Recensioni

📖 S. Bolasco, *L'analisi automatica dei testi. Fare ricerca con il text mining*, Carocci, Roma 2013, 410 pp., € 33,00.

«L'analisi automatica dei testi è ormai un compagno prezioso della conoscenza umana in tutte le sue forme»: così De Mauro apre la Prefazione all'ultimo lavoro di Bolasco, sottolineando il ruolo fondamentale di questa tecnica in moltissimi dei filoni di ricerca contemporanei, grazie alla duplice possibilità di studiare quantitativamente sia le caratteristiche strutturali di un testo sia i fattori (solo in parte espliciti) dei meccanismi individuali che portano alla sua creazione.

In generale, il volume offre una puntuale panoramica sulle potenzialità (oggigiorno verrebbe da dire necessità) di applicazione dell'analisi automatica a ricerche che muovono dai cosiddetti dati "non strutturati" (le parole, i sintagmi, le frasi) riuscendo, al contempo, a delineare uno snello sfondo teorico utile a motivare ed inquadrare con maggiore consapevolezza le tecniche applicative.

Il lavoro di Bolasco, collocato dall'editore nella collana di statistica, è caratterizzato da un taglio fortemente orientato al ricercatore sociale o umanista e riprende, rielaborandole e ampliandole in grande misura, le tematiche già affrontate dall'autore nel settimo capitolo del suo precedente volume *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione* (Carocci, Roma 1999), più chiaramente orientato ad un pubblico con consolidate basi matematico-statistiche.

L'intento del libro, secondo Bolasco, è «dare sistematicità a una materia in continua evoluzione, raccontandone in primo luogo i fondamenti e raccogliendo l'esperienza personale di ricerca degli ultimi quindici anni» (*Introduzione*). La sistematizzazione dei contenuti di base sul tema, quindi, predispone il volume ad un taglio manualistico di indirizzo didattico: per ammissione dell'autore, infatti, il libro include contenuti tratti da lezioni e seminari svolti su un lungo arco temporale, riconducibili all'approccio della scuola statistica francese. Allo stesso tempo l'inclusione di dati scaturiti da analisi pubblicate in articoli e riviste di settore mira ad arricchire la teoria con esempi reali che facilitino la trasmissione dei contenuti. Sempre nell'Introduzione si motiva l'assenza di un esaustivo quadro storico sull'evoluzione dei concetti di linguistica quantitativa, statistica linguistica e lessicale, e quanto attenga alle nozioni proprie del *text mining* per motivi di spazio. Tale assenza, tuttavia, rimane giustificata soprattutto dalla natura a cavallo tra statistica, linguistica e informatica dell'analisi automatica dei testi, la cui indagine storica presupporrebbe *excursus* ed analisi in tutti e tre i campi che di certo non gioverebbero al lettore che aspiri ad un approccio focalizzato su metodi e applicazioni.

Una *prima parte* del volume, che da sola copre poco più della metà del libro, è dedi-

cata a questioni teoriche e metodologiche, fornendo definizioni e concetti generali nonché modelli e soluzioni impiegate nei processi di "scavo o estrazione" dell'informazione dai testi (*text mining*, dall'inglese *to mine*, 'scavare, estrarre in miniera'). Una *seconda parte* è rivolta alla descrizione di risorse e strumenti a disposizione del ricercatore, con particolare riferimento alla lingua italiana. La *terza parte*, infine, consiste di una raccolta di casi di studio e applicazioni che mostrano, nella pratica, le potenzialità dell'analisi automatica. Manca alla fine del volume, o al termine di ciascun capitolo, un sintetico elenco di indicazioni bibliografiche per approfondimenti tematici o letture consigliate a cui potrebbe fare riferimento il lettore che volesse approfondire temi legati ad argomenti specifici lasciati fuori per motivi di spazio (come quelli di cui si è appunto detto poc'anzi). L'ampio insieme di riferimenti presente nelle note a piè pagina riesce, in ogni caso, a indirizzare adeguatamente su fonti esterne chi voglia conoscere più in dettaglio teorie, sviluppi e applicazioni sullo specifico tema.

In generale il volume affronta in maniera chiara e strutturata la parte definitoria. Nelle definizioni dei concetti generali della linguistica computazionale, quali ad esempio *token*, *rango*, *concordanza*, *vocabolario*, *corpus*, si fa ampio riferimento alla consolidata letteratura linguistica in materia, nonostante il quadro generale spesso si confronti necessariamente con sfumature terminologiche di matrice statistica, come nei casi di *bootstrap* o *tabelle di contingenza*. Va sottolineata la presenza di terminologia riconducibile specificamente all'autore o ai lavori del gruppo di ricerca in cui opera, come ad esempio per *lessico peculiare* o *lessico specifico* i quali, anche in forza della loro originalità, vengono definiti e descritti in maniera esaustiva. Un caso particolare è quello dei *polimorfi*, definiti come «segmenti di senso compiuto» (p. 53), che sono perciò riconducibili, per ammissione dell'autore, a ciò che in letteratura conosciamo come unità polirematiche. Va notato, tuttavia, che l'indirizzo applicativo del volume, orientato più all'estrazione di *contenuto* dai testi che alla pura analisi linguistica, permette all'autore di includere nei polimorfi espressioni del tipo *nuovi posti di lavoro* (p. 126) che difficilmente verrebbero considerate unità di senso da un linguista.

Per quanto riguarda i contenuti, gli argomenti risultano di ampia copertura in materia di analisi sia linguistica che statistica. Le problematiche relative alla costituzione di un corpus e al suo trattamento vengono delineate in maniera chiara e ragionata, fornendo puntualmente esempi esplicativi. Le analisi lessicale e testuale vengono affrontate attraverso una serie di argomenti di complessità via via crescente. A partire dalle unità minime di analisi, quali parole, frammenti, sequenze, il discorso si sviluppa sull'importanza dei meta-dati, del comportamento in frequenza e sulle differenze tipologiche, riscontrabili empiricamente, tra diversi tipi testuali (*in primis* scritto e parlato), concedendo anche esaustive panoramiche sulle risorse disponibili per *tasks* specifici (si veda il caso del *tagging* grammaticale per l'italiano, p. 116). Vengono quindi delineate le procedure di estrazione relative ai concetti specifici dell'analisi lessicale, quali il linguaggio peculiare, il linguaggio specifico e, per l'analisi testuale, l'indice di rilevanza *tf-idf*. Grande spazio viene lasciato all'introduzione delle tecniche di interrogazione dei dati (*query*) in maniera automatica e semiautomatica, mostrando sia esempi corretti sia, a volte, errati, segnalati al fine di aiutare il lettore inesperto a focalizzare l'attenzione su dettagli e caratteristiche specifiche dei linguaggi di *query*. La statistica testuale e il *text mining* vengono affrontati con una strategia analoga, privilegiando l'introduzione di elementi di base per poi lasciare spazio alle metodologie relative alle misure di associazione tra parole o testi e all'analisi delle corrispondenze, utile ad estrarre le

relazioni latenti e non esplicite tra le unità di un *corpus*.
La parte relativa alle risorse statistico-linguistiche espone le caratteristiche principali di dizionari elettronici, lessici di frequenza e campioni di lingua di riferimento, problematizzandone aspetti di costruzione e limiti, ma descrivendo anche applicazioni. Come esempi principali vengono riportati studi di *grammatiche locali*, ovvero ricerche di pattern relativi a specifiche categorie di locuzioni che sintetizzino mediante grafi tutti i possibili lessemi e connessioni sintattiche presenti nel corpus e relative alla struttura in esame: una chiara prospettiva, questa, che si ricollega allo sfondo teorico del lessico-grammatica di Gross.
Uno spazio importante della sezione sugli strumenti è riservato ad introdurre il lettore all'uso del software di analisi TALTAC², sviluppato dall'autore, che annovera tra le proprie risorse una parte rilevante delle metodologie esposte nel volume. Gli esempi riportati guidano il lettore a partire dai passi preliminari di importazione di un *corpus* e del suo trattamento, fino agli obiettivi più specifici come l'individuazione di lessie complesse o la categorizzazione automatica dei documenti. Lo stesso software è impiegato in sezioni rilevanti delle analisi per i casi di studio della parte finale, che spaziano da ricerche sul linguaggio eno-gastronomico e politico, all'estrazione di contenuto a partire da testi di varia tipologia (risposte brevi, scambi epistolari). Pur riservando ad altre applicazioni informatiche la possibilità di compiere analisi più prettamente statistiche, come l'estrazione di contenuto semantico latente, TALTAC² rimane l'unico software di esplicito riferimento del volume. Un altro testo di Giuliano e La Rocca (*L'analisi automatica e semi-automatica dei dati testuali. Software e istruzioni per l'uso*, LED, Milano 2008), di approccio fortemente applicativo, può rivelarsi più esaustivo in tal senso, includendo oltre ad esempi d'uso di TALTAC², introduzioni e guide d'uso ad ulteriori tre software (ATLAS.TI5, NVIVO7, LEXICO3) in grado di coprire sia le necessità d'analisi linguistica che quelle più propriamente tipiche della statistica sociale.
È inoltre inevitabile il confronto del volume di Bolasco con un'altra opera dai contenuti simili apparsa ormai da più di dieci anni, vale a dire *L'analisi del contenuto* di A. Tuzzi (Carocci, Roma 2003), che affronta in maniera molto più sintetica alcuni dei punti trattati anche da Bolasco, come la creazione, la suddivisione e il trattamento di un *corpus*, l'importanza del fattore frequenza, nonché l'analisi delle corrispondenze. Il volume di Tuzzi, pur includendo casi di studio significativi, nel confronto risulta oltre che sintetico, di minore copertura a livello di contenuti, nonostante possieda un'ampia sezione metodologica. Esso include, inoltre, numerosi formalismi matematici che, pur garantendo rigore, potrebbero disorientare i non addetti ai lavori. In ogni caso, lo stile piano, la chiarezza dei contenuti e la quantità di esempi e riferimenti (in particolare nei due paragrafi finali, dedicati ad una panoramica su software e bibliografia minima di approfondimento, mancante in Bolasco) hanno reso il lavoro di Tuzzi il naturale punto di riferimento italiano sull'analisi automatica testuale nel periodo successivo all'uscita del già citato volume di Bolasco del 1999. Pur restando validi gli assunti, le definizioni e le metodologie, il testo di Tuzzi sconta inevitabilmente il non aggiornamento sugli studi e le evoluzioni degli ultimi dieci anni, su cui, invece, il nuovo lavoro di Bolasco mostra grande copertura.
Il volume di Bolasco, nel complesso, ha il merito di assolvere l'intento di dare esaustiva sistematizzazione ad un ambito ancor privo, ad oggi, di uno statuto definito. Uno dei suoi maggiori meriti risulta la focalizzazione sulla lingua italiana, carente (se non per il lavoro di Tuzzi) di lavori in volume che tematizzino adeguatamente gli sviluppi delle tecniche in esame. Le descrizioni e le analisi esposte rimangono, tuttavia, spesso orientate al piano della *forma*

lessicale, intesa come occorrenza "grezza" della parola nel testo, pur contemplando il *tagging* grammaticale. Sono assenti applicazioni e risultati che sfruttino il *parsing* sintattico e il riconoscimento dei ruoli funzionali delle unità, sempre più importanti e significativi nelle analisi lessicali, in particolare per l'individuazione di lessemi complessi (si confronti, ad esempio, il testo di V. Seretan, *Syntax-based Collocation Extraction*, Springer, Berlin 2011), che vengono privilegiati, però, ai soli fini della pura analisi linguistica. Tale mancanza, quindi, non limita l'utilità generale del volume, specie per chi si approcci per la prima volta a tali tematiche. Anche il ricercatore più esperto potrà comunque beneficiare dell'ampia varietà di metodi e applicazioni che offrono un quadro di base integrabile, all'occasione, con altre risorse o metodologie. Il maggior pregio del volume rimane, comunque, il suo fungere da ponte tra linguistica e statistica, e rappresentare quindi il varco, per lo studioso dell'uno o dell'altro campo, verso le potenzialità di integrazione con l'altra disciplina. Quanto questa integrazione abbia già avuto successo lo dimostrano gli studi esposti o citati lungo l'intero volume, i quali prospettano sviluppi di ricerca che non potranno che far evolvere in misura maggiore l'analisi automatica dei testi verso traguardi sempre più interessanti.

Luigi Squillante

📖 E. M. Pandolfi, S. Christopher, B. Somenzi, *Capito? Comprendere l'italiano in Svizzera*, Osservatorio linguistico della Svizzera italiana (OLSI), Bellinzona 2014, 305 pp., 40 Fr. (libro + DVD).

Gli intensi contatti tra i parlanti dei paesi europei e l'eterogeneità linguistica del continente pongono nuove sfide per la glottodidattica contemporanea, disciplina interessata allo sviluppo di percorsi di apprendimento che rendano più immediato l'accesso alle altre lingue e lo scambio culturale. All'interno di questo contesto si inserisce il manuale *Capito? Comprendere l'italiano in Svizzera*, frutto di un progetto nato dalla collaborazione tra il Centro scientifico di competenza sul plurilinguismo (CSP, Friburgo), l'Università di Berna e l'Osservatorio linguistico della Svizzera italiana (OLSI, Bellinzona). Il volume propone un percorso glottodidattico finalizzato a stimolare la competenza ricettiva in italiano, permettendo un primo approccio con la lingua e la cultura del Ticino e dei Grigioni italiani. I destinatari sono apprendenti adulti che conoscono il francese come lingua prima o seconda e che utilizzano il manuale sia come risorsa per l'auto-apprendimento sia all'interno di un corso di lingua. In quest'ultimo caso, come consigliano le autrici nell'introduzione, il testo andrebbe accompagnato da esercizi mirati a sviluppare le abilità produttive degli studenti.

Il manuale è suddiviso in sette unità didattiche che includono le soluzioni e le trascrizioni degli esercizi, più alcune interessanti appendici in cui si illustrano le parole che sono presenti nel libro e che non sono in comune tra italiano e francese. Infine, sempre nelle appendici, le autrici elencano i siti e i libri da cui sono stati tratti i testi e che potrebbero fornire materiale di ulteriore approfondimento per lo studente.

La prima unità didattica s'intitola *Chi ben comincia è a metà dell'opera* ed è stata redatta principalmente in francese, venendo incontro alle esigenze degli apprendenti. Le autrici si rivolgono direttamente ai destinatari del libro, spiegando in nove punti le motivazioni della scelta di un approccio non tradizionale all'insegnamento dell'italiano LS. Innanzitutto, si ricorda come la corretta ricezione della lingua costituisca la base per una buona produzione linguistica, soprattutto nel caso di uno studente adulto, il quale possiede già un repertorio linguistico, che include sia la lingua materna sia le altre lingue conosciute. Infatti l'apprendente necessita di uno sforzo inferiore per la comprensione di una lingua

rispetto, invece, alla produzione linguistica. Inoltre, il processo che consente la ricezione della lingua si attiva anche grazie alle conoscenze che ciascun parlante ha del mondo che lo circonda: si tratta di un processo attivo e razionale, ma anche emotivo. La comprensione della nuova lingua è facilitata dalla prossimità lessicale, morfologica, sintattica e fonetica con la lingua nativa dello studente. Per sviluppare tale abilità comparativa sono necessarie strategie specifiche, che attivino le conoscenze degli apprendenti, incluse quelle latenti. Pertanto, se ciascun parlante fosse in grado di comprendere gli enunciati di un'altra lingua, potrebbe esprimersi nella propria senza ostacolare la comunicazione, minimizzando lo sforzo di acquisizione.

Dopo aver spiegato le motivazioni alla base della scelta di un percorso non tradizionale di apprendimento, le autrici propongono alcuni esercizi di simulazione di situazioni comunicative quotidiane, invitando gli studenti a riflettere sugli elementi che potrebbero facilitare la comprensione degli enunciati. Infine, sempre nel primo capitolo, vengono illustrate le «clés pour la compréhension», che costituiscono una legenda dei meccanismi impliciti attivi nel processo di comprensione. Tali chiavi hanno l'obiettivo di stimolare la riflessione metacognitiva e metalinguistica, facilitando i meccanismi di comprensione e rendendo gli studenti coscienti e partecipi del metodo proposto. Infatti le chiavi mettono in evidenza, attraverso colori diversi, le informazioni fornite dal contesto in cui si svolge lo scambio comunicativo, gli elementi precedenti nel discorso (parlato o scritto), il lessico condiviso con il francese e i meccanismi di formazione delle parole (prefissazione e suffissazione).

A partire dalla seconda unità, intitolata *… che volge a mezzogiorno*, le autrici inseriscono le sezioni "temi" e "grammatica". Gli esercizi proposti sono di diverse tipologie: quiz a scelta multipla, *cloze test*, domande a risposta aperta, domande vero/falso. Gli argomenti affrontati nel secondo capitolo riguardano le conoscenze geografiche, economiche, sociali e culturali della Svizzera italiana, con brani, per esempio, sulla costruzione dell'*AlpTransit*, sull'immigrazione e sul fenomeno dei frontalieri. La seconda parte dell'unità contiene indicazioni sulla formazione del presente (indicativo e congiuntivo) dei verbi regolari e di alcuni verbi irregolari e sulle somiglianze morfologiche tra italiano e francese.

Nella terza unità didattica, *L'appetito vien mangiando*, si approfondiscono le tradizioni culinarie della Svizzera italiana, dalla storia dei grotti alle aziende che producono cioccolato. La sezione di grammatica, invece, è incentrata sui pronomi personali, le forme di cortesia, i possessivi e il lessico della gastronomia.

Nella quarta unità, *La pratica è maestra di vita*, gli studenti possono esplorare il vocabolario legato alle professioni e alle abitudini quotidiane. Inoltre, le autrici illustrano le regole per una corretta formazione dell'imperfetto e del passato prossimo e indicano le forme irregolari di alcuni verbi, i giorni della settimana, i mesi dell'anno, le stagioni, alcune espressioni di tempo e i numeri.

La quinta unità didattica, *Non si vive di solo pane*, contiene testi ed esercizi sul tempo libero. Si propone agli studenti una riflessione sul tempo dedicato al lavoro e alla cura della famiglia. Le indicazioni grammaticali riguardano la formazione del condizionale (verbi regolari e irregolari) e del futuro semplice, i verbi riflessivi, alcune espressioni di luogo e le preposizioni articolate.

Canta che ti passa è il titolo della sesta unità, che comprende principalmente interviste ad artisti contemporanei, insieme a nozioni grammaticali su participio passato, gerundio e passato remoto. Infine, la settima unità didattica, dal titolo *Mente sana in corpo sano*, è incentrata sull'importanza dell'attività fisica per una vita sana; le nozioni grammaticali riguardano alcune interiezioni e le regole per la formazione dell'imperativo formale e informale.

L'obiettivo primario di questo lavoro consiste nel promuovere un plurilinguismo autentico in Svizzera, che tenga conto delle identità linguistiche e culturali presenti sul territorio. Inoltre, il percorso proposto ha lo scopo di aiutare lo studente a superare possibili barriere psicologiche e motivazionali che possono emergere nell'apprendimento di una nuova lingua. Pertanto gli esercizi sono costruiti in modo da stimolare il confronto linguistico e culturale, minimizzando lo sforzo di apprendimento ed evitando il rischio di una demotivazione precoce.

La rapida acquisizione di un approccio comparativo consente inoltre di attivare competenze spesso non utilizzate ma fondamentali per l'apprendimento linguistico. Attraverso la riflessione sulle parentele linguistiche e sulla circolazione internazionale del lessico, lo studente si appropria di strumenti che aiutano a sviluppare la competenza ricettiva della nuova lingua. Il confronto tra culture ma anche tra strutture lessicali, morfologiche e sintattiche di due lingue profondamente legate consente di superare un'impostazione troppo spesso fondata su batterie ripetitive di esercizi grammaticali e sulla correzione del singolo errore. Un percorso di apprendimento deduttivo, basato sulla riflessione linguistica e metalinguistica consente di focalizzare lo studente sulla costruzione del processo logico che permette, ad esempio, di comprendere il senso corretto di un testo. In questo modo è possibile affrontare con più tranquillità i timori che spesso ostacolano l'avvicinamento ad una nuova lingua, illustrati nel primo capitolo del manuale (p. 31): 1. *Sono troppo* ***vecchio*** […]; 2. *Non sono dotato per le lingue* […]; 3. *Se imparo ancora una lingua simile,* ***mi confondo*** […]; 4. *Se* ***imparo*** *una nuova lingua rischio di* ***dimenticare*** *l'altra o le altre lingue straniere* […]; 5. *Ho timore di parlare una lingua* ***finché*** *non la parlo correttamente* […].

Il volume è il risultato di riflessioni che travalicano la dimensione linguistica e che includono osservazioni di tipo socio-linguistico che possono agevolare la fase iniziale di apprendimento di una nuova lingua. L'approccio comparativo proietta lo sguardo al di là della singola lingua che si sta studiando e crea un terreno fertile per la conoscenza di nuove lingue imparentate fra loro. Se il libro si rivolge a studenti dell'area romanza, una proposta simile potrebbe essere sviluppata anche per le lingue germaniche o slave. Con *Capito? Comprendere l'italiano in Svizzera* E. M. Pandolfi, S. Christopher e B. Somenzi propongono un percorso innovativo e stimolante nel panorama della glottodidattica italiana e mostrano una forte sensibilità sociolinguistica, che si manifesta nel proposito di salvaguardare l'identità linguistica e culturale di ciascun paese, promuovendo, allo stesso tempo, il plurilinguismo in Europa.

Chiara Gargiulo

Postilla a una recensione

Nel numero XI, fasc. I, 2014, della presente rivista, alle pp. 188-99, è uscita un'ampia e dettagliata recensione di *Dante a Verona nel Settecento. Studi su Giovanni Iacopo Dionisi*, volume da me pubblicato nel 2012. Di questo ringrazio il Comitato scientifico della rivista e in particolar modo il benevolo recensore, Luca Fiorentini, lettore attento e intelligente: è consolante vedere come le proprie opere siano analizzate in maniera così acuta, precisa e dettagliata. Proprio per preservare la recensione da imprecisioni, mi corre l'obbligo di segnalare una svista, un errore che mi viene attribuito, ma che in realtà non ho commesso. A p. 192, si afferma che sarebbe sbagliato il riferimento alle pagine nelle quali Dionisi parla della variante *ingoia* in luogo del vulgato *scuoja* a *Inf.* VI 18. Il rimando bibliografico è il seguente: Dante, *Commedia* (ed. Dionisi), I, pp. II-III (si trova alla n. 34 di p. 54 del mio volume). Questo rimando, come si evince dalla tavola delle abbreviazioni bibliografiche posta alla fine del volume, è riferito all'*Ag-*

giunta critica posta dopo il testo del poema nella Bodoniana, l'edizione della *Commedia* curata da Dionisi, e non, come crede Fiorentini, all'introduzione posta prima del testo del poema, che ho siglato in altro modo (Dionisi 1795). Il riferimento è dunque corretto: Dionisi discute del passo alle pp. II-III dell'*Aggiunta critica del canonico Dionisi alla cantica dell'Inferno*.

Luca Mazzoni