Any book dealing with the use of computers for stylistic analysis shoulders the burden of convincing the reader of such a possibility. The difficulty increases when the targets are any concrete literature and its mass of critics who reject measures, quantitative issues, and patterns produced following the templates of machines. It is therefore remarkable that a prestigious scholar and literary critic, a well-known author in more traditional endeavours, dares to face the inquisitive views of the professional dubitantes, and presents several papers suggesting examples of how to achieve new targets in literary criticism with very unsophisticated computer tools. This is a courageous book based on a simple statement (p. 103): 'Literature is no science, but it is made of words, and those are actually subdued to certain universal and unavoidable laws.'

Although published in 1997, this book collects revised versions of eleven papers published since 1989, together with two unpublished, and follows a methodology basically coincident with that postulated by Oakman in 1984. Therefore, it would be unfair to review it as a book on the state of the art in literary analysis and criticism at the end of the nineties. Computers and Literature in the Eighties would be a more accurate title, probably. The literature dealt with is Spanish literature, both Peninsular and Latin-American. The authors go from Columbus to modern writers, and due attention is paid to women, particularly from the point of view of possible language differences between women's and men's writing. The works of Alonso Ramírez, Carlos de Sigüenza y Gongora, Octavio Paz, Rosario Castellanos, Edgardo Rodriguez-Juliá, Rafael Dieste, R. M. del Valle Inclán, E. F. Granell, B. Pérez Galdós, C. Columbus, A. Buero Vallejo, G. A. Bécquer, E. A. Laguerre, and Concepción Arenal are analysed, searching for stylistic patterns, authorship, and differences in writing according to gender, ordering the chaos, establishing differences in a corpus of thirty Spanish writers, pondering individual evolution, describing the dialogue, resisting deconstruction, graphically viewing the data, or simply judging the evidence. Twelve chapters that answer the question posed by the first paper in the book: 'Why use the computer for literary criticism?'.

The book, in general, reflects a remarkable effort of adaptation to common computer technology by an author who 'had already published eighteen books in the "traditional" fashion. What is important in it is not the introduction of new techniques or theoretical updates in computational linguistics, but the rational and careful application of computers to basic issues in literary criticism. At the same time, it actually helps us individualize those basic issues. It exemplifies low-level statistics for stylometry, shareware, or easily available commercial programs, and an overwhelmingly pedagogical intention.

When I say that this is a book for any literary critic to read, I do not mean that linguists will be disappointed after reading it, only that their concern tends to be more in line with higher numbers, larger corpora, or issues such as tagging, categorizing, and the like, that are not included in a book whose main goal is literature. Being a linguist prone to dealing with that aspect of texts (Marcos-Marin, 1994, 1996), I would have enjoyed some pages devoted to coding, SGML databases, and higher functions, but, at the same time, can be delighted when considering how many possibilities are open for scholars who use the ample panoply of public computer tools. It is not a despicable task: let us not forget that both Technoids and Fluffies (in Irizarry's terms) get lost in the approach to the text by means of the computer.

All this does not mean that there is no criticism to make. The most serious is that the author does not seem to be aware of the typological restrictions imposed by linguistic differences. Apparently, he considers irrelevant the differences among languages for stylistic parametrization. That is why she analyses Valle Inclán's use of punctuation according to the Chicago Manual of Style, or compares phrase structure in Octavio Paz and Rosario Castellanos with the English data and their distribution as described by Mary Hiatt (1975). This confusion repeats itself throughout the book and is undoubtedly clear on p. 47.

Although the point is under discussion, to put the limit of the random samples at 4,000, later 5,000 words, following Ellegård, is not convincing when one tries to deduce from it deep issues such as authorship, or gender differences in writing. It is clear that the author follows a scholarly line, but her conclusions may suffer if her methodological choices are not convincing.

Another drawback is that she compares authors of different periods, separated by more than 200 years, for syntactic purposes, which seems to be too disparate. In language history, linguists tend to consider data separated by more than 200 years as expressions of differences between periods, not just among stylistic choices.

A minor objection is that, due to the fact that the book is composed of papers published independently, there is a reiteration of concepts, expressions, and references, aggravated by the choice of partial bibliographies, where titles are often the same, instead of a common bibliography at the end of the volume. She is also unfamiliar with terminology in Spanish; for instance, she is not aware that the common term, in writing, for both mathematicians and Greek scholars, in Spanish is chi for the letter, and the chi square or 'chi dos' in Spanish, and not ji (albeit pronounced with
the velar spirant). Greek and Latin loan-words ending in atomic vowel + s in Sp. sg. are kept unchanged in the plural, unlike in English, and there is no tradition of reproducing Greek or Latin plural forms, so the plural of corpus (or, respectively, fucus) in Spanish is corpus, not corpora. There is a Spanish translation for tile in Windows (p. 173) it should be mosaico. Missprints are common (‘la mujeres’, p. 42), some of them ugly, such as on page 19, precisely at the beginning of the chapter ‘El uso la computadora’ instead of ‘El uso de la’. Echoes of English are found everywhere, ‘relacionadas a’ instead of ‘relacionadas con’, p. 5. ‘Santa Teresa’ is listed as Teresa, Saint on p. 67. Ian Lancashire appears as Ian Lancaster (p. 166) in the cited books page of that chapter. All of them are minor, but cumulative.

The list of programs used is very illustrative of the power of simple tools: CONSTAT, DYNAKWIC, LEXI, LitStats, Lotus 1-2-3, MCVO, MicroConcord, MicroOCP, Minitab, MTAS, WordCruncher, WordPerfect. All of them are pre-Windows’95 and pre-TACT programs, either MS-DOS or 16bits software. MTAS, for instance, is one of the programs that originated TACT. Scholars in the field of Spanish will be interested in knowing that our Laboratorio de Lingüística Informática at the Universidad Autónoma de Madrid (http://www.lifl.uam.es/) has developed a Spanish TACTweb server working under Windows NT, with textual databases in Spanish, currently used internally for teaching purposes, but projected for the public domain in the near future. Thus, the Laboratorio continues the line of collaboration opened in 1990 between the ADMYTE team and the developers of TACT, including now not only Old Spanish texts, but also Modern and Latin-American.

Estelle Irizarry, with very limited computer tools, achieved a remarkable success, methodologically speaking. Her path has gone along with that of history, and her basic thesis has been demonstrated by the evolution of literary criticism. However, not only for historical reasons does the book deserve to be read, it shows something more important than the adequate use of computers, it shows how a good critic can take advantage of those powerful tools and improve her already impressive scholarly performance.

Francisco A. Marcos-Marín
(Universidad Autónoma de Madrid)

References

DYNAKWIC, CONSTAT, MCVO Louis and Robert Ule, 27 Mustang Road, Rolling Hills, CA 90274.


Given the current popularity of the use of text corpora in both linguistics and language technology, and given the fact that corpora become much more useful when annotated, the time was ripe for a book which describes the full range of corpus annotation practice. Well, Garside et al. are the first to produce such a book. I expect, however, that it will not be the last, for two reasons. First of all, corpus annotation is still so much in flux that regular updates of any general overview are unavoidable. Also—and this really my only major criticism of this book—the Lancaster viewpoint is rather too pervasive. The editors themselves indeed note, both on the back flap and in the preface, that the focus of the book is on UCREL (University Centre for Computer Research on Language) work and that ‘international developments’ are merely ‘surveyed’; however, after reading the book, I feel this focus is so strong that the word ‘Lancaster’ should have been placed in the subtitle.

The book is divided into three parts, which can be classified as being about (i) the nature of linguistic annotation, (ii) software for annotation, and (iii) other aspects of annotation. The first part is really the only one which deserves to be characterized as a general overview of the field. After a general introduction to corpus annotation (Chapter 1), it goes on to describe the individual annotation types, i.e. morpho-syntactic (grammatical tags; Chapter 2), syntactic (Chapter 3), semantic (Chapter 4), discoursal (Chapter 5), prosodic, pragmatic, and stylistic (these last three sharing Chapter 6). In particular, Chapters 1–4 and 6 are an excellent representation of existing practice. Admittedly, and possibly unavoidably, some details are wrong. For example, Table 2.1 lists the TOSCA tagset as the smallest one with thirty-two tags and the TOSCA/ICE tagset as the largest with 270 tags, where in fact the two tagsets are extremely close (270 being the total number of tags and thirty-two the number of major wordclasses). Even so, the overall information is generally correct. Chapter 5, however, is of a completely different nature. It consists mostly of an extremely detailed account of the UCREL scheme for discourse annotation. This could well have been reduced and integrated into Chapter 6, moving any