

# Dizionario elettronico per il latino «moderno»

Tre milioni di vocaboli della lingua colta dei secoli XVII e XVIII sono stati raccolti in una banca dati. Selezionate, schedate e scelte migliaia di opere, da Spinoza a Newton

L'incontro tra informatica e scienze umanistiche ha radici lontane. L'impiego di elaboratori per lo studio di discipline letterarie e filosofiche può infatti farsi risalire addirittura al 1949, anno in cui fu avviata, con l'intervento dell'Ibm, l'impresa dell'Index Thomisticus: un impegno che si è recentemente concluso con la pubblicazione degli oltre 60 volumi contenenti gli indici e le concordanze dell'Opera Omnia di San Tommaso d'Aquino.

In Europa, i primi centri dedicati all'elaborazione elettronica di dati linguistici sorsero a cavallo degli anni '60. Più precisamente, in Italia si arrivò alla costituzione dell'Istituto di linguistica computazionale del Cnr, grazie anche al convergere, presso il Centro nazionale universitario di calcolo elettronico dell'Università di Pisa, dei lavori promossi dall'ateneo di Padova, dall'Accademia della Crusca e dall'Index Thomisticus.

Oggi, a metà degli anni '80, ricerche più o meno sofisticate di questo tipo percorrono in lungo e in largo la nostra penisola culturale. Lo scorso anno, ad esempio, in occasione del bicentenario di Alessandro Manzoni, il centro studi «Franco Falletti» di Vercelli, diretto da Giorgio De Rienzo ed Egidio Del Boca, consegnò alla Fondazione Mondadori un dischetto che registrava in sequenza alfabetica tutte le parole dei «Promessi sposi»: da abate, che apre la fila, a zuffo, che la chiude. Un lavoro che ha riordinato le oltre 223 mila

parole del capolavoro manzoniano in circa 9000 lemmi, con un risultato anche editoriale: cinquemila pagine a stampa raccolte in cinque volumi di «Concordanze». Appena un anno prima, nel 1984, una studentessa di Alessandria aveva elaborato una tesi di laurea (relatore, non a caso, il professor De Rienzo di cui sopra) dal titolo: «Le avventure di Pinocchio. Concordanze». Il testo collodiano fu sezionato nelle sue 40659 parole, con 3371 lemmi sviluppati in 6064 forme.

Ricerche di questo tipo sono state condotte su numerosi altri testi, non esclusa la Divina Commedia. Ma al di là della ormai routinaria memorizzazione e sezionatura di un'opera, cos'altro può ricavare e pretendere lo studioso, l'informatico, l'umanista, da simili applicazioni? Una risposta a questa domanda l'ha offerta il Lessico intellettuale europeo (Lie), un centro di ricerca del Cnr costituito nel 1969, di cui è responsabile e vivace animatore il professor Tullio Gregory. Recentemente, infatti, in occasione e forse un po' troppo ai margini di un convegno organizzato dallo stesso Lie sul tema «Fantasia-immaginazione», svoltosi a Roma presso la facoltà di Filosofia a metà gennaio, il professor Gregory ha dimostrato cosa può fruttare il gemellaggio tra informatica e scienza umanistica: una banca dati del latino «moderno» unica al mondo.

Nei secoli XVII e XVIII il latino era ancora la lingua vivente nelle univer-

sità, nelle scuole, sebbene le lingue nazionali si fossero ormai costituite con la loro autonomia. Da questi ambii linguistici il Lie ha selezionato circa 3000 opere, le ha schedate con una lettura diretta, ne ha estratto tutti i contenuti significativi, memorizzando infine tutta la parte latina dopo un lavoro durato dieci anni.

«Oggi - ci spiega Tullio Gregory - non è così difficile memorizzare un testo. Per quanto ci riguarda abbiamo utilizzato gli strumenti informatici per il lessico degli autori (Bruno, Bacone, Vico, buona parte di Galilei, eccetera) soprattutto per lo spoglio. Non vogliamo infatti pubblicare gli spogli allo stato bruto, quelli che cioè vengono definiti per forma, e questo perché richiede tutte le concordanze secondo le varie parole, una volta memorizzato un testo, aprirebbe un mare enorme di dati. Il primo passo che invece sosteniamo è la lemmatizzazione, ossia riportare le forme sotto quella base che costituisce l'entrata dei dizionari d'uso: amiamo sotto amo, uomini sotto uomo, per fare un esempio. Si tratta comunque di un lungo lavoro, anche perché lingue come latino, greco e tedesco hanno complesse strutture grammaticali».

Così, oggi, il Lie ha memorizzato 49 opere filosofico-scientifiche scritte nella lingua dei colti di quell'epoca: dall'«Ethica» di Spinoza all'«Aesthetica» di Baumgarten, ai «Principia» di Newton. Il che significa possedere una banca dati di circa tre milioni di paro-

le, utilizzabile in varie maniere: applicando programmi particolari è possibile studiare quel lemma o quella parola che interessa, e magari verificare con quali altre parole si unisce: ad esempio, res con res cogitans.

Per questa impresa, il Lie ha adottato dei programmi Sperry, non solo in considerazione del fatto che il Centro di calcolo dell'università di Sperry (un 1100/80 S), ma anche perché la casa americana si è fatta promotrice di tutte le recenti iniziative del Lessico: dalla ricerca vera e propria ai «Colloqui internazionali», che da 12 anni riuniscono tutte le grandi imprese lessicografiche europee attorno a temi di chiave come «Res», «Spiritus», «Ordo», fino al più recente «Phantasia-Imaginatio».

«Da poco - afferma il professor Gregory - stiamo anche avviando dei programmi nostri per ricerca di autenticità d'opere, come l'analisi di strutture sintattico-grammaticali».

Per quanto riguarda il futuro, nei programmi del Lessico intellettuale europeo, rientra l'analisi e la computerizzazione del lessico del 1600 e 1700 europeo anche per quanto riguarda italiano, spagnolo, tedesco, francese e inglese.

«Degno di nota, intanto, è il fatto che qualsiasi studente può utilizzare servizi messi a disposizione dal Lie. Non a caso, il Centro è gemellato con una convenzione in parità assieme alla facoltà di Lettere per il corso di perfezionamento in «Informatica per le scienze umanistiche»: scuola che l'anno scorso aveva solo poche decine di studenti e quest'anno conta già 270 iscritti.

Per il momento, l'augurio è che altre iniziative fioriscano dall'esempio e dalla ormai pluriennale attività dell'équipe di Gregory. A conferma che spesso le buone idee camminano da sole, benché i mezzi scarseggino e i bilanci, come nel caso del Centro, siano fermi da tre anni.

Il discorso su scienze umanistiche e nuove tecnologie è ancora aperto.

g. b.

