Andrea Bozzi

Percorsi di linguistica e di filologia computazionali

a cura di
Maria Sofia Corradini Bozzi e Giacomo Ferrari

testi presentati da
Sylvie Calabretto, Cristina D'Ancona, Giacomo Ferrari,
Valeria Lomanto, Elton Prifti

Edizioni ETS

Edizioni ETS
Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa
info@edizioniets.com
www.edizioniets.com

Distribuzione
Messaggerie Libri SPA
Sede legale: via G. Verdi 8 - 20090 Assago (MI)

Promozione
PDE PROMOZIONE SRL
via Zago 2/2 - 40128 Bologna

ISBN 978-884675551-3

INTRODUZIONE

Maria Sofia Corradini Bozzi
Università di Pisa

Questo libro è nato dall'intento di festeggiare i settant'anni di Andrea Bozzi con la presentazione di una raccolta dei suoi scritti più significativi nel settore delle *digital humanities*. Si tratta di un momento di vita privata, ma anche di un'occasione per mettere in luce gli aspetti principali delle ricerche e dei progetti da lui concepiti nel corso del periodo lavorativo svolto al Consiglio Nazionale delle Ricerche a partire dalla metà degli anni Settanta. Alcuni di questi detengono un carattere innovativo, non solo se considerati in rapporto all'arco temporale in cui hanno visto la luce, ma anche nel contesto contemporaneo, dal momento che riflessioni metodologiche su cui essi si fondavano e ideazione di strumenti informatici che ne rappresentavano la finalità primaria sono indubbiamente la base fondante di alcune linee di ricerca di àmbito linguistico e filologico perseguite attualmente in ambienti differenti.

Senza voler andare indietro nel tempo agli anni Cinquanta e Sessanta del secolo scorso, quando furono avviati i primi lavori di analisi lessicale compiuta con mezzi elettronici[1], è attorno alla metà degli anni Settanta[2] che l'informatica ha incominciato ad imporsi nel campo delle scienze umane come uno strumento di lavoro necessario, diventato ormai imprescindibile, sia che si tratti della costituzione di *corpora* e di biblioteche digitali, oppure dello sviluppo di strumenti per la gestione e l'interrogazione dei dati.

Da principio le attività di base furono essenzialmente quelle legate alla conver-

[1] Nell'àmbito dell'automazione applicata all'analisi lessicale della lingua latina non è possibile non ricordare che già nel 1951 Padre Roberto Busa diede notizia dell'inizio del suo lavoro sull'*Index Thomisticus* (R. BUSA, *Sancti Thomae Aquinatis Himnorum ritualium varia specimina concordantiarum. Primo saggio di indici di parole automaticamente composti e stampati da macchine IBM a schede perforate*, Fratelli Bocca, Milano, 1951); nel settore della francesistica, invece, alla fine degli anni Cinquanta Bernard Quemada si fece promotore del trattamento automatico dei *corpora*, argomento affrontato soprattutto nei «Cahiers de lexicologie», da lui fondati nel 1959. Fondamentale importanza ebbe il colloquio svoltosi nel 1961 a Besançon, dove vennero trattate le problematiche relative alla meccanizzazione delle ricerche lessicografiche («Cahiers de lexicologie» 3, 1962).

[2] Nel 1975, per es., si svolse presso l'École française di Roma una tavola rotonda finalizzata a vagliare le possibilità dell'introduzione di metodi informatici nell'àmbito della medievalista e i cui punti basilari furono formalizzati nel primo numero de «Le Médiéviste et l'Ordinateur» del 1979; si vedano anche le considerazioni espresse da D'Arco Silvio Avalle sull'utilizzo dell'elaboratore elettronico per lavori su opere letterarie in D.S. AVALLE, *Al servizio del vocabolario della lingua italiana*, Accademia Della Crusca, Firenze, 1979.

sione di documenti in forma leggibile da calcolatore al fine di produrre spogli elettronici di un testo o di più testi costituenti un *corpus*; tale utilizzo di sistemi informatici in lessicografia sottintendeva tutta una serie di problematiche[3], fra cui quella della compatibilità fra i differenti formati dei testi da sottoporre ad analisi e dei diversi modi di registrazione, che nel corso degli anni Ottanta fu oggetto di discussione specialmente nell'ambiente della Association for History and Computing, in un gruppo di studio animato da Manfred Thaller[4].

Un settore strettamente connesso col precedente è quello che si rivolge al processo di costituzione dei *corpora* in quanto tali, mediante conversione su supporto digitale del patrimonio documentario. La tecnologia per i beni librari è nata in seno alle istituzioni statunitensi, a partire dai progetti concepiti nell'àmbito della Library of Congress (*Optical Disk Pilot Project*, iniziato nel 1982) e della National Library of Medicine (attività iniziata nei primi anni '80, che continua ancor oggi il processo di digitalizzazione) e poi ancora di NSF (National Science Foundation), NASA (National Aeronautical and Space Administration) e DARPA (Defense Advanced Research Projects Agency) al fine di supportare progetti di biblioteche digitali nelle università. In seguito, anche in contesto europeo sono state promosse iniziative di 'politica digitale' orientate alla valorizzazione del patrimonio documentario di singole biblioteche o di reti di biblioteche con finalità plurime, come accesso ai cataloghi, diffusione di documenti multimediali, educazione a distanza, etc.[5]. Nel caso di edizioni antiche, incunabuli, o addirittura manoscritti, la costituzione dei *corpora* può oggi essere condotta anche utilizzando delle tecniche di riconoscimento di caratteri, che sono in via di affinamento.

Un'altra linea di indagine recente, che sollecita riflessioni nel campo delle discipline filologiche, è relativa alla ricerca del miglior utilizzo degli strumenti computazionali allo scopo di presentare in modalità digitale sia le edizioni critiche che le varianti d'autore[6].

Andrea Bozzi, coerentemente alla propria formazione di linguista classico, ha iniziato l'attività scientifica con lavori di tipo lessicografico relativi al greco antico[7],

[3] Considerazioni a tal proposito sono espresse in B. QUEMADA, *Table ronde sur la lexicographie et l'ordinateur orgnisée par la Fondation Européenne pour la Science*, in «Linguistica computazionale» I, 1981.

[4] A Manfred Thaller si deve la messa a punto al Max Planck Institut für Geschichte di Gottinga del programma KLEIO, indirizzato in modo specifico agli storici.

[5] Per l'àmbito italiano, un panorama dell'applicazione delle tecnologie informatiche a banche dati di varie discipline si legge, per es., in M. MORELLI - M. RICCIARDI (ed.), *Le carte della memoria. Archivi e nuove tecnologie*, Bari, Laterza, 1997.

[6] Si vedano convegni recenti come *Ecdotica digitale e nuovi approcci critici del testo* (Liegi, novembre 2018) e *Congrès International du cinquantenaire de l'Institut des textes et manuscrits modernes. La critique génétique comme processus (1968-2018)* (Paris, École normale supérieure, Bibliothèque nationale de France, octobre 2018).

[7] Si rimanda, per es., ad A. BOZZI, *Il Trattato Ippocratico* Sulle arie, le acque e i luoghi *e la sua traduzione latina tardo-antica. Concordanze contrastive con il calcolatore elettronico e commento linguistico-filologico al lessico tecnico latino*, Giardini Editori e Stampatori, Pisa, 1981.

e si è in seguito interessato ai diversi settori connessi all'utilizzo dell'informatica in àmbito umanistico che via via si andavano definendo nell'ambiente internazionale. Una convinzione che egli ha sempre manifestato fin dall'inizio delle attività, e che costituisce un filo conduttore nello sviluppo delle differenti tematiche, è la consapevolezza del fatto che l'adozione di procedure automatiche in nessun modo deve condurre alla perdita di qualsivoglia tipologia di informazioni contenute nei testi da sottoporre ad analisi, le quali devono essere gestite nella loro totalità; nella pratica corrente questo principio non viene sempre rispettato e, dunque, è necessario «impostare il metodo delle attività computazionali nel settore filologico e letterario su una nuova base»[8].

Gli interessi focalizzati sulle lingue latina e greca considerate in àmbito informatico lo hanno condotto ad assumere fin dal 1981 la responsabilità del *Reparto Lessici Automatici* presso l'ILC ed, in seguito, a formalizzare e guidare progetti di respiro internazionale, come quelli con il Max Planck Institut für Geschichte di Gottinga (*Sistema integrato per la formazione di un archivio testuale latino computerizzato: testi e fonti storiche, lettore ottico e repertorio lessicale*, dal 1986)[9], con l'Accademia Bulgara delle Scienze (*Linguistica computazionale*, *Filologia Classica e Studi Filologici*, dal 1988)[10], con l'Università di Madrid (programmi di ricerca sul *Thesaurus Linguae Graecae* in CD al fine di agevolare il lavoro di preparazione delle citazioni da inserire nel vocabolario Greco-Spagnolo, 1989). In tale contesto di progettualità essenzialmente lessicografica, Andrea si è reso immediatamente conto delle notevoli problematiche con le quali ci si può scontrare qualora si voglia lavorare su un corpus testuale eterogeneo, evidenti sia nel versante grafico, per la frequente mancanza di uniformità, sia nel versante della lemmatizzazione, a causa del diverso trattamento delle forme, delle diverse funzioni che esse possono assumere o della mancanza di indicazioni grammaticali o dei princìpi di base seguiti dai curatori.

L'attenzione verso queste tematiche possedeva una connotazione 'pionieristica' e, di fatto, anticipava l'ampio dibattito che si sarebbe sviluppato da lì a qualche anno sulle varie questioni poste dalla standardizzazione dei testi elettronici e dalla constatazione dei risultati talvolta disastrosi conseguenti all'assenza di uno standard impiegato universalmente[11].

[8] A. BOZZI, *Stazione di lavoro computerizzata per la filologia*, in «Nuova Civiltà delle Macchine» 45, 1994/1, Nuova ERI, Roma, 1994, pp. 43-63 (cfr. testo n. 10, in part. p. 130).

[9] Cfr. anche *supra*, nota 4.

[10] Si veda, fra gli altri lavori prodotti in tale contesto, quello che prende in considerazione il problema del recupero e della memorizzazione delle informazioni presenti nell'apparato critico di un testo: A. BOZZI - A. NIKOLOVA - G. CAPPELLI - G. GIULIANI, *Il trattamento delle varianti nello spoglio elettronico di un testo. Una prova sui Carmina di Claudiano*, in «MD. Materiali e discussioni per l'analisi dei testi classici» 16, Giardini Editori e Stampatori, Pisa, 1986, pp. 155-179.

[11] Tali problematiche sono ampiamente discusse, per es., in J.-PH. GENET (éd.), *Standardisation et échange des bases de données historiques. Actes de la troisième Table Ronde Internationale tenue au L.I.S.H. (CNRS), Paris, 1987*, Centre National de la Recherche Scientifique, Paris, 1988.

Già nei lavori del 1982[12], in effetti, come anche Valeria Lomanto sottolinea nella *Presentazione* al Capitolo I, Andrea aveva proposto una metodologia atta ad ovviare alle difficoltà inerenti alla gestione di una base dati eterogenea, cercando di fornire le soluzioni più opportune per superare la disomogeneità dei testi presi in considerazione e, nello stesso tempo, mantenere tutte le indicazioni che le differenti lemmatizzazioni contenevano. Tali considerazioni metodologiche facevano parte integrante del progetto volto alla realizzazione di un *Repertorio Lessicale Automatico della Lingua Latina* (RELAL), presentato in occasione dell'*International workshop on possibilities and limits of the computer in producing and publishing dictionaries* svoltosi a Pisa nel maggio 1981. La formulazione iniziale è stata successivamente convertita in uno schema di progetto più vasto, il *Sistema Informativo Latino* (SIL), articolato in diversi moduli[13], di cui quelli relativi all'analizzatore morfologico[14] e alla lemmatizzazione automatica sono di particolare rilievo.

Il fine era quello di «costruire un sistema che permetta di accedere a materiali eterogenei e sia in grado di raccordarli»[15] e che, inoltre, si connotasse non solo come un semplice repertorio di lemmi, ma come una struttura articolata, comprendente informazioni di varia tipologia (grammaticali, etimologiche, di evoluzione diacronica, di àmbito semantico, etc.), con la possibilità di consentire all'utente «una pluralità di accessi ai dati contenuti nei documenti»[16]. Un altro aspetto fondamentale è la constatazione che «l'operazione della lemmatizzazione, particolarmente gravosa se eseguita a mano su opere di una certa estensione, risulta agevolata se si dispone di un dizionario di macchina capace di riconoscere le forme prese in considerazione»[17].

In tale contesto Andrea, assieme al prof. Nino Marinone[18] e a Giuseppe Cappelli (che ha curato l'aspetto informatico), ha depositato nel 1992 il marchio LEMLAT (*Analizzatore Morfologico Latino*)[19], frutto di un lavoro che è stato il

[12] Cfr. i testi n. 1 e n. 2

[13] I moduli sono elencati a p. 343 di A. BOZZI - G. CAPPELLI, *Un sistema computerizzato per la produzione di indici lessicali di testi latini*, in «MD. Materiali e Discussioni per l'analisi dei testi classici» 20-21, Giardini Editori e Stampatori, Pisa, 1988, pp. 343-360: «*a*. un archivio di testi organizzato in forma di data base; *b*. una serie di moduli computazionali per l'analisi linguistica delle forme flesse (analizzatore morfologico); *c*. un modulo per la lemmatizzazione automatica; *d*. un modulo per la comparazione di testi già lemmatizzati secondo criteri non omogenei fra loro, in vista della creazione di un unico archivio; *e*. un modulo per la redazione delle voci di un vocabolario della Latinità Media e Recenziore».

[14] Cfr. testo n. 4.

[15] Cfr. testo n. 2, p. 30.

[16] Cfr. A. BOZZI, *Sistema per la redazione semiautomatica delle voci*, in M. FATTORI - M. BIANCHI (ed.), *SPIRITUS. IV Colloquio Internazionale del LIE, Roma, 7-8 gennaio 1983*, Edizioni dell'Ateneo, Roma, 1984, pp. 567-577.

[17] Cfr. testo n. 2, p. 29.

[18] Cfr. N. MARINONE, *A project for a Latin lexical data base*, in «Linguistica computazionale» 3, 1983, pp. 175-187; ID., *A project for Latin Lexicography: I. Automatic Lemmatization and Word-list*, in «Computers and the Humanities» XXIV, 1990, pp. 417-420.

[19] Si tratta del brevetto CNR depositato col n. 564244 in data 18/3/1992.

punto di partenza per la realizzazione di una nuova versione (CHLT LEMLAT) da parte di Marco Passarotti[20].

Il prolungato interesse per le lingue classiche considerate in un contesto informatico hanno condotto Andrea anche ad avviare esperimenti di riconoscimento e di collocazione di frammenti di testi antichi conservati su papiro, sulla base della creazione di un programma di consultazione automatizzato di archivi digitali come, per es., quello del T.L.G. (Thesaurus Linguae Graecae)[21].

Fondate su basi completamente differenti, ma tuttavia sempre nell'ottica del restauro di documenti antichi da condursi mediante tecnologie informatiche, sono le ricerche condotte a partire dal 1994 che lo hanno visto responsabile scientifico nell'àmbito del Progetto finalizzato "Beni Culturali" del CNR (*Sistema integrato grafico-linguistico per il restauro computerizzato di documenti manoscritti o a stampa basato su reti neurali*, 1996-1998), o del Progetto speciale CNR *LAperLA: lettore automatico per libri antichi*. Queste attività, descritte da Elton Prifti nella *Presentazione* al Capitolo II[22], lo hanno condotto alla produzione del brevetto *Metodo ed apparato per il riconoscimento automatico di caratteri*[23], in collaborazione con Giuseppe Fedele e Alfredo Eisinberg (curatori dell'aspetto matematico-informatico), finalizzato alla lettura di documenti testuali manoscritti e a stampa in cattivo stato di conservazione.

A partire dalla fine degli anni Novanta, in conseguenza dell'esperienza maturata e dei lavori prodotti nel settore del *Digital Image Processing*, Andrea ha approfondito aspetti metodologici e tecnologici relativi al processo di costituzione e di valorizzazione dei *corpora.* L'attenzione verso questo settore si è concretizzata nella concezione di progetti nei quali la tecnologia digitale detenesse un ruolo fondamentale soprattutto ai fini della fruizione di un patrimonio di rilevante valore storico-culturale com'è il caso, per esempio, dell'informatizzazione della Biblioteca Pallottino, che costituisce un'applicazione di BIBLOS[24]. Ancora in questo settore, il più recente progetto FAD[25] è stato concepito con una duplice finalità:

[20] Si vedano, per es., i seguenti: G. CAPPELLI - M. PASSAROTTI, *LEMLAT: uno strumento computazionale per l'analisi linguistica del latino - sviluppo e prospettive*, in «Euphrosyne» XXXI, 2003, pp. 519-531; M. PASSAROTTI, *Development and perspectives of the latin morphological analyser LEMLAT*, in A. BOZZI - L. CIGNONI - J.-L. LEBRAVE (ed.), *Digital Technology and Philological Disciplines*, in «Linguistica Computazionale» XX-XXI, 2004, pp. 397-414.

[21] Ci si riferisce all'esperienza documentata in A. BOZZI - R. BINDI - S. FORTUNA, *Nuovi frammenti di* P. OXY. 2181 (Platone, Fedone) *identificati con il ricorso all'archivio computerizzato (T.L.G.)*, in «Studi Classici e Orientali» XXXVII, 1987, Giardini Editori e Stampatori, Pisa, pp. 191-203. La descrizione del sistema IBYCUS, nel quale è confluita una parte dell'archivio T.LG., è in BOZZI 1986 (= testo n. 3).

[22] Oltre ai testi raccolti nel Capitolo II, si vedano i lavori citati da Elton Prifti in nota 2.

[23] Il brevetto è stato depositato dal CNR in data 09/11/2001.

[24] BIBLOS (*Biblioteca virtuale degli Organi appartenenti al Comitato per le Scienze storiche, filosofiche e filologiche del CNR*), è nato con lo scopo di offrire i necessari strumenti hardware e software per la distribuzione sulle reti telematiche internazionali delle informazioni e conoscenze prodotte dal settore umanistico del CNR. Cfr. A. BOZZI, *Il progetto BIBLOS e l'informatizzazione della Biblioteca Massimo Pallottino*, in «Archeologia e Calcolatori» 10, 1999, pp. 305-311 (§§. 1 e 2).

[25] FAD (*Fondi e Archivi Digitali*) è stato presentato nel convegno di Firenze del 14 maggio 2014.

gestire moli considerevoli di dati in formato digitale e, tramite un sistema di moduli software, permetterne la consultazione e la fruizione alla comunità degli studiosi. Il secondo aspetto mostra una particolarità di contro al comune approccio alle basi di dati: la possibilità di utilizzare uno strumento innovativo, e cioè un *bloc notes* telematico posizionato sullo schermo, dove trascrivere parole lette sull'immagine del documento, annotare trascrizioni ed osservazioni personali, e salvare i dati in un apposito server.

Altri progetti concepiti all'interno di questo percorso intrapreso con la finalità di valorizzazione dei *corpora*, come BIBLIOFILO[26] e BAMBI (si veda la *Presentazione* di Sylvie Calabretto, Capitolo IV, e i testi n. 19 e n. 20), possiedono un'articolazione maggiore perché inglobano componenti diverse, concepite nel corso degli anni come moduli indipendenti facilmente integrabili nel sistema di base e volti a costituire la "Stazione filologica multimodulare" (SFM).

Ideata a partire da esperienze svolte su prototipi realizzati su manoscritti medievali in lingua latina ed occitanica, su libri a stampa antichi e su manoscritti di autori moderni e contemporanei, la "Stazione filologica multimodulare" è stata descritta per la prima volta nel 1993, nel numero 29 della «Revue Informatique et Statistique dans les Sciences humaines» dell'Università di Liegi[27]. Le diverse pubblicazioni che la riguardano (Capitolo III) sono presentate da Giacomo Ferrari, che ne mette in evidenza le riflessioni metodologiche ad essa sottese e i cambiamenti occorsi nel tempo. Qui si può ricordare che la SFM comprendeva in un primo momento un modulo di 'concordanza', un modulo di 'apparato critico', contenente le informazioni relative agli elementi extratestuali, ed un modulo 'grafico', progettato per produrre la corrispondenza fra il testo del documento originale e la trascrizione prodotta dallo studioso, con la possibilità di intervenire anche sulla qualità dell'immagine[28]. Tale modulo, incentrato sul rapporto testo/immagine, superava già di fatto la criticità che sarà ancora percepita alcuni anni dopo in occasione della giornata di studi su *La numérisation des manuscrits*

Finanziato dal ministero dei Beni e delle attività culturali e coordinato dalla fondazione Primo Conti di Fiesole, il progetto raccoglie circa 270.000 dati e un sistema di moduli software realizzati dalla società Meta, che permette la consultazione delle raccolte (sovente inedite) del Gabinetto scientifico letterario G.P. Vieusseux (come i preziosi documenti delle avanguardie artistico-letterarie del primo '900), dell'Istituto Papirologico G. Vitelli (manoscritti sugli scavi archeologici nella zona della città di Antinoe in Egitto), delle fondazioni Conti di Firenze e Rosselli di Torino (epistolari dei fratelli antifascisti Carlo e Nello Rosselli).

[26] BIBLIOFILO (*Workstation Filologica Multimodulare*) è un progetto MURST (2000) realizzato nell'àmbito del Programma Nazionale di Ricerca sui Beni Culturali.

[27] A. BOZZI, *Towards a Philological Workstation*, in «Revue Informatique et Statistique dans les Sciences humaines» 29, 1993, Université de Liège, Liège, pp. 33-49.

[28] Si vedano i testi n. 10 e n. 11, ed anche: A. BOZZI - A. SAPUPPO, *Word-Image Concordance in a Philological Workstation Project*, in «Computers & Texts» 8, 1994, Office for Humanities Communication, Oxford University Computing Services, Oxford, pp. 8-10; ID., *Word-image linkage in the computerized analysis of old printed dictionaries*, in O. BOONSTRA - G. COLLENTEUR - B. VAN ELDEREN (eds.), *Structures and Contingencies in Computerized Historical Research. Proceedings of the IXth International Conference of the Association for History and Computing. Nijmegen 1994*, Cahier VGI 9, 1995, Uitgeverij Verloren, Hilversum, pp. 223-230.

médiévaux (Parigi, 2000), nella presentazione della quale si afferma che «s'il est (relativement) facile de numériser, il l'est beaucoup moins de bâtir un système permettant la navigation entre texte et image»[29]. Come Andrea stesso afferma, nella SFM «si tratta, in sostanza, di avere esteso il concetto di concordanza: da semplice elencazione dei passi ove una parola è attestata, si aggiunge la concordanza delle sue varianti e la concordanza delle sue immagini nella riproduzione digitalizzata del documento originale»[30].

Qualora l'àmbito fosse quello dei documenti latini, nella Stazione filologica era prevista l'utilizzazione del lemmatizzatore automatico già prodotto nel 1991 da Andrea stesso[31].

La 'Stazione filologica multimodulare', denominata 'DiPhilos' nel 2003, nel corso del tempo è stata oggetto di modifiche, nuove articolazioni ed ampliamenti dovuti, per esempio, all'introduzione di moduli come il link automatico fra parole ed immagini digitali, il *shortcut module*, e il modulo di indicizzazione, sempre rappresentati da sottoinsiemi indipendenti, integrabili fra loro. Il carattere di 'flessibilità'[32] del sistema ne ha permesso la sperimentazione anche in campo archeologico, oppure in ambienti diversi da quello umanistico come, per es., in quello medico[33].

Il suo impiego, infine, pensato originariamente per un ambiente di lavoro individuale su personal computer, è stato previsto in seguito anche per un'attività su Web, da poter svolgere ugualmente secondo modalità collaborative. È il caso dell'uso nel sistema BAMBI indicato sopra, oppure nei progetti *Greek into Arabic* (descritto da Cristina D'Ancona nella *Presentazione* al Capitolo V) e *Talmud* (descritto ancora da Giacomo Ferrari, Capitolo III). Occorre aggiungere che questo progetto ha anche fornito l'occasione per riflettere sulla possibile costituzione di un ulteriore componente modulare atto a gestire le annotazioni semantiche strutturabili in tassonomie a partire dai dati rilevati sul testo del Talmud. Tale approccio, del resto, era già presente nella progettazione del DiTMAO (*Dictionnaire des Termes Médico-botaniques de l'Ancien Occitan*)[34], al quale Andrea ha portato il

[29] O. GUYOTJEANNIN - E. LALOU, *La numérisation des manuscrits médiévaux*, in «Le médiéviste et l'ordinateur» 40, 2001 (*Actes de la journée d'étude. Paris, 13 octobre 2000*), p. 6.

[30] Cfr. il testo 10, p. 146.

[31] Si veda il testo n. 5.

[32] Sull'aspetto della 'flessibilità' della SFM Andrea ha insisto in più occasioni fin dall'inizio della concezione del sistema; cfr., per es., il testo n. 10, pp. 131-132.

[33] Si indicano, per es., per l'àmbito archeologico, A. BOZZI - E. BRESCIANI - M. MENCHETTI - P. RUFFOLO - A. EISINBERG - G. FEDELE - G. CORRARELLO, *Computational Philology System for demotic texts on Ostraka*, in «XIV Tavola Rotonda Internazionale di Egittologia e Informatica», Pubblicazione su CD, Pisa, 2003; E. BRESCIANI - M. MENCHETTI - A. BOZZI - G. FEDELE, *Sistema di filologia computazionale per testi demotici*, in «Archeologia e Calcolatori» 15, 2004, pp. 267-286. L'esperimento di applicazione in àmbito radiologico è documentato in E. FERDEGHINI - P. MARCHESCHI - A. BOZZI - R. PREDILETTO - A. BENASSI, *Radiologic Image Library for Pathology Related Searches*, in «Computers in Cardiology» 31, 2004, pp. 689-492.

[34] Il progetto DiTMAO è finanziato dalla DFG (Deutsche Forschungsgemeinschaft) (*An XML-based Information System for Old Occitan MedicalTerminology*). Equipe: Università di Cologna: Gerrit

proprio contributo prevedendo l'utilizzo di strutture ontologiche organizzate in dominii di conoscenza differenziati che consentano di recuperare parti del testo che hanno elementi semantici comuni, indipendentemente dalla terminologia adoperata nel corpus di base[35].

Particolare attenzione, al di là delle ricerche via via condotte, è stata rivolta da Andrea nel cercare di determinare il corretto valore da attribuirsi ad alcune espressioni impiegate, non sempre in modo coerente, nelle *digital humanities*. Ciò è evidente, per es., quando, accingendosi a porre le basi metodologiche della SFM[36], egli si sofferma sulla denominazione di 'dizionario di macchina', che può contenere ambiguità concettuali, e su quella di 'testo', sovente utilizzata in modo scorretto nella pratica comune. Tali considerazioni assumono una valenza fondamentale nel momento in cui esse non si esauriscono in precisazioni terminologiche, ma implicano riflessioni metodologiche complesse, come nel caso dell'analisi delle differenti accezioni assunte nel corso del tempo dalle espressioni 'filologia elettronica' ed 'edizione elettronica', e di cosa esse rappresentino. L''edizione elettronica'[37], pur nella variabilità dei dati contenuti, costituisce unicamente un 'archivio' che gestisce informazioni già precostituite; del tutto differente, invece, è la situazione in cui un editore critico, nelle fasi di preparazione, analisi e valutazione dei testimoni di un'opera, voglia essere assistito da uno strumento informatico[38]. Come ben mette in rilievo Giacomo Ferrari nella sua *Presentazione*, l'aver pensato all'uso del computer nell'approccio filologico al testo (ed in tal caso, dunque, parlare di 'filologia computazionale') ha costituito una indubbia innovazione. Del resto, Andrea è convinto che già «il lavoro di spoglio non deve essere concepito esclusivamente come un programma di manipolazione di testi già èditi, ma deve rappresentare uno strumento da attivare durante le fasi del lavoro di edizione. In tal modo le attività computazionali per la filologia possono assumere un ruolo superiore a quello rappresentato dal semplice svolgimento di operazioni di servizio»[39].

Andrea, tuttavia, fedele alla propria matrice di classicista, non ha mai pensato di valicare i limiti entro i quali, invece, è necessario che debbano rimanere conte-

Bos, Veronica Roth; Università Georg August di Gottinga: Guido Mensching, Julia Zwink, Anja Weingart; Università di Pisa: M. Sofia Corradini, Andrea Fiumara; Pisa, ILC-CNR: Andrea Bozzi, Emiliano Giovannetti, Andrea Bellandi.

[35] Cfr. A. Bozzi - D. Liuzzi, *Un'ontologia per il* DiTMAO (Dictionnaire des Termes Médico-botaniques de l'Ancien Occitan), in E. Buchi - J.-P. Chauveau - J.-M. Pierrel (ed.), *Actes du XXVII^e Congrès International de Linguistique et de philologie romanes, Nancy, 15-20 juillet 2013*, ELiPhi, Strasbourg, 2016, II, pp. 1601-1607 ed anche qui il testo n. 15 (§. 4 c. *Classification ontologique des annotations*).

[36] Cfr., per es., il testo n. 10.

[37] Su ciò che si debba intendere con questa espressione si veda il testo n. 13, pp. 193-194, oltre a: A. Bozzi, *Towards a Philological Workstation*, cit.; Id., *Edizione elettronica e filologia computazionale*, in A. Stussi, *Fondamenti di critica testuale*, Il Mulino, Bologna, 2006, pp. 207-232.

[38] Cfr. il testo n. 13, §. 3: «the critical editor needs to be assisted in the various phases of preparation, analysis and evaluation of the witnesses. From this point of view, technology based on hypertextual languages is insufficient».

[39] Cfr. il testo n. 10, p. 131.

nute le attività linguistiche e filologiche. Egli afferma infatti che, «benché la tecnologia odierna offra validi sussidi all'attività del filologo, non si deve assolutamente credere che sia nata una nuova filologia: si tratta solo di nuovi mezzi, di moderni ausilii per una disciplina antica»[40] e, dunque, «l'edizione realizzata grazie ad un sistema di filologia computazionale deriva da una stretta interazione fra dati, strumento informatico e competenza personale dell'editore»[41]. Questa convinzione è evidente anche nell'enunciazione dei principi che sottostanno alla progettazione di quel modulo filologico della SFM che costituisce un elemento aggiuntivo specifico, incentrato sulla classificazione delle varianti in vista della preparazione di un'edizione critica[42]; le componenti software «devono essere in grado di svolgere mansioni ben determinate»[43] ed essere un supporto all'attività del filologo il quale, tuttavia, resta l'unico responsabile delle scelte editoriali. Una particolare attenzione al testo, dunque, sia che si tratti di estrarne e articolarne i dati linguistici, sia che si voglia prepararne l'edizione.

Tali princìpi, che non costituiscono mai un ostacolo all'ideazione di sistemi innovativi, hanno caratterizzato le ricerche di Andrea durante la sua permanenza all'Istituto di Linguistica Computazionale 'Antonio Zampolli', che egli ha guidato dal 2008 al 2013, e sono alla base del significato che egli ha voluto infondere alla *First Euroconference on Philological Disciplines and Digital Technology*, di cui è stato proponente e chairman nel 2003[44].

Le principali pubblicazioni compaiono raggruppate per àmbito tematico in cinque capitoli, introdotte da studiosi i quali, sebbene legati ai diversi campi di ricerca di Andrea per motivi differenti, tutti rappresentano degli amici: amici del primo periodo lavorativo, a partire in particolare da Giacomo Ferrari e poi da Valeria Lomanto e da Sylvie Calabretto, fino ad amici più recenti come Cristina D'Ancona ed Elton Prifti. A loro va il mio ringraziamento per aver reso possibile la preparazione di questa miscellanea e, soprattutto, per aver messo in rilievo, grazie alle loro competenze, gli aspetti più innovativi delle ricerche di Andrea.

Desidero esprimere gratitudine anche a Gloria Borghini, che ha accolto con entusiasmo l'iniziativa, consentendo la pubblicazione per i tipi di ETS.

Giugno 2019

[40] Cfr. il testo n. 11, p. 167.

[41] BOZZI, *Edizione elettronica*, cit., p. 217.

[42] Cfr. i testi n. 12 (§. 9. *Extensions and Particular Use*) e n. 15 (§. 4.4. L'*apparat critique*) e A. BOZZI - M.S. CORRADINI, *New trends in philology: a computational application for textual criticism*, in «Euphrosyne» XXX, 2002, pp. 267-285.

[43] Cfr. BOZZI, *Edizione elettronica*, cit., p. 217.

[44] La *First Euroconference on Philological Disciplines and Digital Technology* si è svolta al Ciocco, Lucca (6-11 settembre 2003) ed è stata patrocinata dalla European Science Foundation, dal CNRS e dalla Regione Toscana. Gli atti sono pubblicati in A. BOZZI - L. CIGNONI - J.L. LEBRAVE (ed.), *op. cit*.

PRESENTAZIONE
I Grammatici latini

Valeria Lomanto
Università di Torino

Non è la sede né ho la competenza per valutare l'apporto di Andrea al perfezionamento e alla diffusione dei metodi computazionali nell'analisi dei testi classici, ma sono testimone diretta di quanto il suo intervento sia stato prezioso per l'analisi mediante computer dei *Grammatici latini*.

Nel lontano 1975 essa è stata avviata dal prof. Nino Marinone dell'università di Torino e io, allora sua assistente, mi ero entusiasmata del progetto, che avrebbe permesso una conoscenza capillare delle grammatiche latine tardo-antiche, edite da Heinrich Keil a Lipsia tra il 1855 e il 1880, e ne avrebbe agevolato tanto lo studio quanto la riedizione. Il professore aveva preso contatto con il prof. Antonio Zampolli, direttore dell'Istituto di Linguistica computazionale di Pisa, l'unico in Italia a dedicarsi a ricerche di questo genere avvalendosi del solo mainframe di cui le università italiane potessero disporre[1]. La circostanza ha comportato per me, che prima ho coordinato il lavoro di un gruppo di giovani ricercatori, poi ho corretto le stampe da nastro magnetico su cui il testo era stato riversato e infine ho seguito le prove di concordanza, un'esperienza di grande utilità sul piano scientifico e un periodo di pendolarismo, graditissimo per la bellezza di Pisa e la simpatia delle persone con cui ho lavorato e con alcune delle quali – in particolare Andrea e la sua splendida famiglia – ho stretto una duratura amicizia.

La registrazione del testo è stata preceduta da una 'pre-edizione' destinata a eliminarne le incoerenze formali e ad adeguarne le segnalazioni alla prassi ecdotica in uso[2]. Sebbene i criteri da adottare, in nessun caso prevaricanti sulle scelte dell'editore, fossero stati definiti prima dell'avvio del lavoro, com'è naturale sono emersi più volte casi imprevisti per cui si è dovuto trovare una soluzione apposita e in sintonia con quelle già assunte. In particolare in queste 'emergenze' è risultato decisivo l'intervento di Andrea che, per la sua competenza nell'ambito sia filologico sia informatico, ha svolto la funzione di raccordo tra le esigenze che avanzavo io, preoccupata di riprodurre il testo con assoluta fedeltà, e la prassi cui i programmatori erano avvezzi operando in genere su lingue moderne. Per merito della sua mediazione le difficoltà di comprensione tra il personale tecnico e me si

[1] Sull'esperienza pisana del prof. Marinone cfr. A. BOZZI, *Nino Marinone e l'Istituto di Linguistica Computazionale*, in A. TRAINA (ed.), *Atti del Convegno di studio: una giornata per Nino Marinone (Vercelli, 28/10/2000)*, Pàtron editore, Bologna, 2001.

[2] BOZZI 2003 (= testo n. 6, in part. pp. 59-60).

sono mano a mano attenuate, finché ogni mia richiesta è giunta a trovare una risposta pertinente e immediata.

La preparazione dei testi per la registrazione e la correzione delle stampe non soltanto hanno richiesto molta pazienza e costante attenzione, ma soprattutto hanno messo in evidenza la necessità di prendere decisioni tassativamente univoche e omogenee: l'assenza di duttilità del computer esalta ogni scelta contraddittoria e, fornendo una preziosa lezione di metodo, impone di porsi di fronte alla lingua con il medesimo rigore da tutti riconosciuto necessario nelle scienze esatte. Ma non soltanto questo mi ha insegnato l'uso, per quanto mediato, del computer; anche la conoscenza del latino ne ha tratto non poco vantaggio. La ricognizione sistematica di tutte le varianti grafiche e morfologiche per scegliere quale motivatamente privilegiare come forma prevalente nel testo in esame e quali segnalare con rinvii permette di acquisire la consapevolezza delle innumerevoli variazioni di una lingua. Questa 'lemmatizzazione grafica' prima che al testo dei grammatici, troppo vasto e complesso, è stata applicata al testo di Simmaco, sottoposto a spoglio, non soltanto per il suo interesse intrinseco, quanto per individuare e saggiare una procedura su di un'opera più breve e relativamente omogenea. E tuttavia tra forme con grafia assimilata o dissimilata (*aggredior / adgredior*), dittongata o monottongata (*caudex / codex*), aspirata o deaspirata (*nihil / nil*), unita o divisa (*eiusmodi / eius modi*) e ancora, sul piano morfologico, con morfema ad es. di genitivo arcaico (*familias / familiae*) o di accusativo alla greca (*Achillen / Achillem*), privilegiarne una come key-word nel cui ordine alfabetico disporre tutte le occorrenze ha comportato un lavoro faticoso e lunghissimo.

Da una parte i tempi richiesti dagli interventi manuali, per quanto circoscritti agli aspetti formali del testo e il carattere inevitabilmente soggettivo di essi, dall'altra la constatazione che materiali di spoglio negli anni sempre più abbondanti non erano in nessun modo confrontabili e tanto meno integrabili in quanto allestiti con modalità difformi (ad es. lemmatizzati o non lemmatizzati e in questo caso secondo criteri diversi)[3] avevano suggerito al prof. Marinone e ad Andrea di progettare un sistema informatico che fosse in grado di assolvere molteplici funzioni[4]. Il sistema è risultato di fatto programmato in modo da agire sull'archivio dei dati, cioè sul materiale di spoglio, mediante un analizzatore morfologico, un modulo per la lemmatizzazione, un modulo per il confronto di testi elaborati con metodi diversi. Una serie di algoritmi e di codici di compatibilità, che consentono il passaggio dalle forme al lemma e dal lemma alle forme e da una all'altra

[3] Si veda BOZZI 1982 (= testo n. 1).

[4] Si veda BOZZI 1988 (= testo n. 4) ed anche: A. BOZZI - G. CAPPELLI, *The Latin Lexical Database and Problems of Standardization in the analysis of Latin Texts*, in F. HAUSMANN *et alii* (eds.), *Data Networks for the historical disciplines*, Graz, 1987, pp. 28-45; IID., *Machine readable textual archive and exchange of data: some experiences at the ILC - Pisa*, in J.-PH. GENET (éd.), *Standardisation et échange des bases de données historiques*, Editions du CNRS, Paris, 1988, pp. 185-190; IID., *Un sistema computerizzato per la produzione di indici lessicali di testi latini*, in «MD. Materiali e Discussioni per l'analisi dei testi classici» 20-21, Giardini Editori e Stampatori, Pisa, 1988, pp. 343-360.

variante, permette tanto la lemmatizzazione automatica quanto l'integrazione dei dati d'archivio. Il sistema presuppone la scomposizione di ogni parola nei suoi elementi costitutivi (base lessicale, prefissi, infissi, suffissi, morfemi, elementi postdesinenziali quali le enclitiche) e soprattutto l'adozione di un criterio rigidamente morfologico: sono considerate entrate lessicali tutte le forme provviste di un'individualità morfolologica, indipendentemente dalla funzione. Peraltro, grazie ai codici grammaticali è possibile, ad es., ricondurre il superlativo di un participio quale *amantissimus* al positivo *amans* o al verbo *amo* da cui la forma, in ultima analisi, deriva o attribuire funzione di lemma al neutro *bonum* usato in funzione di sostantivo, oppure riportarlo all'aggettivo *bonus*.

Il sistema informativo latino è stato dunque applicato da Andrea al testo dei Grammatici[5]. Poiché per l'incompatibilità tra archivi registrati su nastro magnetico e i personal computers sempre più diffusi lo spoglio dei testi raccolti nel *corpus* del Keil sarebbe diventato inservibile, Andrea ha provvisto a riversare i dati su CD, convertendo con non poca fatica il materiale in un formato adatto al nuovo supporto[6]. Questo in primo luogo ha permesso di sostituire le non immediatamente perspicue traslitterazioni dei numerosi passi greci con caratteri greci, rendendo la lettura sia su schermo sia su stampa incomparabilmente più agevole. Il mutamento del formato ha reso necessaria una nuova codificazione di tutti gli interventi operati sul piano formale, dall'omologazione dei segni diacritici e delle scansioni metriche alla segnalazione delle citazioni, della fine dei versi, dell'avvicendamento dei personaggi nei passi dialogici, dei titoli delle opere citate. Ma soprattutto, mentre la prima redazione della concordanza, per così dire grezza, non permetteva altro che la ricerca per forma, affidando all'utente il reperimento di ogni modificazione prodotta dalla flessione e di ogni allografo, il sistema di interrogazione allestito per il CD, cui sono sottesi analizzatore morfologico e modulo di lemmatizzazione, consente di risalire da una qualsiasi forma con qualsiasi grafia a tutte le occorrenze di una parola. Non solo: l'operatore ha una duplice possibilità di selezionare il campo di ricerca. Se, ad es., desidera il repertorio delle occorrenze del solo participio *amans* in tutto il testo dei grammatici, può escludere ogni altra occorrenza del verbo *amare*. Se poi vuole circoscrivere l'indagine a un autore o un argomento, è sufficiente che segnali i codici dell'opera o delle opere di quell'autore e di quell'argomento. In modo analogo è possibile limitare la ricerca alle citazioni, scegliere soltanto quelle in prosa o quelle in versi o escluderle dai risultati servendosi come elemento discriminante delle virgolette e delle sbarre di fine verso. Le innumerevoli combinazioni di queste scelte rendono la ricerca, straordinariamente semplice e duttile, adeguata a ogni esigenza.

[5] Cfr. testo n. 6.

[6] A. BOZZI - V. LOMANTO - A. RAGGIOLI (ed.), *I Grammatici Latini Antichi su CD-ROM*, versione per sistemi Microsoft Windows 95b/98/NT/2000 (prodotto fuori commercio, disponibile gratuitamente su convenzione con l'ILC).

La rapidissima evoluzione della tecnologia ha reso obsoleti anche i CD: ormai si lavora in rete. Il *Laboratoire d'histoire des théories linguistiques*, cui Andrea è stato tanto generoso da donare il CD dei Grammatici latini, ha provvisto ad adattare per la rete l'archivio dei testi, ma non il programma di consultazione. Non posso che augurarmi, per i vincoli non soltanto scientifici ma anche affettivi che mi legano ad Andrea e a questo lavoro, che egli voglia continuare a occuparsi dei grammatici e coordinare gli informatici della Sorbona nella conversione per la rete dei programmi da lui elaborati, tanto sofisticati nella realizzazione quanto agevoli ed efficaci nell'uso.

PRESENTAZIONE

L'informatizzazione del riconoscimento automatico dei caratteri a stampa e manoscritti

Elton Prifti

Universität Wien, condirettore del LEI

Fu il progetto attualmente in corso della digitalizzazione del *Lessico Etimologico Italiano* (LEI) o, per meglio dire, alcune difficoltà e problemi tecnici da risolvere nell'ambito di questa ardua e complessa impresa, iniziata quattro anni fa, a far sì che i nostri percorsi professionali s'incrociassero. Gli inizi del LEI risalgono agli anni Sessanta del secolo scorso, quando il suo illustre fondatore, il compianto Max Pfister, iniziò a creare meticolosamente la base per la sua *opus magnum*, lo schedario del LEI, che tuttora continua a crescere. Il *fichier* del LEI è ora composto da oltre 7 milioni di schede, che rappresentano fisicamente dei fogli di dimensioni A6, sui quali sono stati incollati meccanicamente stralci estratti – precisamente ritagliati a mano – da migliaia di opere, soprattutto da dizionari. Ogni singola scheda è corredata dell'indicazione dell'etimo, appuntata a mano, di un timbro, che di solito contiene la sigla bibliografica abbreviata dell'opera da cui è stata estratta, e delle indicazioni geolinguistiche e cronologiche. L'informatizzazione del LEI, il quale tuttora rappresenta un'impresa di lessicografia storica analogica, consiste molto sommariamente sia nella digitalizzazione delle sue parti pubblicate, che comprendono circa 5100 articoli, inclusi in 15 volumi o 25.000 pagine stampa di formato A4, che – e soprattutto – nell'informatizzazione e automatizzazione del sistema redazionale. Nell'ambito di quest'ultimo punto è collocata anche l'automatizzazione dell'elaborazione del contenuto delle schede sunnominate del LEI. Dopo aver terminato la retrodigitalizzazione dell'intero *fichier* del LEI, impresa molto impegnativa e complicata, durata due anni e realizzata in gran parte grazie ad una stretta e fruttuosissima collaborazione con l'Università per Stranieri di Siena, e dopo aver raccolto per etimo in più di 20.000 documenti in formato PDF le circa 7,5 milioni di singole pagine scansionate, si è passati all'identificazione di un metodo di riconoscimento automatico e di trasformazione altrettanto automatizzata in formato digitale del contenuto delle singole schede.

Ed è durante questo processo che ci siamo imbattuti nei risultati della ricerca pluridecennale nell'ambito del trattamento informatico dei documenti digitali, a scopo di analisi linguistica e filologica, del Festeggiato, partendo da un volume miscellaneo[1] da lui stesso curato. Nei capitoli 7-9 della miscellanea, scritti dal Nostro, si illustra una tecnica di riconoscimento automatico di caratteri a stampa o manoscritti, compresi persino i papiri, tramite l'utilizzo di reti neurali artificiali

[1] A. Bozzi (ed.), *Computer-aided recovery*, cit. Si vedano qui i testi n. 7 e n. 8.

per favorire la trasformazione in forma digitale, utilizzando l'applicazione OCR-Lab, di cui si descrivono le modalità di uso. Di questo argomento Andrea Bozzi ha iniziato a occuparsi già negli anni Ottanta, come testimonia una serie di pubblicazioni[2]. La tecnica può essere utilizzata persino per la ricostruzione di caratteri poco o non leggibili in antichi testi latini a stampa. Ed è proprio in questo ambito che il Nostro ha concepito e coordinato il progetto di durata triennale (1996-1998) *LAperLA: lettore automatico per libri antichi*, finanziato dal Comitato nazionale di consulenza per la Scienza e le Tecnologie dell'informazione del CNR[3].

Le soluzioni innovative e le idee avanzate in questo campo ci sono state utili nel percorso di identificazione di un metodo efficace, pratico e qualitativo per raggiungere il nostro obiettivo.

Circa un anno fa il caso volle poi che ci incontrassimo anche di persona, a Heidelberg, nell'ambito di una giornata di studi di lessicografia storica organizzata dalla redazione del *Dictionnaire Étymologique de l'Ancien Français* (DEAF). Seguì anche una piacevolissima visita presso il Laboratorio LEI dell'Università di Mannheim, dove si sta realizzando la digitalizzazione del LEI, la quale ci ha dato modo di discutere, con grande profitto, varie questioni, allora aperte, inerenti all'informatizzazione del LEI, oramai in fase avanzata.

[2] A. BOZZI - R. BINDI, *Nuovi frammenti di P. OXY. 2181 (Platone, Fedone) identificati con il ricorso all'archivio computerizzato (T.L.G.). Parte II: Procedura semiautomatica per la collocazione dei frammenti*, in «Studi Classici e Orientali» XXXVII, 1987, pp. 198-203; A. BOZZI, *Computer-aided preservation and transcription of ancient manuscripts*, in «ERCIM News» 19, 1994, Imprimerie Barnéoud, Mayenne, pp. 27-28; L. BEDINI - A. BOZZI - A. TONAZZINI, *Digital techniques for character recognition in old documents*, in «ERCIM News» 28, 1997, Imprimerie Barnéoud, Mayenne, p. 24; IID., *Digital techniques for character recognition in old printed books and in modern damaged documents*, in A. GUARINO (ed.), *Proceedings of the 2nd International Congress on Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin (5-9 July 1999, Paris)*, Elsevier, Paris, 2000, pp. 959-962.

[3] Si veda A. BOZZI, LAperLA: *an integrated graphical-linguistic System for old printed Latin* Texts, 2002 (qui testo n. 9).

PRESENTAZIONE
Filologia computazionale, una terza via

Giacomo Ferrari
Università del Piemonte Orientale

L'uso del calcolatore per il trattamento dei dati linguistici è uno dei settori scientifici più antichi. Risale alla fine degli anni '40 del 900, ma fin dall'origine si sono formate due tendenze di ricerca parallele e poco comunicanti tra loro. Da un lato, infatti, si è puntato a costruire programmi che simulano sul calcolatore il comportamento linguistico umano, nella comprensione e generazione di segmenti di linguaggio, siano essi singole frasi o interi testi. Gli inizi di questa tendenza si fanno risalire alla prima proposta di traduzione automatica, avanzata da Warren Weaver nel 1949, e si prosegue, negli anni '60 e primi anni '70, con la ricerca volta a costruire interfacce uomo-macchina in linguaggio naturale che cercano di imitare la capacità umana di comprendere domande su un ambito ristretto e costruire risposte. Dall'altro, invece, iniziando con il lavoro di spoglio elettronico delle opere si San Tommaso, ideato e promosso fin dal fin dal 1949 da Padre Roberto Busa, si è sviluppato un complesso strumentale per la memorizzazione dei testi, in modo da renderli disponibili per operazioni di ricerca come la costruzione di lessici e concordanze.

Il primo modello fa riferimento a paradigmi teorici che fanno capo ai lavori di Turing, di Church e, sul piano più strettamente linguistico, si rifanno alla linguistica chomskiana. Il secondo si fonda su una tradizione linguistica più attenta al dato che non alla facoltà cognitiva del linguaggio, quella tradizione che dà luogo, fin da poco prima dell'epoca dei calcolatori, alla statistica linguistica.

Questo secondo filone di ricerca pone molta cura nei processi di acquisizione e memorizzazione dei testi, che, oltre ad essere predisposti nel migliore dei modi per l'elaborazione, devono essere resi in forma, *machine-readable*, il più possibile riutilizzabile da altri ricercatori. Sarà quindi necessario inserire in fase di acquisizione il maggior numero di informazioni relative al testo che si intende memorizzare, in un formato che non sia orientato unicamente al progetto per cui l'acquisizione viene compiuta. Questa attenzione alla forma del testo, quali edizione, riferimenti (numero pagina, numero riga, capitolo ecc.) e ogni altra informazione aggiuntiva, ha portato spesso ad estendere questo trattamento anche ai testi antichi. Ma la relazione che si può istituire tra filologia e uso del computer è, almeno sul piano concettuale, una delle più complesse e, in un certo senso, indirette. Infatti, mentre al computer si riconosce la capacità di memorizzare, comparare e organizzare grandi quantità di dati, in filologia si richiede la cura per il testo e la sua ricostruzione, una ricostruzione che coinvolge sia l'aspetto fisico delle trascri-

zioni che la ricerca linguistica. Quindi, mentre nella normale prassi degli spogli di testi è sufficiente codificare nel modo più efficace un testo già edito a stampa, in filologia l'attenzione è diretta alla tradizione testuale, che include l'origine del testo stesso e le sue varianti; il generico trattamento di testi deve forzatamente procedere per grandi categorizzazioni e standardizzazioni che non soddisfano gli scopi del filologo.

Sembrerebbero, quindi, due mondi molto diversi e quasi irreconciliabili. La ricerca di Andrea Bozzi va esattamente nella direzione opposta, nel tentativo, a quanto pare riuscito, di conciliare le due anime delle *computational humanities*.

L'articolo del 1994 *Stazione di lavoro computerizzata per la filologia* (qui testo n. 10) offre una panoramica concisa e sintetica sull'uso del computer nel trattamento dei testi, mettendo in evidenza con lucidità le differenze tra il filone principale di memorizzazione ed elaborazione dei testi, quello che Bozzi chiama «operazioni di servizio», e l'uso filologico, giustificando storicamente il minor sviluppo di questo secondo settore. Non sono tanto i limiti tecnologici a rendere, all'epoca, meno sviluppato questo settore, quanto la mancanza di un vero e proprio modello autonomo di filologia computazionale.

Il tratto discriminante è la necessità di usare un programma filologico non solo per il trattamento del testo ma come «uno strumento da attivare durante le fasi di lavoro di edizione»[1]. Per poter soddisfare questo obiettivo occorre implementare una serie di programmi specifici che realizzino le funzioni utili alla critica testuale, ben elencate nell'articolo, sempre del 1994, *Text editing e Text processing: aspetti e problemi di computerizzazione di dati editi ed inediti*[2]. È particolarmente importante insistere su questi primi passi compiuti negli anni '90, perché mettono in luce come il problema della creazione della stazione di lavoro filologica sia un problema di modello di ricerca e di metodologia prima che un problema tecnologico di implementazione di diversi moduli informatici. L'attività filologica è un'attività multidisciplinare, che coinvolge esperti di diversi settori, la codicologia, la paleografia, la papirologia, l'epigrafia che devono interagire con informatici ed esperti di *computer graphics*. Da questa visione nascono le specifiche prime di un sistema formato di moduli diversi che interagiscono tra loro[3].

Dunque, la filologia che voglia trarre vantaggio dall'uso del calcolatore, dovrà avvalersi sì dei programmi di trattamento dei testi, ma dovrà anche integrarli con altri moduli che trattino l'immagine del testo come essa appare al filologo.

Per questo gli articoli di Andrea Bozzi delineano l'architettura di un sistema complesso che integra diversi moduli, seguendo l'evolvere della tecnologia disponibile. Grosso modo il lavoro si divide in due fasi, la prima che potremmo chiamare della Stazione Filologica Multimodulare (la SFM) e quella del sistema Diphilos, che risalgono però allo stesso modello metodologico.

[1] Si veda il testo n. 10, p. 131.
[2] Si veda il testo n. 11, p. 155.
[3] Cfr. p. 157.

La SFM, presentata in articoli degli anni 1993, 1994 è costituita da tre moduli, uno di trascrizione del testo, uno di interpretazione dei codici inseriti con l'utilizzazione del modulo precedente, e un modulo di elaborazione del testo.

Il primo modulo offre una vasta gamma di possibilità di codifica di tutti i dati utili al filologo, che non si limitano al testo edito, ma includono tutta una serie di informazioni sulle varianti e sugli aspetti esterni del testo stesso. Il secondo modulo permette di visualizzare o stampare il testo in modo comodo, interpretando graficamente tutti i codici interni al testo. Il terzo modulo permette di definire i propri criteri di selezione per la visualizzazione del testo e sostituisce le tradizionali funzioni di *query* con funzioni di interattività più completa; permette, inoltre, di elaborare i prodotti tipici della lessicografia computazionale, come le concordanze, gli indici (*Index locorum* e *Index verborum*), la lemmatizzazione, ma anche la ricerca delle varianti. Una funzionalità completamente innovativa è la possibilità di acquisire immagini del testo, fornendo al filologo la possibilità di trascrivere il testo stesso dalla sua rappresentazione grafica, allineando la trascrizione con le corrispondenti regioni dell'immagine. I dati tecnici sono esplicitati attraverso la produzione di Andrea Bozzi e seguono l'evoluzione tecnologica offrendo strumenti sempre più raffinati che vanno a costituire una vera e propria stazione di lavoro per filologi.

La versione evoluta, nota come DiPhiloS, è presentata nell'articolo del 2003 (si veda il testo n. 12), mentre un'idea delle possibilità applicative delle ultime versioni è presentata in *Electronic publishing and computational philology*[4]. I tratti distintivi di DiPhiloS sono il raffinamento del componente grafico e l'aggiunta di numerose funzioni di indicizzazione e di utilizzo delle varianti. I miglioramenti non sono dovuti solo alla normale evoluzione tecnologica, ma ad un ampliamento delle possibili applicazioni che ha portato a integrare molte funzioni proprie della papirologia.

Il percorso delineato in questo capitolo è certamente innovativo, poiché getta le basi di una «terza via» nell'uso del computer nel trattamento del linguaggio, quella dell'approccio filologico al testo.

L'intuizione iniziale giunge in tempi in cui le barriere tecnologiche avrebbero potuto scoraggiare l'avvio di questo tipo di ricerca. Tuttavia l'approccio presentato in questo capitolo definisce i requisiti che sono poi evoluti con il progredire dei mezzi computazionali a disposizione. Oggi molto probabilmente quello che al momento della prima intuizione fu piuttosto avveniristico, può apparire più facilmente raggiungibile, tanto che non manca chi, senza citare il precedente, rivendica il primato di aver pensato ad una «filologia computazionale», come ad es. Jean-Baptiste Camps che ha introdotto il termine E-philologie.

Ma la forza della metodologia presentata e seguita passo per passo in questo

[4] Si veda qui il testo n. 12 e anche A. BOZZI, *Computer-assisted Scholarly Editing of Manuscript Sources*, in P. DÁVIDHÁZI (ed.), *New Publication Cultures in the Humanities. Exploring the Paradigm Shift*, Amsterdam University Press, Amsterdam, 2014, pp. 99-116.

capitolo sta nella riusabilità di certe tecniche e di certi moduli. Un caso per tutti, il software TRADUCO utilizzato nel grande progetto PTTB (si vedano qui i testi n. 16 e n. 17). Il Progetto di Traduzione del Talmud Babilonese (PTTB; vedi https://www.talmud.it/) è il frutto di un accordo tra il MIUR, il CNR, rappresentato dall'Istituto di Linguistica Computazionale «A. Zampolli» e il Consiglio Rabbinico Italiano, siglato nel gennaio del 2011, ed ha per obiettivo la produzione in italiano del Talmud Babilonese, supportata da strumenti informatici. Il sistema di assistenza computazionale alla traduzione[5] è corredato di una serie di moduli che permettono di trattare i problemi di natura filologica posti da un testo complesso come il Talmud, stratificatosi attraverso i secoli e forse i millenni. Si è reso, perciò, necessario utilizzare numerose tecniche di allineamento dei testi, di trattamento delle varianti, di visualizzazione e di annotazione, che risalgono al modello di testuale proposto per la prima volta con il sistema SFM.

Il presente capitolo deve essere letto, quindi, come un percorso metodologico che, partendo dall'identificazione di alcune inadeguatezze della linguistica computazionale nei confronti del trattamento dei testi, fissa un paradigma ed un percorso di ricerca che soddisfa in pieno le esigenze di manipolazione proprie dei filologi, ma va oltre e si integra in numerose altre applicazioni.

[5] Si veda A. Bellandi - D. Albanesi - G. Benotto - E. Giovannetti, *Il sistema* Traduco *nel Progetto Traduzione del Talmud Babilonese*, in «International Journal of Computational Linguistics» 2-2, 2016, pp. 109-126.

PRESENTAZIONE
Le projet BAMBI et d'autres collaborations

Sylvie Calabretto
LIRIS-INSA Lyon

J'ai connu Andrea Bozzi en 1995 dans le cadre du projet européen BAMBI (*Better Access to Manuscripts and Browsing of Images*) du programme européen LIBRARIES, dont il était le responsable dès 1994. La station BAMBI est dédiée aux papyrologues, épigraphistes, paléographes et codicologues, ou plus généralement aux utilisateurs d'une bibliothèque qui souhaitent examiner des sources manuscrites, transcrire et annoter des manuscrits, ainsi que naviguer entre les éléments textuels de la transcription et les portions d'image correspondantes sur le manuscrit scanné. En effet, le projet, s'engageait dans deux buts, qui se manifestent dans sa dénomination : définir des techniques innovatrices de numérisation de manuscrits médiévaux afin de permettre la consultation de bibliothèques digitales (*Better Access to Manuscripts*) et aider les utilisateurs dans les activités de lecture, écriture et indexation du patrimoine manuscrit (*Browsing of Images*). Il est possible d'effectuer les opérations relatives au second aspect grâce à l'introduction de certains modules qui avaient été conçus à l'origine pour la 'Station philologique multimodulaire'[1], une système pour l'étude et la publication de documents manuscrits anciens qui se fonde sur l'emploi de composantes informatiques. L'ensemble des recherches relatives à ce projet ont été réunies dans une publication éditée par Andrea, et dans laquelle il est auteur ou co-auteur de plusieurs chapitres[2].

D'autres publications en ce domaine ont été rédigées conjointement et acceptées dans des revues et des conférences internationales sélectives[3]. Au cours de ce projet, nous avons constaté rapidement la complémentarité de nos compé-

[1] A ce sujet voir ici les textes du chapitre III.

[2] Il s'agit de A. BOZZI, *Better Access to Manuscripts and Browsing of Images. Aims and results of an European Research Project in the field of Digital Libraries (BAMBI LIB-3114)*, Editrice CLUEB, Bologna, 1997. Voir, dans cette oeuvre, A. BOZZI - S. CALABRETTO - F. TARIFFI, *The BAMBI users*, pp. 1-24 ; A. BOZZI - F. TARIFFI, *Manuscripts and microfilms: techniques for digital conversion*, pp. 27-42 et ici les textes n. 19 et n. 20. Le manuel d'utilisation du système a été publié dans un rapport interne CNR : A. BOZZI, *Bambi. Guida per l'utente*, Pisa, 1997.

[3] Voir, par ex. : A. BOZZI - S. CALABRETTO, *The Digital Library and Computational Philology: The BAMBI Project*, dans C. PETERS - C. THANOS (eds.), *Research and Advanced Technology for Digital Libraries*, Springer, Berlin, 1997, pp. 269-285; ID., *The Philological Workstation BAMBI* (Better Access to Manuscripts and Browsing of Images), dans «Journal of Digital Information» I (3), 1998; A. BOZZI - S. CALABRETTO - J.M. PINON, *BAMBI: système de gestion de manuscrits anciens pour historiens*, in «Document numérique» II 2 (3-4), 1998, Hermes, Parigi, pp. 31-50.

tences : Andrea comme spécialiste de linguistique avec une formation classique et moi comme informaticienne avec une formation initiale en mathématiques. Il faut souligner que ce type de collaboration s'intègre dans la thématique Humanités Numériques très porteuse actuellement !

D'autre part, notre collaboration professionnelle s'est vite transformée en amitié. Andrea et sa famille ont effectué deux séjours dans ma famille à Miribel et Lyon et j'ai effectué plusieurs séjours très agréables à Calci et à Pise.

Le projet BAMBI a eu une continuation en 1999, quand nous avons heureusement obtenu un financement pour la version Web de la plateforme, dans le cadre du programme P.A.I. GALILEE 1999[4]. Il s'agissait du projet STEMA (*Station de Travail pour l'Etude des Manuscrits Anciens sur le Web*), dont Andrea était le responsable scientifique pour la partie italienne[5].

Notre amitié s'est consolidée au cours le temps. En septembre 2000 Andrea m'a invité à participer avec une communication[6] à San Millán de la Cogolla au *Segundo Seminario de la Escuela Interlatina de Altos Estudios en Lingüística Aplicada - Matemáticas y Tratamiento de Corpus.* Il était l'organisateur de la *Sesión IV*, qui expose ses objectifs dans le titre : *De la cantidad a la cualidad : técnicas matemáticas para clasificar, visualizar y evaluar los datos filológicos y culturales. Nuevas tendencias para el uso y conservación del patrimonio cultural.*

De plus, en juin 2003 Andrea a participé comme examinateur à mon jury d'Habilitation à Diriger des Recherches (HDR), où j'ai présenté le travail *Modèles de représentation de la sémantique des documents. Application aux bibliothèques numériques ;* la même année, au mois de septembre 2003, Andrea m'a invité à présenter la thèse d'Aurélien Bénel et le logiciel Porphyry (développé en collaboration avec Andrea Iacovella de l'Ecole Française d'Athènes) dans le cadre de la *First Euroconference on Philological Disciplines and Digital Technology*, soutenue par la European Science Foundation, le CNRS et la Regione Toscana, qu'il a proposé et organisé à Castelvecchio Pascoli, Il Ciocco (Lucca)[7].

En 2004, j'ai effectué un séjour très agréable de six mois dans le laboratoire ILC-CNR d'Andrea. Nous avons soumis le projet franco-italien EUMME qui, bien

[4] Projet franco-italien d'une durée de 2 ans : 1999/2000.

[5] Le projet est documenté par les rapports techniques suivants: A. BOZZI - S. CALABRETTO, *Rapport Technique 1: STEMA - Station de Travail pour l' Étude des Manuscrits Anciens sur le WEB,* (1/1/1999 - 31/12/1999), Lyon, 1999, pp. 1-20 ; IID., *Rapport Technique 2: STEMA - Migration d'une station philologique sur le WEB*, (1/1/2000 - 31/12/2000), Lyon, 2000, pp. 1-106.

[6] A. BENEL - S. CALABRETTO, Exploration *de corpus de documents archéologiques à l'aide de théories algébriques*, in *Actas del Segundo seminario de la Escuela Interlatina de Altos Estudios en Lingüística Aplicada. Matemáticas y Tratamiento de Corpus, San Millán de La Cogolla (La Rioja), España, 19-23 de septiembre 2000*, Fundación San Millán de La Cogolla, Logroño, 2002, pp. 343-350. Le texte de Andrea est publié ici avec le n. 21.

[7] Les actes du congrès ont été publiés dans «Euphrosyne» 32, 2004. En particulier, voir les articles : S. CALABRETTO, *Indexation sémantique de corpus documentaires : approche ontologique et approche herméneutique*, pp. 55-74 et A. BOZZI, *Verso una filologia computazionale: la prima Euroconferenza della European Science Foundation*, pp. 127-138.

qu'il n'a malheureusement pas été retenu par le programme VINCI de l'Université Franco-Italienne (UFI), il a toutefois abouti à une publication acceptée à la conférence internationale EP'2005[8]. Ensuite, nous avons participé ensemble au projet de GDR *Européen Plus* HyPERLearning (2004-2007). Enfin, nous avons eu l'occasion de nous revoir à l'UTT de Troyes en 2009 pour le jury de thèse de Chao Zhou dont le Directeur de thèse était Aurélien Bénel.

Je souhaite remercier chaleureusement Andrea pour les fructueuses collaborations professionnelles dans le cadre de projets en Humanités Numériques et pour les moments exceptionnels dans le cadre de séjours privés!

[8] S. CALABRETTO - A. BOZZI - M.S. CORRADINI - B. TELLEZ, *The EUMME project: towards a new philological workstation*, dans *ICCC/IFIP conference on electronic publishing. From author to reader: Challenges for the digital content chain. EP'2005. 8 juin 2005, Heverlee (Belgique) EP'2005. 8 juin 2005*, Heverlee (Belgique), pp. 139-144.

PRESENTAZIONE
G2A: le traduzioni greco-arabe tra passato e futuro

Cristina D'Ancona
Università di Pisa

"Greek into Arabic. Philosophical Concepts and Linguistic Bridges" è un Advanced Grant dell'European Research Council (AdG 249431) che è stato attivo fra il 2010 e il 2015 ed ha visto Andrea Bozzi e l'ILC-CNR - Area della Ricerca di Pisa tra i suoi protagonisti. Dedicato allo studio delle traduzioni greco-arabe di opere filosofiche sia dal punto di vista delle dottrine trasmesse, sia dal punto di vista lessicografico, "Greek into Arabic" ha potuto beneficiare, grazie alla ricerca di Andrea Bozzi e dei suoi collaboratori, di un sistema linguistico-computazionale di straordinarie potenzialità: *G2A*. Una storia per sommi capi della problematica a cui risponde il sistema messo a punto dall'ILC-CNR motiverà, spero, questa affermazione. Il ruolo decisivo svolto dalle traduzioni di opere greche nella nascita della trattatistica filosofica in arabo è unanimemente riconosciuto dagli storici del pensiero medievale e posteriore[1]. Dato che la filosofia araba, nel suo sorgere e svilupparsi sino al suo ultimo grande esponente, Averroè, è intrinsecamente connessa alla recezione e all'adattamento delle fonti greche, l'attenzione per le traduzioni ha accompagnato sino dall'inizio gli studi sistematici della filosofia arabo-islamica e dei suoi grandi autori. Talvolta gli studiosi che si sono occupati per primi di questi filosofi sono gli stessi che hanno anche gettato le basi per lo studio delle traduzioni, come è il caso nel XIX sec. di Moritz Steinschneider[2] o, nel XX sec., di Amélie-Marie Goichon[3].

[1] Fanno eccezione alcuni studiosi secondo i quali la filosofia arabo-islamica si sarebbe sviluppata a partire dal Corano. Per una sintesi delle acquisitzioni della storiografia e per una panoramica dei testi tradotti si può vedere la voce *Greek Sources in Arabic and Islamic Philosophy* https://plato.stanford.edu/entries/arabic-islamic-greek/.

[2] Steinschneider ha scritto alla fine del XIX secolo la prima monografia su al-Fārābī (m. 950 d.C.): M. STEINSCHNEIDER, *Al-Farabi (Alpharabius) des Arabischen Philosophen Leben und Schriften mit besonderer Rücksicht auf die Geschichte der griechischen Wissenschaft unter den Arabern, nebst Anhängen Joh. Philoponus bei den Arabern, Leben und Testament des Aristoteles von Ptolemaeus, Darstellung der Philosophie Plato's, grösstentheils nach handschriftlichen Quellen*, in *Mémoires de l'Académie Impériale des Sciences de Saint Petersbourg*, VIII[e] série, tome XIII, No. 4, 1869 (rist.: Philo Press, Amsterdam, 1966) e due opere monumentali, tuttora consultate, che documentano le traduzioni medievali dal greco in arabo in ebraico: *Die arabische Übersetzungen aus dem Grieschischen*, Akademische Druck und Verlagsanstalt, Graz, 1960 (ristampa di una serie di articoli apparsi tra il 1889 e il 1896 nelle riviste «Beihefte zum Centralblatt für Bibliothekswesen, Zeitschrift für Deutschen Morgenländischen Gesellschaft» e «Archiv für pathologische Anatomie und Physiologie und für klinische Medizin», e *Die hebraischen Übersetzungen des Mittelalters und die Juden als Dolmetscher*, Kommissionsverlag des Bibliographischen Bureaus, Berlin, 1893 (rist. Graz, 1956).

[3] A.-M. Goichon ha tradotto Avicenna (Ibn Sīnā, m. 1037 d.C.), su cui ha scritto anche una monografia

La storia dei primi tentativi di lessici filosofici greco-arabi è stata tracciata da Soheil M. Afnan in un libro del 1964, *Philosophical Terminology in Arabic and Persian*, non senza un rilievo critico[4] sul quale tornerò tra poco, in questa breve introduzione ai decisivi contributi di Andrea Bozzi nel campo della linguistica computazionale applicata alle traduzioni greco-arabe. Annessi in modo via via più sistematico alle edizioni europee di traduzioni medievali arabe – soprattutto di opere aristoteliche – i glossari bilingui e gli indici terminologici delle occorrenze dei termini arabi con i loro corrispondenti greci hanno costituito il punto di partenza di una grande impresa lessicografica, confluita nel progetto "Greek into Arabic": il *Greek and Arabic Lexicon* (*GALex*) ideato e diretto da Gerhard Endress – protagonista anch'egli, assieme ai suoi collaboratori, di "Greek into Arabic".

Nel 1985, in occasione del primo *Symposium graeco-arabicum*[5], Gerhard Endress, assieme a Hans-Hinrich Biesterfeldt e a Dimitri Gutas, presentò quello che sarebbe in seguito divenuto il *GALex*: una raccolta sistematica e strutturata di materiali destinati a rendere "readily available to scholars the direct information which the Graeco-Arabic translations of the eighth to the tenth century contain for several areas of research. […] It is a first attempt to present in a rationalized and systematic way the lexical results of Graeco-Arabic studies during the past hundred years, and it should be viewed as a workbook containing methods and materials toward the compilation of a comprehensive Graeco-Arabic thesaurus in the future"[6]. È così – e si potrebbe dire solo così – che si può ovviare al difetto delle comparazioni "a priori", che era stato rilevato da Afnan nei lessici che non fornivano il contesto. Su che basi asserire che il termine arabo *ǧawhar* traduce *οὐσία*? Sempre, o solo talvolta? Solo questo termine, o anche altri? E quindi, generalizzando l'esempio: che valore hanno le affermazioni degli storici della filosofia o dei linguisti, se sono fondate su una corrispondenza stabilita in modo episodico, non circostanziato, e quindi in definitiva poco oggettivo?

tuttora importante: *La distinction de l'essence et de l'existence d'après Ibn Sīnā*, Desclée de Brouwer, Paris, 1937, e le dobbiamo anche il *Lexique de la langue philosophique d'Ibn Sīnā*, Desclée de Brouwer, Paris, 1938, con il supplemento: *Vocabulaires comparés d'Aristote d'Ibn Sīnā*, Desclée de Brouwer, Paris, 1939.

[4] S.M. AFNAN, *Philosophical Terminology in Arabic and Persian*, Brill, Leiden, 1964, pp. 2-3, critica la mancanza di contesto nei *Vocabulaires comparés* di A.-M. GOICHON nei seguenti termini, dopo averne lodato lo studio della terminologia filosofica di Avicenna: «But when in a subsequent work – cioè i *Vocabulaires comparés* – she undertook to supply the Greek equivalents, the *a priori* method was carried to the extreme. With no other authority save the *Index* of Bonitz, she indulged in a series of guesses with results hat are sometimes far from happy». Per evitare i rischi dell'"*a priori* method", Afnan seleziona un certo numero di termini filosofici significativi e mette in parallelo passi greci e versioni arabe, mostrando la traduzione del termine nel suo specifico contesto.

[5] H.-H. BIESTERFELDT - G. ENDRESS - D. GUTAS, *The Glossarium Graeco-Arabicum*, in P.L. SCHOONHEIM - G. ENDRESS (eds.), *Symposium graeco-arabicum I. The Transmission of Greek Texts in Mediaeval Islam and the West*, in *Proceedings of a Conference held at the Netherlands Institute for Advanced Study, Wassenaar, 19-21 February, 1985*, Studienverlag Brockmeier, Bochum, 1986.

[6] G. ENDRESS - D. GUTAS (eds.), *A Greek and Arabic Lexicon (GALex). Materials for a Dictionary of the Mediaeval Translations from Greek into Arabic*, Volume One: ا to ى, Brill, Leiden-New York-Köln, 2002 (Handbook of Oriental Studies, Section 1, vol. 11), p. 1*.

Il progresso degli studi nell'ambito delle traduzioni ha reso evidente che il grandioso fenomeno storico della trasmissione del sapere dal mondo antico al mondo medievale, nelle varie lingue dell'area euro-mediterranea, può essere conosciuto nel suo insieme solo a partire da conoscenze parziali acquisite con una metodologia scientifica, ossia con ricerche strutturate che danno origine ad acquisizioni controllabili e perciò oggettive. Nel caso delle traduzioni dal greco all'arabo, ciò ha determinato la formazione – appunto con il *GALex* – di un modello che prevede l'assegnazione esatta non soltanto dei contesti rispettivi (indicazione univoca della frase araba in cui si trova l'occorrenza desiderata, con riferimento a pagina e linea dell'edizione, e altrettanto univoca indicazione del passo greco tradotto) ma anche di una serie di parametri linguistici atti a presentare in modo chiaro l'equivalenza delle espressioni, a fronte della diversa struttura delle due lingue.

Elaborato per studiare le traduzioni greco-arabe nel senso stretto del termine 'traduzione', ossia quelle nelle quali lo scostamento della frase araba dalla frase greca è dovuto solo alla diversità a cui ho appena accennato, il modello rapresentato dal *GALex* è meno adatto per strutturare la conoscenza di un altro tipo di traduzioni: quelle parafrastiche. Traduzioni di questo tipo non solo sono esistite, ma sono state determinanti nella formazione della filosofia araba delle origini: in esse l'opera è stata resa in arabo con degli adattamenti, o all'interno stesso delle singole unità di significato (espressioni, frasi o pericopi più lunghe di una frase), o addirittura nella struttura dell'opera. È in questo modo che nel IX sec. dell'era cristiana sono stati costruiti alcuni testi decisivi per la formazione del pensiero filosofico arabo-islamico: attribuiti ad Aristotele, scritti di questa natura hanno talvolta modificato opere autenticamente aristoteliche, come i trattati di zoologia[7] o i *Meteorologica*[8]; talvolta invece, con una procedura analoga, opere non-aristoteliche sono state adattate – nel lessico, nella dottrina e nella struttura – ed attribuite ad Aristotele, come nel caso della cosiddetta *Teologia*, che deriva in realtà da una selezione di parti delle *Enneadi* di Plotino[9], o come nel caso del *Liber de causis*, che deriva in realtà dalla selezione e dall'adattamento degli *Elementi di teologia* di Proclo[10].

Discussioni lontane nel tempo e continui scambi successivi con Gerhard Endress – che è anche l'editore del *Proclus Arabus*[11] – sul modo in cui fosse oppor-

[7] Per la storia degli studi e per la composizione araba del *Libro degli animali* di Aristotele a partire da vari trattati zoologici cfr. E. CODA, *Il* Libro degli animali *(K. al-Ḥayawān). Materiali di studio sulla zoologia aristotelica nel medioevo arabo ed ebraico*, in M.M. SASSI (ed.), *La zoologia di Aristotele e la sua ricezione dall'età ellenistica e romana alle culture medievali*. Atti della X Settimana di formazione del Centro GrAL, Pisa U.P., Pisa, 2017 (Greco, Arabo, Latino. Le vie del sapere. Studi, 6).

[8] P.L. SCHOONHEIM, *Aristotle's Meteorology in the Arabico-Latin Tradition*, Brill, Leiden, 2000 (Aristoteles Semitico-Latinus, 12).

[9] G. ENDRESS, *Proclus Arabus. Zwanzig Abschnitte aus der Institutio Theologica in arabischer Übersetzung*, Imprimerie Catholique, Wiesbaden-Beirut, 1973.

[10] 'A. BADAWĪ, *Aflūṭīn 'inda l-'arab. Plotinus apud Arabes. Theologia Aristotelis et fragmenta quae supersunt*, Dār al-Nahḍa al-Miṣriyya, Cairo, 1966.

[11] O. BARDENHEWER, *Die pseudo-aristotelische Schrift ueber das reine Gute bekannt unter dem Namen Liber de causis*, Freiburg im Breisgau, 1882.

tuno strutturare la conoscenza di opere di questo tipo hanno trovato una soluzione nell'incontro del 2009 con Andrea Bozzi e con il suo approccio alla linguistica computazionale. Ne è nato *G2A*.

Chiunque abbia provato a servirsi per una ricerca scientifica dei vari esempi presenti nel web di traduzioni greco-arabe che si presentano come 'allineate' saprà che tale allineamento si perde in fretta nel corso dell'opera, spesso poco dopo l'inizio. La diversa struttura delle frasi rende difficile la consultazione, e avere i due testi greco e arabo aperti sul tavolo si rivela più semplice rispetto al far scorrere due colonne parallele, anche laddove – ad imitazione di *G2A* – essi siano divisi in caselle. Ma questa anticipazione rischia di non essere chiara, e andiamo quindi con ordine.

Due tipi di problemi sono stati sottoposti ad Andrea Bozzi quando abbiamo valutato insieme la fattibilità di una proposta ERC. Da un lato, il problema del trattamento nel quadro della linguistica computazionale di testi scritti in due alfabeti non latini, il greco e l'arabo, con l'obiettivo di produrre un sistema per la ricerca terminologica degli equivalenti. D'altro lato, il problema dell'allineamento di testi nei quali il disallineamento si presenta molto rilevante, e addirittura deliberato. *G2A* è stato studiato e realizzato da Andrea Bozzi e dai suoi collaboratori per permettere di analizzare e 'ricercare' – nel senso specializzato del verbo – in pericopi che possono essere viste affiancate o partendo dall'ordine della frase in greco, o partendo dall'ordine nel quale si presenta il testo arabo. Le caselle di G2A non sono un artificio grafico ma un potente strumento di ricerca: il testo che esse permettono di leggere in parallelo è ricercabile, e se si cerca un termine in arabo si è condotti su tutte le pericopi affiancate che permettono di vedere a partire da quale termine greco (con il suo contesto) esso è stato scelto dal traduttore; se si cerca un termine greco, si è condotti a vedere tutti i passi in cui esso compare, e a quali rese arabe ha dato origine (ciascuna con il suo contesto).

I problemi che Andrea Bozzi ha dovuto affrontare con i suoi collaboratori, le soluzioni innovative perseguite e trovate, la duttilità e le potenzialità di *G2A* non saranno descritte in questa breve presentazione, che non deve essere appesantita di punti di dettaglio. Concludo esprimendo, oltre alla gratitudine personale per degli anni di lavoro estremamente positivi su molti piani, la gratitudine che è appropriate nella scienza,cioè quella operativa. *G2A* non è usato soltanto per le ricerche lessicali, qualora cioè un ricercatore voglia sapere dove l'Uno di Plotino diviene l'"Essere puro, uno e vero, Dio benedetto e sublime" della pseudo-*Teologia di Aristotele*, oppure dove τὸ εἶναι è tradotto con *anniyya*e dove invece è tradotto con il verbo *kāna* nella *Metafisica* di Aristotele. Grazie ai suoi campi di commento e alla sua struttura duttile, *G2A* si sta dotando di un nuovo strumento per la comparazione dei testi greco-arabi con le traduzioni latine. È in questo modo che i ricercatori all'opera nel grande ambito della trasmissione del sapere filosofico e scientifico dal mondo antico al mondo moderno, attraverso il Medioevo, mettono a frutto il sapere e l'impegno profusi da Andrea Bozzi nel contesto di "Greek into Arabic".

INDICE

III. *Filologia del testo assistita da calcolatore: la stazione modulare*

IV. *Biblioteche digitali e beni librari*

V. *Un'applicazione di linguistica computazionale per testi antichi e loro traduzioni antiche*

Edizioni ETS
Palazzo Roncioni - Lungarno Mediceo, 16, I-56127 Pisa
info@edizioniets.com - www.edizioniets.com
Finito di stampare nel mese di giugno 2019