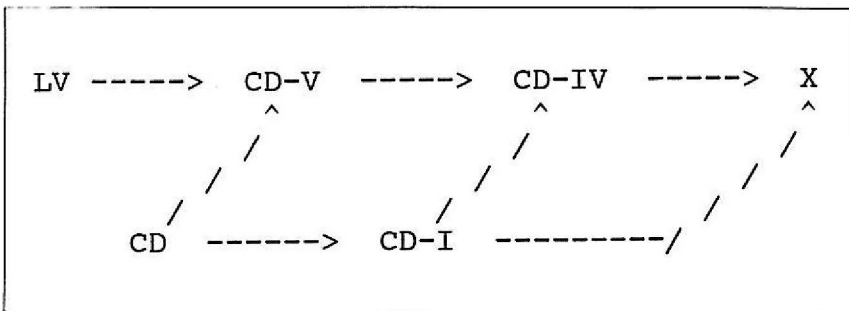
CD-ROM vil sannsynligvis ha to forholdsvis adskilte markeder, og utstyr tilpasset disse. I «profesjonelle» anvendelser (næringsliv og institusjonsbruk) står CD-ROM-spillere som periferutstyr for datamaskin, og begrepet CD-ROM vil stort sett være knyttet til slik bruk. For underholdning/hjemmemarkedet vil det komme integrert utstyr som kan kobles opp mot TV og stereoanlegg, og navnet på dette konseptet er *CD-I — CD-Interactive*.

CD-I, som ble presentert som en foreløpig standard i 1985, har vakt meget stor interesse. Ikke alle er glade for dette. De mener at all publisiteten omkring CD-I har forsinket det «profesjonelle» CD-ROM-markedet med ett år, fordi mange har ventet på CD-I istedenfor å ta i bruk CD-ROM. Mange mener også at det ikke er kommet godt nok fram at CD-I først og fremst er et forbrukerprodukt.

En må nok regne med en viss tid før CD-I har etablert seg. Det må eksistere et visst tilbud av plater før en får et gjennomslag på forbrukermarkedet, og «fullverdige» CD-I-produksjoner vil være ressurskrevende. Samarbeid om utgivelser er imidlertid i gang, og en venter de første «forfatterverktøy» (hjelpemiddel for skriving av interaktiv multimedia programvare) i slutten av året.

CD-I-spilleren vil ha sin egen innebygde datamaskin — en 16-bits 8 MHz prosessor — Motorola 68070, og et eget operativsystem, det OS9-baserte CD-RTOS. Utstyret vil ventelig være tilgjengelig i år, men ikke på det kommersielle marked før tidligst neste år.

CD-I vil kunne ha levende video, men i første utgave begrenset til en fjerdedel av skjermen, og med bare 12-15 bilder i sekundet. Senere versjoner vil kunne få fullskjerm digital video i full hastighet. På «Optica '87» skisserte en representant for Philips dette firmaets planer for utviklingen fra dagens CD og LaserVision til en fremtidig digital plate med alle typer multimediainformasjon:



CD-V (CD-Video) vil bli introdusert som en slags video-singel, i CD-størrelse med 20 minutter digital CD-lyd og 5 minutter levende video i analog form. Det vil også komme 8" og 12" plater («EP-er» og «LP-er»), og CD-V kan dermed også sees på som en etterfølger til LaserVision videoplater. LaserVision har allerede lenge hatt muligheter for digital informasjon, og for amerikanske plater (NTSC) finnes det en standard for digital lyd. To umiddelbare ulemper med CD-V er at platene ikke vil kunne spilles på vanlige CD-spillere, og at en pga. den analoge videoen vil få et skille mellom amerikanske og europeiske plater. Til gjengjeld kan en regne med at det kommer spillere for både CD, LaserVision og CD-V.

Fra Storbritannia meldes det nå at Philips og Polygram planlegger å sende et begrenset antall CD-V-spillere ut på markedet ved juletider. Prisen vil bli omkring GBP 500. Dette framstøtet skal teste markedet for en større lansering i 1988-89. Ifølge kilden (New Musical Express 2.5.87) vil det være tilgjengelig ca. 500 platetitler med emner som sport, musikk og spillefilm. (Det er da kanskje ikke urimelig å tro at mange av titlene er allerede utgitte Laservision videoplater, og at CD-V-spillerne kan spille disse?)

CD vil bli mer direkte videreført i CD-I, som altså vil være fullstendig digital og bare mangle (fullskjerm, full hastighet) levende video i forhold til den fremtidige «ønskeplaten». CD-I-spillere skal også kunne spille CD, men ikke plater med analog video (LV og CD-V).

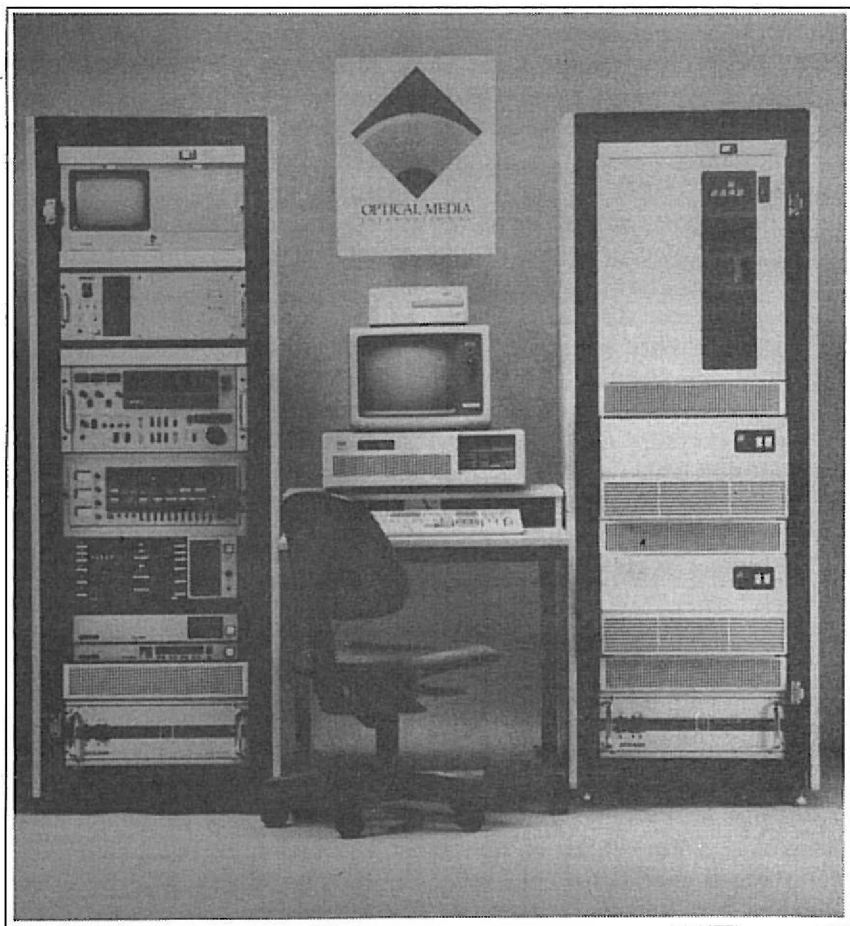
Ved å «krysse» CD-V og CD-I skal en så også få en CD-I-plate med levende video (CD-IV), men da i analog form. En tenker seg her muligheten for utstyr som kan spille alle plater. Etter CD-IV gjenstår bare ett steg, nemlig å erstatte analog levende video med digital.

Nå trenger ikke digital levende video på CD ligge så langt inn i fremtiden som dette skulle tyde på. Den store sensasjonen på Microsofts CD-ROM-konferanse i mars var General Electric/RCA's demonstrasjon av DVI (Digital Video Interactive). DVI tillater opptil én time digital video på CD-ROM. Dette er mulig gjort av nye kraftige komprimeringsteknikker som baserer seg på å lagre forskjeller mellom bilder (så kapasiteten kan dessverre ikke omregnes i stillbilder). Dekomprimeringen foretas av et sett VLSI-brikker som ikke trenger ta mer enn to kortplasser i en PC (under demonstrasjonen på konferansen stakk riktignok kortene ut av et hull sagt i dekslet på maskinen). DVI krever ingen modifikasjoner i selve CD-ROM-spilleren. CD-ROM med DVI kan bli en farlig konkurrent til CD-I, og også til videoplater.

DIVERSE CD-ROM-NYHETER

Maksimumstørrelsen for filer på PC er ikke noe problem etter at Microsoft i mars annonserte sin MS-DOS Extension for CD-ROM. En kan nå ha DOS-filer på over 32 Mbyte på CD-ROM.

Det finnes nå utstyr for anvendelser der man trenger tilgang til flere plater samtidig. Eksempelvis leverer Online Computer Systems kort som kan styre åtte CD-ROM-spillere på én gang. Opptil fire slike



CD-ROM «premastering» system fra Optical Media International. Systemet tar data og lager ferdige bånd for produksjon av master.

kontrollkort kan settes i en mikromaskin, og maskinen kan dermed ha hele 32 CD-ROM-spillere tilkoblet på én gang.

Prisene på CD-ROM-spillere har sunket det siste året. En spiller med interfacekort for PC koster i Norge i dag ca. 10.000 kroner + moms. (Philips oppgir en pris på 10.595 for sin modell CM100, Hitachi ca. 10.000 for sin 1503S, med mulighet for forsknings/universitetsrabatt. 1503S har også audio-utgang.)

Mesteparten av aktiviteten rundt CD-ROM foregår i USA, men Europa kommer etter. En spesielt interessant nykommer er firmaet EIKON — et samarbeid mellom Microsoft og italienske Olivetti og SEAT. EIKON skal bl.a. utvikle programvare for det europeiske markedet. EIKON arrangerer også en CD-ROM-konferanse i høst — «The Business of CD-ROM» i Roma 27.-29. oktober i år.

Et nytt og meget interessant CD-ROM-produkt er «the Microsoft Bookshelf», selv om det er rettet mot det engelskspråklige marked, og spesielt USA. Bookshelf består av ti kjente referanseverk, som kan nås direkte fra tekstbehandling. Brukeren får dermed muligheten til å finne informasjon og eventuelt kopiere denne rett inn i sitt eget dokument. Bookshelf inneholder bl.a. orddefinisjoner og synonymer, navne- og stedsopplysninger, sitater, stavekontroll, postkoder, osv.

Eksempler på europeiske produksjoner er «Who supplies what» («Wer liefert was») — en oversikt over leverandører av en rekke produkter), databasen «Mikropharm II» og bibelen, alle fra tyske Bertelsmann. Norges første CD-ROM bør selvfølgelig også nevnes. Norsk Regnesentral har gjennom britiske Archetype Systems Ltd. fått laget en plate med statistiske data, kart og tekst.

Selv om CD-ROM er både robust og lagringsholdbar i forhold til magnetiske datamedier, finnes det anvendelser der den vil komme til kort. Derfor vil firmaet Digipress lansere sin «Space-CD» basert på metall eller glass istedenfor plast. Space-CD skal bl.a. tåle fuktighet og ekstreme temperaturer og temperatursvingninger. Den vil kunne brukes i romfart og til militære formål. Med en garantert levetid ved lagring på 100 år, vil den også være godt egnet som arkivmedium. Dagens plater garanteres foreløpig kun i 10 år.

«*THE USE OF COMPUTERS IN THE TEACHING OF LANGUAGE AND LANGUAGES*»

RAPPORT FRA EIN KONFERANSE VED UNIVERSITY OF HULL 22.-23. APRIL

Per Vestbøstad

Konferansen var organisert av «The School of Modern Languages and Cultures» ved University of Hull, og var støtta av «The Computers in Teaching Initiative Support Service».

Målet med å senda ein representant frå NAVFs edb-senter til denne konferansen var å kunna rapportera tilbake til interesserte miljø heime om det som finst av interessant programvare på dette feltet i Storbritannia.

DELTAKARAR, PROGRAM OG LOKALITETAR

Konferansen samla kring 130 deltakarar, dei fleste forskarar og pedagogar, og alle var britar utanom ein italiendar, ein franskmann og eg.

Foredraga var samla i tre plenums-sesjonar og seks sesjonar med 4 parallelle presentasjonar. Ein av desse var ein øvingssesjon med følgjande emne: «Basic Concepts», «Working with Authoring Packages» og «Introduction to Concordances».

Universitetet hadde relativt gode forelesingslokale, så eg hadde ingen vanskar med å få tak i det som vart presentert. Annleis var det med ei utstilling av programvare og tilknytta datasystem, som var samla på eit altfor lite område og berre hadde fått tildelt 70 min. i tidsskjemaet. Det var såleis vanskeleg å få eit grundig inntrykk av meir enn eit par produkt.

PLENUMSFØREDRAGA

Gabriel Jacobs frå Swansea (deltakarlista oppgjev diverre ikkje tittel

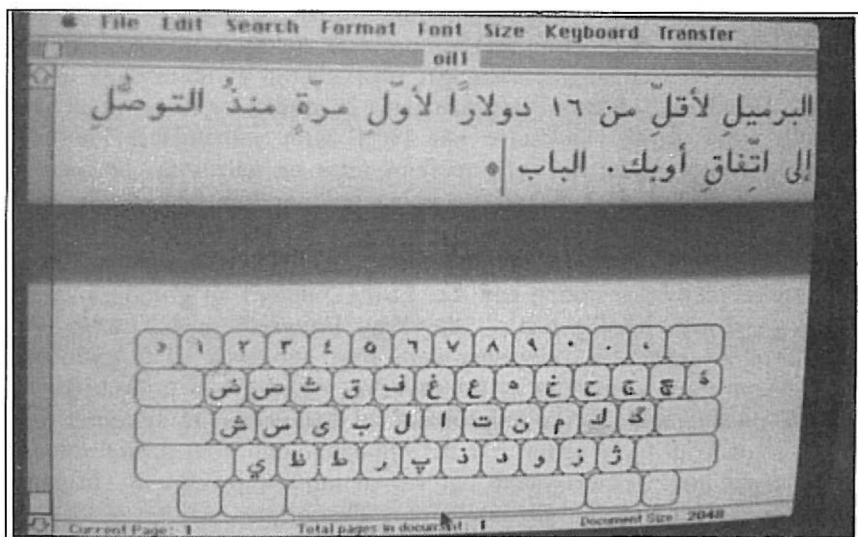
eller institusjon) tok i første plenumsesjonen «Multichoice questions in CALL» opp pedagogiske metodespørsmål ved bruk av CALL (Computer Assisted Language Learning). Han meinte at feilrettinga er ein sterkt hemmande faktor i innlæring av 2. språk, og viste til at feilretting skjer mykje sjeldnare når barn lærer morsmålet. Fleirvals-spørsmål, der eleven får velja mellom rette og gale svar, og der ein oppmodar eleven til å finna/retta eigne feil når han svarar gale, gjev høgare motivering. Sjølv CALL-programmet bør også vera feiltolerant, slik at eleven ikkje går av sporet på enkle og banale feil.

I andre plenumsesjonen gav *Rex Last* (Dundee) eit grundig oversyn over «Artificial Intelligence — the Way Forward for CALL?». Han definerte AI slik: Eit programsystem som har evne til å læra, dvs. modifisera eigen kunnskap ut frå røynsler, og som kan ta kunnskapen i bruk på nye felt. Eit hovudproblem vil her vera å få systemet til å finna attende til trygg grunn når det oppdagar feil. Han peikte elles på at arbeidet med ekspertsystem har vist at kunnskap ikkje er ein generell prosess i menneske sin intelligens. Ekspertane tenkjer nemleg på ulike måtar om ulike emne, slik at ikkje berre fakta, men også slutningsreglane må tilpassast emnet.

Sjølv om desse problema kan gjera oss motlaus, så er gevinstante med delvise løysingar så store at det bør vera råd å laga nyttige program også på CALL-feltet. Rex Last har sjølv laga noko han kalla eit «Inexpert System», som lærer studentane visse tyske setningsmønster.

Medan Rex Last la vekt på det problematiske ved å få eit datasytem til å handtera lingvistiske kunnskapar, gav *Christopher Butler* (Nottingham) ei nyttig orientering om eksisterande programvare for språklege formål. Utanom programvare for tekstbehandling og tilrettelegging av tekstar for databehandling delte Butler den språkleg interessante programvaren i tre grupper:

1. For statistiske formål vil mange statistikkprogram kunna brukast. SPSS er stor og dekkjer det meste, medan Minitab er enklare for studentane. Ei lærebok på høgt nivå er Woods: «Statistics for Linguists».
2. For analyseformål finst t.d. Oxford Concordance Program, som er stort og ressurskrevjande, medan WordCruncher (frå Brigham Young University) er hendigare på mikromaskiner. Clock-programmet (skrive av Alan Pee i Birmingham) er særleg velegna til å studera kollokasjonar (ord-sekvensar).
3. Innom den tredje gruppa, simulering av menneskeleg språkleg aktivitet, la han vekt på ulike parsing-teknikkar (lingvistisk analyse av tekst, der ein siktar mot å skjøna innhaldet). Her fekk det gamle



Frå utstillinga på konferansen: Tekstbehandlingsprogram for arabisk på Macintosh. Programmet er utvikla av Arabic Software Ass. Inc., 240 East Center St., Provo, Utah, USA.

programmeringsspråket Snobol heiderleg omtale som undervisningsspråk for studentar. Parsing-teknikkar er viktige for at CALL-program kan verta meir feiltolerante og på lengre sikt skjøna naturleg språk. Butler meinte at etter kvart som maskinkrafta aukar, vil parsing verta vanleg i CALL-program. Maskinlagra ordbøker og annan bakgrunnskunnskap («real world knowledge») vil også vera naudsynt i mange samanhengar.

Av konkret programvare nemnde han Pigins Grammarland (vokabulartrening på Commodore 64) og Faringtons omsetjingsprogram (engelsk-fransk på BBC-maskina).

Butler kom også inn på programvare for datamaskinstøtta skriveopplæring, der den klassiske Writer's Workbench er innhenta av Epistle og Critique, som begge nyttar parsing.

Noel Williams (Sheffield) gjekk grundigare inn på programvare for skriveopplæring på bakgrunn av eit større prosjekt ved Sheffield City Polytechnic, der dei har evaluert denne typen programvare generelt og Writer's Workbench spesielt. På bakgrunn av dei manglane dei fann i WW, vil dei no utvikla ny programvare til bruk i England. Når det gjeld grundigare omtale av slik programvare, viser eg til artikkelen om «Forfattarens høvelbenk» i HD 1-87.

PRESENTASJONAR FRÅ DEI PÅRALLELLE SESJONANE

Dei fleste foredraga her rapporterte frå konkret prosjektarbeid og forsøksverksemd i framandspråkopplæringa ved universitet og høgskular. Franskfaget dominerte, men også tysk, spansk, italiensk, latin, arabisk og fonetikk var representerte. Ikkje alt var like interessant, og ved eit par høve måtte eg spørja meg sjølv om prosjektet først og fremst var sett i gang for di det var lettare å få midlar til arbeid der ein brukte datamaskiner.

Fleire av deltakarane rapporterte røynsler frå bruken av datamaskinsentralar for studentar og tilsette. *W.S. Dodd* frå School of Modern Language i Exeter la fram statistikk som synte at brorparten av bruken gjekk på teksthandtering, medan resten (30%) delte seg på databasebruk, tekstanalyse og CALL. *Gordon Burgess* frå Aberdeen la vekt på dei mange positive biverknadane ved å gje studentane tilgang til datamaskiner: Det verkar motiverande ved at det er annleis og utfordrande, det fremjar samarbeid og lag-innsats, det oppsedar til streng logikk, og studentane får betre sans for fleirtydige ord og konstruksjonar. Begge desse stadane hadde studentane tilgang til datamaskinene 24 timar i døgret.

Konferansen gav eit breitt inntrykk av aktivitetane på CALL-feltet i England, og eg knytte fleire interessante kontaktar med tanke på det arbeidet som vert gjort på feltet her heime.

ICAME 8th

Knut Hofland

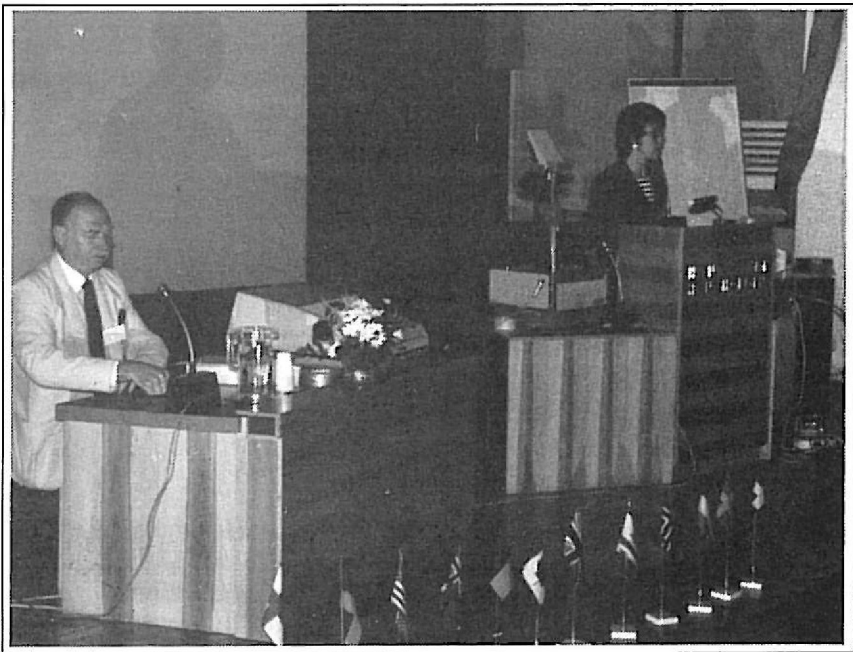
Den åttende ICAME konferansen ble holdt på Hanaholmen utenfor Helsinki 21.-24. mai 1987 med over 70 deltakere fra 12 land.

Som ved tidligere ICAME konferanser var det en blanding av rapporter fra miljøer som driver med korpuslingvistik og foredrag der en hadde gjort studier basert på ett eller flere tekstkorpora.

I Nijmegen har en gående analyseprosjekter for engelsk, arabisk

og spansk. Korpusstørrelsene varierer fra 0.5-1 mill. ord og tekstene er i hovedsak fra 1980-årene. Tekstene analyseres ved hjelp av formelle grammatikker som bygges opp parallelt og lagres i det lingvistiske databasesystemet LDB. I tillegg til analyseprosjektene har en prosjektet LDB II. Dette er dels en overføring av LDB til MS-DOS og dels en tilpassing av systemet til bruk innen datamaskinstøttet læring (setningsanalyse). Det pågår også arbeid innen formelle grammatikker for korpuslingvistikk og hvorledes disse er relatert til modeller innen teoretisk lingvistikk. LDB er tilgjengelig for forskere for et nominelt beløp (utgaver for IBM VM/CMS, VAX VMS/Unix og MS-DOS). Kontakt: *TOSCA Work Group, Dept. of English, University of Nijmegen, P.O. Box 9103, NL-6500 HD Nijmegen, Nederland (EARN/BITNET: COR_HVH@HNYKUN52).* (Jan Aarts, Nelleke Oostdijk).

I Amsterdam har en avsluttet prosjektet ASCOT. Det er bygget opp et datamaskinelt leksikon (ASLEX) basert på Longmans Dictionary of Contemporary English (LDOCE). Det er dessuten startet et nytt prosjekt, LINKS, som går ut på å bygge opp en leksikalsk kunnskaps-



Sesjonsleder Nils Erik Enqvist og Antoinette Renouf.

base basert på LDOCE. Denne skal brukes i semantisk analyse. (*Eric Akkerman, Willem Meijs, Piek Vossen*).

Også i Lund er en i avslutningsfasen for et prosjekt som har pågått i flere år, TESS, «Text Segmentation for Speech». Det ble demonstrert et program for PC/AT som startet med skrevet tekst (eksempel fra Brown Corpus). Programmet fant frasestrukturen og ut ifra et sett med regler, segmenterte det setningen i tone-enheter. Dette ble bl.a. brukt til å generere pauser i forbindelse med bruk av Voxtrax talemaskin. Det ble også demonstrert et program for en grafisk representasjon av tone-enheter. Arbeidet siste året har konsentrert seg om regler for segmentering og uttesting av disse bl.a. med bruk av programvare i Turbo Prolog. Prolog er et kraftig og fleksibelt verkøy i systemer som kan beskrives med et sett regler. Men en er avhengig av regnekraft, på en PC tar behandlingen av en tekstprøve fra Brown korpus (2000 ord) 12 timer til parsing og 4 timer til segmentering. (*Jan Svartvik, Bengt Altenberg, Anna-Brita Stenström, Mats Eeg-Olofsson*).

I Birmingham er en nå ferdig med den store ordboken som ble finansiert og utgitt av forlaget Collins. Utviklingsgruppen innen prosjektet fortsetter på selvfinansierende basis. I tillegg til de ca. 20 millioner ord som ble bygget opp i ordboksprosjektet (hvorav 7 millioner ord som Collins har rettighetene til), arbeider en med oppbygging av to mindre tekstsamlinger. Det ene er en samling med NATO tekster (uklassifisert!, 3/4 skrevet og 1/4 tale), det andre er materiale for testing av engelskkunnskaper. En jobber også med anvendelsesorienterte grammatikkmodeller for bruk i behandling av større tekstsamlinger. Kjernen er her den leksikalske databasen som er bygget opp og der grammatiske opplysninger blir knyttet til den leksikalske enhet. (*Antoinette Renouf, Jeremy Clear*).

Taleprosessering var emne for foredragene fra Lancaster. Oppbyggingen av et talespråk-korpus som ble startet i 1984 har to formål, det ene er selve oppbyggingen av korpuset som er på 100.000 ord. Korpuset skal også være utgangspunkt for et program for automatisk intonasjonsmarkering. Omfanget av korpuset er nå redusert til 50.000 ord. Medarbeiderne har krav til materialet som god lyd kvalitet og man ønsker ikke pauser, avbrytelser eller uintelligent tale. Dette har gjort at det er brukt opptak fra BBC med en delvis forberedt og mer formell karakter enn vanlig dialog. Korpuset vil være tilgjengelig fra Lancaster høsten 1987. Det er også i gang et prosjekt innen talegjenkjenning. Dette er et samarbeid med IBMs forskningssentra i New York og Winchester. I IBMs talegjenkjenningssystem brukes det en enkel lingvistisk modell basert på statistikk over trigram, bigram og

unigram (3, 2 og 1 ord). Det ble vist med konkrete feileksempel hvorledes en forbedret lingvistisk modell som innebar bruk av ordklasseopp-
slag, syntaks og semantikk, kan minske feilprosenten, som i dag er på
2-3 prosent. (*Lita Taylor, Geoffrey Leech*).

Ved universitetet i Sydney er det i gang arbeid med et australsk
tekstkorpus. En vil ta utgangspunkt i tekster fra 1986, men ellers føl-
ge oppbyggingen av Brown og LOB korpus for sammenligningens
skyld. Parallelt vil det bli bygget opp en større tekstsamling med også
andre tekstkategorier enn de 15 som er i LOB/Brown, bl.a. elektronis-
ke medier. Det ble presentert noen resultater fra en pilotstudie basert
på avismateriale. (*Peter Collins*)

Årets konferanse understreket hvorledes PC'er i økende grad også
blir tatt i bruk i korpusarbeid. *Randall Jones* fra Brigham Young
University presenterte bruk av WordCruncher (tidligere Brigham Young
Concordance Program) på Brown Corpus. *Raymond Hickey*, universite-
tet i Bonn, hadde laget en programpakke for bruk i leksikologisk
arbeid, basert på dBase/Clipper. En PC-utgave av LDB ble annonsert
fra Nijmegen, og i Lund og Helsinki gjør en utstrakt bruk av PC'er
i prosjektarbeidet. Oxford Concordance Program (OCP) er også nå
tilgjengelig for PC.

Det ble gitt en oversikt over arbeidet med å bygge opp et program-
bibliotek i tilknytning til ICAME. Til distribusjon av program og
informasjoner vedrørende ICAME har en nå tatt i bruk elektronisk
post. Filtjeneren ved EARN-noden i Bergen ble vist via oppkalling fra
en EARN-node i Helsinki. I en diskusjon om programbiblioteket kom
vanskene frem med å gi fra seg programmer. En konklusjon var at
en innen ICAME bør få frem informasjoner om utstyr og program
som er i bruk og også en presentasjon/diskusjon av de metoder som
ligger til grunn for programmene. Dette vil være utgangspunkt for en
direkte kontakt mellom programsjåper og potensielle brukere. (*Knut
Hofland*).

Som ved to tidligere konferanser vil foredragene bli utgitt på Rodopi
forlag, Amsterdam. I tillegg vil det bli gitt resymé av foredragene
i neste nummer av ICAME Journal.

Neste konferanse vil bli holdt i Birmingham i tiden 18.-21. mai
1987. For informasjon kontakt *Antoinette Renouf, English Language
Research, 357 Bristol Rd., Edgbaston, Birmingham B5 7SW (Janet:
RENOUF@BIRMINGHAM)*.

LITTERATUR

ICAME Journal (tidligere ICAME News, utgitt av NAVFs edb-senter for humanistisk
forskning)

Aarts, J. & Willem Meijs (eds.), 1984, 1986: *Corpus Linguistics I+II*, Amsterdam: Rodopi.
Meijs, Willem (ed.), 1986: *Corpus Linguistics and beyond*, Amsterdam: Rodopi.

XIV ALLC CONFERENCE

Espen S. Ore

Konferansen ble holdt i Göteborg 1.-5. juni. Det var 65 påmeldte deltagere fra 15 land, de fleste fra Europa. Foredragene spente over hele spekteret av språklig og tekstlig databehandling, selv om det var en viss konsentrasjon om tekstsamlinger og leksikalske databaser. Et poeng som flere kom inn på i forskjellige kontekster, var datautveksling mellom ulike kodeformater og behovet for standardiserte utvekslingskoder.

Nicoletta Calzolari fra Università di Pisa var invitert til å holde konferansens innledningsforedrag. Hun tok opp leksikalske databaser, spesielt slike som var beregnet til maskinoversettelse etc. Hun viste til at deres struktur og informasjonsinnhold ville avspeile de grammatiske og språkvitenskapelige modellene som lå under programmene der de skulle brukes. Dette kunne i verste fall føre til at de samme språkdata måtte registreres flere ganger: én gang for hver modell som brukes. En slik dobbeltregistrering betyr sløsing med tid og penger, og for å unngå dette foreslo hun at leksikalske databaser bør bygges opp slik at dataene kan brukes innen de fleste systemer som arbeider med naturlig språk.

Lignende tanker kom frem i et foredrag skrevet av *D.T. Barnard*, *C.A. Fraser* og *G.M. Logan*. Der ble det gjort rede for arbeidet med å utvikle merkingssystemer basert på SGML-rammen (SGML = Standard Generalized Markup Language, en standard utviklet innen ISO) Kodene som brukes her, skal både gi språklig informasjon (ordklasse e.l.) og typografisk informasjon (skrifttype etc). SGML er ikke i seg selv et kodesystem, men en standard for hvorledes slike systemer skal bygges opp.

I et foredrag av *M.A. Heather* og *B.N. Rossiter* ble det vist hvorledes eksisterende typografisk koding i en tekst kan brukes i oppbyggingen av et gjenfinningssystem. I den engelske *Osborn's Concise Law Dictionary* brukes skifte av skrifttyper (fet, kursiv) og forekomst av forskjellige spesialtegn til å angi informasjonselementer i en artikkel (lovreferanser, henvisninger til andre oppslagsord etc). Et magnetbånd fra setteriet har disse kodene, og det kan derfor brukes som rådata for en halvautomatisk innlesing av oppslagsverket i en strukturert database. (Det vil alltid gjenstå noen tvilstilfeller når det gjelder hvorledes spesielle skrifttyper skal tolkes, og her må en avgjørelse tas manuelt.)

Blant foredrag som tok opp andre temaer, kan et par nevnes: *J.C. Lejosne* fra Université de Metz viste hvorledes metoder fra datalingvistikk/maskinoversettelse kunne brukes til å tolke et manuskript som var blitt forvansket i flere ledd. (Originalmanuskriptet hadde ligget noen måneder i Mekong, deretter var det blitt skrevet av av tre forskjellige personer som ikke helt klarte å lese det. Teksten var på middel-hollandsk med utstrakt bruk av partisipper, og gjennom forvanskningen av materialet var det blitt vanskelig å bestemme hva de forskjellige partisippene refererte til.) *L. Borin* fra universitetet i Uppsala ga et eksempel på bruk av Benny Broddas BETA-system til å simulere lydoverganger fra proto-indoeuropeisk til kirkeslavisk.

UTVIKLINGSSEMINARET 1987

GEILO, 10.-12. JUNI

***Kristin Natvig, Øystein Reigem og
Per Vestbøstad***

Sammen med Holger Hagan, edb-konsulent ved Inst. for språk og litteratur, Universitetet i Tromsø arrangerte NAVFs edb-senter for humanistisk forskning det årlige utviklingsseminaret for viderekomne edb-medarbeidere på Geilo i juni. Årets seminar samlet vel 30 deltakere og omfattet tre tema som ble behandlet i delvis parallelle sesjoner: CD-ROM, digitalisert bildebehandling og desktop publishing.

CD-ROM

Onsdag etter lunsj sto CD-ROM på programmet. Forsker *Erling Maartmann-Moe* fra Norsk regnesentral gav først en orientering om Regnesentralen og plasserte CD-ROM-arbeidet, som han leder, innenfor det totale aktivitetsfeltet deres.

Han gav videre en oversikt over det som skiller CD-ROM fra andre datalagringsmedier og gikk grundig inn på den såkalte «High Sierra»-standarden for slike plater.

Størsteparten av sesjonen brukte Maartmann-Moe naturlig nok på temaet: «Gjennomgang av erfaringer en har gjort under produksjon av en CD med digitaliserte kart, statistiske data og tekst». Han gav bl.a. følgende råd til mulige CD-produsenter: Vent ikke til du har data som fyller hele plata (550 Mbyte). Lagre data mest mulig ferdig bearbeidet, gjerne i flere ulike sorteringer, slik at prosessida på mikro-maskinen blir minst mulig. Tenk plate-geografi slik at data som brukes sammen også lagres sammen. Les store blokker for å motvirke den relativt høye tilgangstida.

Maartmann-Moe gjorde også greie for programmeringsmessige og juridiske sider ved bruken av tekstsøkesystem i «Norsk CD-ROM nr. 1», som Regnesentralen har laget i samarbeid med Cappelens forlag og Statistisk sentralbyrå. Den logiske organiseringen av stoffet er også viktig.

Produksjonen av denne CD-plata tok 10 mnd. og kostet 150.000 kr i direkte utgifter, men Maartmann-Moe regnet med at totalkostnaden inklusive databearbeiding var om lag 650.000 kr.

LAGRING OG GJENFINNING AV DIGITALISERTE BILDER OG DOKUMENTER PÅ EN PC AT — INNLEDNING

Denne innledningen besto av et foredrag og en demonstrasjon. Foredraget ble holdt av *Eric Swane* fra SINTEF, Trondheim, og demonstrasjonen av *Otto Milvang*, Bildebehandlingslaboratoriet, Institutt for Informatikk, UiO.

Swane startet sitt foredrag med å klargjøre en del begreper før han gikk videre med en oversikt over bildebehandlingsfeltet. Han gav også en rask oversikt over norske bildebehandlingsmiljøer, og nevnte kort NOBIM (NOorsk forening for Bildebehandling og Mønster-gjenkjenning).

Swane brukte så mesteparten av sin tid til å forklare om digitalisering av bilder, dokumentprosessering, bildeforbedring og komprimering,

samt redegjøre for diverse utstyr — CCD-kameraer og scannere, grafikk- og digitaliseringskort, og skrivere. Under demonstrasjonen etterpå viste Milvang digitalisering av opptak fra et videokamera vha. et digitaliseringskort for video.

LAGRING OG GJENFINNING AV DIGITALISERTE BILDER OG DOKUMENTER PÅ EN PC AT — INNLEDNING

Denne fordypningsseksjonen fikk et løsere preg. Det ble holdt innlegg av tre av fire «ressurspersoner», og så fikk det hele utvikle seg fritt. Det ble stilt spørsmål fra de andre deltakerne, og det ble eksperimentert noe med medbrakt utstyr og programvare. De fire ressurspersonene var *Nils Jacob Berland* fra Christian Michelsens Institutt, Bergen, *Ole-Hjalmar Kristensen* fra RUNIT, Trondheim, samt de tidligere nevnte *Otto Milvang* og *Eric Swane*.

Berland fortalte om CMIs virksomhet med scanning/OCR av dokumentmateriale/tekniske tegninger. Han fortalte om relevante teknikker i den forbindelse, og kom inn på egnede datastrukturer for bildebehandling. Til sist tok han for seg programmeringsspråk, nærmere bestemt C og Modula-2.

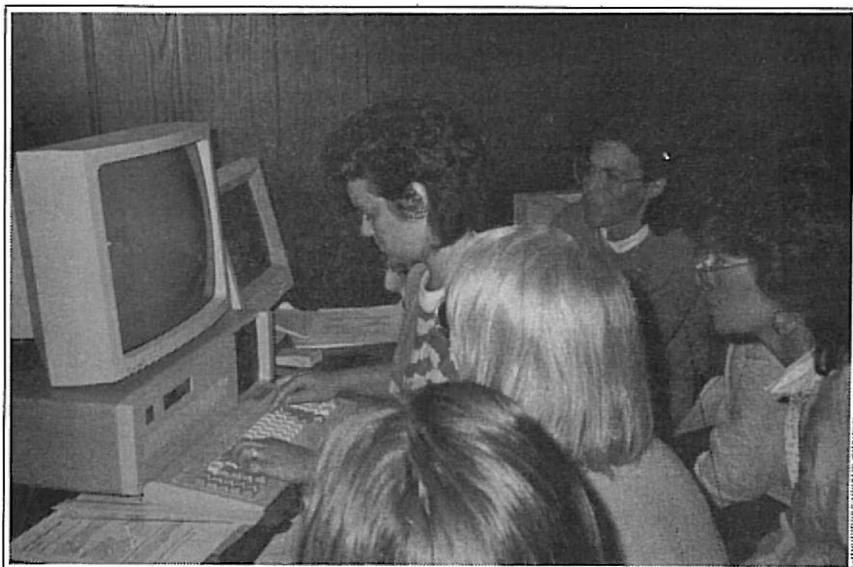
Kristensen konsentrerte seg om et system for lagring av bilder på skrivbare digitale optiske plater utviklet ved Institutt for databehandling ved UiT — NTH. Han demonstrerte også en «browser» beregnet for søking og navigering i databaser med grafisk informasjon.

Milvang fortalte om Bildebehandlingslaboratoriet (som har NAVF-støtte t.o.m. 1987), og en del prosjekter som er utført der, bl.a. et prosjekt om restaurering av en samling tilsynelatende ødelagte fotos.

DESKTOP PUBLISHING

Per H. Jacobsen, Universitetets sentrale edb-tjeneste, Universitetet i Oslo holdt et foredrag med tittelen «Desktop Publishing for Macintosh.» Foredraget var så å si identisk med artikkelen av Jacobsen i dette nummeret av HD. Jacobsen demonstrerte lay-out-programmet Pagemaker for deltakerne, som fikk anledning til selv å prøve seg fram.

En annen innledning om desktop publishing ble holdt av *Lars Ericsson* fra firmaet CINETs Bergens-avdeling. Ericsson la spesiell vekt på lay-out-programmet Ventura. Ventura kan kjøres på IBM-kompatible maskiner og bygger bru mellom sideorienterte programmer og bokombrekingsprogrammer. *Sidsel Røthe*, også fra CINET, demonstrerte



Sidsel Røthe fra CINETs Bergens-avdeling demonstrerer lay-out-programmet Ventura for noen av seminardeltakerne.

Ventura på både vanlig skjerm og A4-skjerm. Seminardeltakerne fikk øve seg i bruken av programmet på medbrakte tekstfiler.

MELDINGER

SLUTTRAPPORT OM SENTERETS VIDEOPLATEPROSJEKT

Sluttrapporten om Senterets videoplateprosjekt (se HD 1-87) er nå utgitt som Senterrapport nr. 40: *Videoplateteknologi i humanistiske fag*. Rapporten er skrevet av de prosjektansvarlige: Roger Erlandsen (Sekretariatet for fotoregistrering, Oslo), Claus Huitfeldt og Øystein Reigem (begge Senteret).

Rapporten inneholder kapitler om bakgrunnen for prosjektet, målsetting, forberedende arbeid, administrasjon og organisering, tilrettelegging av bildemateriale og referansedata, programvare, økonomi, vurderinger av de ulike sidene ved prosjektet, råd til andre som vil produsere videoplater, og arbeid i forlengelse av prosjektet (i alt 31 sider). De åtte vedleggene består av bl.a. referater fra møter og det avsluttende seminaret med de deltakende institusjonene, samt innholdsfortegnelse til plata.

Rapporten koster kr 55,- og kan bestilles fra: *NAVFs edb-senter for humanistisk forskning, Boks 53 - Universitetet, 5027 Bergen.*

HUMANISTISKA FAKULTETENS DATASERVICE, GÖTEBORG

Hum fak:s dataservice ved Göteborgs universitet ble opprettet i 1985 som et service- og utviklingsorgan for forskning og forskerutdanning

innenfor fakultetets arbeidsfelt. Tjenesten samarbeider med Institutionen för språkvetenskaplig databehandling (Språkdata).

Formålet med virksomheten er å være hjelpelig med rådgivning og utdanning innenfor databehandling og å drive tjenestens minimaskin (Data General Eclipse MV/10000). De fire medarbeiderne gir også informasjon om tilgjengelig programvare innen humaniora og holder seg å jour med utviklingen på feltet.

80 terminaler kan være tilknyttet til minimaskinen samtidig. Via et datanettverk er det mulig å kommunisere med både andre svenske og utenlandske maskiner. Tjenesten har også to IBM PC'er, hvorav den ene lånes ut for kortere tidsrom. Tre terminalrom er tilgjengelig for fakultetets brukere.

På minimaskinen fins de fleste vanlige programmeringsspråkene, i tillegg til programvare for kontorautomasjon, presentasjonsgrafikk, statistikk, databasehandtering og lingvistikk.

Flere opplysninger kan fås fra: *Hum fak:s dataservice, Humanisthuset, Renströmsparken, 412 98 Göteborg, Sverige.*

THE MUSEUM DOCUMENTATION ASSOCIATION

The Museum Documentation Association (MDA) tilbyr omfattende konsulenttjenester til museer som ønsker å revidere sine rutiner for dokumentasjon og automatisering. Organisasjonen har utviklet retningslinjer for rådgivning av denne art og anvendt disse ved oppdrag for ulike typer/størrelser museer i Storbritannia. Assistanse gis i forbindelse med systemspesifikasjon, valg av programvare og utstyr, systeminstallasjon, driftsoppstart, opplæring, utvikling av programvare og registrering/konvertering av data.

MDA har utviklet en programpakke — Museum Object Data Entry System (MODES) — for dokumentasjonsarbeid og administrasjon av samlinger. Systemet er basert på MDAs egen dokumentasjonsstandard. MODES har fasiliteter for strukturering av poster, valideringsregler og utskriftstyper som kan tilpasses det enkelte museums behov.

MDA har også utviklet andre typer edb-systemer for museumsfaglige formål. I tillegg gir organisasjonen ut bøker og et meldingsblad, *MDA Information* (omtalt i HD 1-87).

Flere opplysninger fra: *The Museum Documentation Association, Building 0, 347 Cherry Hinton Road, Cambridge CB1 4DH, United Kingdom.*

EDB-BRUK I BRITISK ARKEOLOGI

Ei arbeidsgruppe om edb-bruk nedsatt av Institute of Field Archaeology/Royal Commission on Historical Monuments of England (IFA/RCHME) har utgitt en rapport med tittelen *IFA Occasional Paper Number 1*. Rapporten inneholder en omfattende katalog over utstyr og programvare i bruk innenfor britisk arkeologi, arbeidsgruppas anbefalinger og en fullstendig bibliografi over publikasjoner på engelsk om edb-anvendelser i arkeologi. Den inkluderer også et stort antall artikler om ulike anvendelsesområder. Rapporten koster £3 og kan bestilles fra: *Institute of Field Archaeology, Minerals Engineering Building, University of Birmingham, PO Box 363, Birmingham B15 2TT, England.*

Utgivelsen av denne rapporten markerer lanseringen av *The IFA Computer Advisory Service*. Tjenestens formål er å gi råd angående utvikling og implementering av edb-systemer og om valg av utstyr og programvare. Styringsgruppa skal være ansvarlig for å etablere brukergrupper og å drive og oppdatere databasen om edb-bruk opprettet av RCHME.

1987 DIRECTORY OF COMPUTER ASSISTED RESEARCH IN MUSICOLOGY

1987-utgaven av dette oppslagsverket inneholder informasjon om bl.a. musikkrepresentasjon, konferanser, avhandlinger, pågående teknologisk forskning (inkl. optisk lesing), anvendelser, kurs, tjenester, en flerspråklig begrepsordliste (engelsk, fransk, tysk og italiensk) og adresselister.

Verket er på 151 sider, har 54 illustrasjoner og er redigert av Walter Hewlett og Eleanor Selfidge-Field. Pris: \$10 + \$5 porto (luftpost). Adresse: *Center for Computer Assisted Research in the Humanities, 525 Middlefield Road, Suite 120, Menlo Park, California 94025, USA.*

AKTUELLE TIDSSKRIFTER

THE COMPUTERS AND PHILOSOPHY NEWSLETTER

The Computers and Philosophy Newsletter utgis av Center for Design of Educational Computing, Carnegie Mellon University, USA. Bladet består for det meste av faglige artikler, men inneholder også korte presentasjoner av ny programvare og meldinger av diverse typer. I 1987 skal det utgis tre nummer, til en abonnementspris på \$9. Adresse: *Leslie Burkholder, Center for Design of Educational Computing, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890, USA.*

MUSLETTER

IBM-Los Angeles Scientific Center og the USC School of Music har startet utgivelse av meldingsbladet *MUSLETTER* hvert kvartal. Tema for bladet er utviklingen innenfor edb-assistert musikkundervisning. Bladet skal fungere som et forum for utveksling av informasjon, erfaringer, problemer og forskningsresultater. Hovedredaktøren tar gjerne imot bidrag fra leserne i form av spørsmål, kommentarer og artikler. To andre redaktører er tilgjengelige for å hjelpe leserne med å løse tekniske problemer.

Selv om *MUSLETTER* kalles et meldingsblad, er det hovedsaklig artikler som fyller sidene. Bladet kan fås gratis ved henvendelse til: *Linda B. Sorisio, Editor, IBM-Los Angeles Scientific Center, 11601 Wilshire Blvd., 4th floor, Los Angeles, California 90025, USA.*

THE CTISS FILE

Dette tidsskriftet utgis 3-4 ganger pr. år av The Computers in Teaching Initiative Support Service i Storbritannia. Hvert nummer inneholder en mengde artikler og rapporter om bruk av edb på universitetsnivå — generelle teoretiske og praktiske problemer, programvare,

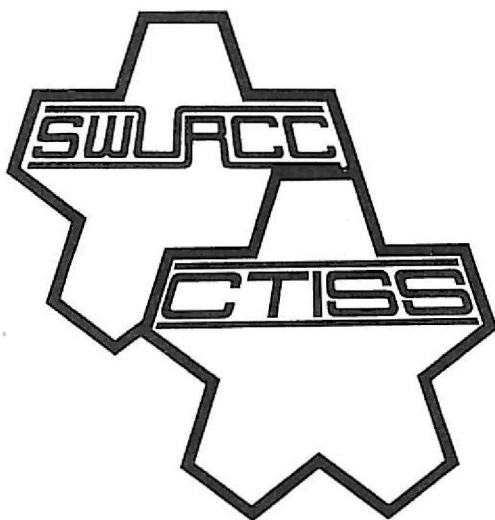
prosjekter og kurs for studenter. Det gis også informasjon om tidskrifter, bøker, organisasjoner, kommende kurs og konferanser m.m. Mye av stoffet omhandler/er relevant for humaniora.

The CTISS File distribueres gratis og fås fra: *Computers in Teaching Initiative Support Service, South West Universities Regional Computer Centre, University of Bath, Claverton Down, Bath BA2 7AY, England.*

THE CTISS FILE

Number 2

November 1986



Computers in Teaching
Initiative Support Service

NYTT I BIBLIOTEKET

EDB-EMNER

- Albregtsen, Fritz (red.): En oversikt over norske fagmiljøer innen bildebehandling og mønstergjenkjenning. Oslo — UiO, 1987. 96 s.
- Chellappa, Rama/Sawchuk, Alexander A.: Digital image processing and analysis. Volume 1: Digital image processing. Los Angeles, CA: IEEE Computer Society Press, 1985. 721 s.
- Lambert, Steve/Ropiequet, Suzanne (eds.): CD-ROM: the new papyrus: the current and future state of the art. Redmond, Washington: Microsoft, 1986. 619 s.
- Norton, Peter: Inside the IBM PC. New York: Prentice Hall, 1986. 387 s.
- Norton, Peter/Socha, John: Peter Norton's assembly language book for the IBM PC. New York: Prentice Hall, 1986. 413 s.
- Ropiequet, Suzanne/Einberger, John/Zoellick, Bill (eds.): CD-ROM: optical publishing: a practical approach to developing CD-ROM applications. Redmond, Washington: Microsoft, 1987. 358 s.
- Schwerin, Julie B. et. al.: CD-ROM standards: The book. Oxford: Learned Information 1986. flere pag.
- Tanenbaum, Andrew S.: Operating systems: design and implementation. London: Prentice-Hall International, 1987. 719 s.

KONFERANSERAPPORTER

- Moberg, Thomas F. (ed.): Databases in the humanities and social sciences 3, 1985. Osprey, Florida: Paradigm Press, 1987. 533 s.
- Raben, Joseph/Sugita, Shigeharu/Kobo, Masatoshi: Toward a computer ethnology. Osaka: National Museum of Ethnology, 1987. 179 s.

SPRÅK

- Association for Computational Linguistics. European Chapter. (utg.): Conference of the European chapter of the Association for Computational Linguistics. 2. Geneva 1985. ACL, 1985. 276 s.
- Butler, C.S.: Computers in linguistics. Oxford: Basil Blackwell, 1985. 266 s.
- Humanistisk-samhällsvetenskapliga forskningsrådet (utg.): Språk kommunikation och teknologi. Rapport från en arbetsgrupp inom HSRF. HSRF, 1987. 94 s.

Meijs, Willem (red.): Corpus linguistics and beyond. International conference on English language research on computerized corpora. 7. (ICAME 7th) Amsterdam 1986. Amsterdam: Rodopi, 1987. 302 s.

KONFERANSER

CALICO '88

CALICOs femte årlige internasjonale symposium skal holdes i Salt Lake City, Utah, 23.-27. februar 1988. Denne gangen blir det anledning til å delta i såkalte «Special Interest Groups» om kunstig intelligens, CD-ROM, edb i klasserommet, utvikling av undervisningsprogrammer, fonter for fremmedspråklige alfabeter/asiatiske språk, og interaktiv video/audio.

Deltakeravgiften er \$75 for medlemmer av CALICO, \$90 for ikke-medlemmer (25.-27. februar). «Workshops» (23.-24. februar) koster \$95 pr. tre-timers sesjon.

Flere opplysninger fås fra: *CALICO, 3078 JKHB, BYU, Provo, Utah 84602, USA.*

LINGUISTIC APPROACHES TO ARTIFICIAL INTELLIGENCE

I perioden 23.-26. mars 1988 skal Linguistic Agency ved Universitat Duisburg (LAUD) i Vest-Tyskland arrangere dette internasjonale symposiet. Halvparten av programmet skal vies til foredrag gitt av to eksperter pa feltet. Den andre halvparten skal besta av en-times foredrag (inkl. diskusjoner) knyttet til sprakororientert kunstig intelligens. Emnemessig skal symposiet fokusere pa problemer som kan loses ved et tverrfaglig samarbeid mellom forskere i lingvistikk, psykolingvistikk, kognitiv vitenskap og kunstig intelligens (f.eks. grammatisk teori, grammatikkbeskrivelse, parsing, sprakgenerering, naturlig spraksystemer, kunnskapsrepresentasjon, situasjonssemantikk, teori om diskursrepresentasjon).

Flere opplysninger fas fra: *Prof. Dr. R. Dirven, Universitat-Gesamthochschule-Duisburg, Postfach 10 16 29, D-4100 Duisburg 1, Vest-Tyskland.*

COMPUTERS AND MUSIC RESEARCH CONFERENCE

Denne konferansen skal finne sted 11.-14. april 1988 ved University of Lancaster, England. Konferansen er rettet mot de som bruker, eller har planer om å bruke, edb i musikkforskning. Arrangementet skal fungere som et forum for utveksling av ideer, metoder og erfaringer heller enn for presentasjon av forskningsresultater. Arrangørene håper at konferansen skal dekke alle feltene innen musikkforskning (f. eks. analyse, teori, pedagogikk, redskaper for komposisjon og syntese). En av konferansens sesjoner skal vies til demonstrasjoner av program- og maskinvare. Konferanseavgift pluss opphold kommer sannsynligvis på ca. £75.

De som ønsker å holde et foredrag, bes sende et sammendrag (250 ord) til nedenstående adresse innen 2. november.

Flere opplysninger fås fra: *Allan Marsden, Centre for Research into the Applications of Computers to Music, Department of Music, University of Lancaster, Bailrigg, Lancaster LA1 4YW, England.*

WOMEN, WORK AND COMPUTERIZATION

En internasjonal konferanse om dette temaet skal holdes i Amsterdam 27.-29. april 1988. Dette er den tredje konferansen av denne art som blir organisert av International Federation of Information Processing (IFIP). Formålet med konferansen er å: utvikle strategier som gjør kvinner i stand til å utøve større kontroll over anvendelsen av ny teknologi i deres arbeid; legge fram forslag til tiltak som kan forbedre kvinners stilling i arbeidsmarkedet innenfor rammen av introduksjonen av ny teknologi; utforske den framtidige utviklingen innen feltet arbeid og datamaskiner og mulighetene disse kan gi kvinner, også i den tredje verden. Et annet formål med konferansen er å etablere grunnlaget for en permanent arbeidsgruppe innenfor IFIP.

Konferansen skal bestå av to deler: én plenumsdag og en to-dagers «workshop». På plenumsdagen skal eksisterende kunnskaper om konferansens tema samt analyser av den framtidige utviklingen legges fram av inviterte foredragsholdere fra både Nederland og andre land. «Workshop»en er beregnet på forskere og andre eksperter.

Adgang til plenumsdagen skal koste ca. 50 Dfls. Deltakelse på «Workshop»en kommer på ca. 500 Dfls. (inkl. kost og losji i tre døgn).

Flere opplysninger kan fås fra: *IFIP Conference «Women, Work and Computerization», SIC, Paulus Potterstraat 40, 1071 DB Amsterdam, The Netherlands.*

ALLC — AIBI JOINT CONFERENCES

The Association for Literary and Linguistic Computing (ALLC) og Association Internationale Bible et Informatique (AIBI) skal holde internasjonale konferanser i juni 1988 i Jerusalem. Den 15. ALLC-konferansen skal finne sted 5.-9. juni, den andre AIBI-konferansen 9.-13. juni. Konferansene har en felles arrangementskomite, men registrering m.m. blir separat. Betydelig rabatt kan oppnås for deltakere på begge konferansene.

Programmene skal bestå av foredrag, paneldiskusjoner, sesjoner med produktvurderinger og demonstrasjoner. Det skal i tillegg arrangeres utstillinger over maskin- og programvare, bøker og tjenester. Utvalgte foredrag skal publiseres i to separate bind.

Til ALLC-konferansen ønskes spesielt velkommen foredrag som presenterer spesifikke teoretiske modeller koblet til nye forskningsresultater, men også bidrag som omhandler kritiske evalueringer kan komme i betraktning. Sammendrag av foredrag må være arrangørene i hende innen 15. desember 1987. Adresse: *Yaacov Choueka, Department of Mathematics and Computer Science, Bar-Ilan University, Ramat-Gan, Israel 52100.*

Også generelle opplysninger om begge konferansene kan fås fra ovenstående adresse.

CONFERENCE ON TEACHING COMPUTING AND HUMANITIES COURSES

The Association for Computing and the Humanities arrangerer dette kurset i tidsrommet 9.-11. juni 1988 ved Oberlin College, Ohio, USA. Konferansen er beregnet på medarbeidere som holder eller er i ferd med å utvikle kurs for humaniorastudenter (i litteratur, språk, historie, filosofi, kunst og musikk) om bruk av edb som redskap innenfor disse fagene. Et annet tema blir opplæring av både studenter og undervisningspersonale i edb-metoder. Konferansen skal *ikke* omhandle undervisning i humaniora ved hjelp av edb. Konferansearrangørene søker både foredrag og forslag til paneldiskusjoner. Sammendrag må være arrangørene i hende innen 30. november 1987. Flere opplysninger fås fra: *Professor Robert S. Tannenbaum, Chair, Program Committee, ACH Conference on Teaching Computers and the Humanities Courses, Department of Computer Science, Hunter College CUNY, 695 Park Avenue, New York, New York 10021, USA.*

Konferansen «Computers and Teaching in the Humanities — Redefining the Humanities?» skal holdes ved University of Southampton, England 13.-15. desember 1988. På konferansen skal det legges vekt på faglige spørsmål vedr. introduksjonen av edb i humanistiske fag i høyere utdanning. Søkelyset skal settes på grensesnittet mellom datamaskinen og humaniora. Hoveddelen av konferansen skal vies til «workshops» og seminarer, men det skal også være mulighet for uformelle demonstrasjoner.

Ytterligere opplysninger vil bli gitt i et senere nummer av *Humanistiske Data*. De som ønsker å få tilsendt framtidig informasjon direkte fra konferansearrangørene, kan henvende seg til: *Dr. May Katzen, Office for Humanities Communication, University of Leicester, LE1 7RH, England.*

SUMMARY

TRADISJON OG TEKNOLOGI

TRADITION AND TECHNOLOGY

In April Information Officer Kristin Natvig of the Centre visited the University of Oxford in order to report on projects and recent developments within computing in the humanities. Natvig got the latest news on the activities of Oxford University Computing Service (OUCS), Computing in the Arts, from Catherine Griffin and Lou Burnard.

OUCS has recently purchased a laser printer that is connected to the service's new VAX Cluster. A micro version of Oxford Concordance Program is now available, and a new mainframe version will be launched later this year.

A Content Adressable File Store (CAFS) has been temporarily installed in connection with OUCS' ICL mainframe. Lou Burnard has written a general retrieval program that exploits CAFS' potential for improved utilization of the corpora of the Oxford Text Archive.

Work has just started on a two-year project financed by the Computers in Teaching Initiative. The aim of the project is to make the set texts in undergraduate language and literary subjects available for on-line searching and the production of concordances and indexes via any IBM-compatible workstation within the university. The project is a co-operative venture and employs two people.

OUCS has recently been granted a research assistantship from the British Library from January 1st, 1988. The assistant will be responsible for a one-year investigation of current and potential applications of machine-readable texts.

Griffin and Burnard also arranged interviews with several researchers at Oxford. *John Cooper*, Institute for Oriental Studies, St. John's College, typesets Arabic texts using his own mapping system for handling certain vowels.

Elaine Matthews is one of the editors of *Lexicon of Greek Personal Names*. This project was started in 1972 and is funded by The British Academy. The aim is to produce a database/lexicon of all Greek personal names from the earliest historical times up to the Islamic

conquest of the Near East in the mid-7th century A.D. Half a million names have been collected, comprising some 13,000 different names. The material will be published by Oxford University Press in six volumes. The first volume will appear later this year. OUCS is responsible for the development of the software for phototypesetting the material.

DESKTOP PUBLISHING

Per H. Jacobsen, Manager of the Users' Support Service at the University of Oslo's central computing service discusses some aspects of desktop publishing. Jacobsen gives an introduction to the term and proceeds to concentrate on the Macintosh concept. A list is provided of the hardware and software available for this concept, including fonts, a device called Switcher (for sharing memory), and PostScript.

Characteristics of DTP systems in relation to typography and aesthetic issues in connection with the use of DTP are also discussed. Advice is given on preparing documents for publication and the use of layout programs. Problems that arise when processing texts in foreign languages are also dealt with.

Jacobsen winds up by discussing the pros and cons of the Macintosh concept for DTP, which in his view still is one of the best available.

TEKSTBEHANDLING OG TEGNSETT

WORD PROCESSING AND CHARACTER SETS

The subject of this article by Senior Computing Officer Espen S. Ore of the Centre is the problems that arise when texts written in other languages than English are data processed. These problems occur when keying in, printing and processing texts. Philologists have often had to resign themselves to ad hoc solutions that result in inefficient exchange of data and programs.

Ore gives an introduction to the technicalities of coding and representing characters and also to different character standards on microcomputers. The technical reasons for the difficulties connected with processing and printing non-English texts are accounted for. In this connection Apple's MacIntosh is much more flexible than other PC's, if not perfect.

Ore's suggested solution to these problems is that a joint exchange format be developed that retains all information on text formatting, character sets etc., that will allow researchers to employ the local

software and hardware that gives the best results. This would be possible if all producers of word processing systems at the same time design transferral routines to such a joint format. Perhaps this format could be based on the Standard Generalized Markup Language.

EI FREKVENSORDBOK FOR NYNORSK

A FREQUENCY DICTIONARY FOR NEW NORWEGIAN

In Norway there are two (very similar) official languages: «bokmål» and «nynorsk» (New Norwegian). Under the leadership of Senior Computing Officer Per Vestbøstad the Centre has developed a frequency dictionary for New Norwegian. In contrast to existing frequency dictionaries for «bokmål», the basic texts for this up-coming publication have been homograph separated and lemmatized.

60% of the texts derive from newspapers (1980-1983), 20% from a journal (1978 and 1983), and 20% from novels (1977-1982). This material comprises about 1 million words. Vestbøstad gives an account of the work involved in the different phases of processing the texts.

Special programs have been written for the production of a list sorted according to frequency with rank numbers and total frequency for each lemma, a final alphabetized list of lemmas, and an alphabetized list of all lemmas with both total frequency and frequency for each form. These lists will be published in book form at a later date.

FEM UNDERVISNINGSPROGRAMMER FRA DATASEKRETARIATET

FIVE CAL PROGRAMS FROM THE COMPUTER SECRETARIAT

The Centre receives CAL software from the Norwegian Computer Secretariat (a temporary government agency that promotes computing in schools) for test purposes. Senior Computing Officer Espen S. Ore describes five such programs intended for various age groups and types of schools. Rather than reviewing the pedagogical contents of these programs, Ore concentrates on user-friendliness and technical aspects.

«Artist» is a general drawing/painting program, roughly corresponding to «MacPaint». «Artist» is a strong program with many good possibilities, but not particularly suited to classroom needs.

«Ballong» («Balloon») simulates a trip with a hot air balloon and

functions more as a game than a program for learning physics. «Dataflora» is a very user-friendly, if rather over-simplified, tool (rather than CAL program) that helps pupils to identify plants. A physics program called «Newton» is in Ore's opinion utterly useless.

«Vevplan» also functions as a tool, in this case for designing woven patterns. «Vevplan» offers the pupils many possibilities, but is fairly difficult to learn.

POLITISKE HOLDNINGER

POLITICAL ATTITUDES

The Norwegian Social Sciences Data Service in Bergen has developed a combined workbook, statistics program and set of data — called «NSD-STAT» — for use in teaching high school-level social studies. In the opinion of Senior University Lecturer Einar-Arne Drivenes, the University of Tromsø, this package is well-suited as both an introduction to social scientific analysis and the methods and aids (including computers) used by social scientists.

The data set the package is based on has been extracted from an election poll in which people's attitudes to various societal questions were investigated. The workbook gives a good introduction to the data set and the questions it can help shed light on. On the other hand, the instructions for using the computer program are on the scanty side. The program itself functions satisfactorily.

Despite some weaknesses, NSD-STAT demonstrates that it is possible to design teaching packages on a high methodological and pedagogical level.

CALICO '87

The conference CALICO '87 was arranged in Monterey, California in April. Director Jostein H. Hauge of the Centre reports that several hundred people, mainly from the U.S., took part in the conference. 70 papers were presented within a wide range of topics — the development of CD-ROM, interactive video, etc.; the use of authoring languages; specialized program systems for language teaching on PC's; and a number of interactive language projects in various national languages.

In several papers basic pedagogical problems were discussed. Analyses were given of the possibilities of the new electronic media for adequate language teaching and which elements of the more traditional teacher/classroom-based instruction should be included in the interactive learning situation. However, new educational technology is both expensive and extremely time-consuming to develop.

Several speakers dwelled on the enormous potential of systems combining PC's, CD-ROM and interactive video. A demonstration of «The Visual Dictionary CD-ROM» bore witness to this. A presentation was given of the ATHENA Language Learning Project at MIT, Boston. The aim of the project, which started in 1985, is to create «the language lab of the future» in the form of a centre containing electronic audiovisual aids and AI-based software. All language students at MIT are supposed to be able to utilize the centre's facilities via a network of a potential several thousand powerful PC's. The project is to last for three years.

COMPUTERS AND TEACHING IN THE HUMANITIES

This is the title of a conference that was arranged in Southampton in April by the Office for Humanities Communication, University of Leicester, and the Department of Electronics and Computer Science, University of Southampton. The conference was one of the first of its kind in the U.K. and was attended by about 100 people, mainly from institutions of higher education in the United Kingdom and Europe.

Information Officer Kristin Natvig of the Centre reports that Richard Ennals, Kingston College of Further Education, gave the introductory speech. Ennals concentrated on the *relationship* between computers and the humanities.

In sessions on language teaching MICROCONC, a concordance package for teaching technical English, was presented, in addition to a computing course for students of German that was established at the University of Aberdeen in 1981.

In another session fundamental questions were raised concerning the role of computer methods in language teaching. Two demonstrations were held: of CONCORDE, an authoring system for testing verb conjugation, and «Text Jumbler», for practice in text construction.

A session on computing in the teaching of literature dealt with problems that arise when computer applications are to be adapted to literary studies. In the session on history teaching the impressive «DISH Project» at the University of Glasgow was presented.

J.-C. Gardin, Ecole des Hautes Etudes, Paris gave a final paper that went way beyond the scope of the conference.

A lack of pedagogical theory was evident at the conference, and some speakers seemed to consider computers as toys rather than pedagogical aids. In spite of this the conference was highly interesting, and the participants got a good impression of the variety and scope of computing methods in teaching humanities in the U.K.

OPTICA '87

Senior Computing Officer Øystein Reigem of the Centre took part in Learned Information's conference «Optica '87» in Amsterdam in April. Here Reigem gives a combination of a conference report, news, and his own comments, all concerning CD-ROM and off-shoots.

For the entertainment market integrated equipment will be launched next year that can be connected to television and hi-fi sets — CD-I (CD-Interactive). CD-I will be entirely digital. It will allow live video, but in the first version limited to 1/4 of the screen and at a speed of a mere 12-15 images per second. A CD-I player will have its own built-in computer and operating system, and will also be able to play CD's.

CD-V (CD-Video) will be introduced as a kind of video single record, in the size of a CD with 20 minutes of digital CD sound and 5 minutes of live analog video. 8" and 12" discs will also be launched, so CD-V can be considered a successor to LaserVision videodiscs. A disadvantage is that CD players will not be able to play CD-V discs. However, it is likely that combined players for CD, Laservision and CD-V discs will be developed.

By crossing CD-V and CD-I a CD-I disc with live video (CD-IV) will be possible, but only with analog video.

Digital live video on CD is not necessarily a thing of the future. At a conference in March General Electric/RCA demonstrated the so-called DVI (Digital Video Interactive). DVI allows up to one hour of digital video on CD-ROM, and does not require modifications of the CD-ROM player. CD-ROM with DVI can be a serious competitor to both CD-I and videodiscs.

CD-ROM news: Microsoft have announced their MS-DOS Extension for CD-ROM files. Equipment is now available for managing several discs simultaneously. A new firm, EIKON, is going to develop software specifically for the European market. A new CD-ROM production

is «The Microsoft Bookshelf», which consists of 10 well-known reference works that can be accessed directly from word-processing systems. Digipress is planning on launching their «Space-CD» based on metal or glass rather than plastic.

«THE USE OF COMPUTING IN THE TEACHING OF LANGUAGE AND LANGUAGES»

A report is given on this conference by Senior Computing Officer Per Vestbøstad of the Centre. The conference was attended by 130 researchers and teachers, all of whom apart from three were British.

Papers were given in three plenary sessions and six sessions of four parallel presentations. One of these was a workshop on the topics «Basic Concepts», «Working with Authoring Packages» and «Introduction to Concordances».

In the first plenary session Gabriel Jacobs from Swansea discussed methodological questions connected to the use of CALL programs. In the second plenary session Rex Last (Dundee) gave an in-depth overview of artificial intelligence in connection with CALL, with emphasis on the problems involved when trying to make a computer system handle linguistic knowledge.

Christopher Butler (Nottingham) gave a useful list of existing software for language teaching. Writing aids were dealt with in detail by Noel Williams, on the basis of a major project at Sheffield City Polytechnic, where this type of program has been evaluated.

Most of the papers dealt with on-going projects, especially in French, but German, Spanish, Italian, Latin, and phonetics were also represented. Several participants reported on their experience with the use of computing centres for students and teachers.

ICAME 8th

Senior Computing Officer Knut Hofland of the Centre reports on the 8th ICAME conference, which was held near Helsinki in May. More than 70 people from 12 countries took part in the conference.

As previously the conference consisted of a combination of reports on projects concerning corpus linguistics and papers presenting studies based on one or more text corpora. This year's conference bore evidence of the increasing use of PC's in corpus work.

In Nijmegen work is currently taking place on projects concerning the analysis of English, Arabic and Spanish. The project LDB II deals

partly with a transferral of the linguistic database system LDB to the MS-DOS operating system, partly with an adaptation of the system for use in computer-assisted learning.

In Amsterdam the project ASCOT has been concluded. A computer-stored lexicon (ASLEX), based on Longmans Dictionary of Contemporary English (LDOCE) has been devised. The aim of a new project, LINKS, is to develop a lexical knowledge base based on LDOCE.

In Lund the project TESS, «Text Segmentation for Speech», will be brought to a close soon. In Birmingham a new dictionary funded and published by Collins has been completed. At the University of Sydney work has commenced on an Australian text corpus.

Speech processing was the subject of the papers given by speakers from Lancaster. A corpus of spoken English will be available from this autumn. A project on speech recognition has also been started at Lancaster.

A program library in connection with ICAME is presently being built up in Bergen. Electronic mail is now being used in order to distribute information and programs.

The proceedings of the conference will be published as usual by Rodopi, Amsterdam. In addition, abstracts will be included in the next issue of ICAME Journal.

XIV ALLC CONFERENCE

The 14th conference of the Association for Literary and Linguistic Computing was held in Gothenburg in June. Senior Computing Officer Espen S. Ore of the Centre reports that it was attended by 65 delegates from 15 countries. The papers that were given covered the whole range of literary and linguistic computing, with a certain emphasis on text corpora and lexical databases. An issue several speakers dwelled on was the exchange of data between various coding formats and the need for standardization.

Nicoletta Calzolari, Università di Pisa, gave the introductory paper on lexical databases, especially those intended for machine translation. In a paper written by D.T. Barnard, C.A. Fraser, and G.M. Logan an account was given of work on the development of tagging systems based on the SGML framework.

M.A. Heather and B.N. Rossiter showed how existing typographical coding in a text can be utilized in the construction of a retrieval system. J.C. Lejosne from Université de Metz demonstrated how

methods from computational linguistics/machine translation can be used in the interpretation of a manuscript that has been distorted several times over.

UTVIKLINGSSEMINARET 1987

THE 1987 DEVELOPMENTAL SEMINAR

In cooperation with Holger Hagan of the University of Tromsø the Centre arranged the annual developmental seminar for advanced computing staff at Geilo in June. Three members of the Centre's staff report that the seminar had 30 participants and dealt with three fields: CD-ROM, digitalized image processing, and desktop publishing.

A researcher at the Norwegian Computing Centre in Oslo described CD-ROM in relation to other optical storage media and discussed the «High Sierra» standard in detail. He also gave an overview of the work involved in the production of the first Norwegian CD-ROM, which his institution has produced in cooperation with a publishing firm and the Central Bureau of Statistics. This disc contains digitalized maps, statistical data, and texts.

The session on storage and retrieval of digitalized images and documents on a PC AT was divided in two. First a theoretical lecture on digitalization, document processing, picture improvement and compression, and hardware was given, followed by a demonstration. In the second part of the session practical issues were discussed between four lecturers and the participants, and some experimentation with equipment and software took place.

In the session on desktop publishing presentations and demonstrations were given of the layout programs Pagemaker and Ventura. The seminar's participants had the opportunity to try these programs out for themselves.

MELDINGER

NEWS

A lengthy report on the Centre's videodisc project has recently been published. The report contains chapters on the background of the project, preparatory work, administration and organization, preparation

of the pictorial material and reference data, software, finances, evaluations of the various aspects of the project, advice to those who wish to produce a videodisc, and further work on optical media.

The computing service at the Faculty of Arts, University of Gothenburg, was established in 1985. The service cooperates with the Department of Linguistic Computing. Its aim is to advise, support, and educate researchers and students in computing, besides to run the service's minicomputer. The staff of four also provides information on available software for the humanities and keeps abreast of developments in the field.

The Museum Documentation Association (MDA) offers extensive consultancy services to museums that wish to revise their routines for documentation and automation. The association has developed guidelines for advice of this kind and applied them in commissions for different types and sizes of museums in the U.K. MDA has developed software for various purposes, including a program package — Museum Object Data Entry System (MODES) — for documentation work and collection management. The association also publishes books and a newsletter, *MDA Information*.

A working group initiated by the Institute of Field Archaeology/Royal Commission on Historical Monuments of England (IFA/RCHME) has published the report *IFA Occasional Paper Number 1*. The report contains a comprehensive catalog of hardware and software in use within British archaeology, recommendations of the working group, and a complete bibliography of publications in English on computer applications in archaeology. The publication of this report marks the launch of The IFA Computer Advisory Service.

The 1987 Directory of Computer Assisted Research in Musicology has been published. The directory comprises 151 pages and is edited by Walter Hewlett and Eleanor Selfridge-Field.

Journals presented in this issue are: 1. *The CTISS File* — published 3-4 times per year by The Computers in Teaching Initiative Support Service in the U.K. Free of charge. 2. *The Computers and Philosophy Newsletter* — published by The Center for Design of Educational Computing, Carnegie Mellon University, USA. Three issues will be published in 1987 for \$9. 3. *Musletter* — published quarterly by IBM-Los Angeles Scientific Center and the USC School of Music. Deals with developments within computer-assisted music teaching. Free of charge.

Forthcoming conferences:

— CALICO '88 — Salt Lake City, Utah, February 23-27

- Linguistic Approaches to Artificial Intelligence — Duisburg, West Germany, March 23-26
- Computers and Music Research Conference — Lancaster, April 11-14
- Women, Work and Computerization — Amsterdam, April 27-29
- ALLC — — AIBI Joint Conferences — Jerusalem, June 5-13
- Teaching Computing and Humanities Courses — Oberlin College, Ohio, USA, June 9-11
- CATH 88 (Computers and Teaching in the Humanities) — Southampton, December 13-15

Forts. fra 2. omslagsside.

RAPPORT nr. 29, 30, 31, 32: *Stig Welinder et al.: STAR I-IV. A program package for archaeological use.* Bergen 1983.

Nr. 29 *STAR I. Introduction and STAR manual.* ISBN 82-7283-033-7 Pris kr. 50.

Nr. 30 *STAR II. Student textbook and STAR examples.* ISBN 82-7283-034-5 Pris kr. 60.

Nr. 31 *STAR III. Archaeology for statisticians.* ISBN 82-7283-035-3 Pris kr. 60.

Nr. 32 *STAR IV. STAR algorithms.* ISBN 82-7283-036-1 Pris kr. 30.

RAPPORT nr. 33. *Årsmelding 1983.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-038-8 Gratis.

RAPPORT nr. 34. *Jostein H. Hauge: Tutorial on Machine Translation.* Rapport fra en konferanse i Lugano 2.-6. april 1984. Juni 1984. ISBN 82-7283-039-6 Pris kr. 60.

RAPPORT nr. 35. *Ole Lauvskar: Bruk av statistiske metoder i språk- og litteraturforskninga. Rapport frå ei spørjeundersøking.* September 1984. ISBN 82-7283-041-8 Pris kr. 50.

RAPPORT nr. 36. *Årsmelding 1984.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-042-6 Gratis.

RAPPORT nr. 37. *Årsmelding 1985.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-043-4 Gratis.

RAPPORT nr. 38. *Jon Birger Østby: Edb-metoder for kunst- og kulturhistoriske museer.* ISBN 82-7283-045-0 Pris kr. 60.

RAPPORT nr. 39. *Årsmelding 1986.* NAVFs edb-senter for humanistisk forskning. ISBN 82-7283-047 Gratis.

RAPPORT nr. 40. *Roger Erlandsen, Claus Huitfeldt, Øystein Reigem: Videoplateteknologi i humanistiske fag.* ISBN 82-7283-048-5 Pris kr. 55.

Øvrige rapporter er utsolgt.

C

ARTIKLER I DETTE NUMMER:

*J.J.N. Palmer: The Hull Domesday
Database Project*

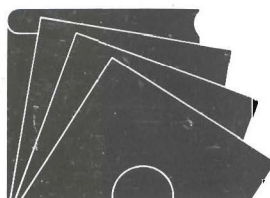
*Daniel Greenstein: Programming in
SPITBOL for Historians*

Kristin Natvig: Tradisjon og teknologi

Per H. Jacobsen: Desktop Publishing

*Espen S. Ore: Tekstbehandling og
tegnsett*

*Per Vestbøstad: Ei frekvensordbok for
nynorsk*



**NAVF'S EDB-SENTER
FOR HUMANISTISK FORSKNING**
NORWEGIAN COMPUTING CENTRE
FOR THE HUMANITIES

RETURADRESSE:

NAVF's edb-senter for humanistisk forskning, Boks 53 — Universitetet,