Book Reviews

The Humanities Computing Yearbook 1989–90

Ian Lancashire (editor)
(University of Toronto)

Oxford: Clarendon Press, 1991,
xviii + 701 pp.
Hardbound, ISBN 0-19-824253-0,
$125.00, £60.00

Reviewed by
Rosanne G. Potter
Iowa State University

This excellent reference book should be on the shelves of all scholars interested in computational approaches to the humanities disciplines. The international Editorial Board of the *Humanities Computing Yearbook (HCY) 1989–90* construes the term "humanities disciplines" broadly; i.e., to include traditional and computational linguistics. Since the selling price is high, I give extensive details from the table of contents and include the number of pages per topic to assist readers of this review in making a purchasing decision.

The 701-page volume is organized into three unequal parts. Part I (380 pp.) covers 16 disciplines: Archaelogy (8 pp.), Art History (8 pp.), Computational Linguistics (34 pp.), Creative Writing (4 pp.), Dance (1 p.), Drama (2 pp.), English Language Instruction (19 pp.), Folklore Studies (1 p.), Historical Studies (34 pp.), Law (13 pp.), Lexicography (11 pp.), Linguistics (24 pp.), Musicology (9 pp.), Natural Languages and Literatures (155 pp.—27 languages or language groups), and Philosophy (30 pp.). Many disciplines are divided into sub-disciplines: Computational Linguistics has Grammar Development Systems, Machine Translation, Morphological and Syntactic Analysis, Question-Answering Systems, Semantic Analysis and World Knowledge, Text and Discourse Analysis, and Text Generation; Linguistics contains Corpus Linguistics, Dialectology and Dialectometry, Modeling, Phonemic and Phonetic Transcription, Speech Analysis, and Speech-to-Text, Text-to-Speech. Each disciplinary division contains a discursive introduction and a thoroughly annotated listing of software and data; if there are sub-disciplines, each also has an introduction and a software and data section.

Part II (137 pp.) covers eight kinds of methods or tools: Bibliographic Databases: Online and CD-ROM (16 pp.), Editing and Publishing (27 pp.), Information Management (21 pp.), Programming Languages (9 pp.), Second-Language Instruction (13 pp.), Statistics (4 pp.), Text Analysis (30 pp.), Text Processing Techniques (10 pp.). Part III (182 pp.), on Resources, consists of seven sections: Bibliographies (6 pp.), Electronic Texts (9 pp.), General Guides and History (2 pp.), OCR (3 pp.), People and Places (26 pp.), Reference Bibliography (25 pp.), and Index (104 pp.).

Some sections are so brief (Dance, Drama, Folklore) that one must wonder why they were included; some sub-sections are also quite incomplete (Stylistic Analysis in the Text Analysis section springs to mind). In general the coverage is broad and detailed.

The discursive introductions are inevitably variable in quality since over half were done by members of the international Advisory Board, but generally they are very good. Those done by the editor are full of thoughtful generalizations about what is going on in the specific area and how it relates to work in other related fields. Lancashire often also points out the principal researchers, data collectors, and software developers.

The sections and sub-sections designed by Lancashire and his team are intuitively on target though, surely, some scholars and researchers will wish their material had appeared under different headings.[1] The overall organizational method of this collection (listings of bibliographic citations, software, and data collections under multiple headings and sub-headings) may strike different bibliographers as either insufficiently or excessively analyzed. I found the mix of essays, books, commercial, and noncommercial products perfectly fine. The fields are still small enough that separating out the genres would be pedantic, while a purely alphabetical listing would be unusable. Eventually, the bibliographic materials will drop out of the *HCY*, as increases in software and data push the citations into subject bibliographies, but for the time being it is fascinating to look at these different products of computational scholarship cheek-by-jowl with one another.

Rather than generalize about what a typical software description looks like, I quote a short example:

> **GRAMmarCRACKER.** English sentence parser (vers. 5.1) for MS/PC-DOS applications in computational research that introduces the beginner or supports the expert in natural language processing. GRAMmarCRACKER parses in three steps and learns a fourth step. Each of the first three steps is a self-contained parser. The first parser, the lowest level, assembles one sentence's characteristics into a list of words. The second step classifies those words by part of speech using a dictionary and a database of sentence structures. The third step (the diagram sentence step) collects words into phrases and then groups these phrases into sentences. The fourth step learns by asking the operator if the output is correct and then adds to the dictionary. The knowledge base consists of a dictionary and a collection of sentence typologies. This program recognizes a majority of common sentence forms and most English language tokens including numbers, contractions, and abbreviations. It includes pruning algorithms for speed and has well-documented source files. A users' manual is included. (p. 37)

The description of the software is followed by citations of essay reports on, and reviews of, the software, if any. Each entry ends with a listing of the name(s) of developer, designer, programmers, publisher, and distributors with addresses and phone and FAX numbers; and details on availability, cost of disks, manuals, etc., appear. As Lancashire points out, the computer industry is volatile, so much information will be out of date when published in a yearbook. The *HCY 1989–90* gives enough names and addresses that one is likely to be able to track down someone who knows something.

Citations reporting on the applications of the software to specific research projects (by the developers and others) point to more developed software; reviews in the major journals enable readers to read independent evaluations; notes about publishers

1 Possibly these wishes (at least about previously reported citations) will be taken into consideration in subsequent volumes of the *HCY*.

and distributors help readers to find generally useful software (as opposed to that developed for one project and never picked up by a publisher). In general, the entries vary in length from a couple of lines to a couple of pages; most are one long paragraph, followed by details on how to get a copy.

My most serious complaint as a user is the lack of boldfacing for main entries in the index. Searching for primary definitions or descriptions, whether unpackings of acronyms or details on what software does, can be rather like a treasure hunt; the information is definitely here, but the path to it is not direct. I provide here a sample following-my-nose search from initial questions aroused by a citation to answers, using the signposts provided. In my readings about data available in English, I discovered that "the 1880 Diary of Mary Ann Pierpont (CCAT)" can be found on the "PHI/CCAT Demonstration CD-ROM #1" and that it runs "on Macintosh (with *Pandora*, etc.)" (p. 243). Let's say I don't recognize CCAT, but am intrigued by the Macintosh, and want details on *Pandora*. I wonder what kind of output the *Pandora* software would produce from the 1880 Diary and am particularly interested as I knew of no concordance package for the Mac.[2] First, I looked in the reference bibliography, which unpacks the meanings of many acronyms, to no avail because CCAT is not an abbreviation for a book or serial. The index contains 16 references; the primary entry on CCAT appears on page 533, the 14th page reference! Luckily, the second reference points to skipping the index and going straight to Part III—Text and Data Archives. Boldfaced page numbers in the index (for the main descriptions as opposed to secondary citations) would have made this, and many other searches, easier for users of the index. In this field, dominated by alphabet-soup names of programs, archives, organizations, and centers, one needs more transparent aids for navigating, especially when references to bibliographic citations are also abbreviated to acronyms.

The complete description of *Pandora*, i.e., who the developer is (Elli Mylones), where she is (Harvard), what the software was designed to do (search multiple texts in the *TLG* and display the results as a list of references or surrounded by lines of context) is found on page 205, the fourth page reference in the index. Again, if I'm not familiar with what the *TLG* is, I must look up six page references before I come to the main entry under Classical Languages—Greek. The answer is always there, but is much more quickly available to readers who already know that CCAT is a text archive than to those who don't. There are 48 abbreviations (listed on the first page of the index) that identify what a software product in the index is, e.g., a **db** (database) or a **ta** (text archive), but they are not applied consistently, e.g., CCAT is not identified as a text archive.

My question about what *Pandora* does in the way of searching and displaying search results for the 1880 Diary of Mary Ann Pierpont is, after all this, answered by implication only; i.e., it probably works, but *Pandora* was designed for searching Greek texts in the *TLG* format. On the way to answering it, however, my knowledge has increased (about text archives, Macintosh concordancing, and text retrieval software), and my use of the index to the *HCY* has improved.

However, there are inconsistencies, both trivial and serious, between the Introduction and the text itself, which represent sloppy work on the part of Clarendon Press's copyeditors. They should have caught references to Parts A, B, and C in the Introduction: the book is divided into Parts I, II, and III. More seriously, someone should have made sure that when Lancashire thanks Nancy Ide for her excellent introduction to the general subject of "artificial intelligence methodology, classical and connectionist"

2 The index led me to *MacConcordance* by Stephen Clausing.

(p. xiii), that these terms actually appear in the text referred to. My interest piqued by Lancashire's comments in the Introduction, I wanted to read what Ide had to say, but it wasn't simple to find. When my initial scan for Ide and Artificial Intelligence came up dry, I wrote down all the references to "Artificial Intelligence" (listed as "AI" in the Index), "connectionism," and "Ide" and compared the page numbers; eventually I found one "close enough" combination ("connectionism" on 60 and Ide on 61, but nothing about AI nearby). The section called "Text and Discourse Analysis" begins on page 60, but there is no credit line for Ide. There is, however, an Ide citation on page 61. Maybe I've found the introduction Lancashire was pointing to, but maybe not. The shifting terminology and inconsistency in crediting the work of collaborating editors makes it difficult to follow up the editor's good pointers. One does not expect to see this type of error in a book copyedited by a major university press.

Although I've mentioned a number of minor difficulties in this reference tool, my general experience in using it has been quite pleasant. I am very impressed by what Ian Lancashire has accomplished with the aid of his international team of sub-editors and the support staff at the Centre for Computing in the Humanities at the University of Toronto. The more one uses the tool, the clearer it becomes that the *HCY 1989–90* has been carefully constructed, and remarkably well edited and proofread.

This yearbook provides an admirably detailed snapshot of how this large baggy monster, computational studies in the humanities (especially languages and literatures), looked at the end of 1990. The *HCY 1989–90* both stimulates interest and satisfies it; as a one-volume overview to this complex net of interwoven research, data, and applications, it cannot be beaten.

Rosanne G. Potter is the editor of *Literary Computing and Literary Criticism: Theoretical and Practical Essays on Theme and Rhetoric* (University of Pennsylvania Press, 1989) and is a member of the Literary Texts Group and the Performance Texts Workgroup of the Text Encoding Initiative. Her research interests include pragmatic studies of reader responses to literary texts, character definition through assigned syntax in plays, and Women's Studies computing. Potter's address is: Department of English, Iowa State University, Ames, Iowa 50011; e-mail: s1.rgp@isumvs.bitnet