



**New Media in the Humanities: Research and
Applications**
**Proceedings of the first seminar on
Computers, Literature and Philology, Edinburgh, 9-11
September 1998**

Edited by Lou Burnard, Domenico Fiormonte, and Jon Usher

Foreword

The process of globalisation, now under way thanks to the wider use of computer science, has increasingly influenced the field of literature, not so much in literary creation itself as in the dissemination of literature throughout the world. As Pierre Levy has recently pointed out in his writings, the Internet has altered space and time; the literary dimension has been transformed by the web, and we are building a humanity which seems closer to itself.

The publication of the Proceedings of the first conference on *Computers, Literature and Philology* undoubtedly constitutes a significant stage in understanding the relationship between literary disciplines and new technologies. This first international event linking Italian Studies with European trends in Humanities Computing establishes stable relationships between Italian research centres and prestigious institutions abroad.

We hear more and more often of 'Information Technologies' but little is known about the role humanists, philosophers and writers play in the planning of the new information designs. The essays in this volume prove how substantial the contribution of these 'knowledge managers' are to the theoretical elaboration and implementation of tools and resources.

This link between 'tools' and forms of knowledge brings us back to another revolutionary age in the field of communication: the Renaissance. The great Humanists, by creating and disseminating a new cultural model based on new instruments (for example libraries), rediscovered the indissoluble bond between support theories and content.

Today, we find ourselves in a similar situation. Even in the daily use of the cellular telephone (consider the problems of the delivery of documents in the applications of the future, like WAP and UMTS) there is a backing operation of 'restructuring of knowledge'.

Men of letters cannot, and must not, stay marginalized. Their place is in the centre of these revolutionary restructuring processes.

In conclusion to this Foreword, I would like to remind you that the fourth conference of CLP will take place in December 2001 in Germany, at the University Gerhard-Mercator, Duisburg.

In thanking the organizers and participants of CLP, may I express the hope that this series of events will be continued, and that these conferences will help reinforce the influence of the literary and humanistic tradition also within the public institutions.

Dante Marianacci (Director, Italian Cultural Institute, Edinburgh)

Contents

| | |
|---|-----------|
| 1 Poem and algorithm: humanities computing in the life and place of the mind | ?? |
| <i>Willard L. McCarty</i> | |
| 2 Where is electronic philology going? The present and future of a discipline. | ?? |
| <i>Francisco A. Marcos Marín</i> | |
| 3 Literal transcription — Can the text ontologist Help? | ?? |
| <i>Allen Renear</i> | |
| 4 On the hermeneutic implications of text encoding | ?? |
| <i>Lou Burnard</i> | |
| 5 Text Encoding as a Theoretical Language for Text Analysis | ?? |
| <i>Fabio Ciotti</i> | |
| 6 ‘Reports of my death have been greatly exaggerated’ Scholarly editing in the digital age | ?? |
| <i>Claire Warwick</i> | |
| 7 Hypertext as a critical discourse: from representation to pragmeme | ?? |
| <i>Federico Pellizzi</i> | |
| 8 Language resources: the current situation and opportunities for coo-operation between Computational Linguistics and Humanities Computing | ?? |
| <i>Antonio Zampolli</i> | |
| 9 Teaching Romance Linguistics with Online French, Italian and Spanish Corpora | ?? |
| <i>Elisabeth Burr</i> | |
| 10 Researching and teaching Italian literature in the digital era: the CRILet project | ?? |
| <i>Giuseppe Gigliozi</i> | |
| 11 Sounds and their structure in Italian narrative poetry | ?? |
| <i>David Robey</i> | |
| 12 Per una edizione informatica dei <i>Mottetti</i> di Eugenio Montale:varianti e analisi statistica. | ?? |
| <i>Massimo Guerrieri</i> | |
| 13 Exploring the Literary Web: The Digital Variants Browser | ?? |
| <i>Staffan Björk, Lars Erik Holmquist</i> | |

14 The Postmodern Web: An Experimental Setting

??

Licia Calvi

Acknowledgements

Finally, after more than two years, thanks to Oxford University's Humanities Computing Unit, the proceedings of the first seminar on *Computers, Literature and Philology* are published, shortly after the third such seminar was celebrated at the University of Alicante, 16-17 October 2000. Therefore, with some pride, we might say today that the intellectual gamble originally presented by CLiP has been a successful one. But apart from the scientific intuitions, the style chosen also proved effective (few papers and a great deal of discussion) as did the atmosphere, that the delegates were able to create in three days of open, sometimes passionate, debate. In Edinburgh, numerous schools and cultures (Italy, Germany, the United Kingdom, Spain, the United States, Sweden) worked together, and that which emerged, other than vitality, was a clear interdisciplinary and intercultural vocation: Humanities Computing.

The essays we present in this volume recall those days and go further. The long work of gathering and editing meant that most of the authors could take up their own papers again. Thus, the volume emerges as an updated contribution to the HC field.

Many people have contributed to both the success of the seminar and the collection and translation of these essays. Anna Middleton was a matchless organiser in terms of her efficiency, kindness and style. The help of Thomas Kerridge, Colin Swift, and Sally Owens was important in the revision of all the articles of the non-native speakers. It goes without saying that responsibility for any remaining errors are entirely the responsibility of the editors. Prof. Giuseppe Gigliozzi of the Department of Linguistic and Literary Studies at the University of Rome "La Sapienza", with his department and the project **Testi Italiani in Linea**, made a substantial contribution to the publication of the acts. Finally, particular thanks go to Prof. Dante Marianacci, director of the Italian Institute of Culture in Edinburgh, who never ceased to provide the initiative with his support and financial help. Without his personal commitment this volume would never have been published.

Introduction: Where Lachmann and Von Neumann meet

Edinburgh was an appropriate host for a conference examining the impact of the new technologies on old forms of communication, for it was here in this city that Napier devised his 'logarithmic bones' (*rabdologia*) to revolutionise the mechanical application of mathematics in a great conceptual leap forwards from the abacus. Such complex interplay between the material medium of record and the meta-understanding of the uses it is put to (in other words cultural technology) is now recognised as an exciting branch, not just of history, but also of the cartography of a rapidly evolving future. This alliance of the very new, indeed hardly invented, and the re-examination of a tradition-encrusted past is a powerful intellectual stimulus to fresh conceptualisations and daring transfers across disciplines. Recent work on the textual transmission of Chaucer, for instance, have made use of software solutions devised to keep track of serial modifications in DNA. The freedom of movement between disciplines, between media, and between stratifications of culture is both beneficial and disturbing: beneficial because of the creative ferment, but confusing because it calls into question both the status of established demarcations of disciplines and begs the question of what kind of common framework this pioneering community shares. In other words, what constitutes a definition for humanities computing (though the more radical, technology-driven question might be: what constitutes the humanities)?

The words we use today hide in their etymologies unsuspected early technologies: both the English ‘book’ and the Latin ‘liber’ refer to the practice of writing on tree bark. ‘Writing’, like Latin *scribere* and Greek *graphèin*, originally meant to score or incise. Text is something woven together. To read, in English, was to guess from mysterious signs, as in the closely related word a ‘riddle’. The connections between technology, literature and magic should not be overlooked. All literature is to some extent magic, but sometimes technology can increase its spell. One of the most ambitious attempts was that of Giulio Camillo, whose *Idea del teatro* a miniature wooden amphitheatre filled with drawers full of phrases arranged analogically, attempted to provide a mechanical generator of unquenchable, always appropriate, purportedly Ciceronian eloquence, powered by astrological conjunctions and financed, appropriately, by the king of France, François premier. Despite great claims made on its behalf, and massive subsidies which would make some of today’s humanities computing pioneers green with envy, the device was lost before being able to prove its worth. Humorous Ariosto and serious Erasmus both caught a whiff of its passing. Its true successor was the industrial-scale literature machine laboriously hand-cranked by the inhabitants of Laputa Jonathan Swift’s *Gulliver’s Travels*, an unconscious mechanical realisation of the great combinatory wheels of the medieval Catalan thinker Ramón Lull.

Camillo and Swift were part of a long line of innovators, real and imaginary. Changes in such material ‘technology’ can of course have substantial intellectual consequences. The move from clay writing tablet to papyrus not only meant that records could for the first time be transported through geographical space as well as time, it also provoked the evolution of new forms and rhythms in the act of writing itself; forms to which we are still indebted today. The subsequent move from the scroll (with its continuous ribbon of papyrus) to the book (with its discrete pages) meant that text could now be accessed out of sequence - the first crucial step, material but also mental - towards hypertext. The advent of printing, too, was not merely a matter of economic progress, making books available to a wider audience; it was also a spur to standardisation - of orthography; of syntax and vocabulary - but also, crucially, of texts. It was no accident that many of the early printers such as Aldo Manuzio were also textual scholars. Much of the current debate about norms for markup languages could be effortlessly translated back to the concerns of these early printers, editors and linguistic legislators.

But such history also has to recognise countervailing tendencies. When printing became available as a cutting edge technology during the renaissance, the typesets first adopted respectfully aped the manuscript tradition. Similarly, priority was given to printing older texts: the first printed book in Italian was an edition of Petrarch, who had been dead a century, rather than of a contemporary like Poliziano. Today, with the exponential development of the computer we live with exactly the same ambiguities. New technologies cohabit paradoxically with old expectations; traditional disciplines such as classics and patristics have embraced the power of the database with enthusiasm, whilst deconstruction and post- post-modernism stick largely with the biro.

The terms *Literature and philology*, in the title of this collection, imply fundamentally different attitudes to text, to production and reception, to aesthetic and material structures of writing, to the subjectivity and objectivity of the observer. The papers cover the gamut of current activity from the study of philosophical implications, through the technical aspects of encoding, the practical difficulties of applying standardised procedures to individual projects, to the crucial organisation of informatics communities in the humanities, and the use of computers in the teaching of literature. The range of topics may seem dispersive, but that is what the new technology beckons us to do - to explore, to question and to re-negotiate. This conference was just the first phase in a ‘stock-taking’ exercise in a rapidly

developing field. The paper by Willard McCarty does just that, both backwards (a concise, and hard-hitting history of humanities computing), and forwards (the description of an ambitious programme of exploration into the challenging throwaway onomastics of Ovid: at the limits of 'fuzziness' computer analysis can be programmed to manage). Inevitably, further exercises in stocktaking will have to take place, and with ever-increasing frequency. The era of the isolated solution is coming to an end, and we are witnessing a gradual emergence of a canon of questions, if not answers, the tentative mapping (and creative overlapping) of a constellation of disciplines.

With the offer of ever more powerful and capable devices and exponentially expanding corpora of texts, how do we know what issues are best addressed, and what common ground and methodology to maintain with others working in the field. And, crucially, what relationship to advertise with the mass of work, both past and present, performed without the aid of the computer. Where do Lachmann and Von Neumann meet? The essential models are the simile and the metaphor: at one end of the spectrum, substantial resources and great ingenuity are being invested in trying to make the 'virtual' presentation as much like the original as possible. This is, in effect, the 'Philology' part of the title. Standards of accuracy and their representation and transmission through common platforms are not as easy as they sound, as ultimately we have to make a decision as to whether objects are unique or whether there can be thresholds of taxonomic equivalence. Comparisons with the world of print are not entirely helpful as, even if the end-user is still human, there are intermediate stages of non-human operation and processing where rigorous and logically reliable norms are essential. The other approach (not in opposition, but frequently complementary to the above), is to regard the electronic capabilities of text processing as essentially metaphoric; in other words their primary function is the generation and presentation of two or more notionally unrelated objects linked via a third, given but not necessarily declared term. This is where the 'Literature' of the title comes in. Literary texts are essentially dissimulatory, but the strategies of dissimulation are organised and are capable of being analysed because of their regularity. At the literary-metaphoric end of the spectrum, humanities computing is discussing text twice removed, whereas at the philological-simile end the text itself and its ontological status is of primary importance.

This theoretical division, though useful, is in practice very hard to maintain. Allen Renear (??) suggests caution when talking about the entity of the text in the humanities computing environment, and outlines a series of questions about the theory of text and its relevance (or indeed centrality) to the task of transcription. Lou Burnard (??) rightly points out that encoding, even of the most apparently objective kind, is not value-free; the hierarchy of what to encode implies that hermeneutic activity has taken place. But within such a pecking order there are major issues, and Fabio Ciotti's observations (??) describe this situation well. If the *physical* subdivisions of a literary text respond well to encoding, what about the *literary* structures. To what extent, and with what status, can cultural and interpretative features share the same platform with objective textual phenomena? The problem is not new, of course; it has been around as long as criticism itself. But it is a major conceptual hurdle if we are to take full advantage of the pattern-seeking capabilities of digital record analysis. Answers, if there are any out there waiting to be found, will probably come from interdisciplinary groups working on the gamut of issues ranging from encoding itself, through textual analysis, and into the practical applications both in media presentation and in criticism. An important part of the proceedings was taken up with reports of such groups, each of which has something to tell us about the evolving environment in which humanities computing has to find and maintain its place. Giuseppe Gigliozzi's account (??) of the CRILET initiative should be read as a dynamic exploration of the difficulties and rewards of

forming, maintaining and in particular guiding just such an interdisciplinary group. One of the key challenges, which needs addressing urgently, is how to train (and particularly train for employment) the new, post-individual pioneer generation, and CRILet is potentially a model for such a centre; though the institutional framework in all countries has a long way to go to catch up with reality. As to the typology of such activities, and their rightful if still disputed place within the intellectual role of academe, Francisco Marcos Marín (??) suggests, from his experience in initiating and running large-scale international projects, a set of model definitions which has the great advantage of translating, with exemplary clarity, the language of the enthusiastic and ‘technologically converted’ into the everyday discourse of institutions. It is a potentially powerful weapon to be used when trying to squeeze humanities computing into reluctant faculties.

A pioneer of large-scale projects, where persuasive abilities of a high order have been necessary both for collaboration and finance, has been Antonio Zampolli (??). His account of the interaction between corpus and computational linguistics, and the strategic and commercial spin-offs which are beginning to result from it, is both an exemplary story and an object lesson in the potential for humanities computing to make a contribution to society.

Humanities computing is not just large-scale, however important strategically and culturally the creation of corpora and virtual libraries are. Behind such universalising programmes of cataloguing and publishing lies the individual textual scholarship of those producing critical editions. The hierarchical display abilities of computers make the fruits of such work much more user-friendly. At the diplomatic end of the spectrum, Staffan Björk and Erik Holmquist (??) explain how techniques of ‘flip-zooming’ elaborated for navigating amongst hierarchies of graphic objects (such as pictures and details of pictures in art galleries) could be potently adapted for use in the graphic display of complicated manuscript traditions or specific features, combined with purely text-format screens. But even at the critical end of the spectrum, the potential is enormous. Massimo Guerrieri (??) demonstrated conclusively that an electronic edition of Montale, a poet with a propensity for multiple variants, can demonstrate facets of a writer’s compositional process which would be well-nigh impossible to present by traditional means. Such excitement comes at a price, and Claire Warwick warns (??) that we have to be objective about the merits and demerits of both the old and the new philology, and that a strategy of persuasion needs to be put in place if the new methods are to gain widespread acceptance in the wider academic community.

When it comes to the application of computing to the ‘metaphoric’ end of the spectrum, the examination of the literary construction and interpretation of texts, the range of approaches is equally wide. Federico Pellizzi (??) points out that the very conceptualisation of hypertext as a model for digital presentation has consequences for the reading and even the writing of traditional text, and that there is a risk of circularity of argument. To escape from such an impasse, he suggests the notion of a ‘pragmasphere’ in which the architecture and discourse of works co-exist and co-operate at the same hierarchical level. David Robey (), faced with a history of essentially subjective, impressionistic and even opportunistic accounts of the relationship between prosody and meaning in the *Divine Comedy*, offers the model of a statistically inclusive survey of Dante’s way with hendecasyllables; once this primary survey is established, the secondary task is still interpretative, however, but at least the critic now has the advantage of basing judgements on verifiable, quantitative data.

As applications move towards the interpretative, the challenge both to systems and to critics becomes ever greater; Teaching, with its high degree of pre-determined outcome, is indeed one of the most immediately promising avenues. Licia Calvi (??) describes the creative tensions involved in designing a web-environment for teaching purposes, where the freedom of the individual user, and the guiding input of teachers, has to be brought to an

ideal but difficult synergy. Finally, Elisabeth Burr (??) takes an apparently less interventionist line, describing the construction of a teaching database of linguistic material from the Romance languages which allows the students to construe and model in their own way. In such an approach, the process of modelling and consequent experimentation with manipulating and interpreting data is as valuable a learning experience, in intellectual terms, as the specific results of individual projects.

We have to conclude this introduction by mentioning CLiP's next challenge: the project for a European MA in Humanities Computing, which will be discussed in Duisburg at the next edition of the seminar (December 2001). In Europe, according to surveys carried out by Microsoft and Cisco, there is a skills shortage of 600,000 people in the ITs area. Of these 600,000 there are not only technicians, programmers or engineers but also editors, journalists, e-literacy teachers, digital archives managers, experts in Web research, Net Semiotologists, etc.

Teaching these skills is the task of the Humanities rather than Science faculties. The creation of a European curriculum of Humanities Computing is the first step towards the education of that new type of humanist which not only the market but **above all** human sciences need. It is to be stressed here that we are not flirting with the new economy or injecting massive doses of technology into the faculty. Perhaps these things are needed to attract students, but what is at stake is something greater: the future set-up of the educational system of the Old Continent.

The humanities faculties have to understand that digital technologies constitute the greatest opportunity we have had to redraw the map of knowledge since the times of Erasmus and Leon Battista Alberti. As in those days, the indissoluble ensemble of theories and 'techniques' has a central role — think of the redefinition of the standards, **methods** and **vehicles** of teaching. Like the gurus of Web usability, people like Piccolomini or Manuzio struggled for 'clarity' in writing and printing, so that they could reach a wider public. Was not this the ideal of the first pioneers of the Internet?

Jonathan Usher and Domenico Fiormonte
University of Edinburgh

Teaching Romance Linguistics with Online French, Italian and Spanish Corpora

Elisabeth Burr

Gerhard-Mercator-Universität, Duisburg

Was allein geeignet ist, als Leitstern, durch das ganze Labyrinth der Sprachkunde hindurchzuführen, findet auch hier Anwendung. Die Sprache liegt nur in der verbundenen Rede, Grammatik und Wörterbuch sind kaum ihrem toten Gerippe vergleichbar. Die bloße Vergleichung selbst dürftiger und nicht durchaus zweckmäßig gewählter Sprachproben lehrt daher viel besser den Totaleindruck des Charakters einer Sprache auffassen, als das gewöhnliche Studium der grammatischen Hilfsmittel. [...] Freilich führt dies in eine mühevollen, oft ins Kleinliche gehende Elementaruntersuchung, es sind aber lauter in sich kleinliche Einzelheiten, auf welchen der Totaleindruck der Sprachen beruht, und nichts ist mit dem Studium derselben so unverträglich, als bloss in ihnen das Grosse, Geistige, Vorherrschende aufsuchen zu wollen. (Humboldt 1827-1829/1963, 186 & 200).

1. Introduction

Since 1990 a computer corpus-based course for students of Romance linguistics has been gradually established at Duisburg University. When the course was started, there was only an Italian newspaper corpus, but now there are newspaper corpora for each of the three languages: French, Italian and Spanish. All of these have been enriched with tags which represent the structural features of newspapers and text. To date, the tagging is in the COCOA-Format. These corpora can be used to study grammatical and stylistic features in general or the variation present inside each corpus. Furthermore, as the newspapers in the different corpora were published at the same time, synchronic comparative studies of the three languages can be undertaken. The Italian corpus, being composed of two subcorpora commenced at different times, can also be studied under a diachronic perspective. But why should corpora be used to teach Romance linguistics?

2. Romance linguistics

2.1. Traditional synchronic Romance linguistics

Linguistics is the scientific study of language. Thus, Romance linguistics, if we just concentrate on the synchronic perspective, studies Romance languages in a systematic way. Up to now, Romance linguistics has been, however, very heavily theory-driven. Language is, in fact, looked upon from an exclusively theoretical point of view, so to speak from above, and is forced into models constructed a priori. These models themselves are based on a very small number of isolated phrases, which in the majority of cases have been invented. Indeed, invention is very often considered to be the only effective way to reach a comprehensive description of language.

Such is, for example, the concept behind one of the most thorough studies of the Italian verbal system. Bertinetto, in fact, invented himself most of the examples which appear in his work. In his view, it would have been an act of pure optimism to rely on naturally occurring speech in order to find the most adequate example for every variation of meaning, given the marginality of some uses ¹:

¹Geoffrey Sampson, however, is of quite the opposite opinion: "If the linguist relies on data invented by himself [sic] in his role of native speaker of the language [...], then it is near-inevitable that the linguist will focus on a limited range of phenomena which the research community has picked out as posing interesting problems, while overlooking many other phenomena that happen never to have struck anyone as noteworthy." (Sampson 1993, 268).

La maggior parte degli enunciati illustrativi contenuti in questo lavoro sono frutto di libera invenzione; [...] non nego affatto che l'uso concreto offra formulazioni più ricche e più duttili di quelle che in genere sono partorite dalla mente di chi se le costruisce da sé secondo le esigenze del proprio argomentare; ma è indubbio che la speranza di trovare già bell'e confezionato l'esempio più adatto per ogni minima sfumatura di senso [...], appare un tantino ottimistica, data la marginalità di certi impieghi. (Bertinetto 1986, 13).

He justifies such a procedure with Chomsky's concept of the native speaker, who as Bertinetto says is capable of producing an unlimited number of phrases in their own language by drawing on the registers and styles which a certain type of education has put at their disposal: "Nessuno mette in dubbio che il parlante nativo sia in grado di produrre infiniti enunciati nella propria lingua, sfruttando tutti quei registri e quegli stili discorsivi che il suo livello d'istruzione gli consente di padroneggiare." (Bertinetto 1986, 13).

Thus using naturally occurring language data would mean restricting one's own native speaker competence to the stylistic registers present in a corpus. Such a restriction, can, however, according to Bertinetto, only be justified in cases where languages are studied by linguists who are not native speakers: "[...] questa autolimitazione mi pare assennata solo nel caso di studiosi che si propongano di descrivere una lingua straniera, non certo nel caso di chi possieda in proprio la 'competenza' del parlante nativo." (Bertinetto 1986, 14).

A similar principle governs teaching. Students are, in fact, trained to look at models developed on the basis of de-contextualised and invented language 'data' and then to apply these models to equally invented isolated phrases. Hence, Michael Stubbs' evaluation of modern synchronic linguistics applies, likewise, to Romance linguistics in more specific terms: "[...] it is so easy to be blind to the very small amount of data on which contemporary linguistics is based, and this is fundamental to the whole intellectual organisation of the discipline: the theories which linguists develop, the types of corroboration they claim, the methods they use, and the ways in which students are trained." (Stubbs 1993, 10).

Furthermore, if the language in question itself is studied at all, analysis is carried out on the basis of very small amounts of spoken or written text or, and this is much more often the case, on the basis of dictionaries and grammars. Dictionaries and grammars are themselves, however, static abstractions of the language in question and present linguistic elements in an atomistic way. Speech, i.e. the only reality of language at our disposal is represented there either via quotations, mostly from literary works, or via artificial examples which have been constructed in such a way that they illustrate precisely the rules in question, or as Bertinetto admits in the quote given above, "secondo le esigenze del proprio argomentare" (Bertinetto 1986, 13).²

2.2. Data-driven synchronic Romance linguistics

According to Wilhelm von Humboldt, though, whom I have quoted extensively above, language as such can only be studied "in der verbundenen Rede", i.e. in naturally occurring speech, compared to which dictionaries and grammars are, at best, a dead skeleton. Even a comparative study, he says, of unsubstantial and not necessarily well chosen samples of speech allow for more insight into the overall character of a language than the traditional study of tools like dictionaries and grammars. Charles J. Fillmore puts it in similar terms: "every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way." (Fillmore 1992, 35).

It is precisely this insight into the characteristics of languages like French, Italian and Spanish which students are supposed to gain in courses devoted to synchronic Romance linguistics. Gaining insight does not mean, however, that the study of theory, models,

²Wallace Chafe, occupying himself with the methods used when studying language, comes to a similar conclusion: "The techniques that have most dominated the field of linguistics [...] have been techniques focused on artificial rather than naturally occurring data." (Chafe 1992, 85).

dictionaries and grammars has to be excluded from courses devoted to it. On the contrary, they can and should be integrated and critically explored by comparing their assertions to actual findings.

Studying samples of naturally occurring speech leads, as Humboldt points out, to a “mühevoller, oft ins Kleinliche gehende Elementaruntersuchung”, to a tedious and cumbersome study of the single elements of language. But it is precisely these possibly mediocre details which Humboldt claims make up the overall character of the individual languages and, thus, nothing is more of an obstacle to the study of individual languages than trying to find in them only the spiritual, the dominant or the grand. It is obvious, I think, that the type of language study Humboldt is advocating here, would, today be called *data-driven* linguistics, using a term coined by John Sinclair and his school (see f. ex. Sinclair 1992).

At a first glance, then, the type of linguistics which constitutes the framework of part of my research and which I am trying to teach in some of my classes, could be described as data-driven Romance linguistics aimed at allowing students to gain insight into the complexity of naturally occurring contemporary French, Italian and Spanish and to examine critically their more traditional and theoretical descriptions, on the basis of their own findings.

3. The conception of language

Being data-driven is, however, not enough to define the type of linguistics in question. Instead, it has to be differentiated from the dominant type of linguistics, which in this century and up to now has not only been very much theory-driven, but has also been characterised by an essentially dualistic conception of language.

3.1. The dualistic conception of language

This dualism does not just concern the functional oppositions which are seen as binary oppositions, i.e. *homme / uomo / hombre / man* [+ masculine] - *femme / donna / mujer / woman* [- masculine] (cf. for example Renzi 1989, 318), but language itself. In fact, Saussure's dichotomy *langue* and *parole* is still today the starting point for most approaches. Its endurance is, in a way, sustained by Chomsky's *competence - performance* differentiation, which likewise remains central to modern linguistics. The two entities *langue* and *parole* or *competence* and *performance*, are, however, not of equal importance for modern main-stream linguistics. On the contrary, whereas *competence* or *langue* is really the centre of attraction around which gravitate the linguistic models which have been created and the discussions carried out about them, *performance* or *parole* is looked upon as a secondary manifestation of language. For this reason elaboration of a theory of speech is not even seriously envisaged. This means that the situation in linguistics has not really changed since the time of Dell Hymes and that his characterisation of modern linguistics is still valid today: “A major characteristic of modern linguistics has been that it takes structure as primary end in itself, and tends to depreciate use” (Hymes 1972, 272).

It is true that since the beginning of the 'eighties, above all in Britain and in northern Europe a new type of linguistics has evolved and gradually established itself. This type of linguistics, now generally called corpus linguistics or even computer corpus linguistics, tries to turn *performance* into its starting point and to describe, in a systematic way, languages on the basis of natural speech. This does not necessarily mean, however, as the following remark from Jan Svartvik shows, that it has overcome the *competence - performance* dualism: “Linguistic competence and performance are too complex to be adequately described by introspection and elicitation alone.” (Svartvik 1992, 8). Instead, it aims, more than anything else, at the mere substitution of introspection and elicitation by data or their integration with

data. The reason for this might be that although speech is considered to be the only reality at our disposal, this type of linguistics has not yet, as far as I can see, developed its own coherent theory of speaking. Because of this, it seems to be constrained either to defend itself against the assertions of the theory-driven type of linguistics and its fundamental opposition to the analysis of speech, or to fill the binary distinction with new content but without ever questioning the dualistic way of looking at language. Halliday's concept of *system* and *instance* (cf. Halliday 1992, 66) seems to me a good example of this tendency.

3.2. Social implications of the dualistic conception of language

The dualistic conception constitutes in my eyes a serious problem, however. First of all, an activity like speaking is much too complex to be studied in binary terms, i.e. as the realisation of systematic possibilities or relations.³ Secondly, this view has far reaching implications with respect to the type of speech which is studied at all and thus for corpus linguistics as well. In fact, as the primary sources of corpus linguistics are not just any portions of speech but corpora constructed on the basis of a particular conception of language, conceiving language in dualistic terms and thus basing corpus construction on the *core* and *periphery* distinction⁴ will lead to the exclusion of vital parts of natural speech and, what is more important, of whole categories of speakers.⁵

Obviously, such an exclusion has always been the case. Over the centuries the dominant view of what is a language or of what it means to know a language, has in fact worked against a whole range of expressions, language varieties and strata of the population. But there have been historical events which more than others sanctioned such an exclusion. Just think of what happened to the Romance languages and others when printing established itself. Heavy pressure towards normalisation was brought to bear on the language in favour of the economic success of the new technology, which, although fostering progress in cultural and economic terms, led to a considerable reduction of possibilities of expression. The purist dictionaries which were edited in the 17th and 18th centuries by the famous academies *Accademia della Crusca*, *Académie Française* and *Real Academia Española* (cf. Nencioni 1990, 345-346) are just one outcome of this process. In our century, the background against which in Italy the *Dieci tesi di una educazione linguistica democratica* were elaborated (cf. De Mauro 1975, 1981, 138-151) is a good example, instead, of what happens when on the official level of education only one type of language is allowed to exist and the people who do not know it are pushed aside.

If these sanctions could take place without the technology which we have at our disposal, we have to ask ourselves what will happen in times when the power of deciding what is possible or correct is no longer in the hands of lexicographers or printers who in many cases were at the same time literary scholars, but is handed over to a computer by the speakers themselves, or to be more precise, by those speakers who have access to the technology and its control. That the computer is really made into the arbiter of language use is shown by the following type of arguments you might come across nearly every day, above all in institutional

³See also John Sinclair's opinion on this: "Language is an abstract system; it is realized in text, which is a collection of instances. This is clearly an inadequate point of view, because we do not end up with anything like text by 'generating' word strings from grammars." (Sinclair 1991, 102).

⁴The concept of core was originally developed by Charles F. Hockett in order to account for the fact that people with different idiolects could nevertheless communicate together (cf. Hockett 1958, 332). Later on, however, it was extended to all the different varieties of a language: "however esoteric or remote a variety may be, it has running through it a set of grammatical and other characteristics that are common to all." (Quirk and Greenbaum 1975, 1).

⁵For a discussion of the core on which the BNC is built and the exclusions it entails see for example Atkins; Clear; Ostler (1992) and Clear (1992). According to Gerhard Leitner the concept of core is also at the basis of the International Corpus of English (cf. Leitner 1992, 37 & 44-46).

and administrative environments: “the computer can’t accept this, this is too long for the computer, it only accepts *il vient* and not *elle vient*, *professor* and not *professora*”.

Another important factor, today, is the so-called language industries. Leaving aside, at this point, the problems involved for languages which will either not be able to take part in the development in this field or are not yet taking part efficiently enough,⁶ I should point out, that there could be considerable danger involved for those languages, too, which are already at the centre of this process. In order to be more easily processable, these languages might, in fact, be reduced, in the long run, to a presumed *core*.

If we as linguists, given this background and the ongoing collaboration between branches of linguistics and the language industries, continue to maintain a view of language and of linguistic knowledge which does not correspond to its complexity or just takes the traditional hierarchies for granted, we, too, will bear responsibility, should an even more incisive ideological normalisation of language come about in the near future and if in teaching and wide areas of communication, language varieties differing from a supposed *core* are assigned to the *periphery* together with the people who use them.⁷

3.3. A non-dualistic theory of speech

What we need for our studies, teaching and corpus building, instead, is a theory of speech or, more precisely, a theory of what the speakers know when they speak, which takes into account precisely the complexity of language. Obviously, by speaking I do not just mean oral expression, but written expression as well. Such a theory is not at all new. On the contrary, it has already been proposed by Dell Hymes (1972) and by Eugenio Coseriu (1988b).⁸ Both linguists, in fact, base their considerations on language as it manifests itself in reality and not on an idealisation, and develop their theory in opposition to Chomsky’s dualistic conception of language and in opposition to the idea of a *common core*. In both theories, speaking is considered to be a very complex activity based on a whole range of different spheres of knowledge. Chomsky’s ideal speaker-listener who acts in a homogeneous speech community is confronted with the heterogeneous character of real speech communities and with the socio-culturally determined knowledge of languages and language varieties.

The starting point for both theories is, thus, the observation that in most speech communities people are not mono- but plurilingual, i.e. that they speak more than one historical language: “[...] in much of our world, the ideally fluent speaker-listener is multilingual.” (Hymes 1972, 274), or, at least, more than one variety of such a language: “Even an ideally fluent monolingual of course is master of functional varieties within the one language.” (Hymes 1972, 274). In similar terms Coseriu says that different norms and rules of different language systems are not only applied in natural speech of a speech community but also of individual persons (cf. Coseriu 1988b, 263).

If we want to take this reality into account, we have to, as Hymes says, “break with the tradition of thought which simply equates one language, one culture, and takes a set of functions for granted.” (Hymes 1972, 289).

According to Hymes, we have to realise, furthermore, that not even the different language varieties of a certain language can be related to a common grammar, let alone the different languages spoken in a speech community (cf. Hymes 1972, 275). This fact is perhaps best explained by using two concepts which play an important role in Coseriu’s theory and which

⁶See Francisco Marcos Marín (1994, esp. 49-62) for a discussion of this topic.

⁷According to Michael Barlow the concept of common core has, in fact, already had negative consequences not only for grammatical description but for the teaching of foreign languages, as well (cf. Barlow 1996, 4).

⁸Coseriu has proposed elements of such a theory of speech in many of his works. The book I am referring to, here, has the advantage, however, that these elements are drawn together and elaborated into a coherent theory of linguistic knowledge.

take languages into account in two different dimensions, i.e. in their heterogeneity and their homogeneity. It will be useful, furthermore, to call to mind the linguistic situation in countries like Italy, Germany or Spain, where dialects which are by no means mutually understandable are still alive.

3.4. The historical language

The concept which tries to cover the heterogeneous dimension of a language is the concept of the *historical language* and its external structure or architecture (cf. Coseriu 1988b, 24-25 & 139-148). A historical language constitutes, in fact, an autonomous combination of linguistic traditions. As such it is recognised not only by its own speakers but also by speakers of other historical languages, who identify it with adjectives like English, French, Italian, or Spanish. Regarding the linguistic traditions of which such a language is composed, we can distinguish three different types on the basis of diatopic, diastratic and diaphasic differentiation.

With respect to the diatopic differentiation a historical language like Italian is composed of primary dialects like Sicilian, Venetian or Tuscan and a historical language like Spanish of Asturian, Castilian, or Aragonese etc. They are primary in the sense that they existed as independent languages long before one of them, – in the case of Italian the Florentine variety of Tuscan, in the case of Spanish, Castilian, – developed into the communal language of the respective country. When the use of such a communal language is spreading all over the country and is thus becoming a sort of *lingua franca* for the speakers or a *roofing language* for its former co-languages, diatopical differences will normally lead to its own internal differentiation and bring about secondary dialectal traditions. In Italian, these secondary dialects are the so-called *italiani regionali* in Italy or Italian in Switzerland, the Spanish secondary dialects are Andalusian, Canarian or the different forms of Spanish in America etc., all of which are distinguished inside the communal language.⁹ Diatopical differences might, however, also be present inside the socio-cultural norm, which in a linguistic community is recognised as an exemplary model or standard. These tertiary types of dialects can be discerned, for example, when listening to the speeches Italian or Spanish politicians make in parliament, i.e. in a formal speech situation where the standard language is normally used¹⁰. The formal type of Italian in Switzerland or the Mexican, Peruvian, Chilean, Rioplatense etc. form of the Spanish standard are to be considered tertiary dialects, too¹¹.

By diastratically characterised traditions we refer, instead, to the languages spoken by the various socio-cultural classes or strata. These traditions can be called sociolects or *niveaus*. With respect to Italian we normally distinguish, in fact, between *italiano colto*, *italiano dell'uso medio* and *italiano popolare*, and with respect to Spanish between *español culto / habla culta* - *español medio / habla mediana* - *español vulgar / popular*. Such traditions can, by the way, also be observed inside the various types of dialects themselves. A sociolect or *niveau* is, thus, a variety which characterises a specific socio-cultural class. As such it can also be differentiated further in diatopical terms.

The last type of differentiation derives from the different communicative situations. Diaphasic varieties or styles can be distinguished within the various types of dialects and within the sociolects, as well. As the various social groups like football fans, women, men,

⁹Obviously, diatopical differentiation exists inside communal English, as well: In fact, British, American or Indian English are all to be considered secondary dialects of the historical language English.

¹⁰This does not mean, however, that tertiary dialects are characterised only by differences in pronunciation. There are usually also some differences in vocabulary and grammar.

¹¹Another good example for a diatopically differentiated standard language is German, where there are not just tertiary dialects in the standard spoken in Germany itself, but different standard languages in the German speaking countries, as well, with their own dictionaries and grammars. The same goes for English where different standard forms have at least developed in the different countries where English is spoken if not inside the individual countries themselves.

professional people and youths etc., are delimited by the respective point of observation the type of language they speak as a group falls into the category of style, too. When considered as autonomous entities, these varieties show, in their own right, diatopic and diastratic differentiation. With respect to Italy or Spain, this means that the language spoken, for example, inside the family is neither the same in all regions nor in all socio-cultural strata.

Hymes' point that the different language varieties cannot be related to a common grammar, becomes even clearer when we consider that dialects can themselves function as sociolects or styles. This is the case when a certain dialect is normally spoken by a specific socio-cultural class whereas the other socio-cultural classes speak the communal language. The Milanese upper classes speaking Milanese in order to distinguish themselves from the Italian-speaking lower classes may serve here as an example. Dialects function, instead, as styles, when in a certain communicative situation a certain dialect is normally used, whereas in another situation a different dialect or the communal or standard language is normally spoken. An example could be, that inside a family which originally comes from Aragón but lives now in Madrid, Aragonese is normally spoken whereas at work the Spanish communal language is used. It should be obvious, thus, that the varieties of the historical languages Italian or Spanish can by no means all be related to the same system ¹².

The functional substitution of certain varieties by other varieties, which is in question here, is synchronically, however, possible only in one direction. That is why Koch and Oesterreicher call it a *Varietätenkette*, that is a variety chain (Koch and Oesterreicher 1990, 14), which can be presented in the following way:

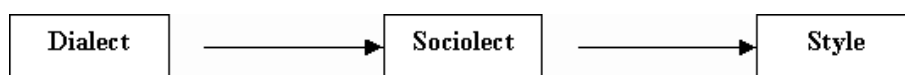


Figure 1. Variety chain (cf. for example Coseriu 1988b, 146)

This combination of three types of dialects, of niveaus or sociolects and of styles together with the possibilities offered by the variety chain take into account the complex heterogeneity of a historical language and constitute, in fact, its architecture or external structure, which can be visualised as in the following figure:

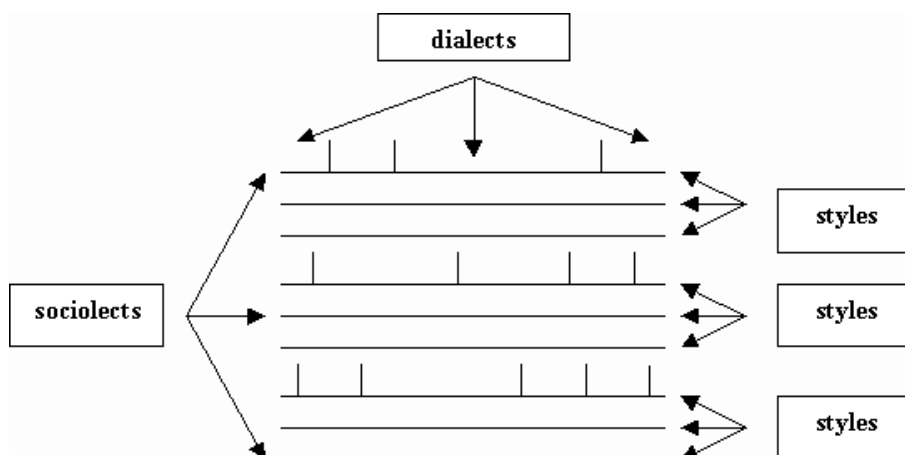


Figure 2. The architecture of a historical language (cf. Coseriu 1988a, 283)

¹²There are also speech communities where a different historical language functions as sociolect or style. In Italy and Spain the situation in Alto Adige or in the Basque country can be cited here.

3.5. The functional language

The concept which tries to cover the homogeneous dimension of a language, instead, is the concept of *functional language* with its internal structure (cf. Coseriu 1988b, 25-27 & 266-278). In this dimension we find those varieties which are to be considered synchronic, syntopic, synstratic and synphasic entities. In more concrete terms, what is envisaged is a certain dialect or the communal language spoken by a certain socio-cultural class in a specific communicative situation.

With respect to such an entity alone, we can talk about *system* and *norm*, i.e. the two structural levels distinguished by Coseriu within the ‘virtual technique’¹³. The *system* in this theory is understood as a very abstract entity which contains only the functional oppositions, the elements and procedures. The *norm*, instead, is a social entity which comprises the already traditional realisations. Whether these realisations are functional or not is in this case of no importance. Whereas the *system* is understood, above all, as a system of possibilities, which includes also the virtual side of a variety, the social norm is seen as mainly a system of traditional constraints which have developed through history, not least with the help of influential people or institutions. Through these constraints, the possibilities offered by the system are reduced to those traditionally already realised. It is this social norm, and not the abstract system, which is then realised in a concrete discourse:

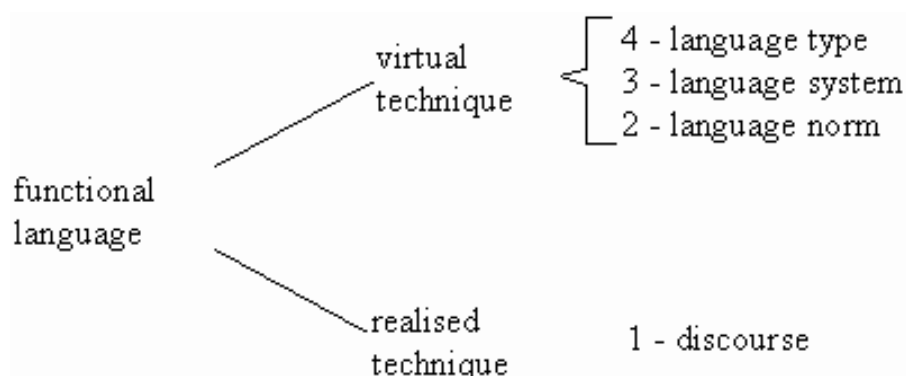


Figure 3. The structure of a functional language - virtual and realised technique (cf. Coseriu 1988a, 293)

3.6. The knowledge of the speakers

The speakers' linguistic knowledge of such a variety, however, does not correspond to the variety's synchrony. Instead, the speakers' linguistic knowledge is made up at any point in time of linguistic facts which like *thou* in English do not function in the variety's present synchrony but are functional entities only in a former state of the language. Other linguistic facts, on the contrary, will be valued by the speakers in a diachronic perspective, i.e. some might say, for example that *udire* is not used any more in Italian today, others might still consider *sentire* as an innovation. This diachronic consciousness determines the speakers' own attitude towards the linguistic phenomena and the application of their own linguistic knowledge (cf. Coseriu 1988b, 134-135). In this context it is not really important whether the judgments of the speakers correspond to the objective situation. Instead, the speakers are always right, because it is precisely their own opinion which influences their speaking (cf.

¹³As can be seen in the figure below, Coseriu distinguishes a third level, i.e. the language type. In our context, this level is, however, not of immediate importance.

Coseriu 1988b, 178). Such a diachronic consciousness exists, according to Coseriu, above all, in linguistic communities with a written literary tradition. In such a community, older forms can even be used deliberately, in order to refer to the respective state of the language or to the respective historical moment. Thus, the individual forms and the norms applicable to them remain present even when these forms no longer belong to the actual synchrony¹⁴. They can even be revitalised and will then be part of the actual synchrony again (cf. Coseriu 1988b, 135-137)¹⁵.

In addition, the speakers' linguistic knowledge does not comprise just one variety of the respective historical language. Instead, in the normal course of things all speakers know several diatopic, diastratic and diaphasic varieties in various ways and they use them, too. Furthermore, they have a rudimentary (correct or incorrect) knowledge of other varieties of their own language or of other historical languages. This knowledge might very well be confined to the so-called languages of imitation, i.e. the traditions which exist in speech communities with respect to the imitation of varieties of their own historical language or of foreign languages which are not spoken by the members of the speech community itself. Italian communities, for example, which do not speak Tuscan, imitate this dialect via an exaggerated use of the *gorgia*, in Spain, *seseo* and *yeísmo* can be used in the same way, and German is generally imitated by putting an *-en* on every single Italian or English word, which then leads to forms like *spaghetten*, *mangiaren* and so on¹⁶. As these examples show, the individual imitations do not necessarily have to correspond to the realisations which are traditional in the variety or historical language in question. On the contrary, such languages of imitation very often contain elements which do not exist in the variety or languages which are imitated or would even be impossible there (cf. Coseriu 1988b, 148-152). In German, for example, there exist expressions like *uno momento*, *picco bello*, *dalli dalli*, *alles paletti* which are supposed to be Italian, but are not Italian at all.

The speakers' knowledge however, does not correspond to the historical language either. This means that no speaker knows all the varieties a historical language is composed of and not all the speakers know the same varieties or do not know them in the same way. Furthermore, the languages or varieties speakers do know, are not put to use indiscriminately in all the situations of communication. Instead, the speakers possess, according to Hymes, a *differential competence*, which allows them to discern the communicative adequacy of each variety:

fluent members of communities often regard their languages, or functional varieties, as not identical in communicative adequacy. It is not only that one variety is obligatory or preferred for some uses, another for others (as is often the case, say, as between public occasions and personal relationships). Such intuitions reflect experience and self-evaluation as to what one can in fact do with a given variety.

(Hymes 1972, 274).

Apart from that, speakers possess also a *competence for production* and a *competence for reception* which are both socio-culturally determined, but do not coincide with each other in the majority of cases (cf. Hymes 1972, 275).

4. Teaching data-driven linguistics of speech with corpora

Parts of the theory I have tried to outline above and which is at the basis of my research,

¹⁴Thus, in England the language of the King James's Bible, of the Shakespearean plays, of Dickens' novels or at least some elements of these all belong to the actual knowledge of present day speakers.

¹⁵The example given by Coseriu of the Spanish Suffix *-i* might suffice here. This suffix, which had been used in older times to form adjectives for persons and facts related to the Orient, had been unproductive for a long time. It was revitalised, however, in our century for the formation of adjectives like *pakistaní* etc. (cf. Coseriu 1988b, 137).

¹⁶See, for a literary use of such imitation Shakespeare's *All's Well That Ends Well*, Act IV Scenes I & II.

teaching and corpus-building, have in the meantime become an integral part of Romance linguistics. These are above all the concept of the *historical language* and the distinction between *system* and *norm*. The reason for this is that for historical reasons and because of its object, Romance linguistics possesses, in many respects, superior insight into the complexity of historical languages than the linguistics of many other languages and is, thus, less inclined to reduce languages to a *core*. What is generally missing, however, is a systematic investigation into the knowledge of the speakers and an evaluation of the many theories and models which have been elaborated within variational linguistics by confronting them with the reality of speech. Thus, even today, as a recent article by Raffaele Simone (1996) shows, Italian linguistics, despite its very elaborate linguistics of varieties, does not yet really know what is spoken in Italy, when and by whom. Instead, it has to rely on the statistics of institutions like ISTAT or on subjective impressions. To solve this problem it seems necessary to me to combine the above theory of speech with corpus or data-driven linguistics as outlined for example in Sinclair (1991). In order to do this we need, however, corpora which respect the linguistic knowledge of speakers.

5. The corpus of Romance newspaper languages

Without entering into the discussion about reference or sublanguages corpora here ¹⁷, I shall now present the corpus I am currently using for teaching. The corpus is composed of whole editions of French, Italian and Spanish newspapers which were published around the time of the last European elections in 1994. This period was selected for two different reasons. One reason was that distribution of newspapers is expected to be much larger at the moment of an important event. The other reason was the aim to achieve a high level of correspondence between the newspapers in question with respect to the themes handled, in order to allow for comparative studies of certain styles ¹⁸. With the *European Elections* taking place everywhere in Europe at the same time, it could be expected that every newspaper in every country would report extensively about this event.

5.1. The composition of the corpus

At the moment, the corpus consists of the three subcorpora *Le Monde*, *Corriere della Sera* and *La Vanguardia*, each of which is composed of three editions of the same paper ¹⁹. The size of the subcorpora and of their individual components are given in the following table:

Table 1. Size of the subcorpora and their components

| | | |
|---------------------|---------|--------|
| Le Monde | tokens | types |
| total | 236.236 | 22.545 |
| 12./13.06.1994 | 89.786 | 13.449 |
| 14.06.1994 | 70.936 | 10.399 |
| 15.06.1994 | 75.514 | 11.328 |
| Corriere della Sera | tokens | types |

¹⁷Let me just say that neither corpus as such really meets the requirements of the theory of speech, the first type by excess and the second type by deficiency. However, a discussion of this has to be left for a later contribution.

¹⁸Theme is after all one of the parameters which delimit speech situations.

¹⁹Altogether, much more material has been collected which, however, in its present state cannot be called a corpus as yet.

| | | |
|---------------|---------|--------|
| total | 303.641 | 31.285 |
| 13.06.1994 | 102.976 | 16.400 |
| 14.06.1994 | 102.441 | 15.652 |
| 15.06.1994 | 98.224 | 16.498 |
| La Vanguardia | tokens | types |
| total | 261.133 | 31.330 |
| 13.06.1994 | 86.468 | 14.367 |
| 14.06.1994 | 94.251 | 17.844 |
| 15.06.1994 | 80.414 | 16.234 |

Although on account of its size and composition this corpus can by no means represent the whole knowledge of the speakers who make up the reading public of the individual papers, its components stand for a “complete experience, that permits distinction but excludes selection” which according to Giovanni Nencioni (1990, 351) has to be taken account of when constructing a corpus. That this is really the case can be justified in the following way. The individual editions of the papers are retained as they presented themselves to the readers the day of their publication. Thus all texts which appeared have been retained in their entirety and phenomena which would disturb the picture if the aim was to represent a *core*²⁰ have not been excluded, on the contrary they have been assumed to belong to the experience of the readers and are, therefore, part of their knowledge. Such phenomena are the elements of foreign languages and dialects present in probably all the papers and the Catalan elements in the Spanish paper as well as elements which might not function in the present synchrony. Furthermore, the components do not represent a homogeneous language variety but are, in fact, a portrait of the real combination of stylistic and socio-cultural varieties peculiar to the paper.

There are, however, a few limitations (still) with respect to the congruence between the printed editions and their counterparts in the corpus. The reason is that CD-ROM editions and thus the sources of the French and Italian component do not mirror exactly the printed editions²¹ and even from the magnetic tape *La Vanguardia* put at my disposal, every kind of publicity had for obvious reasons been excluded. In spite of these exclusions, the newspapers can nevertheless be understood as a quite reliable representation of what the actual readers are presented with and thus of the knowledge expected of them.

As not every reader will read all the paper, some might even just read one special section or type of article, this knowledge as such must be regarded as an abstraction, however. In order to achieve a more realistic notion, the different socio-cultural strata which make up the readership of the individual papers would have to be related using the demographic data at our disposal, to the various parts of the subcorpus²². Such a distinction is made possible by the markup of the corpus.

²⁰That to represent a core of language is also the aim of the LIP can be understood when considering the following remarks: ‘Abbiamo deciso di escludere dalla raccolta e dalla trascrizione tutti i documenti di parlato orientati in modo dominante verso il dialetto.’ (De Mauro 1993, 31) or “Il corpus del LIP non presenta dunque una varietà specifica dell’italiano, ma raccoglie testi di italiano tendenzialmente comune e unitario parlato in tutto il territorio nazionale.” (Voghera 1987, 34).

²¹For example, picture captions do not appear in both the CD-ROM editions, the weather report is missing on the CD of *Le Monde* and TV- and radio programs are not retained on the CD of *Il Corriere della Sera*.

²²When doing this, the BNC and the demographic parameters used when creating it could serve as a model. See for example Crowdy (1993).

5.2. The Markup System

In the header the following information about the corpus and its elaboration is retained:

```
<<|Spanish Newspaper Corpus "European Elections 1994" Component1>>
<<|Original text copyright La Vanguardia, Barcelona>>
<<|Text source magnetic tape, ATEX format>>
<<|Data extraction Gert-jan Burggraaf (NL)>>
<<|Translation ATEX format to DOS codepage 850 Gert-jan Burggraaf (NL)>>
<<|with support of Steward Whitelaw & Roger Tripper, Daily Telegraph,London>>
<<|Extraction & Translation Period 1995>>
<<|by extraction of this edition first letter of every line was lost>>
<<|has been corrected collating electronic with printed version>>
<<|Format MS-DOS Text>>
<<|not lemmatised>>
<<|preposition-article contraction like "del" to "de+el", "al" to a+el>>
<<|verb-enclitics contraction like "veder+la" to veder+la">>
<<|composed words with dash are retained as one form>>
<<|Apostrophe is followed by blank, i.e. "del' autovia">>
<<|Hyphenation, line breaks correspond to printed version>>
<<|Disambiguation: punctuation and digits, i.e. # = decimal point, \=decimalcomma>>
<<|Disambiguation: Quotes and minutes / seconds, i.e. &#x00A3; =minutes, $ =second>>
<<|Mistakes not corrected, but annotated {sic}>>
<<|Markup textual features>>
<<|Markup format COCOA>>
<<|Values of variables given in Italian>>
<<|Occhiello = CINTILLO, Titolo = TITOLO, Sottotitolo = SUBTITOLO>>
<<|Sommario = ENTRADILLA, Catenaccio = DESTACADO>>
<<|Markup of comments { } and <<>>>>
<<|Correction, Disambiguation, Markup Elisabeth Burr>>
<<|Elaboration Period 1995-1997>>
<<|Copyright Elisabeth Burr>>
<<|The June 13, 1994 issue of the newspaper La Vanguardia, Barcelona, Spain>>
```

Every component of the corpus has been enriched with the following series of tags in the COCOA-format, which allow for a whole range of differentiations in terms of readership or language varieties ²³

Table 2. The Markup System

| Reference | Code | Example |
|--------------------|------|-------------------|
| paper | <Z> | <Z La Vanguardia> |
| edition | <E> | <E 130694> |
| section | <S> | <S Politica> |
| origin of the text | <A> | |
| signed | | <A firmato> |
| anonymous | | <A non firmato> |
| name of author | <N> | <N Tapia Juan> |
| page | <C> | <C 01> |
| language | <L> | <L Inglese> |

²³At the moment a research project is under way which should allow the corpus to be enriched with tags for parts of speech.

| | |
|--------------------------------|-----------------|
| texttype | <T> |
| head-line | <T Occhiello> |
| slugline | <T Titolo> |
| sub-title | <T Sottotitolo> |
| abstract | <T Sommario> |
| in between title | <T Catenaccio> |
| announcement | <T Civetta> |
| article | <T Articolo> |
| front-page story | <T Spalla> |
| TV-, cinema program | <T Programma> |
| film content | <T Film> |
| commentary | <T Corsivo> |
| interview | <T Intervista> |
| column | <T Rubrica> |
| criticism | <T Critica> |
| stop press | <T Flash> |
| news in brief | <T Breve> |
| leading article | <T Fondo> |
| letters to the editor | <T Lettera> |
| listings | <T Elenco> |
| news | <T Notizia> |
| weather report | <T Tempo> |
| title of book, film, song etc. | <T Nome> |
| picture caption | <T Foto> |
| type of speech | <P> |
| running text | <P Prosa> |
| quote from written source | <P Citazione> |
| quote from oral source | <P Discorso> |
| interview question | <P Domanda> |
| interview response | <P Risposta> |

5.3. Corpora and TACTWeb

When corpora had been used in the years before, quite a few problems arose with the installation of TACT, the textual analysis program to be used, either on the central university server or when students wanted to install TACT at home. The most critical point was, however, that the time needed by inexperienced students to learn to use the program did not really allow for the teaching of a course of Romance linguistics where students would gain insight into the characteristics of languages, learn how to do empirical research and to critically compare their findings to theoretical assertions. Former courses were, thus, devoted

mainly to textual analysis as such. The availability of TACTWeb has altered the situation radically.

Part of the corpus, i.e. the edition of the 15.06.1994 of *Le Monde*, *Il Corriere della Sera* and of *La Vanguardia* is now on-line in the form of a database ²⁴ into which it has been converted by means of *TACT 2.1*²⁵. This database is installed on a PC in my office. With the aid of a trial version of *O'Riley Software's WebSite* this PC functions as a server for *TACTWeb*²⁶, which itself supports the querying of TACT-databases over the Web by operating with the server software. This has the advantage that students can work with it whenever they want or have time, the setup can be organised in a user-friendly way, and it can even be adapted when needed and can be enriched with information which is directly relevant to the course. This corpus can be accessed via the following page <http://www.uni-duisburg.de/FB3/ROMANISTIK/PERSONAL/Burr/humcomp/home.htm>.

5.4. An example

The example I am going to present is taken from the first course where TACTWeb was actually used by the students to retrieve the necessary data for their own research projects. As this course was taught at Duisburg University during the summer semester 1998 its results can be evaluated. I will be using for this evaluation a concrete research project which has been carried through by one of my students. Although the course itself was quite experimental in nature this project together with the results achieved should be able to show some of the advantages of courses where an integration of linguistic theory or descriptions of languages and empirical research is sought by using the new technology, text analysis software and electronic corpora now at our disposal.

5.4.1. The course

The course was part of the program for undergraduate students and was devoted to the study of lexical phenomena in the three Romance languages French, Italian and Spanish. Variation was, where possible, to be taken into account, too.

The topics which were proposed for study during the course included:

- like English *leader*, French *main propres*, German *hinterland*, Italian *a contrario*, Latin *mea culpa*;
- either with the help of prefixes like anti-, archi-, ex-, dis- or suffixes like -ista / -iste, -ione / -ión / -ion, via composition as in *molino de viento* / *macchina da scrivere* / *moulin à vent* or juxtaposition as for example *hombre-rana*, *café-teatro*, or in more specific terms, terminologies like *direc-trice* / -teur, *vinci-trice* / -tore, *ministr-a* / -o;
- for example *más loco que una cabra*, *bueno como el pan*, *lancia in resta*, *povero in canna*, *son dos caras de la misma moneda*, *un chien vivant vaut mieux qu'un chien mort*;
- as *temps mort*, *natures mortes*, *salto mortale*, *trappola mortale*, *errore mortale*, *abbraccio mortale*, *peccato mortale*, *muerte política*, *muerte social*;
- like prepositions, conjunctions or adverbs.

Students were, however, free to propose different topics, as well. In order to be able to judge whether the study was feasible with the means at our disposal, students had to hand

²⁴In the meantime, part of my Italian newspaper corpus 1989, where finite verb forms have been tagged according to the categories of tense, mode and aspect, is equally accessible over the web and is used for teaching a course on the verbal system of Italian.

²⁵TACT was developed originally by John Bradley and Lidio Presutti at the University of Toronto from 1984 onwards.

²⁶TACTWeb is a project developed by John Bradley and Geoffrey Rockwell.

in first of all a project proposal, giving a short description of why they wanted to study a certain topic and what they were interested in finding out. Then, at some later point, a draft of the structure of the paper they were to write and the steps they planned for their empirical research was requested. All the way through the course there was discussion of the project with the individual students.

5.4.2. A study

In the framework of this course a particularly interesting study was done by a part-time student, a secondary teacher, who had only just started to use a PC (cf. Steinert-Schmitz 1998). By looking at the denominations of and the terminology for persons she had, in fact, become interested in the question if and how women are present in the *Corriere della Sera*. This concern lead her to formulate the following questions:

- How many women?
- Which are the linguistic means employed to make them visible?
- In which sections do they appear and how?
- What is their position?

5.4.2.1. How many women?

In order to find an answer to the first question of how many women appear in the newspaper she retrieved the occurrences of the following pairs and of a few (gender) specific denominations: *lei / lui, ella / egli, donna / uomo, madre / padre, moglie / marito, mamma / papà, figlia / figlio, femmina / maschio, ragazza / ragazzo, nonna / nonno, cugina / cugino, suocera / suocero, zia / zio, cognata / cognato, amico / amica, bambina / bambino, signora / signore, padrona / padrone, compagna / compagno, impiegata / impiegato, femminile / maschile, prostituta, capo, boss, leader*.

Whenever it was sensible the wildcard . * was used, in order to retrieve plural forms like *ragazze / ragazzi* and diminutives like *figliola / figliolo, ragazzina / ragazzino*, as well. This step led to 440 occurrences of denominations referring to men, against 170 which refer to women.

5.4.2.2. Which are the linguistic means employed to make them visible?

The second question concerning the linguistic means with which women are made visible was studied first of all on the basis of the suffixes which according to grammars and studies like my own (cf. Burr 1995) are offered by the system for the creation of terms which denominate persons from a perspective of their activity or profession: *-aio, -aro, -ario, -ino, -tore, -sore, -one, -iere, -aiulo, -aiolo, -ano, -ato, -vendolo, -alaio, -ante, -ista, -ente, -fice, -trice, -aia, -iera, -aiola, -ana, -ara, -aria, -ina, -ona, -ora, -essa, -ata*.

The explicitly feminine terms which resulted from this query are given in the table below. Obviously, also denominations like *bambina* and *ragazzina* turned up because of the suffix used for their formation but then had to be discarded because they do not belong to the terminology in question:

Table 3.

| | | | |
|--------------------|-------------------|-------------------|---------------|
| attrice (3) | baronessa (1) | croceroassina (1) | operaia (1) |
| danzatrice (1) | dottoressa (1) | madrina (1) | astrologa (2) |
| direttrice (1) | presidentessa (1) | regina (5) | casalinga (1) |
| nuotatrice (2) | poetessa (1) | sovrana (2) | collega (1) |
| organizzatrice (1) | studentessa (2) | suora (1) | candidata (7) |

scrittrice (1)

senatrice(1)

sostenitrice (1)

A few more denominations which refer to women like *socia*(1) and *diplomatica*(1) were found by going through the material, others by looking at the co-text of the individual occurrences of formations with – *ista* which by themselves do not indicate the sex of the person. It was found that the terms in the following table actually refer to a woman:

Table 4.

| | | | |
|----------------|-------------------|---------------|-----------------|
| registra (1) | inserzionista (1) | artista (1) | giornalista (1) |
| brigatista (1) | cronista (1) | legghista (1) | |

In this way, a total of 51 instances referring to women was put together which compared with the ca. 850 occurrences retrieved for men by using the above given list of suffixes shows that women are clearly underrepresented in the newspaper. Men, on the contrary, are not only present in large numbers but also as performing a very large range of activities. As these would have made up a huge table, for the table below only those terms are retained which appear relatively frequently:

Table 5.

| | | | |
|------------------|------------------|-------------------|------------------|
| autore (24) | procuratore (17) | presidente (71) | ministro (44) |
| direttore (42) | scrittore (12) | protagonista (11) | segretario (47) |
| fondatore (6) | assessore (15) | registra (18) | bersagliere (13) |
| imprenditore (9) | | | carabiniere (17) |

The next step the student took in order to answer the present question was a co-occurrence search for the combination of the feminine article *la* with non sex-specific terms like *presidente* or *leader*. This resulted in the following two examples:

la presidente della Camera, Irene Pivetti
la leader dei Verdi

which led her on to the question whether there might be some women hidden also behind denominations which are supposed to be masculine because of their ending or because of the article used together with them. In order to find out about this she carried out another query using the following list of denominations: *presidente, ministro, deputato, sindaco, magistrato, giudice, procuratore, ambasciatore, funzionario, senatore, consigliere, parlamentare, portavoce, ingegnere, avvocato, pubblicista, interprete, architetto, docente*

The result was that *consigliere, sindaco, presidente*, and *avvocato* each referred on one occasion to a woman:

il presidente della Camera Irene Pivetti anche la moglie è un avvocato di successo
il sindaco Ariana Cavicchioli il consigliere comunale Luisa Camelli

As by then three different ways of denominating a presiding woman had been found, it was only logical that she should ask herself whether there was a difference in meaning between *il / la presidente Pivetti* and *presidentessa* in the following example:

Questa mostra - spiega Carla Marino, da un anno presidentessa del Comitato provinciale della Croce rossa -, è un'occasione importante per rendere visibile il nostro operato.

The conclusion she came to was that, as some linguists had argued, the different terms were used in accord with the higher or lower prestige of the respective activity ²⁷.

As *donna* in conjunction with a masculine term in descriptions of Italian is treated as another way to make women discernible and as it was argued in a study that foreign terms were used in most cases in a sex specific way and thus, when they referred to women were either combined with the feminine article *la* or with *donna* (cf. Burr 1995, 153), she then analysed the occurrences of foreign terms like *designer*, *leader* and *star* and of *donna*. She could, in fact, confirm that foreign terms were mostly used with the feminine article, whereas *donna* appears in those cases where, like in the following example, the exceptionality of the presence of a woman is stressed:

E di un leader donna che ne dice?

The last step to be undertaken was an investigation into the usage of the different types of names which appear together with the women in the samples retrieved by means of all the afore-mentioned queries. The results obtained confirmed, according to the author, another study carried out on the basis of a larger corpus of Italian newspapers (cf. Burr 1997), where it was argued that in most cases where women appear either their name or their name and surname is used. If the surname is used on its own, the feminine article appears. Her findings were, in fact:

Table 6.

| | | |
|----------------|----|--|
| name + surname | 30 | |
| name | 7 | |
| la + surname | 4 | la Thatcher, la Pivetti (2), la Serafino |

5.4.2.3. In which sections do Women appear and how?

In order to find an answer to the question concerning the sections of the paper women were present in above all, she had to make use of the differentiations made possible by the markup system with which the corpus had been enriched and which, therefore, can be made visible next to each retrieved occurrence. This led to the following table:

Table 7.

| | |
|--------------------|-----|
| Cultura/Spettacoli | 40% |
| Politica | 23% |
| Cronaca | 31% |
| Sport | 6% |

Thus women are above all present in the section *Cultura* of the *Corriere della Sera* and much less in sections like *Politica* and *Sport*.

By then analysing more thoroughly the data, she found that female politicians in the section *Politica* are mostly talked about in quite neutral terms. In the section *Cultura*, instead, the appearance and character of women are of great importance and the ideal woman seems, in fact, to possess professionalism, intelligence and beauty as in the following example:

l'inserzionista laureata, 50enne, manager e donna attraente, intelligente e colta

Due to the film- and theatre program written about in the section *Spettacoli*, the image of women there is predominantly stereotyped and full of clichés. Women are presented as victims of violence and of the sexual appetites of others or as victims of their own irrationality

²⁷I personally would be tempted to see even a further differentiation in the following terms: il presidente @ la presidente @ presidentessa.

and unaccountable emotions. A similar picture of women can be found in the *Cronaca*. In the section *Sport*, however, where very few women appear, there are a few cases in which a new consciousness of women is allowed to transcend the reports:

Le ragazze sono in prima fila; mezzo milione di loro partecipano ai campionati regionali della Federazione Giovanile e si prepara il varo della Lega nel '95.

5.4.2.4. What is their position?

A query of *moglie* / *mogli* will result in altogether 27 occurrences, whereas *marito* will turn up only 9 times. As Steinert-Schmitz recognised, more than twenty times in this context alone women were portrayed as ornaments of their husbands, indicated by the expression *insieme alla moglie* or variations thereof. The following examples may suffice:

Mancavano soltanto le mogli e le fidanzate dei giocatori
 Freddie Francis ama andare al cinema, incontrare gente di ogni razza e cultura, in compagnia della moglie
 L'ultimo sfogo con al fianco la moglie Alessandra
 E ieri Carlo Sama, nella conferenza stampa presso lo studio del legale Francesco Galgano, "esterna" l'ultima difesa. Al suo fianco la moglie
 Ken Follett, lo scrittore di bestseller miliardario che insieme alla moglie Barbara ha aperto uno dei salotti più ambiti di Londra.
 Philip Gould, insieme alla moglie Gail Rebuck
 Luigi Vecchini, in compagnia della moglie, stava aprendo il portone del palazzo di via Ucelli di Nemi
 al momento della morte, nella casa di Beverly Hills, era con lui la moglie Ginny signori di mezza età che tengono per mano le loro mogli

Apart from this there is also a tendency to define the authority of a woman by means of the position of her husband:

'La guerra è in città. Non solo in Bosnia o in Ruanda' - continua Carla Martino, che è sorella del ministro degli Esteri -.

5.5. Conclusion

Language can only be studied "in der verbundenen Rede" and even a relatively small sample of speech will, as this study of women in the *Corriere della Sera* shows, allow for more insight than dictionaries and grammars. With her "mühevoller[n], oft ins Kleinliche gehende[n] Elementaruntersuchung" the student was, in fact, able to find out quite a lot about the overall character of the language. As her study shows, the possibilities of the system are, in fact, not just realised in concrete discourse, but they are restricted by the social norm which is itself bound up with a certain world-view, which plays a more or less dominant role according to the domain where language is used.

I am convinced that a study like this could not have been done in the framework of a traditional undergraduate course of Romance linguistics. Instead, it was made possible by the new means at our disposal, in this case an electronic corpus which can be systematically queried in a satisfactorily user-friendly way. This allows students to become curious in the first place and to enter whatever forms come to mind. The first results of a query where a wildcard is used will show them, however, that next to the occurrences they had in mind, forms they had never thought of will be retrieved as well, and that they, therefore, have to redefine their original definition if they intend to control the complexity of speech. Even if in a traditional course of synchronic linguistics the interest of the students in a certain form, category or usage can be raised, the sheer amount of work implied in a manual study of real speech on the basis of a printed newspaper, for example, would prevent them from following up their curiosity and from defining and redefining certain queries.

The same goes for questions which only come up when working on a research project. As the study presented here shows, quite a few questions were only asked because of a certain result. In order to answer these questions the student actually went through the whole corpus several times. She could only do this, however, because she did not have to reread the whole newspaper. Had she worked on a printed newspaper, instead, it would have been only natural if she had given up asking questions and nobody could have blamed her for that, the research project being part of an undergraduate course and not of a doctoral thesis. In the present case, on the contrary, she was quite prepared to retrieve the data she needed to follow through her own ideas and in the process she learnt how to interpret them.

Apart from making it possible to carry out research projects like the one used here as an example, courses which integrate theory with practice allow for much more participation of the individual student. In fact, every student can look at what the others are doing, ask for help or exchange experiences with others. Whereas in traditional courses all the work for a paper is done in the library or at home and the other members of a course are only presented with the results, here the process of defining and retrieving data is open to everybody and problems and findings can be directly discussed. Thus such a project involves the collective elements of a 'workshop' with individual reflection.

Last but not least, teaching also profits very much from such an organisation of the course. The teacher has, in fact, more time for the individual student. This can be used not only for problem-solving but also for finding out about their reading of the literature which is relevant for their topic. Furthermore, discussing the project or the methodology with them while they are actually working on their research allows for much more insight into the problems involved or the progress they are making, than the traditional counselling during reception hours normally permits.

Appendix A.

Bibliography

- Aarts, J., and Meijs, W. (eds.), (1986), *Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora*, Amsterdam, Rodopi.
- Accademia Nazionale dei Lincei, (1975), *Colloquio sul tema: Tecniche di classificazione e loro applicazione linguistica* (Florence, 1972), Rome.
- Acevedo, H., (1995), 'Biblioteca Nacional de Argentina', in Moreno de Alba, J. G., and Ramírez Leyva, E. M. (eds.), *Historia de las Bibliotecas Nacionales de Iberoamérica: pasado y presente. Asociación de Bibliotecas Nacionales de Iberoamérica (ABINIA)*. Mexico, UNAM, 1995 (2nd ed.), pp. 3-24.
- Actes du Colloque International sur la Mécanisation des Recherches Lexicologiques*, Besançon 1961, *Cahiers de Lexicologie*, 3, 1961.
- Aijmer, K., and Altenberg, B., (1991), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, London – New York, Longman.
- Almanacco Letterario Bompiani 1961*, Milan, 1961.
- ALPAC Report, (1966), *Language and Machine: Computers in Translation and Linguistics*, Washington, DC: National Research Council. Automatic Language Processing Advisory Committee.
- ANITE, (1998), *Language Engineering. Progress and Prospect '98*, Luxembourg.
- Antoni-Lay, M.H., Francopoulo G., and Zaysser, L., (1994), 'A Generic Model for Reusable Lexicons: The Genelex Project', *Literary and Linguistic Computing*, 9, 1, pp. 47-54.
- Atkins, B.T.S., Clear, J., and Ostler, N., (1992), 'Corpus Design Criteria', *Literary and Linguistic Computing*, 7, pp. 1-16.
- Atkins, B.T.S., Levin, B., and Zampolli, A., (1994), 'Computational Approaches to the Lexicon: An Overview', in Atkins, B.T.S., and Zampolli, A. (eds., 1994), pp. 17-45.
- Atkins, B.T.S., and Zampolli, A., (1994), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press.
- Ausiello, G. et al., (1991), *Modelli e linguaggi dell'informatica*, Milan, McGraw-Hill Italia.
- Bakhtin, M., (1963), *Problemy poetiki Dostoevskogo*, Moscow, Sovetskii Pisatel'; Eng. trans. *Problems of Dostoevsky's poetics*, Ardis, 1973.
- Bakhtin, M., (1975), 'Slovo v romane', in *Voprosy literatury i estetiki: Issledovaniia raznykh let*, Moscow, Khudozhestvennaia literatura; Eng. trans. in Holquist, M. (ed.), *The Dialogic Imagination*, Austin, University of Texas Press, 1981, pp. 259-422.
- Bakhtin, M., (1979a), 'Problema rechevykh zhanrov', in Bocharov, S. G. (ed.), *Estetika slovesnogo tvorchestva*, Moscow, Iskusstvo; Eng. trans. in Emerson, C., and Holquist, M. (eds.), *Speech Genres and Other Late Essays*, Austin, University of Texas Press, 1986, pp. 60-102.
- Bakhtin, M., (1979b), 'Iz zapisej 1970-1971 godov', in Bocharov, S. G. (ed.), *Estetika slovesnogo tvorchestva*, Moscow, Iskusstvo; Eng. trans. in Emerson, C. and Holquist, M. (eds.), *Speech Genres and Other Late Essays*, Austin, University of Texas Press, 1986.
- Bakhtin, M., (1994), *Problemy tvorchestva poetiki Dostoevskogo*, Kiev, Next.
- Barlow, M., (1996), 'Corpora for Theory and Practice', *International Journal of Corpus Linguistics* 1, pp. 1-37.
- Barnard, D. T. et al., (1988), 'SGML-Based Markup for Literary Texts: Two Problems and Some Solutions', *Computers and the Humanities*, 22, pp. 265-276.

- Barthes, R., (1975), *S/Z*, (trans. R. Miller), London, Jonathan Cape, p. 5.
- Bartram, L., Ho, A., Dill, J., and Henigman, F., (1995), 'The Continuous Zoom: A Constrained Fisheye Technique for Viewing and Navigating Large Information Spaces', in *Proceedings of ACM UIST '95*, New York, ACM Press, pp. 207-215.
- Bazell, C., Catford, J., Halliday, M. A. K., and Robins, R.H. (eds.), (1986), *In Memory of J.R. Firth*, London – New York, Longman.
- Beccaria, G.L., (1975), 'Allitterazioni dantesche' and 'Figure dantesche' in *L'Autonomia del significante*, Turin, Einaudi, 1975, pp. 90-113 and 114-135.
- Bederson, B., and Hollan, J., (1994), 'Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics', in *Proceedings of ACM UIST '94*, New York, ACM Press, pp. 17-26.
- Bederson, B., Hollan, J. D., Stewart, J., Rogers, D., Vick, D., Ring, L. T., Grose, E., and Forsythe, C., (1998a), 'A Zooming Web Browser', in Forsythe, C., Ratner, J., and Grose, E. (eds.), *Human Factors in Web Development*, (Mahwah, NJ: Lawrence Erlbaum, 255-266.
- Bederson, B., Meyer, J., (1998b), 'Implementing a Zooming User Interface: Experience Building Pad++' in *Software: Practice and Experience*, 28, 10, pp. 1101-1135.
- Bertinetto, P. M., (1986), *Tempo, aspetto e azione nel verbo italiano: Il sistema dell'indicativo*, Florence, Accademia della Crusca.
- Bettetini, G., (1993), *La simulazione visiva. Inganno, finzione, poesia, computer graphics*, Milan, Bompiani
- Biber, D., and Finegan, E., (1991), *On the exploitation of computerized corpora in variation studies*, in Aijmer, K., and Altenberg, B. (eds., 1991), pp. 204-220.
- Biggs, M., and Huitfeldt, C., (1996), 'Discussion of 'Theory and Metatheory in the Development of Text Encoding'', Edited version of electronic discussion held in the Spring of 1996 under the auspices of *The Monist*.
<http://hhobel.phl.univie.ac.at/mii/pesp.html>.
- Bindi, R., Monachini, M., and Orsolini, P., (1989), *Italian reference Corpus*, Pisa, Istituto di Linguistica Computazionale.
- Björk, S. and Holmquist, L. E., (1998), 'Formative Evaluation of a Focus + Context Visualisation Technique'. Poster at *Human-Computer Interaction '98*.
<http://www.viktoria.informatik.gu.se/groups/play/publications/1998/formative.pdf>.
- Blackmans, W., (1998), 'Editorial. The Humanities and European Integration', *Reflections, The Newsletter of the Standing Committee for the Humanities*, ESF 2, 1998, pp. 2-3.
- Boguraev, B., and Briscoe, T., (1989), *Computational Lexicography for Natural Language Processing*, London – New York, Longman.
- Boguraev, B., Briscoe, T., Calzolari, N., Cater, A., Meijs, W., and Zampolli, A., (1988), *Acquisition of Lexical Knowledge for Natural Language Processing Systems*, Acquilex Technical Annex, ESPRIT Basic Research Action No. 3030 (unpublished).
- Bolter, J. D., (1991), *Writing Space: The Computer, Hypertext, and the History of Writing*, Mahwah (NJ), Lawrence Erlbaum Associates.
- Bolter, J. D., (1992), 'Literature in the Electronic Writing Space', in Tuman, M. (ed., 1992), pp 19-42.
- Bolter, J. D., (1996), 'Ekphrasis, virtual reality and the future of writing', in Nunberg, G. (ed., 1996), pp. 253-72.

- Booth, A.D., Cleave, J.P., and Brandwood, L., (1958), *Mechanical Resolution of Linguistic Problems*, London, Butterworths Scientific Publications.
- Bortolini, U., Tagliavini, C., and Zampolli, A., (1971), *Lessico di Frequenza della Lingua Italiana Contemporanea*, IBM Italia.
- Brown, M. H., and Weihl, W. E., (1996), 'Zippers: A Focus + Context Display of Web Page', in *Proceedings of World Conference of the Web Society*, Association for the Advancement of Computing in Education (AACE). Also available as DEC-SRC Tech Report 140.
- Brusilovsky, P., (1996), 'Methods and Techniques of Adaptive Hypermedia', *User Modeling and User-Adapted Interaction*, 6, pp. 87-129.
- Bryan, M., (1988), *SGML: An Author's Guide to the Standard Generalized Markup Language*, Wokingham/Reading/New York, Addison-Wesley.
- Burnard, L., (1995), 'Text Encoding for Information Interchange. An Introduction to the Text Encoding Initiative' (document TEI J31), paper presented at the *Second Language Engineering Conference*, London, October 1995.
<http://www.tei-c.org/Vault/SC/J31/>
- Burr, E., (1995), 'Agentivi e sessi in un corpus di giornali italiani', in Marcato, G. (ed.), *Donna e Linguaggio: Convegno Internazionale di Studi, Sappada / Plodn (Belluno)*, 26-30 giugno 1995, Padua, CLEUP, pp. 141-157.
- Burr, E., (1997), 'Neutral oder stereotyp. Referenz auf Frauen und Männer in der italienischen Tagespresse', in Dahmen, Wolfgang; Holtus, Günter; Kramer, Johannes; Metzeltin, Michael; Schweickard, Wolfgang, and Winkelmann, O. (eds.), *Sprache und Geschlecht in der Romania: Romanistisches Kolloquium X*, TBL 417, Tübingen, Narr, pp. 133-179.
- Burrows, J. F., (1992), 'Computers and the Study of Literature', in Butler, C. S. (ed., 1992), pp. 167-204.
- Burton, D. M., (1981), 'Automated Concordances and Word Indexes: The early sixties and the Early centers', *Computers and the Humanities*, 15, pp. 83-100.
- Busa, R., (1951), 'Sancti Thomas Aquinatis hymnorum ritualium: Varia specimina concordantiarum', in *Primo saggio di indici di parole automaticamente composti e stampati da macchine IBM a schede perforate*, Milan, Bocca.
- Busa, R., (1962), 'L'analisi linguistica nell'evoluzione mondiale dei mezzi di informazione', *Almanacco Letterario Bompiani 1962*, pp. 103-117.
- Busa, R. (ed.), (1968), 'Actes du Séminaire International sur le dictionnaire latin de machine', *Calcolo*. Supplement 2, V.
- Busa, R., (1980), 'The Annals of Humanities Computing: The *Index Thomisticus*', *Computers and the Humanities*, 15, pp. 1-14.
- Butler, C. S. (ed.), (1992), *Computers and Written Texts*, Oxford, Blackwell.
- Buzzetti, D., and Tabarroni, A., (1991), 'Informatica e critica del testo: il caso di una tradizione fluida', *Schede umanistiche*, 2, pp. 185-193.
- Buzzetti, D., (1995), 'Image Processing and the Study of Manuscript Textual Traditions', *Historical Methods*, 28 (1995), pp. 145-163.
- Calvi, L., and De Bra, P., (1997), 'Proficiency-Adapted Information Browsing and Filtering in Hypermedia Educational Systems', *International Journal of User Modelling and User-Adapted Interaction*, 7, pp. 257-277.
- Calvino, I., (1980), 'Cibernetica e fantasmi', in Calvino, I., *Una pietra sopra*, Turin, Einaudi, pp. 164-181.

- Calzolari, N., and McNaught, J., (1994), *Editors' Introduction*, EAGLES document EAG-EB-IR-2.
- Calzolari, N., and Picchi, E., (1994), *A Lexical Workstation: from Textual Data to Structured Database*, in Atkins, B.T.S., and Zampolli, A. (eds., 1994), pp. 439-467.
- Card, S. K., Mackinlay, J. D., and Shneiderman, B. (eds.), (1999), *Readings in Information Visualization: Using Vision to Think*, San Francisco (Ca.), Morgan Kaufmann Publishers.
- Card, S. K., Robertson, G., and York, W., (1996), 'The WebBook and the Web Forager: An information workspace for the World Wide Web', in *Proceedings of ACM SIGCHI '96*, New York, ACM Press, pp. 111-117.
- Chafe, W., (1992), 'The importance of corpus linguistics to understanding the nature of language', in Svartvik, J. (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, Berlin and New York, Mouton de Gruyter, pp. 79-97.
- Chernaik, W., Davis, C., and Deegan, M. (eds.), (1993), *The Politics of the Electronic Text*, Oxford, Office of Humanities Communication, 3.
- Chernaik, W., Deegan, M., and Gibson, A. (eds.), (1996), *Beyond the Book: Theory, Culture and the Politics of Cyberspace*, Oxford, Office of Humanities Communication, 7.
- Cignoni, L., Peters, C., and Rossi, S., (1983), *European Science Foundation Survey of Lexicographical Projects*, Pisa, Istituto di Linguistica Computazionale.
- Ciotti, F., (1994), 'Il testo elettronico. Memorizzazione, codifica ed edizione', in Leonardi, C., Morelli, M., and Santi F. (eds.), *Macchine per leggere. Tradizioni e nuove tecnologie per comprendere i testi*, Florence, Fondazione Ezio Franceschini – Spoleto, Centro Studi sull'Alto Medioevo, pp. 213-230.
- Ciotti, F., (1995), 'Testi elettronici e banche dati testuali: problemi teorici e tecnologie', *Schede Umanistiche*, 2, pp. 147-178.
- Ciotti, F., (1997), 'Cosa è la codifica informatica dei testi?', in Gruber, D., and Paoletto, P. (eds.), *Umanesimo e Informatica*, Pesaro, Metauro edizioni.
- Clear, J., (1992), 'Corpus sampling', in Leitner, G. (ed.), *New Directions in English Language Corpora: Methodology, Results, Software Developments*, Topics in English Linguistics 9, Berlin and New York, Mouton de Gruyter, pp. 21-31.
- Cole, R., (1997), 'Foreword of the Chief Editor', in Varile G.B., and Zampolli, A., (managing eds., 1997), pp. xi-xii.
- Coombs, J. S., Renear, A., and De Rose, S. J., (1987), 'Markup Systems and the Future of Scholarly Text Processing', *Communications of the Association for Computing Machinery*, 30, pp. 933-47.
- Coseriu, E., (1988a), *Einführung in die allgemeine Sprachwissenschaft*, UTB 1372, Tübingen, Francke.
- Coseriu, E., (1988b), *Sprachkompetenz: Grundzüge der Theorie des Sprechens*, UTB 1481, Tübingen, Francke.
- Crowdy, S., (1993), 'Spoken Corpus Design', *Literary and Linguistic Computing*, 8, pp. 259-265.
- Cumming, S., (1995), 'The Lexicon in Text Generation: Progress and Prospects', in Walker, D. et al., (1995), pp. 171-206.
- Danzin, A., (1992), Groupe de réflexion stratégique pour la Commission des Communautés Européennes (DG XIII), *Vers une infrastructure linguistique européenne*, Document available from DG XIII-E, Luxembourg.

- De Bra, P., and Calvi, L., (1998), 'Towards a Generic Adaptive Hypermedia System', in Brusilovsky, P., and De Bra, P. (eds.), *Proceedings of the Second Workshop on Adaptive Hypertext and Hypermedia*, Computing Science Reports, Eindhoven, Eindhoven University of Technology, pp. 5-10.
- De Bruïne, F., (1998), 'Foreword', in ANITE (1998).
- De Gerolamo, C., (1999), 'Tutto Shakespeare in 10 grammi', *Panorama*, 1, 1999, pp. 50-51.
- De Kerckhove, D., (1995) *The Skin of Culture*, Toronto, Somerville House Books.
- De Mauro, T., (1975 – 1981), *Scuola e linguaggio: Questioni di educazione linguistica*, Rome, Editori Riuniti.
- De Mauro, T., (1993), 'Le scelte per la costituzione del corpus', in De Mauro, T., Mancini, F., Vedovelli, M., and Voghera, M., *Lessico di frequenza dell'italiano parlato: Ricerca a cura dell'Osservatorio linguistico e culturale italiano OLCI dell'Università di Roma "La Sapienza"*, Milan, Fondazione IBM Italia – Etas, pp. 29-32.
- De Tollenaere, F., (1963), *Nieuwe wegen in der lexicologie*, Verb. Konink. Ned. Akad. Wetensch. A fit Letterkunde, N.R., Deel XX, N.1.1.
- Debray, R., (1996), *Media Manifestos: On the Technological Transmission of Cultural Forms*, (trans. E. Rauth), London, Verso.
- Dennet, D., (1990), *The Interpretation of Texts, People and Other Artifacts*, 'Philosophy and Phenomenological Research', L, Supplement, Fall 1990, pp. 177-194 (Reprinted in Losonsky, M., ed., *Language and Mind: Contemporary Readings in Philosophy and Cognitive Science*, Oxford, Blackwell, 1995).
- <http://ase.tufts.edu/cogstud/papers/intrptxt.htm>.
- De Rose, S. J., Durand, D., Mylonas, E., and Renear, A., (1990), 'What is Text, Really?', *Journal of Computing in Higher Education*, 1, 2, pp. 3-26.
- Donald, M., (1991), *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*, Cambridge (Mass.), Harvard University Press.
- Donaldson, P. S., (1997), 'Digital Archive as Expanded Text: Shakespeare and Electronic Textuality', in Sutherland, K. (ed., 1997), pp. 173-98.
- Doss, P. E., (1996), 'Traditional Theory and Innovative Practice: The Electronic Editor as Poststructuralist Reader', in Nunberg, G. (ed., 1996), pp. 213-224.
- Dyer, R.R., (1973), 'The Measurement of Individual Style', in Zampolli, A. (ed., 1973), pp. 325-348.
- 'EAGLES Update', (1993), *Elsnews Bulletin*, 2, 1.
- Eco, U., (1995), *Sei passeggiate nei boschi narrativi*, Milan, Bompiani.
- Eco, U., (1976), *A Theory of Semiotics*, Bloomington (Ind.), Indiana University Press, 1976, pp. 264-268.
- Eco, U., (1990), *I limiti dell'interpretazione*, Milan, Bompiani.
- Estoup, J. B., (1907), *Gammes Sténographiques*, Paris.
- European Commission DG-XIII, (1994), *Telematics Applications Programme (1994-1998), Work Programme*, Luxembourg.
- Fattori, M., Bianchi, M. (eds.), (1976), *I° Colloquio Internazionale del Lessico Intellettuale Europeo*, Rome, Edizioni dell'Ateneo.
- Fezzi, P., (1994), 'Gli ipertesti: un nuovo media?', in Ricciardi, M. (ed.), *Oltre il testo: gli ipertesti*, Milan, Franco Angeli, 1994, pp. 175-188.
- Fillmore, C. J., (1992), 'Corpus linguistics' or 'Computer-aided armchair linguistics', in Svartvik, J. (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, Berlin and New York, Mouton de Gruyter, pp. 35-60.

- Finneran, R. J. (ed.), (1996), *The Literary Text in the Digital Age*, Ann Arbor, University of Michigan Press.
- Fisher, B., Agelidis, M., Dill, J., Tan, P., Collaud, G., and Jones, C., (1997), 'CZWeb: Fish-eye Views for Visualizing the World Wide Web', in Smith, M., Salvendy, G., and Koubek, R. J. (eds.), *Design of Computing Systems: Social and Ergonomics Considerations*, Amsterdam, Elsevier, pp. 719-722.
- Flanders, J., (1997), 'The Body Encoded: Questions of Gender and the Electronic Text', in Sutherland, K. (ed., 1997), pp. 127-144.
- Foucault, M., (1969), *L'archéologie du savoir*, Paris, Gallimard, 1969; Eng. trans. *The Archaeology of Knowledge*, London, Tavistock Publications, 1972.
- Fourcin, A., and Gibbon, D., (1993), "Spoken Language Assessment in the European Context", *Literary and Linguistic Computing*, 9, 1, pp. 79-86.
- Francis, W., and Kucera, H., (1964; revised 1971 and 1979), *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Providence (R.I.), Department of Linguistics, Brown University.
- Furnas, G. W., (1986), 'Generalized Fisheye Views', in *Proceedings of ACM SIGCHI '86*, New York, ACM Press, pp. 16-23.
- Furnas, G. W., (1997), 'Effective View Navigation', in *Proceedings of ACM SIGCHI '97*, New York, ACM Press, pp. 367-374.
- Galison, P., (1997), *Image and Logic: A Material Culture of Microphysics*, Chicago, University of Chicago Press.
- Gardin, J. C., (1991), *Le calcul et la raison. Essai sur la formalisation du discours savant*, Paris, Édition de l'École des Hautes Études en Science Sociale.
- Genet, J.-P., and Zampolli, A., (1992), *Computers and Humanities*, Aldershot, Dartmouth Publishing – European Science Foundation.
- Genette, G., (1980), *Narrative Discourse*, Oxford, Blackwell.
- Gigliozzi, G. (ed.), (1987a), *Studi di codifica e trattamento automatico dei testi*, Rome, Bulzoni.
- Gigliozzi, G., (1987b), 'Codice, testo e interpretazione', in Gigliozzi, G. (ed., 1987a), pp. 65-84.
- Gigliozzi, G., (1992), 'Modellizzazione delle strutture narrative', in *Calcolatori e scienze umane*, Milan, Fondazione IBM Italia – Etas, pp. 303-314.
- Gigliozzi, G., (1993), *Letteratura, modelli e computer. Manuale teorico pratico per l'applicazione dell'informatica al lavoro letterario*, Rome, EUROMA-La Goliardica.
- Gigliozzi, G., (1997), *Il testo e il computer*, Milan, Bruno Mondadori.
- Ginzburg, C., (1998), 'Distanza e prospettiva: due metafore', in *Occhiacci di legno. Nove riflessioni sulla distanza*, Milan, Feltrinelli, pp. 171-193.
- Goldfarb, C. F., (1981), 'A Generalized Approach to Document Markup', in *Proceedings of the ACM SIGPLAN—SIGOA Symposium on Text Manipulation*, New York, ACM Press.
- Goldfarb, C. F., (1990), *The SGML Handbook*, Oxford, Oxford University Press.
- Goodman, N., (1968), *Languages of Art: An Approach to a Theory of Symbols*, Indianapolis, Bobbs-Merrill.
- Greenstein, D., and Porter, S., (1998), 'Scholars' Information Requirements in a Digital Age', AHDS, Kings College London, 24 October 1998.
<http://www.ahds.ac.uk/public/uneeds/un0.html>.

- Greetham, D., (1992), 'Textual Scholarship', in Gibaldi, J. (ed.), *Introduction to Scholarship in Modern Languages and Literatures*, New York, The Modern Language Association.
- Guiraud, P., (1954), *Bibliographie critique de la statistique linguistique*, Utrecht-Anvers, Mouton.
- Guiraud, P., (1960), *Problèmes et méthodes de la statistique linguistique*, Paris, P.U.F.
- Halliday, M. A. K., (1986), 'Lexis as a linguistic level', in Bazell, C., Catford, J., Halliday, M. A. K., and Robins, R.H. (eds., 1986).
- Halliday, M. A. K., (1992), 'Language as system and language as instance: The corpus as a theoretical construct', in Svartvik, J. (ed.), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991*, Berlin and New York, Mouton de Gruyter, pp. 61-77.
- Harkin, D., (1957), 'The History of Word Counts', *Babel*, 3, pp. 113-124.
- Hays, D. G., (1967), *Introduction to Computational Linguistics*, New York, American Elsevier.
- Hays, D. G., (1969), 'Computational Linguistics: Introduction', in Meetham, A. R., and Hudson, R.A. (eds., 1969), pp. 49-51.
- Hays, D. G., (1976), 'The field and scope of computational linguistics', in Papp, F., and Szépe, G. (eds., 1976), pp. 21-25.
- Heid, H., and McNaught, J. (eds.), (1991), 'Eurotra-7 Study: Feasibility and Project Definition Study on the Reusability of Lexical and Terminological Resources in Computerised Applications', *Eurotra-7 Final Report*, Stuttgart.
- Heilmann, L., (1963), 'Considerazioni statistico-matematiche e contenuto semantico', *Quaderni dell'Istituto di Glottologia*, VII, Università di Bologna, pp. 34-45.
- Herdan, G., (1964), 'Quantitative Linguistics or Generative Grammar', *Linguistics*, 4, pp. 56-65.
- Herwijnen, E. van, (1994), *Practical SGML*, Dordrecht, Kluwer Academic Publishers.
- Hockett, C. F., (1958), *A Course in Modern Linguistics*, New York, Macmillan.
- Holland, P., (1993), 'Authorship and Collaboration: The Problem of Editing Shakespeare', in Chernaik, W., Davis, C., and Deegan, M. (eds., 1993), pp. 17-24.
- Hollands, J. G., Carey, T. T., Matthews, M. L., and McCann, C. A., (1989), 'Presenting a graphical network: a comparison of performance using fisheye and scrolling views', in Salvendy, G., and Smith, M. (eds.), *Designing and Using Human-Computer Interfaces and Knowledge Based Systems*, Amsterdam, Elsevier, pp. 313-320.
- Holmquist, L. E., (1998), 'The Zoom Browser: Showing Simultaneous Detail and Overview in Large Documents', *Human IT*, (3), 2, 1, Borås (Sweden), ITH, pp. 131-150.
- Holmquist, L. E., and Björk, S., (1998a), 'A Hierarchical Focus + Context Method for Image Browsing', in *Proceedings of SIGGRAPH '98*, New York, ACM Press, pp. 276-284.
- Holmquist, L. E., and Björk, S., (1998b), 'Formative Evaluation of Flip Zooming: Towards Effective Integration of Detail and Context on Small Displays', in *Viktoria Research Report VRR-98-1*, Gotenburgh, Viktoria Institute.
<http://www.viktoria.informatics.gu.se/publications/>.
- Huitfeldt, C., (1992), 'MECS – A Multi-Element Code System', in *Working Papers from the Wittgenstein Archives at the University of Bergen*, No 3. (Seen in draft in October 1992).
- Huitfeldt, C., (1994), 'Multi-Dimensional Texts in a One-Dimensional Medium', *Computers and the Humanities*, 4/5, 28, pp. 235-241.

- Humboldt, W. von, (1827-1829 / 1963) 'Über die Verschiedenheiten des menschlichen Sprachbaus', in Humboldt, W. von, *Schriften zur Sprachphilosophie*, Werke in fünf Bänden III, Darmstadt, Wissenschaftliche Buchgesellschaft, pp. 144-367.
- Hymes, D. H., (1972), 'On Communicative Competence', in Pride, J. B., and Holmes, J. (eds.), *Sociolinguistics*, Harmondsworth, Penguin, pp. 269-293.
- IBM, (1964), *Literary Data Processing Conference Proceedings*, September 9-11 1964, Washington D.C.
- Ide, N., and Véronis, J., (1996), *Corpus Encoding Standard*, EAGLES document EAG-CWG/CES.
- Ide, N., and Sperberg-McQueen, C. M., (1995), 'The Text Encoding Initiative: Its History, Goals and Future Development', in Ide, N., and Véronis, J. (eds., 1995), pp. 5-16.
- Ide, N., and Véronis, J. (eds.), (1995), *The Text Encoding Initiative: Background and Context*, Dordrecht, Kluwer Academic Publishers (reprinted from *Computers and the Humanities*, 29, 1, 2 and 3, 1995).
- Ingria, J.P., (1995), 'Lexical Information for Parsing Systems: Points of Convergence and Divergence', in Walker, D. et al. 1995, pp. 93-170.
- International Organization for Standardization (ISO), (1986), *Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML)*, ISO8879-1986, International Organization for Standardization (ISO).
- IST - Information Society Technologies. *General Outlines, Scientific and Technological Objectives and Priorities*, (1998), Brussels.
- Jakobson, R., (1978), *Saggi di linguistica generale*, Milan, Feltrinelli (*Essais de linguistique générale*, Paris, Éditions de Minuit, 1963).
- Johansson, S., (1980), 'The LOB Corpus of British English Texts: Presentation and Comments', *ALLC Journal I*.
- Johansson, S., Leech, G.N., and Goodluck, H., (1978), *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers*, Oslo, Department of English, University of Oslo, pp. 25-36.
- Juilland, A., and Traversa, V., (1973), *Frequency dictionary of italian words*, The Hague, Mouton.
- Juilland, A., Edwards, P. M. H., and Juilland, I., (1965), *Frequency dictionary of rumanian words*, The Hague, Mouton.
- Kaeding, F. W., (1898), *Häufigkeitwörterbuch der deutschen Sprache*, Berlin, Steglitz.
- Kalgreen, H., (1973), *Foreword*, in Zampolli, A., and Calzolari, N. (eds., 1973-1977), pp. xiii-xiv.
- Kandogan, E., and Shneiderman, B., (1997), 'Elastic Windows: Evaluation of Multi-Window Operations', in *Proceedings of ACM SIGCHI '97*, New York, ACM Press, pp. 250-257.
- Katzen, M. (ed.), (1991), *Scholarship and Technology in the Humanities*, London, British Library Research.
- Kay, M., (1964), 'Report of Informal Meeting on Standards Formats for Machine-Readable Texts', in IBM (1964), pp. 327-328.
- Kay, M., (1967), 'Standards for Encoding Data in Natural Language', *Computers and Humanities*, 1, 5, pp. 170-177.
- Kay, M., (1983), 'The Dictionary of the Future and the Future of the Dictionary', in Zampolli, A., and Cappelli, A. (eds., 1983), pp. 161-174.

- Kendall, R., (1996), 'Hypertextual Dynamics in *A Life Set for Two*', *Proceedings of the Seventh ACM Conference on Hypertext*, New York, ACM Press, pp. 74-83.
- Khatchadourian, H., and Modiano, N., (1994), 'Use and Importance of Standards in Electronic Dictionaries: The Compilation Approach for Lexical Resources', *Literary and Linguistic Computing*, 9, 1, pp. 55-64.
- Koch, P., and Oesterreicher, Wulf, (1990), *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*, Romanistische Arbeitshefte 31, Tübingen, Niemeyer.
- Kowalski, G., (1996), *Information Retrieval Systems - Theory and Implementation*, Dordrecht, Kluwer Academic Publisher.
- Kristeva, J., (1974), *La révolution du langage poétique*, Paris, Seuil.
- Kucera, H., and Francis, W. N., (1967), *Computational Analysis of Present-Day American English*, Providence, Brown University Press.
- 'La Biblioteca Nacional durante el quinquenio 1932-1936', *Revista de la Biblioteca Nacional*, I.1 (January-March 1937).
- Lamping, J., Rao, R., and Piroli, P., (1995), 'A focus+context technique based on hyperbolic geometry for viewing large hierarchies', in *Proceedings of ACM SIGCHI '95*, New York, ACM Press, pp. 401-408.
- Landow, G. P., (1992a), 'Hypertext, Metatext and the Electronic Canon', in Tuman, M. (ed., 1992), pp. 67-94.
- Landow, G. P., (1992b), *Hypertext: The Convergence of Contemporary Critical Theory and Technology*, Baltimore, The Johns Hopkins University Press; Ital. trans, *Iper testo. Il futuro della scrittura*, Bologna, Baskerville, 1993.
- Landow, G. P., and Delany, P. (eds.), (1991), *Hypermedia and Literary Studies*, Cambridge-London, The MIT Press.
- Landow, G. P., and Delany, P. (eds.), (1993), *The Digital Word: Text-Based Computing in the Humanities*, Cambridge - London, The MIT Press.
- Laurel, B., (1991), *Computer as Theater*, Reading (Mass.), Addison-Wesley.
- Laurillard, D., (1992), 'Principles for Computer-Based Software Design for Language Learning', *Computer Assisted Language Learning*, 4, 3, pp. 141-152.
- Laver, J., 'Research for the Next Generation', *The Japan Society for the Promotion of Science Symposium*, Churchill College, University of Cambridge, 15-16 April 1998.
- Leech, G., (1991), *The state of the art in Corpus Linguistics*, in Aijmer K., and Altenberg, B. (eds. 1991), pp. 8-29.
- Leech, G. (ed.), (1990), *Proceedings of a Workshop on Corpus Resources*, Oxford, January, 1990, London, DTI Speech and Language Technology Club.
- Leech, G., and Wilson, A., (1994), *Morphosyntactic Annotation*, EAGLES document EAG-CWG/Annotate.
- Leitner, G., (1992), 'International Corpus of English: Corpus design - problems and suggested solutions', in Leitner, G. (ed.), *New Directions in English Language Corpora: Methodology, Results, Software Developments*, Berlin and New York, Mouton de Gruyter, pp. 33-64.
- Les Machines dans la Linguistique*, (1968), Prague.
- Leung, Y. K., and Apperley, M. D., (1994), 'A Review and Taxonomy of Distortion-Oriented Presentation Techniques', *ACM Transactions on Computer-Human Interaction*, 1, 2, New York, ACM Press, pp. 126-160.

- Levi, C., (1975), 'Il 'Tristram Shandy' di Sterne', in De Donato, G. (ed.), *Coraggio dei miti. Scritti contemporanei 1922-1974*, Bari, De Donato.
- Llisterri, J., (1994), *Spoken Texts*, EAGLES document EAG-CWG/Spokentx.
- Locke, W.N., and Booth, A. D., (1955), *Machine Translation of Languages*, New York, The Technology Press of the MIT – J. Wiley & Sons.
- Lotman, J., (1990), *Struttura del testo poetico*, II ed., Milan, Mursia (*Struktura chudozestvennogo teksta*, Moskva, Iskustvo, 1970).
- Lughi, G., (1993), 'Ipertesti letterari e labirinti narrativi', *Igitur*, V (2). <http://www.univ.trieste.it/~niritallughi/infomum/inform/saggi/lughiper.html>.
- Maccanico, A., (1997), *Sessione di apertura*, in Paola, R., and Piraino, R. (eds. 1998), pp. 17-19.
- Mackinlay, J. D., Robertson, G. G., and Card, S. K., (1991), 'The Perspective Wall: Detail and Context Smoothly Integrated', in *Proceedings of ACM SIGCHI '91*, New York, ACM Press, pp. 173-179.
- Maegaard, B., (1988), 'EUROTRA, The Machine Translation Project of the European Communities', *Literary and Linguistic Computing*, 3, 2, pp. 61-65.
- Malmberg, B., (1966), *Les nouvelles tendances de la linguistique*, Paris, P.U.F.
- Marcos Marín, F., (1985), 'Computer-Assisted Philology: Towards a Unified Edition of OSp. Libro de Alexandre', in *Proceedings of the E(uropean) L(anguage) S(ervices) Conference on Natural-Language Applications*, section 16, Copenhagen, IBM Denmark.
- Marcos Marín, F. (ed.), (1987), *Libro de Alexandre. Estudio y edición*, Madrid, Alianza Universidad.
- Marcos Marín, F., (1991a), 'ADMYTE (Archivo Digital De Manuscritos y Textos Españoles); The Digital Archive of Spanish Manuscripts and Texts' *Literary and Linguistic Computing*, 6/3, pp. 221-224.
- Marcos Marín, F., (1991b), 'Computers and Text Editing: A Review of Tools, an Introduction to UNITE and Some Observations Concerning its Application to Old Spanish Texts', *Romance Philology*, XLV/1, pp. 102-122, (Bibliography, pp. 205-237).
- Marcos Marín, F., (1994), *Informática y Humanidades*, Madrid, Gredos.
- Marcos Marín, F., (1996), *El Comentario Filológico con Apoyo Informático*, Madrid, Síntesis.
- Marrotti, A. F., (1986), *John Donne, Coterie Poet*, Madison (Wisc.), University of Wisconsin Press.
- McCarty, W., (1993), 'Encoding persons and places in the *Metamorphoses* of Ovid: 1. Engineering the text', *Texte: Revue de critique et de théorie littéraire*, 13/14.
- McCarty, W., (1994), 'Encoding persons and places in the *Metamorphoses* of Ovid. Part 2: The metatextual translation', *Texte: Revue de critique et de théorie littéraire*, 15/16.
- McCarty, W., (1998), 'The shape of things to come is continuous change: fundamental problems in electronic publishing', 3 July 1998. <http://www.kcl.ac.uk/humanities/cch/ohc/shape.html>
- McCullough, M., (1998), *Abstracting Craft: The Practiced Digital Hand*, Cambridge-London, The MIT Press.
- McGann, J., (1991), *The Textual Condition*, Princeton, Princeton University Press.
- McGann, J., (1997a), 'The Rationale of Hypertext', in Sutherland, K. (ed., 1997), pp 19-46.
- McGann, J. (ed.), (1997b), *The Complete Writings and Pictures of Dante Gabriel Rossetti: A Hypermedia Research Archive*, IATH, University of Virginia. <http://www.iath.virginia.edu/rossetti>

- McKerrow, R. B., (1927), *An Introduction to Bibliography for Literary Students*, Oxford, Oxford University Press.
- McLaughlin, M., and Robey, D. (1997), 'Tasso's Epic Style: Changes in Theory and Changes in Practice', *Journal of the Institute of Romance Studies*, V, pp. 23-46.
- Meetham, A. R., and Hudson, R. A., (1969), *Encyclopaedia of Linguistics, Information and Control*, Oxford, Pergamon Press.
- Miall, D., (1995), 'Representing and Interpreting literature by computer', in *Yearbook of English Studies* 25, 1995, pp. 199-212.
<http://www.ualberta.ca/~dmiall/complit.htm>.
- Michea, R., (1964), 'Les Vocabulaires fondamentaux', in *Recherche et techniques nouvelles au service de l'enseignement des langues vivantes*, Strasbourg, Université de Strasbourg, pp. 21-36.
- Minsky, M. (1968-1995), 'Matter, Mind and Models.' Reprinted in Minsky, M. (ed.), (1968), *Semantic Information Processing*, Cambridge-London, The MIT Press.
<http://www.ai.mit.edu/people/minsky/papers/MatterMindModels.html>.
- Monachini, M., and Calzolari, N., (1996), *Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons and Corpora. A Common Proposal and Applications to European Languages*, EAGLES document EAG-LWG/Morphsyn.
- Moreau, R., (1962), 'Au sujet de l'utilisation de la notion de fréquence en linguistique', *Cahiers de Lexicologie*, 3, pp. 140-158.
- Muller, Ch., (1964), *Initiation à la Statistique Linguistique*, Paris, Larousse.
- Nagao, M., (1989), *Machine Translation: How Far can it Go?*, Oxford, Oxford University Press.
- Nelson, T. H., (1992), 'Opening Hypertext: A Memoir', in Tuman, M. C. (ed., 1992), pp. 43-58.
- Nencioni, G., (1990), 'The Accademia della Crusca: New Perspectives in Lexicography', *Computers and the Humanities*, 24, pp. 345-352.
- NERC Final Report, (1993), Pisa, Istituto di Linguistica Computazionale.
- NERC2 Report, (1993), Pisa, Istituto di Linguistica Computazionale.
- Norling-Christensen, O., (1996), *Design and Composition of Reusable Harmonized Written Language Reference Corpora for European Languages*, PAROLE MLAP: 63-386 document, WP 4 - Task 1.1.
- Norman, D. A., (1993), *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*, Reading (Mass.), Addison-Wesley.
- Nunberg, G. (ed.), (1996), *The Future of the Book*, Berkeley, University of California Press.
- O'Donnell, W., and Thrush, E. A., (1996), 'Designing a hypertext Edition of a Modern Poem', in Finneran, R. J. (ed., 1996), pp. 193-212.
- Olivetto, G., (1998), 'Las ediciones de *Celestina* de la colección Foulché-Delbosc en la Biblioteca Nacional de la República Argentina', *Celestinesca*, 22.1, pp. 67-74.
- Ong, W. J., (1982), *Orality and Literacy: The Technologizing of the Word*, London-New York, Methuen.
- Orlandi, T., (1987), 'Informatica Umanistica. Riflessioni storiche e metodologiche, con due esempi', in Gigliozzi, G. (ed., 1987), pp. 1-37.
- Orlandi, T., (1990), *Informatica umanistica*, Rome, La Nuova Italia Scientifica.
- Orlandi, T., (1998), 'Teoria e prassi della codifica dei manoscritti', in Picone, M., and Cazalé Bérard, C. (eds.), *Gli Zibaldoni di Boccaccio. Memoria, scrittura, riscrittura*,

- Proceedings of the International Seminar, Florence-Certaldo 26-28 April 1996*, Florence, Franco Cesati Editore.
- Panofsky, E. (1927), 'Die Perspektive als "symbolische Form"', *Vorträge der Bibliothek Warburg*, Leipzig-Berlin, Teubner.
- Papp, F., and Szépe, G., (eds.), (1976), 'Papers in Computational Linguistics', *Proceedings of the 3rd International Meeting on Computational Linguistics, Debrecen, 1971*, Budapest, Akadémiai Kiado.
- Paxson, J. J., (1994), *The poetics of personification*, in Macksey, R., and Sprinker, M. (eds.), *Literature, Culture, Theory*, Cambridge, Cambridge University Press.
- Penge, S., (1996), *Storia di un ipertesto. Leggere, scrivere, pensare in forma di rete*, Florence, La Nuova Italia, 1996, p. 105.
- Perrin, M., (1998), 'What Cognitive Science Tells us About the Use of New Technologies', *Conference on Languages for Specific Purposes and Academic Purposes – Integrating Theory into Practice*, Dublin, 6-8 March.
- Picchi, E., (1997), *DBT 3 Data Base Testuale*, Roma - Pisa, Lexis Progetti Editoriali – Consiglio Nazionale delle Ricerche.
- Pichler, A., (1996a), 'Advantages of a Machine-Readable Version of Wittgenstein's Nachlass', in Johannessen, K. S., and Nordenstam, T. (eds.), *Wittgenstein and the Philosophy of Culture, Proceedings of the 18th International Wittgenstein-Symposium, Kirchberg am Wechsel*, Vienna, The Austrian Ludwig Wittgenstein Society.
- Pichler, A., (1996b), 'Transcriptions, Texts and Interpretations', in Johannessen, K. S., and Nordenstam, T. (eds.), *Wittgenstein and the Philosophy of Culture, Proceedings of the 18th International Wittgenstein-Symposium, Kirchberg am Wechsel*, Vienna, The Austrian Ludwig Wittgenstein Society.
- Putnam, H., (1981), *Reason, Truth and History*, Cambridge, Cambridge University Press.
- Putnam, H., (1988), *Representation and Reality*, Cambridge-London, The MIT Press.
- Quemada, B., (1961), 'Introduction', *Cahiers de Lexicologie*, 3, 1, pp. 13-18.
- Quemada, B., (1987), 'Notes sur la lexicographie et la dictionnaire', *Cahiers de Lexicologie*, 51, 2, pp. 229-242.
- Quine, W. van O., (1953), 'On What There Is', in *From a Logical Point of View*, Cambridge (Mass.), Harvard University Press.
- Quirk, R., and Greenbaum, S., (1975), *A University Grammar of English*, London, Longman.
- Reid, B., (1980), 'A High-Level Approach to Computer Document Formatting', in *Proceedings of the 7th Annual ACM Symposium on Programming Languages*, New York, ACM Press.
- 'RELATing to Resources', (1993), *Elsnews Bulletin*, 2, p. 5.
- Remarks prepared for VicePresident Al Gore*, (1998), 15th International ITU Conference, October 1998.
- Renear, A., (1992), 'Representing Text on the Computer: Lessons for and from Philosophy', *Bulletin of the John Rylands University Library of Manchester*, 74, pp. 221-48.
- Renear, A., (1994), 'Theory and Practice: The Textbase Methodology of the Brown Women Writers Project', *SCMLA The Journal of the South Central Division of the Modern Language Association*, 11, 3.
- Renear, A., (1996a), 'Practical Ontology: The Case of Written Communication', in Johannessen, K. S., and Nordenstam, T. (eds.), *Wittgenstein and the Philosophy of*

- Culture, Proceedings of the 18th International Wittgenstein-Symposium, Kirchberg am Wechsel*, Vienna, The Austrian Ludwig Wittgenstein Society.
- Renear, A., (1996b), 'Theory and Metatheory in the Development of Text Encoding', target paper discussed electronically in the Spring of 1996 under the auspices of *The Monist*. <http://hhobel.phl.univie.ac.at/mii/pesp.html>.
- Renear, A., Durand, D., and Mylonas, E., (1996), 'Refining Our Notion of What Text Really Is', in Ide, N., and Hockey, S. (eds.), *Research in Humanities Computing*, Oxford, Oxford University Press. Originally presented at ACH/ALLC 1992, at Christ Church Oxford.
- Renzi, L. (ed.), (1989), *Grande grammatica italiana di consultazione I: La frase. I sintagmi nominale e preposizionale*, Bologna, Il Mulino.
- Ridolfi, P., and Piraino, P. (eds.), (1997), *Trattamento automatico della lingua nella Società dell'Informazione*, Roma, ISPT.
- Riffaterre, M., (1994), 'How Do Images Signify?', *diacritics*, 24.1 (Spring 1994), pp. 3-15.
- Riva, M., (1997), *The Decameron Web Newsletter*, 0. http://www.brown.edu/Departments/Italian_Studies/dweb/dweb.shtml
- Robertson, G. G., and Mackinlay, J. D., (1993), 'The Document Lens', in *Proceedings of ACM UIST '93*, New York, ACM Press, pp. 101-108.
- Robey, D., (1980), 'The Stylistic Analysis of the *Divine Comedy* in Modern Literary Scholarship', in Grayson, C. (ed.), *The World of Dante*, Oxford, Oxford University Press, pp. 198-237.
- Robey, D., (1984), 'Language and Style in the *Divine Comedy*', *Romance Studies*, V, pp. 112-227.
- Robey, D., (1985), 'Literary theory and critical practice in Italy: formalist, structuralist and semiotic approaches to the *Divine Comedy*', *Comparative Criticism*, VII, pp. 73-103.
- Robey, D., (1987), 'Sound and sense in the *Divine Comedy*', *Literary and Linguistic Computing*, 2, 2, pp. 108-115.
- Robey, D., (1988), 'Alliterations in Dante, Petrarch, and Tasso: a computer analysis', in Hainsworth, P. R. J., Lucchesi, V., Roaf, E. C. M., Robey, D., and Woodhouse, J. R. (eds.), *The Languages of Literature in Renaissance Italy*, Oxford University Press, pp. 169-89.
- Robey, D., (1990), 'Rhymes in the Renaissance Epic: A Computer Analysis of Pulci, Boiardo, Ariosto and Tasso', *Romance Studies*, XVII, pp. 97-111.
- Robey, D., (1997a), 'The *Fiore* and the *Comedy*: some computerized comparisons', in Baranski, Z., and Boyde, P. (eds.), *The 'Fiore' in Context. Dante, France, Tuscany*, Notre Dame (Ind.), University of Notre Dame Press, pp. 109-134.
- Robey, D., (1997b), 'Rhythm and Metre in the *Divine Comedy*', in Baranski, Z., and Pertile, L. (eds.), *In amicizia. Essays in honour of Giulio Lepschy*, special supplement to *The Italianist*, XVII, pp. 100-116.
- Robinson, P. M. W., (1997), 'New Directions in Critical Editing', in Sutherland, K. (ed. 1997), pp. 145-172.
- Ross, C. L., (1996), 'The Electronic Text and the Death of the Critical Edition', in Finneran, R. J. (ed., 1996), pp. 224-249.
- Rowald, P., (1914), *Repertorium lateinischer Wörterzeichnisse und Speziallexica*, Leipzig.
- Salas, H., (1997), *Biblioteca Nacional Argentina*, Buenos Aires, Manrique Zago ediciones, pp. 80-82.
- Salgado, O. N., (1992), 'Buenos Aires, Bibliothèque Nationale (Mexico 564, 1097 Buenos Aires, Argentine). Fonds Raymond Foulché-Delbosc', *Nouvelles du Livre Ancien*, 71, pp. 5-6.

- Salton, G., (1989), *Automatic Text Processing - The Transformation, Analysis, and Retrieval of Information by Computer*, Reading (Mass.), Addison-Wesley.
- Sampson, G., (1993), 'The Need for Grammatical Stocktaking', *Literary and Linguistic Computing*, 8, pp. 267-273.
- Sarkar, M., Snibbe, S. S., Tversky, O. J., and Reiss, S. P., (1993), 'Stretching the Rubber Sheet: A Metaphor for Viewing Large Layouts on Small Screens', in *Proceedings of ACM UIST '93*, New York, ACM Press, pp. 81-91.
- Sarkar, M., Brown, M. H., (1994), 'Graphical Fisheye Views', *Communications of the ACM*, 37, 12, New York, ACM Press, pp. 73-84.
- Schaffer, D., Zuo, Z., Greenberg, S., Bartram, S., Dill, J., Dubs, S., and Roseman, M., (1996), 'Navigating hierarchically clustered networks through fisheye and full-zoom methods', *ACM Transactions on Computer-Human Interaction*, 3, 2, pp. 162-188.
- Schöne, H., (1907), *Repertorium griechischer Wörterzeichnisse und Speziallexica*, Leipzig.
- Shannon, C. E., and Weaver, W., (1993), *La teoria matematica delle comunicazioni*, Milan, Etas (*The mathematical theory of communication*, Urbana, University of Illinois Press, 1949).
- Shillingsburg, P., (1996), *Scholarly Editing in the Computer Age: Theory and Practice*, Ann Arbor, University of Michigan Press.
- Shklovsky, V., (1925), *O teorii prozy*, Moscow; It. trans. *Teoria della prosa*, Turin, Einaudi, 1976.
- Simone, R., (1996), 'Italiano sì, ma quale?', *Italiano & Oltre*, XI, pp. 260-261.
- Sinclair, J. M., et al., (1987a), *Cobuild Dictionary of English Language*, London, Collins.
- Sinclair, J. M. (ed.), (1987b), *Looking-up: An Account of the COBUILD Project*, London, Collins.
- Sinclair, J. M., (1991), *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- Sinclair, J. M., (1992), 'The automatic analysis of corpora', in Svartvik, J. (ed.), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991*, Berlin and New York, Mouton de Gruyter, pp. 379-397.
- Sinclair, J. M., (1994), *Corpus Typology*, EAGLES document EAG-CWG/Corptyp.
- Sinclair, J. M., Hoelter, M., and Peters, C. (eds.), (1995), *The Languages of Definition: The Formalization of Dictionary Definitions for Natural Language Processing*, Studies in Machine Translation and Natural Language Processing, Brussels-Luxembourg, Office for Official Publications of the European Communities.
- Smith, J., (1973), 'Ideals Versus Practicalities in Linguistic Data Processing', in Zampolli, A., and Calzolari, N. (eds., 1973), V, II, 2, pp. 895-898.
- Spence, R., and Apperley, M., (1982), 'Data base navigation: an office environment for the professional', *Behaviour and Information Technology*, 1, 1, pp. 43-54.
- Sperberg-McQueen, C. M., and Burnard, L. (eds.), (1994), *Guidelines for the Encoding and Interchange of Electronic Texts (TEI P3)*, Chicago and Oxford, ACH, ACL, ALLC.
- Sperberg-McQueen, C. M., (1991), 'Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts', *Literary and Linguistic Computing*, 6, 1, pp. 34-46.
- Sperberg-McQueen, C. M., (1994), *Textual Criticism and the Text Encoding Initiative*, paper presented at *MLA 1994 conference*, San Diego, December 1994.
<http://www.tei-c.org/Vault/XX/mla94.html>.

- Steinert-Schmitz, S., (1998), 'Analyse sprachlicher und inhaltlicher Besonderheiten bei der Referenz auf Frauen im *Corriere della Sera*', Seminararbeit, Gerhard-Mercator-Universität GH Duisburg.
- Sterne, L., (1997), *The Life and Opinions of Tristram Shandy, Gentleman*, London, Penguin.
- Stokes, R. (ed.), (1931-1967), *Esdaile's Manual of Bibliography*, London, Allen & Unwin, pp. 32-49.
- Stoppelli, P., and Picchi, E. (eds.), (1998), *LIZ. Letteratura Italiana Zanichelli*, Bologna, Zanichelli. [CD-ROM]
- Stubbs, M., (1993), 'British Traditions in Text Analysis. From Firth to Sinclair', in Baker, M., Francis, G., Tognini-Bonelli, E. (eds.), *Text and Technology: In Honour of John Sinclair*, Philadelphia and Amsterdam, John Benjamins, pp. 1-33.
- Sutherland, K., (1996), 'Looking and Knowing: Textual Encounters of a Postponed Kind', in Chernaik, W., Deegan, M., and Gibson, A. (eds.), (1996), pp. 11-22.
- Sutherland, K. (ed.), (1997), *Electronic Text: Investigations in Method and Theory*, Oxford, Oxford Clarendon Press.
- Svartvik, J., (1992), 'Corpus linguistics comes of age', in Svartvik, J. (ed.), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991*, Berlin and New York, Mouton de Gruyter, pp. 7-13.
- Table ronde sur les grands dictionnaires historiques*, (1973), Florence, Olschki.
- Tanselle, T. G., (1981), 'Textual Scholarship', in Gibaldi, J. (ed.), *Introduction to Scholarship in Modern Languages and Literatures*, New York, The Modern Language Association.
- Thaller, M., (1992), *Images and Manuscripts in Historical Computing*, St. Katharinen, Max Planck-Institute für Geschichte, Scripta Mercature Verlag.
- Thaller, M., (1993), *Kleio: a database system*, St. Katharinen, Max Planck-Institute für Geschichte, Scripta Mercature Verlag.
- The Chicago Manual of Style*, (1982), Chicago, University of Chicago Press.
- Thüring, M., Hannemann, J., and Haake, J., (1995), 'Hypermedia and Cognition: Designing for Comprehension', *Communications of the ACM*, 38, 8, pp. 57-66.
- Tip top. Il mistero dei libri scomparsi*, (1996), Milan, Rizzoli New Media – MediAround. [CD ROM]
- 'Toward the Fifth Framework Programme', (1998), in ANITE (1998).
- Trubetzkoy, N. S., (1968), 'Grundzüge der Phonologie', *Travaux du Circle Linguistique de Prague*, 7.
- Tuman, M. C. (ed.), (1992), *Literacy Online: The Promise (and Peril) of Reading and Writing with Computers*, Pittsburgh and London, University of Pittsburgh Press.
- Varile, G.B., and Zampolli, A. (managing eds.); Cole, R., Mariani, J., Uszkoreit, H., Zaenen, A., and Zue, V. (editorial board), (1997), *Survey of the State of the Art in written and Spoken Language Processing*. Sponsored by the Commission of the European Union and the National Science Foundation of the USA, Pisa, Giardina.
- Varile, G.B., and Zampolli, A. (eds.), (1992), *Synopses of American, European and Japanese Projects Presented at the International Projects Day at COLING 1992*, (Linguistica Computazionale VIII), Pisa, Giardini Editori.
- Vauquois, B., (1975), *La Traduction automatique à Grenoble*, Paris, Dunod.
- Vidal-Beneyto, J., (1986), 'Presentation', *Encrages. Special issue dedicated to 'Les Industries de la langue enjeux pour l'Europe'*, *Actes du colloque de Tours*, 5-7, Saint-Denis, Université Paris VIII-Vincennes, pp. i-vii.

- Vidal-Beneyto, J. (ed.), (1991), *Las industrias de la lengua*, Madrid, Ediciones Pirámide.
- Voghera, M., (1993), 'Le variabili testuali e pragmatiche', in De Mauro, T., Mancini, F., Vedovelli, M., Voghera, M., *Lessico di frequenza dell'italiano parlato: Ricerca a cura dell'Osservatorio linguistico e culturale italiano OLCI dell'Università di Roma "La Sapienza"*, Milan, Fondazione IBM Italia – Etas, pp. 32-38.
- Walker, D., and Zampolli, A., (1989), 'Foreword', in Boguraev, B., and Briscoe, T. (eds., 1989), pp. xiii-xiv.
- Walker, D., Zampolli, A., and Calzolari, N. (eds.), (1987), *Towards a Polytheoretical Lexical Data Base*, Pisa, Istituto di Linguistica Computazionale.
- Walker, D., Zampolli, A., and Calzolari, N. (eds.), (1995), *Automating the Lexicon. Research and Practice in a Multilingual Environment*, Oxford, Oxford University Press.
- Weiner, E., (1994), 'The Lexicographic Workstation and the Scholarly Dictionary', in Atkins, B. T. S., and Zampolli, A. (eds., 1994), pp. 413-438.
- Weissberg, J.-L., (1992), 'Réel et virtuel', *Futur Antérieur*, 11, pp. 18-28.
- Wellek, R., and Warren, A., (1956), *Theory of Literature*, New York, Harcourt Brace.
- Whitelock, P., Wood, M., Somers, H., Johnson, E., and Bennett, P. (eds.), (1987), *Linguistic Theory and Computer Applications*, New York, Academic Press.
- Wistrand, E., (1996), 'Creating Image Context using ImageTrees', in *ACM SIGCHI '96 Conference Companion*, New York, ACM Press, pp. 281-282.
- Zampolli, A., (1968), 'Projet pour un lexique électronique de l'italien', in Busa, R. (ed., 1968), pp. 109-126.
- Zampolli, A., (1970), 'Cronaca: La terza *International Conference on Computational Linguistics*', Stoccolma, 1969, *Archivio Glottologico Italiano*, LV, 1-2, pp. 272-279.
- Zampolli, A., (1973a), 'Humanities Computing in Italy', *Computers and the Humanities*, 7, 6, pp. 343-360.
- Zampolli, A. (1973b), 'Introduction', in Zampolli A., Calzolari, N. (eds., 1973-1977), pp. xix-xxviii.
- Zampolli, A., (1973c), 'L'Automatisation de la recherche lexicologique: état actuel et tendances nouvelles', *META* 18, 1-2, pp. 101-136.
- Zampolli, A., (1975), 'L'elaborazione elettronica dei dati linguistici: stato delle ricerche e prospettive', in *Accademia Nazionale dei Lincei* (1975), pp. 23-107.
- Zampolli, A., (1976), 'Les dépouillements électroniques: quelques problèmes de méthode et d'organisation', in Fattori, M., Bianchi, M. (eds., 1976), pp. 173-178.
- Zampolli, A., (1983), 'Lexicological and Lexicographical Activities at the Istituto di Linguistica Computazionale', in Zampolli, A., and Cappelli, A. (eds., 1983), 237-278.
- Zampolli, A., (1987), 'Perspectives for an Italian Multifunctional Lexical Database', in Zampolli, A., Cappelli, A., Cignoni, L., and Peters, C. (eds., 1987), pp. 304-41.
- Zampolli, A., (1989), 'Introduction to the Special Section on Machine Translation', *Literary and Linguistic Computing*, 4, 3, pp. 182-184.
- Zampolli, A., (1991a), 'Bases multifuncionales de datos léxicos', in Vidal-Beneyto, J. (ed., 1991), pp. 127-146.
- Zampolli, A., (1991b), 'Corpora de Referencia', in Vidal-Beneyto, J. (ed., 1991), pp. 119-124.
- Zampolli, A., (1991c), *Preliminary Considerations on the Constitution of an ELTA* (European Language Technology Agency), Pisa, Document prepared for the DG XIII.

- Zampolli, A. (1991d), 'Technology and Linguistic Research', in Katzen, M. (ed., 1991), pp. 21-51.
- Zampolli, A., (1994a), 'Introduction', in Atkins B. T. S., and Zampolli, A. (eds., 1994), pp. 3-15.
- Zampolli, A., (1994b), 'Preface', in Zampolli, A., Calzolari, N., and Palmer, M. (eds., 1994), pp. ix-xv.
- Zampolli, A., (1997), 'Introduzione', in Ridolfi, P., Piraino, R. (eds., 1997), pp. 7-14.
- Zampolli, A. (ed.), (1973), *Linguistica Matematica e Calcolatori, Atti del Convegno e della Prima Scuola Internazionale, Pisa, 16 agosto - 6 settembre 1970*, Florence, Olschki.
- Zampolli, A. (ed.), (1977), *Linguistic Structures Processing*, Amsterdam, North-Holland.
- Zampolli, A., and Calzolari, N. (eds.), (1973-1977), *Computational and Mathematical Linguistics, Proceedings of the International Conference on Computational Linguistics 1973*, Florence, Olschki.
- Zampolli, A., Calzolari, N., Baker, M., and Krut, J. G., (1995), *Toward a Network of European Reference Corpora*, (Special Issue of *Linguistica Computazionale*, XI), Pisa, Giardini.
- Zampolli, A., Calzolari, N., Palmer, M. (eds., 1994), *Current Issues in Computational Linguistics: In Honour of Don Walker*, (*Linguistica Computazionale IX-X*), Pisa, Giardini – Dordrecht, Kluwer Academic Publisher.
- Zampolli, A., and Cappelli A. (eds.), (1983), *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries, Proceedings of the European Science Foundation Workshop, Pisa, 1981*, (*Linguistica Computazionale III*), Pisa, Giardini Editori.
- Zampolli, A., Cappelli, A., Cignoni, L., and Peters, C. (eds.), (1987), *Studies in Honour of Roberto Busa S.J.*, (*Linguistica Computazionale IV-V*), Pisa, Giardini.
- Zampolli, A., and Quemada, B., (1981), *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries*, Final Report on the ESF Pisa Workshop, presented to the ESF (unpublished).
- Zampolli, A., and Walker, D., (1986), *Multilingual Lexicology and Lexicography: New Directions*, paper presented at the CG12-TSC of the EC (unpublished).
- Zampolli, A., and Walker, D., (1987), *Report on the Grosseto Workshop*, Pisa (unpublished).
- Zampolli, A., with consultation of Spang-Hansen, H., Perschke, S., Cerdà, R., (1988), *Reusability of lexical resources*, Document CG12/159/88/b presented to the CGC-12 (*Comité de Gestion et Consultation*), Luxembourg (unpublished).
- Zipf, G. K., (1935), *The Psycho-biology of Language: An Introduction to Dynamic Biology*, Cambridge (Mass.), The MIT Press.

Appendix B. Notes on Contributors

Staffan Björk works at the PLAY studio of the Interactive Institute in Sweden. He has conducted research on user interfaces since 1998, specializing in the research field of Information Visualization, and presented the Ph.D. thesis *Flip Zooming* in the autumn of 2000. His current research interests includes information visualization, human-computer interfaces, context-aware computing, ubiquitous computing, and interactive narratives.

Lou Burnard is manager of the Humanities Computing Unit at Oxford University and European editor of the Text Encoding Initiative. He has worked at the interface between literature, linguistics, and communications and information technologies since the late seventies. His most recent publication is *The BNC Handbook*, jointly written with Guy Aston.

Elisabeth Burr is Senior Lecturer in Romance Linguistics at the Universität Duisburg. Her main research areas are corpus linguistics of Romance languages, grammaticography (French, Italian, Portuguese, Spanish), feminist linguistics, language policy and normalisation. Her most recent book is her Habilschrift, *Wiederholte Rede und idiomatische Kompetenz. Französisch, Italienisch, Spanisch*. She is currently working on a corpus of Romance newspaper languages (Italian, French, Spanish).

Licia Calvi is Assistant Professor at the Department of Computing Science at the Eindhoven University of Technology. Her research research interests include multimedia for cultural heritage, adaptive hypermedia systems for distant learning, philosophy of culture and digital media and art.

Fabio Ciotti is Lecturer in Humanities Computing at the University of Parma. Since 1996, together with Marco Calvo, Gino Roncaglia and Marco Zela, he has been producing the most popular Italian guides to the Internet. The latest volume in the series is *Frontiere di rete*, published by Laterza, Rome, 2001.

Domenico Fiormonte teaches Teoria e tecniche della comunicazione at the University of Bologna, Faculty of Conservazione dei Beni Culturali. With Ferdinanda Cremascoli, he is the author of the volume *Manuale di scrittura*, Turin, Bollati Boringhieri, 1998.

Giuseppe Gigliozzi is Professor of Italian literature at the University of Rome La Sapienza and Director of the Centro Ricerche Informatica e Letteratura. His research interests centre on narrativity, text analysis and encoding, and literary theory. He is author of *Il testo e il computer*, Milano, Bruno Mondadori, 1997.

Massimo Guerrieri works as project assistant at the Centro Ricerche Informatica e Letteratura (University of Rome La Sapienza) and at the University of Rome III. His main research interest is the application of new technologies to the analysis, studying, and teaching of contemporary Italian literature.

Francisco Marcos Marín is Professor of Linguistics at the Universidad Autónoma of Madrid. He has published numerous works in the field of Computers and the Humanities since 1971. He is also editor of Admyte®, a series of advanced Cd-Roms of digitalised manuscripts and incunabula of medieval Spanish literature (1992, 1993, 1998).

Willard McCarty is Senior lecturer at the King's College Centre for Computing in the Humanities. He is editor of *Humanist*, and Vice-President of the Association for Computers and the Humanities. Among his various digital projects there is the Analytical Onomasticon, a printed and electronic reference work to all devices of language by which persons are named in the *Metamorphoses* of Ovid.

Federico Pellizzi teaches Literature and computing at the University of Bologna, and Literary theory and methodology at the Iulm University (Milan). He is the editor of *Bollettino '900* and has organised three international computers and literature conferences at the University of Bologna. He collaborates with several online research projects, including *The Narrated Dream in Modern Literature* and *Italian Culture on the Net*.

Allen Renear is Associate Professor at the Graduate School of Library and Information Science, University of Illinois at Urbana/Champaign. He

was Director of the Scholarly Technology Group (Brown University) from its founding and maintains an active affiliation with STG as Senior Technology Consultant. His publications are focused on the interplay between technology, methodology, and theory in the humanities.

David Robey is Professor of Italian and Head of the School of Modern Languages at the University of Reading. He has published on 15th-century humanism, language and style in Dante and Renaissance narrative poetry, the computer analysis of literature, and modern critical theory. He has recently completed a computer-based study on Sound and Structure in Dante's *Divine Comedy* (Oxford University Press, 2000), and is joint editor of the Oxford Companion to Italian Literature.

Jonathan Usher is Professor of Italian at the University of Edinburgh. He teaches in the medieval, renaissance and modern periods, and researches narrative technique in medieval and modern literature. He is author of the chapter *Origins and duecento* in the Cambridge History of Italian Literature.

Claire Warwick is a lecturer in Electronic Publishing at the University of Sheffield Department of Information Studies, and is programme co-ordinator for the MA in electronic Journalism, run jointly with the Department of Journalism. Her research interests are concerned with electronic text ending and humanities computing, with particular emphasis on a user perspective. She has previously worked for the Humanities Computing Unit and the Faculty of English at Oxford University.

Antonio Zampolli is Professor of Computational Linguistics at the University of Pisa and Director of the Istituto di Linguistica Computazionale of the CNR. He has been working in the field of Computational Linguistics since 1967 and is responsible for a number of European projects related to digital language resources.