HICKEY R., M. KYTÖ, I. LANCASHIRE & M. RISSANEN eds. 1997. *Tracing the trail of time. Proceedings from the Second Diachronic Corpora Workshop*. Amsterdam - Atlanta G.A: Rodopi (vi + 241) [ISBN: 90-420-0214-X (bound); ISBN: 90-420-0209-3 (paper)]

The suggestiveness of Peter Laslett's (1983) title *The world we have lost further explored* serves us as a good point of departure to comment on this excellent and encouraging book. Fortunately the past is no longer a dead and sacred world one can or must respectfully approach, but an adequate interlocutor to present scientific concerns. *Tracing the trail of time* comes to enliven a dynamic dialogue, and like any vigorous conversation it exhibits a naturally 'modal' nature. In fact we could safely say that this 18th volume of the Studies in Practical Linguistics (Language and Computers) Series constitutes a 'modal book'. In it the reader will not find objective final statements, round conclusions, stylized arguments. Very much to the contrary, the emphasis of the volume clearly falls on the purposes, intentions, fears, doubts and hopes that colour present and mainly future research, that is, the articles collected represent the outcome of a real, successful workshop. In my opinion, the book's deep freshness and singularity lie precisely in the selection of these modal aspects as relevant. From this approach almost any reader may feel encouraged to take part in what reveals itself as an ongoing, attractive activity: the further exploration of a world we have lost, or even worse, may have thought lost.

By the turn of the century no one will doubt that electronic data retrieval and computer-aided linguistics has been given due credit or that, lacking the evidence of the native speaker, the historical linguist may widely benefit from a rapid and friendly access to the texts, as Kytö and Rissanen observe. Substantial text samples constitute a solid alternative to impressionistic studies and smooth the way to a comprehensive and intellectually more satisfying view of language according to which the boundaries between synchrony and diachrony fade away.

Formally, the book is structured as a collection of fifteen papers stemming from the Second Diachronic Corpora Workshop that took place in Toronto in 1995, to which the introductory article by M. Kytö and M. Rissanen 'Language analysis and diachronic corpora' has been added. The articles are divided into two sections: (a) nine contributions devoted to the description of corpora development, and (b) studies based on corpora themselves. Many of the papers offer appendices with sample pages of ongoing work or examples of working software.

After the pioneering work of the Helsinki Corpus and the Old English Dictionary of the University of Toronto in the 1980s, what seems to be a forest of historical corpora has emerged and corpus coverage of the History of English keeps constantly improving under their influence. In fact, most of the scholars convened to that Conference belong to or cooperate with these well-known teams. In the book twelve corpora are discussed, ranging from Old to 18th c. English. The projects, however, exhibit a clearly heterogeneous nature: we are presented text-only vs annotated compilations; completed and at the researcher's disposal or in preparation, more or less ambitious in the time-span selected for analysis, more or less refined in their tagging, more or less comprehensive in their sampling.

As for the contents, the reader learns both encouraging results obtained so far and tentative solutions to yet unsolved problems. The book constitutes, thus, an idealistic

catalogue of questions and desired goals, and of the problems and difficulties still pending to achieve them, either technical or theoretical. It materializes, thus, a shared and enthusiastic reflection on the qualitative limits of computerizing.

All researchers agree that the aim of corpora is to provide material and background information in the most accessible form so as to allow the corpus-user to draw empirical conclusions. However, this is no easy task. Since many of the corpus-specific problems can be of general interest, problems of a varied sort are put forward through the articles: i.e. how to get further and better tagging and parsing (so that grammatical and syntactic issues can be retrieved), how to attain the full linking of information and resources, how to unify coding conventions (although TEI 'Text Encoding Initiative' is recognised to be a great step forward), how to offer helpful user-friendly retrieval tools, etc. The reader moves from the particular -warnings about the dangerous and misleading risk of typographical errors when dealing with medieval texts- to the most general reflections: will the amount of information provided as complete as possible be helpful to users or would it "bespeak an errant pedantry"? (Speirs: 142). The book makes clear too that advances in computer technology and new options force corpora to undergo transformations and constant reshaping, which turns out to be often slow and retarding, although it is also this dynamic character that allows for both on-going correction of errors (not without risks, however) and the incorporation of scholarly advances.

As corpora should become not only more informative and more adequately so, but also more internationally available, more portable and long-living, in this rapid overview of common concerns, the reader becomes aware of problems regarding the international distribution, publicization and even commercialization of products. Complaints about copy-right difficulties and slow achievements are scattered through the book.

Apart from these common technical hindrances, *Tracing the trail of time* makes the reader familiar with theoretical difficulties that emerge in corpus ellaboration. I will only point out two of these theoretical concerns, which are broadly shared by historical linguists of any kind: genre definition and editorial criteria.

Authors complain about the lack of a satisfactory and generally accepted classification of text types. This question of vital importance has been more cared for since the emergence of sociohistorical and text linguistics. Functional and situational factors as well as factors of linguistic form should be used in the definition of text-types as types of social activity. Researchers are, thus, led into the social embedding of long-term language change and into the consideration of broad cultural-pragmatic parameters. However, conventions and culture have proved to be concepts dependent on time and place and, therefore, subject to change. In fact, many of the projects present detailed documentation of changing relations between language and its users. In the prevailing dynamic approaches to historical linguistics neither intralinguistic nor extralinguistic variables can be taken as constant and there lies the complexity to define genres or text-types. At the same time, as different genres are affected differently by linguistic change, it is especially important for historical corpora to code genre in some way. Despite interesting proposals and categorizations that have proved partially ad hoc, however, we seem far from reaching a satisfactory logical and unique system of genre labels, and further discussion and research in the grouping of texts is

called for. Up to now, some text-types are better represented than others in the corpora available. Literary prose works and, more specifically, letters have been traditionally considered, but the near future will undoubtedly give way to new genre-based corpora, as Kohnen (189) suggests. It will also be open to specific selections of non-central areas of literature. Works on non-literary, non-fictional material will certainly proliferate, although the decision to apply those labels to texts before the 17th century is recognisably risky (Schmiedt & Claridge: 125). Regional corpora recently begun to be compiled that is, Irish, Older Scots, Early American English among others, will obviously find their proper place. In short, we can expect a new stage of corpus-compilation in which text-selection of a narrower scope will be based mainly on genre, author or regional variety.

A second topic in the book deserves our attention, namely the treatment that should be given to older texts. As is well known, the new relations that hold between language and history have increased academic sensitiveness towards historical documents. A new epistemology has arisen where the objectivity, purity and neutrality of traditional scientific procedures are questioned, both ideologically and philologically (cf. Edwards 1987, Pearsall 1987, Lass 1993, Robinson 1994, Lass 1997, etc.). We have been made increasingly aware of the artificial and mediating character of modern editions of older texts (cf. Griffiths 1994 and Richards 1994) and, consequently, editorial criteria have been notably changing in the recent past. Rejecting hidden and excessive interpretation of 19th c. practice, the reader is invited to trace the normalizing paths of editors and to savour the details of the original text through bilingual critical editions, which have turned more and more frequent even in handbooks (cf. Bravo 1982, Burnley 1992 or Cusack 1998, to mention only a few). This philological sensitiveness towards documents is also reflected in computer-aided research. Far from the older editing practice of constructing the 'best text' from available versions, the attitude behind computer-aided work opts for the more recent trend of respect. The optimal mode of manuscript editing for academic purposes, as Markus (66) remarks, is a dynamic one through which the reader is provided with different text versions so that s/he can decide on the subset of data s/he is interested in and the trajectory of his/her investigation. That is, the responsible corpus-compiler must concentrate on the relations among texts, manuscripts and editions, and should signal secondary or other sources where relevant, avoiding ready-made solutions. The goal would be to reach facsimile editions or digitized images of manuscripts carefully tagged. This formal treatment will not only offer us the pleasure of looking at documents, but allow improved readings of damaged parts. The marking would include characters indicating fire damage to particular lines, different scribal hands, subsequent additions to the text, indications of the different manuscripts used for the basic edition, etc. From the point of view of this pluralistic, non-intervening attitude toward editing several projects seem to be noteworthy: *The Canterbury Tales* and that of *Beowulf* -just referred to in the book-, and those of *Machyn's Diary* and *Middle English Prose* carried out under the supervision of Bailey and Markus, respectively.

Tracing the trail of time offers many more refined descriptions and subtle commentaries on the subject that will suit the needs of different readers. However I would not miss the chance to mention what I consider to be the book's most important contribution: a constant and often explicit warning against electronic data abuse.

Although development of corpora and databases has enlivened the historical study of language, not all studies should be thought conclusive or even worth-while, in that they do not contribute to the advance of learning. As Meurmann Solin clearly states, the linguist must feel constantly challenged by structural decisions s/he must take when building or using corpora (200). We cannot but deeply agree with her call to get rid of earlier assumptions regarding the linguistic features established as 'diagnostic' in the reconstruction of past stages or varieties, that is, in the reconstruction of history. Many researchers follow Meurman-Solin in reminding us of (Rissanen's 1989) words: "Computerized corpora are just a tool, a further technique for analyzing data. It will never replace the study of primary sources and thorough knowledge of the past stages of English. This simple fact should be kept in mind both in carrying out research and in planning graduate programmes of tuition". Unsurprisingly, after the productive fashion of computer-aided studies, theory building seems to be needed more than ever; more emphasis is demanded on the research part of the project over the sophistication of the corpus work, and further critique on information, so that data may not be interpreted as final statements. Electronic research cannot justify previous uniform assumptions or sidetrack us from the real goal of actual progression of learning. In short, throughout the book we openly confirm the impression that we need the intuitive and bravely creative linguist more than ever.

The reading of *Tracing the trail of time* proves that corpus-compilation turns out to be much more than translating older texts into machine-readable versions. It demonstrably constitutes the task of a heterogeneous team of scholars, from computer specialists to strict philologists, all of whom recognise the generosity of colleagues suggesting improvements. One must also acknowledge the intellectual vitality exhibited by these long term projects of concentrated effort, frequent updating, and tight close-up of partial results. The enthusiasm underlying the whole enterprise is reflected in the fact that pilot studies and new corpora are announced (Meurman-Solin 212), as applications of this kind of study are tested, from profound interpretations of a single text or language change to contrastive linguistics of any kind.

I would like to end this review by praising the institutional assistance some researchers count on to build their data-bases. Toronto Senior High School students are invited to take part in voluntary programmes as part-time research assistants. Projects of this kind may certainly encourage a fruitful institutional integration of teachers and students from different environments and of different levels of experience, which would render substantial benefits in the long run. We can only regret that these policies still remain but a dream in our country.

Paloma Tejada Caller
Departamento de Filología Inglesa
Universidad Complutense de Madrid

REFERENCES

Bravo A. (1982). *Literatura anglosajona y antología bilingüe del antiguo inglés*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Burnley J.D. (1992). *The history of the English language. A source book*. Londres: Longman.

Cusack B. (1998). *Everyday English 1500-1700*. Edinburgh: Edinburgh University Press.

Edwards A.S.G. (1987). Observations on the history of Middle English editing, in Pearsall: 34-48.

Griffiths, J. & P. Beal. (1994). *English Manuscripts Studies 1000-1700*. Toronto: University of Toronto Press.

Laslett P. (1983). *The world we have lost further explored*. London: Methuen.

Lass R. (1993). Interpreting vs disappearing. On texts as historical objects. *Selim* 3: 7-26.

Lass R. (1997). *Historical linguistics and language change*. Cambridge: Cambridge University Press.

Pearsall D. (1987). *Manuscripts and texts*. Cambridge: Brewer.

Richards, M.P. (1994). *Anglo-Saxon Manuscripts. Basic Readings*. Nueva York: Garland.

Robinson, F. (1994). *The editing of Old English*. Oxford: Blackwell.