RECENSIONI E SEGNALAZIONI

ESTRATTO

da

NUOVI ANNALI DELLA SCUOLA SPECIALE
PER ARCHIVISTI E BIBLIOTECARI
2018 ~ a. 32

Leo S. Olschki Editore
Firenze

NUOVI ANNALI DELLA SCUOLA SPECIALE PER ARCHIVISTI E BIBLIOTECARI

NUOVI
ANNALI
DELLA SCUOLA
SPECIALE PER
ARCHIVISTI E
BIBLIOTECARI

SAPIENZA UNIVERSITÀ DI ROMA

NUOVI ANNALI DELLA SCUOLA SPECIALE PER ARCHIVISTI E BIBLIOTECARI

Anno XXXII, 2018

LEO S. OLSCHKI EDITORE

«Nuovi annali della Scuola speciale per archivisti e bibliotecari»
is a peer-reviewed journal

Direttore
Alberto Petrucciani

Comitato di direzione
Paola Castellucci (coordinamento redazionale), Giovanni Paoloni,
Marina Raffaeli, Francesca Santoni

Hanno collaborato a questo volume:
Enrico Pio Ardolino, Eleonora De Longis, Lorenzo Mancini,
Paola Massa, Simona Turbanti

Comitato scientifico • Editorial Board
Alberto Bartola, Sapienza Università di Roma
Maria Teresa Biagetti, Sapienza Università di Roma
Italo Birocchi, Sapienza Università di Roma
Giorgetta Bonfiglio Dosio, già Università degli studi di Padova
Rosa Marisa Borraccini, Università degli studi di Macerata
Paola Carucci, già sovrintendente dell'Archivio storico della Presidenza della Repubblica
Flavia Cristiano, direttrice del Centro per il libro e la lettura del MiBACT
Attilio De Luca, già Sapienza Università di Roma
Luciana Duranti, University of British Columbia, Vancouver
Linda Giuva, Sapienza Università di Roma
Maria Guercio, già Sapienza Università di Roma
Elio Lodolini, professore emerito Sapienza Università di Roma
Antonio Manfredi, Biblioteca Apostolica Vaticana
Guido Melis, Sapienza Università di Roma
Antonella Meniconi, Sapienza Università di Roma
Outi Merisalo, University of Jyväskylä
Massimo Oldoni, già Sapienza Università di Roma
Fermín de los Reyes Gómez, Universidad Complutense de Madrid
Antonella Rovere, Università degli studi di Genova
Pedro Rueda Ramírez, Universitat de Barcelona
Deanna Shemek, University of California Santa Cruz
Marc Smith, École nationale des chartes, Paris
Giovanni Solimine, Sapienza Università di Roma
Paul Gabriele Weston, Università degli studi di Pavia

Sapienza
Università di Roma

La rivista è pubblicata con il contributo
della Sapienza Università di Roma

INDICE

Francesca Santoni, *AVA e «l'arte di decifrare le antiche scritture»* . . p. 5
Corinna Drago Tedeschini, *I* libri instrumentorum *della sezione LXVI dell'Archivio notarile generale urbano* . . . . . » 29
Federica Formiga, *Con Francisk Skorina l'arte della stampa arriva in Bielorussia* . . . . . » 53
Maria Cardillo, *Giuseppe Antonio Guattani e la genesi della stampa periodica illustrata a Roma a cavallo tra Settecento e Ottocento* . . » 69
Alberto Petrucciani, *Scrittori in biblioteca: Dino Campana alla Nazionale di Firenze (1905-1909)* . . . . . » 83
Federica Olivotto, *Il catalogo di libri cinesi di Lodovico Nocentini e la sua raccolta nella Biblioteca di Studi orientali della Sapienza* . . » 111
Sara Dinotola, *Teorie e pratiche dello sviluppo delle collezioni in area anglo-americana, tedesca e italiana: una rassegna comparata (dalla seconda metà del XIX alla fine del XX secolo)* . . . . . » 131
Mariangela Roselli, *Les enquêtes ethnographiques en bibliothèque: analyser les relations pour comprendre les barrières à la lecture publique* . . . . . » 171
Giovanni Paoloni, *In biblioteca e in archivio: una riflessione sulla natura documentaria dei bandi* . . . . . » 197
Salvatore Corasaniti, *Un thesaurus per l'indicizzazione di fonti radiofoniche: il caso dell'archivio di Radio Onda Rossa* . . . . . » 207
Emanuele Atzori – Francesca Nemore, *Vero, verosimile, immaginario: archivi e archivisti e letteratura* . . . . . » 229

RECENSIONI E SEGNALAZIONI

Luciano Canfora, *Per una storia delle biblioteche*, a cura e con una postfazione di Ugo Fantasia (Enrico Pio Ardolino) . . . . » 255
Konrad Hirschler, *Leggere e scrivere nell'Islam medievale*, edizione italiana a cura di Arianna D'Ottone Rambach (Giulia Villani) » 258
Itinerari del libro nella storia: per Anna Giulia Cavagna a trent'anni dalla prima lezione, a cura di Francesca Nepori, Fiammetta Sabba e Paolo Tinti (Roberta Cesana) . . . . . » 260

Antonio Iurilli, *Quinto Orazio Flacco: annali delle edizioni a stampa, secoli XV-XVIII* (Alberto Petrucciani) . . . . . p. 265

Collectanea Manutiana: studi critici su Aldo Manuzio, a cura di Pier Davide Accendere e Stefano U. Baldassarri (Eleonora de Longis) » 267

Antonella Barzazi, *Collezioni librarie in una capitale d'antico regime: Venezia secoli XVI-XVIII* (Lorenzo Mancini) . . . . . » 269

Before the public library: reading, community, and identity in the Atlantic world, 1650-1850, edited by Mark Towsey, Kyle B. Roberts (Alessandra Toschi) . . . . . » 271

Lettere di Theodor Mommsen agli italiani, a cura di Marco Buonocore (Chiara Cambrai) . . . . . » 273

Paolo Traniello, *Contributi per una storia delle biblioteche in età contemporanea* (Giovanni Paoloni) . . . . . » 277

The Library of Congress, *The card catalog: books, cards, and literary treasures* (Elisa Pederzoli) . . . . . » 279

Gabriele Turi, *Libri e lettori nell'Italia repubblicana* (Maria Iolanda Palazzolo) . . . . . » 283

Reading and the First World War: readers, texts, archives, edited by Shafquat Towheed, Edmund G.C. King (Eleonora de Longis) » 286

Federico Valacchi, *Archivio: concetti e parole* (Giulia Villani) . . » 288

Inventari e censimento delle fonti archivistiche degli Agostiniani in Toscana, a cura di Emanuele Atzori (Francesca Nemore) . . . . » 290

La musica e l'Arma, a cura di Flavio Carbone (Rovena Sakja) . . » 292

Milena Farina – Luciano Villani, *Borgate romane: storia e forma urbana* (Flavio Conia) . . . . . » 294

Dalla tutela al restauro del patrimonio librario e archivistico: storia, esperienze, interdisciplinarietà, a cura di Melania Zanetti (Francesca Nemore) . . . . . » 296

The net: la rete come fonte e strumento di accesso alle fonti: atti del convegno, Firenze, 25 febbraio 2016, a cura di Andrea Becherucci, Francesca Capetta (Giovanni Paoloni – Francesca Nemore) . . » 299

Mirko Tavosanis, *Lingue e intelligenza artificiale* (Simona Turbanti) » 301

Gino Roncaglia, *L'età della frammentazione: cultura del libro e scuola digitale* (Chiara Faggiolani) . . . . . » 305

Vita della Scuola . . . . . » 311

Mirko Tavosanis, *Lingue e intelligenza artificiale*, Roma, Carocci, 2018, 126 pp. (Le bussole, 555).

Dopo *L'italiano del web* (2011), Mirko Tavosanis, docente di Linguistica italiana all'Università di Pisa, si dedica al tema dell'elaborazione automatica del linguaggio che tocca discipline diverse. Nei sette capitoli del volume vengono affrontate, per passi successivi, le principali questioni connesse all'argomento. A un'agile introduzione (*Storia e contesto*) il compito di illustrare le caratteristiche e le 'anomalie' della ricerca condotta facendo presente che cosa si intenda oggi con «intelligenza artificiale», quale sia la differenza tra il riconoscimento e la comprensione del linguaggio umano, il tipo di problematiche legate al riconoscimento della lingua parlata, i modelli matematico-statistici alla base.

Il secondo capitolo, *Gli strumenti disponibili*, è dedicato alla presentazione dei software esistenti per l'elaborazione automatica del linguaggio. Tavosanis chiarisce innanzitutto le difficoltà, a partire dalla numerosità delle lingue parlate nel mondo (circa 6.000); se uno strumento come Google traduttore supporta oltre cento lingue, anche poco diffuse, è pur vero che esse sono comunque una minoranza rispetto al totale. Oltre a ciò, per molte lingue tra le cento previste, le 'prestazioni', incluse le funzioni base per il riconoscimento vocale, sono ridotte; le lingue maggiormente rappresentate dopo l'inglese sono le lingue nazionali europee e alcune fuori dall'Europa (cinese, giapponese, arabo, ecc.). Come dimostra il caso dell'italiano – ventunesima lingua al mondo per numero di parlanti madrelingua, ma assai supportata sotto il profilo dei sistemi di elaborazione automatica – il tipo di assistenza fornita non è influenzata solo da un elemento. Se si prende in considerazione l'aspetto economico, fattore rilevante per le aziende produttrici di questi sistemi, rispetto all'italiano sono soltanto sette le lingue più parlate da madrelingua e, al tempo stesso, incardinate in nazioni con realtà produttive fiorenti (cinese mandarino, spagnolo, inglese, giapponese, tedesco, francese, portoghese). L'italiano e le grandi lingue nazionali europee sono, quindi, le lingue che godono di tecnologie vocali avanzate; a ciò si aggiunge un'alta qualità degli strumenti informatici, come i dizionari e i corpora elettronici.

L'autore passa a illustrare le caratteristiche dei prodotti esistenti per la traduzione automatica, da sistemi di sintesi vocale, *Text to speech*, a strumenti con cui il computer riesce ad elaborare quanto pronunciato (*Speech to text*). Il secondo tipo di *tool* può prevedere o meno «una fase di addestramento alla voce del singolo parlante» (p. 35), usata a lungo per aumentare la qualità del riconosci-

mento vocale e, più recentemente, utile ai fini del riconoscimento dell'utente in procedure commerciali o burocratiche online. Pur di peso minore rispetto al passato, esistono ancora oggi vari limiti dal punto di vista lessicale; le uniche due aree in cui sono stati sviluppati strumenti *ad hoc* sono quelle relative alla trascrizione medica e giuridica. Occorre, inoltre, tenere presente come questi sistemi per la comprensione e trascrizione della lingua parlata funzionino quasi sempre in remoto, necessitando dunque di un collegamento in rete ben funzionante (e portandosi dietro, inevitabilmente, i problemi legati al trattamento dei dati sensibili). Vengono, infine, descritti i principali programmi di sintesi vocale disponibili.

Al centro del terzo capitolo (*Il riconoscimento del parlato*) la questione complessa della trascrizione da parte delle 'macchine' delle parole umane. Come ricorda Tavosanis, «Nel Ventesimo secolo, gli essere umani si sono abituati a sentire voci che escono da apparecchi elettrici ed elettronici: non si sono però abituati ad avere apparecchi che, senza l'intervento di esseri umani, rispondono o compiono azioni in base a ciò che viene detto» (p. 41). Il riconoscimento del parlato «relativamente controllato», come nel caso della richiesta vocale di avvio di una telefonata a un cellulare o la dettatura di testi a un computer, è ormai una realtà consolidata; ciò che invece continua ad essere problematico è ottenere buoni risultati in ambito di conversazioni umane 'normali'. Dopo aver elencato i principali sistemi esistenti, l'autore affronta il tema della gestione delle lingue: in linguistica si parla di «enunciazione mistilingue o *code mixing*» quando parole appartenenti ad almeno due lingue diverse sono utilizzate nella medesima frase, mentre con «commutazione o *code switching*» si intende l'uso di frasi in almeno due lingue differenti all'interno di uno stesso discorso. I sistemi di dettatura non sono in grado di identificare la lingua che non sia stata impostata – o impostata non correttamente – dall'utente, come invece accade nei principali sistemi di videoscrittura. Un ulteriore limite è rappresentato, inoltre, dal mancato inserimento, in automatico, di segni di punteggiatura che possono essere aggiunti soltanto ordinandolo esplicitamente al sistema. In chiusura del capitolo viene descritto uno degli indicatori più diffusi con cui valutare i sistemi di riconoscimento del parlato, ossia la percentuale di errori presenti in una trascrizione di un discorso (*Word error rate*, WER), e vengono analizzati esempi concreti.

Come già anticipato, i problemi maggiori di riconoscimento del parlato riguardano il registro informale, oggetto del quarto capitolo del libro, *La trascrizione delle conversazioni*. Tavosanis evidenzia l'importanza di questa attività nella società, per esempio in campo giuridico o nelle discusse intercettazioni telefoniche o ambientali – senza pensare ad altri possibili impieghi, come la produzione automatica di sottotitoli per audio e video – per la quale sono attive ditte e industrie. Le difficoltà nascono a causa dell'impossibilità di riprodurre nella scrittura alcune caratteristiche del parlato (tratti indicali, intonazione, segni paralinguistici). Nella scrittura, sia in quella tradizionale che in quella elettronica, si è tentato di trovare alcune alternative a questi elementi della lingua

parlata che 'non rientrano' nell'alfabeto. Le differenze tra il registro scritto e il parlato dipendono sia dalle caratteristiche del mezzo sia da vincoli temporali; esistono, infatti, «usi linguistici» tipici della lingua scritta o di quella orale. Con italiano «neostandard» (o «italiano dell'uso comune») si intende l'italiano, parlato e scritto, che si colloca in una posizione mediana in quanto a formalità e si diversifica in vari punti dalla lingua delle grammatiche; possono dunque definirsi «neostandard» gli elementi che contraddistinguono il parlato e lo scritto. All'atto pratico, le differenze, spesso semplificazioni, esistenti tra la lingua scritta e quella parlata rendono difficile trasporre sulla pagina un'«imitazione convincente del parlato spontaneo» (p. 63), nonostante questo costituisca un intento di vari scrittori moderni. Dopo aver delineato le componenti linguistiche tipiche di una conversazione, Tavosanis illustra i risultati ottenuti dagli attuali sistemi di riconoscimento automatico sottolineando una delle principali difficoltà riscontrate, vale a dire la mancata distinzione dei diversi soggetti che parlano, e la presenza di un fattore che influenza fortemente la qualità del riconoscimento (la velocità nell'enunciazione).

Nel quinto capitolo – *Il dialogo con le macchine* – viene affrontato il tema dell'interazione tra uomo e macchina; si parla di *computer talk*, una modalità di parlato semplificato per i dispositivi informatici (al pari del *baby talk* e del *foreigner talk*), benché sinora nessuna ricerca specifica in merito sia stata condotta. Attualmente gli assistenti vocali, come Cortana per Microsoft, Siri di Apple e le funzioni di Google (Google now, Google assistant, ecc.), difettano nell'utilizzo dei dati correttamente registrati e non sono in grado di garantire un vero dialogo. Non riescono, infatti, a gestire i riferimenti a parti precedenti della conversazione, versante sul quale molto probabilmente si giocherà la sfida dei futuri sistemi. L'autore fa poi notare l'enorme diffusione in tutto il mondo della conversazione testuale, mediante sistemi come WhatsApp o Messenger di Facebook, e ricostruisce brevemente l'evoluzione storica, a partire dai «primi tentativi di creare 'conversazione' con i sistemi informatici» (p. 86), condotti negli anni Sessanta e Settanta sulla scia del test di Turing del 1950. Al giorno d'oggi questi programmi sono definiti *chatbot* (neologismo derivante da *chat* + *robot*) e, benché abbiano per lo più un intento legato all'intrattenimento e allo svago, sono oggetto d'attenzione da parte dei 'colossi' come Google e Facebook; pur slegati dall'azione di riconoscimento del parlato e dai problemi che essa comporta, essi presentano le medesime restrizioni degli assistenti vocali nell'interpretazione degli *input* forniti loro, ossia «i limiti dell''intelligenza artificiale' intesa come capacità di agire autonomamente su compiti complessi» (p. 89).

La traduzione automatica, una delle finalità dell'elaborazione automatica del linguaggio, è il *focus* del sesto capitolo; come noto, la traduzione di un testo da una lingua a un'altra non è mai un'operazione meccanica. Tavosanis ricorda come lo spunto di usare la tecnologia per la traduzione nasca negli anni Quaranta del secolo scorso, ma i primi tentativi realmente efficaci siano molto posteriori e risalgano agli inizi del nuovo millennio. Al giorno d'oggi la traduzione automatica rappresenta anche un importante supporto per la tra-

duzione fatta dagli uomini: programmi di *computer assisted translation* (*CAT*) vengono utilizzati da traduttori professionisti, per esempio al fine di comparare e uniformare parti di testo ricorrenti più volte. Non è possibile, dunque, delimitare nettamente quanto discende dall'attività umana e quanto dalla macchina. Dopo una descrizione dei più diffusi sistemi di traduzione automatica della lingua scritta e del parlato – Google traduttore e, meno usato, Microsoft Translator – si sottolinea un aspetto essenziale nelle traduzioni, vale a dire la difficoltà di valutarne la qualità. Il modo più «intuitivo» di valutazione delle traduzioni automatiche consiste nel giudizio proveniente da professionisti 'in carne ed ossa', ma non esistendo criteri univoci e sempre validi per stabilire la superiorità di una traduzione rispetto a un'altra, neanche nel caso delle traduzioni umane, la procedura non è affatto banale. Inoltre, il costo delle valutazioni operate da esseri umani ha indotto a trovare altre strade, per esempio il confronto della traduzione automatica con una traduzione umana di riferimento, nella convinzione che «anche se le traduzioni possono essere parecchio diverse fra di loro, una buona traduzione abbia numerosi punti di contatto con altre buone traduzioni» (p. 100). BLEU, *Bilingual evaluation understudy*, è lo standard valutativo maggiormente adottato in questo metodo. I risultati di test effettuati nell'ambito del corso di Linguistica italiana II presso l'Università di Pisa, su traduzioni disponibili in rete di articoli di giornale in lingue diverse dall'italiano, mostrano due tratti ricorrenti: la qualità assai inferiore delle traduzioni automatiche rispetto alle traduzioni umane e un'equivalenza delle prestazioni tra i due principali sistemi a disposizione (Google traduttore e Microsoft Translator). Tavosanis si sofferma infine sulle differenze tra la traduzione della lingua scritta operata da traduttori e la stessa operazione condotta, in tempo reale, sul parlato da parte di interpreti; pur 'deboli', esistono anche sistemi di interpretariato automatico, ad esempio Microsoft Skype translator.

Nel settimo e ultimo capitolo, *Gli sviluppi prevedibili*, l'autore ipotizza quali possano essere le novità in futuro. Posto il limite insito nei sistemi di «intelligenza artificiale» (per i quali, come già chiarito, l'intelligenza è da intendersi in modo diverso dal significato comune), esiste «la possibilità di sviluppi imprevisti e interessanti» (p. 110). Già adesso queste tecnologie consentono di effettuare automaticamente operazioni prima assolvibili solo dagli uomini (basti pensare ai riconoscimenti delle persone presenti nelle fotografie caricate in rete eseguiti da Google foto); occorre, quindi, comprendere le ricadute dell'uso di tali strumenti sul piano linguistico facendo attenzione a non lasciarsi andare in rischiosi voli pindarici. Riallacciandosi a *L'italiano del web*, Tavosanis ricorda, infatti, come quanto accaduto con la comunicazione elettronica – il cui avvento non ha comportato alcun cambiamento degno di nota a livello di struttura delle lingue, ma ha permesso lo sviluppo della comunicazione scritta – debba invitare alla prudenza nelle previsioni. Secondo l'autore l'effetto maggiore dell'intelligenza artificiale, sempre sul piano linguistico, potrebbe essere in relazione con la flessibilità nei canali di scrittura. Anche le funzioni e gli assistenti vocali è prevedibile che non porteranno una rivoluzione nel

campo comunicativo e sociale, limitandosi ad offrire nuove opportunità, per esempio a livello di accessibilità. In chiusura, viene avanzata l'ipotesi, tanto futuristica quanto affascinante, che le conseguenze più evidenti potrebbero avvenire nello studio delle lingue straniere, soprattutto in quelle utilizzate per la comunicazione a fini pratici. Il ricorso alle tecnologie nella sfera didattica potrebbe condurre, infatti, alla fine delle lingue franche, come l'inglese nell'epoca attuale.

Questo volume costituisce una lettura stimolante anche per i non specialisti addentrandosi, con fare agile, in una serie di temi assai complessi ma i cui effetti pratici sono facilmente constatabili da ognuno di noi nella vita quotidiana. Al lettore vengono forniti numerosi spunti su cui riflettere alla luce della propria esperienza concreta, anche in futuro, a distanza di qualche anno.

Vorrei concludere riportando le parole usate dallo stesso Tavosanis durante un'intervista pubblicata sul sito *Letture.org* in merito ad un'ipotetica perfezione raggiungibile dalla traduzione automatica. «La perfezione probabilmente non si raggiungerà mai... Le lingue naturali sono strumenti di comunicazione molto imperfetti e imprecisi. Però oggi le traduzioni fatte da bravi esseri umani, anche se lasciano sempre fuori qualcosa, sono comunque uno strumento che consideriamo "abbastanza buono". Mi sembra quindi utile riformulare un po' la domanda di partenza: quand'è che le traduzioni automatiche saranno migliori di quelle fatte da un bravo traduttore umano? Forse solo quando esisterà una vera intelligenza artificiale simile a quella umana. Cosa che potrebbe avvenire nel giro di qualche decina d'anni, o forse mai: anche su questo gli esperti sono divisi. Anche prima del raggiungimento di questo obiettivo, però, la traduzione automatica può avere un impatto notevole. Oggi Google riesce a tradurre molte lingue meglio di quanto si possa fare con diversi anni di studio. Naturalmente, per gli esseri umani imparare lingue è importante di per sé, ed è anche un atto culturale rilevante. Però, per chi è interessato solo alla comunicazione pratica, perfino la traduzione di Google o Microsoft fa risparmiare parecchia fatica» (<https://www.letture.org/lingue-e-intelligenza-artificiale-mirko-tavosanis/>).

Simona Turbanti