

Giulia Venturi, Andrea Cimino,
Felice Dell'Orletta

La fede dichiarata

**Un'analisi
linguistico-computazionale**

Laboratorio Sociologico

Ricerca empirica
ed intervento sociale

FRANCOANGELI

Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con **Adobe Acrobat Reader**



La versione completa dell'e-book (a pagamento) è leggibile **con Adobe Digital Editions**.

Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta [cliccando qui](#) le nostre F.A.Q.

Direttore Scientifico: Costantino Cipolla

Laboratorio Sociologico approfondisce e discute criticamente tematiche epistemologiche, questioni metodologiche e fenomeni sociali attraverso le lenti della sociologia. Particolare attenzione è posta agli strumenti di analisi, che vengono utilizzati secondo i canoni della scientificità delle scienze sociali. Partendo dall'assunto della tolleranza epistemologica di ogni posizione scientifica argomentata, Laboratorio Sociologico si fonda su alcuni principi interconnessi. Tra questi vanno menzionati: la combinazione creativa, ma rigorosa, di induzione, deduzione e adduzione; la referenzialità storico-geografica; l'integrazione dei vari contesti osservativi; l'attenzione alle diverse forme di conoscenze, con particolare attenzione ai prodotti delle nuove tecnologie di rete; la valorizzazione dei nessi e dei fili che legano fra loro le persone, senza che queste ne vengano assorbite e – ultimo ma primo – la capacità di cogliere l'alterità a partire dalle sue categorie "altre". Coerentemente con tale impostazione, Laboratorio Sociologico articola la sua pubblicistica in sei sezioni: *Teoria, Epistemologia, Metodo; Ricerca empirica ed Intervento sociale; Manualistica, Didattica, Divulgazione; Sociologia e Storia; Diritto, Sicurezza e Processi di vittimizzazione; Sociologia e storia della Croce Rossa.*

Comitato Scientifico: Natale Ammaturo†; Ugo Ascoli (Ancona); Claudio Baraldi (Modena e Reggio Emilia); Leonardo Benvenuti, Ezio Sciarra (Chieti); Danila Bertasio (Parma); Giovanni Bertin (Venezia); Rita Biancheri (Pisa); Annamaria Campanini (Milano Bicocca); Gianpaolo Catelli (Catania); Bernardo Cattarinussi (Udine); Roberto Cipriani (Roma III); Ivo Colozzi, Stefano Martelli (Bologna); Celestino Colucci (Pavia); Raffaele De Giorgi (Lecce); Paola Di Nicola (Verona); Roberto De Vita (Siena); Maurizio Esposito (Cassino); Antonio Fadda (Sassari); Pietro Fantozzi (Cosenza); Maria Caterina Federici (Perugia); Franco Garelli (Torino); Guido Giarelli (Catanzaro); Guido Gili (Campobasso); Antonio La Spina (Palermo); Clemente Lanzetti (Cattolica, Milano); Emiliana Mangone (Salerno); Giuseppe Mastroeni (Messina); Rosanna Memoli (La Sapienza, Roma); Everardo Minardi (Teramo); Giuseppe Moro (Bari); Giacomo Mulè (Enna); Giorgio Osti (Trieste); Mauro Palumbo (Genova); Jacinta Paroni Rumi (Brescia); Antonio Scaglia (Trento); Silvio Scanagatta (Padova); Francesco Sidoti (L'Aquila); Donatella Simon (Torino); Bernardo Valli (Urbino); Francesco Vespasiano (Benevento); Angela Zanotti (Ferrara).

Corrispondenti internazionali: Coordinatore: Antonio Maturo (Università di Bologna) Roland J. Campiche (Università di Losanna, Svizzera); Jorge Gonzales (Università di Colima, Messico); Douglas A. Harper (Dquesne University, Pittsburgh, USA); Juergen Kaube (Accademia Brandeburghese delle Scienze, Berlino, Germania); André Kieserling (Università di Bielefeld, Germania); Michael King (University of Reading, Regno Unito); Donald N. Levine (Università di Chicago, USA); Christine Castelain Meunier (Casa delle Scienze Umane, Parigi, Francia); Maria Cecilia de Souza Minayo (Escola Nacional de Saúde Pública, Rio de Janeiro, Brasile); Everardo Duarte Nunes (Universidade Estadual de Campinas, São Paulo, Brasile); Furio Radin (Università di Zagabria, Croazia); Joseph Wu (Università di Taiwan, Taipei, Taiwan).

Coordinamento Editoriale delle Sezioni: Giuseppe Masullo

Ogni sezione della Collana nel suo complesso prevede per ciascun testo la valutazione anticipata di due referee anonimi, esperti nel campo tematico affrontato dal volume. Alcuni testi di questa collana sono disponibili in commercio nella versione e-book. Tali volumi sono sottoposti allo stesso controllo scientifico (doppio cieco) di quelli presentati in versione a stampa e, pertanto, ne posseggono lo stesso livello di qualità scientifica.

Sezione *Teoria, Epistemologia, Metodo* (attiva dal 1992). *Responsabile Editoriale*: Leonardo Altieri. *Comitato Editoriale*: Agnese Accorsi; Gianmarco Cifaldi; Francesca Cremonini; Davide Galesi; Francesco Gandellini; Ivo Germano; Maura Gobbi; Francesca Guarino; Silvia Lolli jr.; Alessia Manca; Emmanuele Morandi†; Alessandra Rota; Barbara Sena.

Sezione *Ricerca empirica ed Intervento sociale* (attiva dal 1992). *Coordinatore Scientifico*: Andrea Bassi; *Responsabile Editoriale*: Sara Sbaragli. *Comitato Editoriale*: Sara Capizzi; Teresa Carbone; Paola Canestrini; Carmine Clemente; David Donfrancesco; Laura Farneti; Ilaria Iseppato; Lorella Molteni; Paolo Poletti; Elisa Porcu; Francesca Rossetti; Alessandra Sannella.

Sezione *Manualistica, Didattica, Divulgazione* (attiva dal 1995). *Coordinatore Scientifico*: Linda Lombi. *Responsabile Editoriale*: Arianna Marastoni. *Comitato Editoriale*: Veronica Agnoletti; Flavia Atzori; Alessia Bertolazzi; Barbara Calderone; Raffaella Cavallo; Carmela Anna Esposito; Laura Gemini; Silvia Lolli sr.; Ilaria Milandri; Annamaria Perino; Fabio Piccoli.

Sezione *Sociologia e Storia* (attiva dal 2008). *Coordinatore Scientifico*: Nicola Strizzolo (Università di Udine) *Consiglio Scientifico*: Nico Bortoletto (Università di Teramo); Alessandro Bosi (Parma); Camillo Brezzi (Arezzo); Luciano Cavalli, Pietro De Marco, Paolo Vanni (Firenze); Sergio Onger, Alessandro Porro (Brescia); Carlo Prandi (Fondazione Kessler – Istituto Trentino di Cultura); Adriano Prosperi (Scuola Normale Superiore di Pisa); Renata Salvarani (Cattolica, Milano); Paul-André Turcotte (Institut Catholique de Paris). *Responsabile Editoriale*: Alessandro Fabbri. *Comitato Editoriale*: Barbara Baccarini; Roberta Benedusi; Elena Bittasi; Emanuele Cerutti; Pia Dusi; Giancarlo Ganzerla; Nicoletta Iannino; Riccardo Maffei; Vittorio Nichilo; Ugo Pavan Dalla Torre; Alessandra Pignatta; Ronald Salzer; Stefano Siliberti†; Paola Sposetti.

Sezione *Diritto, Sicurezza e processi di vittimizzazione* (attiva dal 2011). *Coordinamento Scientifico*: Carlo Pennisi (Catania); Franco Prina (Torino); Annamaria Rufino (Napoli); Francesco Sidoti (L'Aquila). *Consiglio Scientifico*: Bruno Bertelli (Trento); Teresa Consoli (Catania); Maurizio Esposito (Cassino); Armando Saponaro (Bari); Chiara Scivoletto (Parma). *Responsabili Editoriali*: Andrea Antonilli e Susanna Vezzadini. *Comitato Editoriale*: Flavio Amadori; Christian Arnoldi; Michele Bonazzi; Rose Marie Callà; Teresa Carbone; Dafne Chitos; Gian Marco Cifaldi; Maria Teresa Gammona; Veronica Moretti; Annalisa Plava; Antonia Roberta Siino.

Sezione *Sociologia e storia della Croce Rossa* (attiva dal 2013). *Direttori*: Costantino Cipolla (Bologna) e Paolo Vanni (Firenze). *Consiglio Scientifico*: François Bugnion (*presidente* - CICR), Roger Durand (*presidente* - Société "Henry Dunant"), Giuseppe Armocida (Varese), Stefania Bartoloni (Roma III), Paolo Benvenuti (Roma III), Fabio Bertini (Firenze), Paola Binetti (Campus Bio-Medico, Roma), Ettore Calzolari (Roma I), Giovanni Cipriani (Firenze), Franco A. Fava (Torino), Carlo Focarelli (Perugia; LUISS), Edoardo Greppi (Torino), Gianni Iacovelli (Accademia di Storia dell'Arte Sanitaria, Roma), Giuseppe Palasciano (Bari), Jean-François Pitteloud (già CICR), Alessandro Porro (Brescia), Duccio Vanni (Firenze), Giorgio Zanchin (Padova). *Comitato Editoriale*: Filippo Lombardi (coordinatore), Massimo Aliverti, Nico Bortoletto, Luca Bottero, Virginia Brayda, Carolina David, Antonella Del Chiaro, Renato Del Mastro, Gerardo Di Ruocco, Boris Dubini, Alberto Galazzetti, Livia Giuliano, Laura Grassi, Veronica Grillo, Riccardo Romeo Jasinski, Pier Francesco Liguori, Maurizio Menarini, Maria Enrica Monaco, Gianluigi Nava, Marisella Notarnicola, Marcello Giovanni Novello, Raimonda Ottaviani, Isabella Pascucci, Francesco Ranaldi, Piero Ridolfi, Anastasia Siena, Calogera Tavormina, Silvana Valcavi Menozzi. *Segreteria Scientifica*: Alberto Ardissona (responsabile), Alessandro Fabbri (responsabile), Barbara Baccarini, Elena Branca, Giovanni Cerino Badone, Emanuele Cerutti, Alessandro D'Angelo, Carmela Anna Esposito, Simona Galasi, Sara Moggi, Paola Sposetti.

Giulia Venturi, Andrea Cimino,
Felice Dell'Orletta

La fede dichiarata

**Un'analisi
linguistico-computazionale**

LABORATORIO SOCIOLOGICO



FRANCOANGELI

Ricerca empirica
ed intervento sociale

La cura redazionale ed editoriale del testo è stata realizzata da Barbara Baccarini.

Il coordinamento editoriale e i referenti di “Laboratorio Sociologico online” sono indicati nel box a chiusura del volume.

Copyright © 2022 by FrancoAngeli s.r.l., Milano, Italy.

Ristampa	Anno
0 1 2 3 4 5 6 7 8 9	2022 2023 2024 2025 2026 2027 2028 2029 2030 2031

L’opera, comprese tutte le sue parti, è tutelata dalla legge sui diritti d’autore.

Sono vietate e sanzionate (se non espressamente autorizzate) la riproduzione in ogni modo e forma (comprese le fotocopie, la scansione, la memorizzazione elettronica) e la comunicazione (ivi inclusi a titolo esemplificativo ma non esaustivo: la distribuzione, l’adattamento, la traduzione e la rielaborazione, anche a mezzo di canali digitali interattivi e con qualsiasi modalità attualmente nota od in futuro sviluppata).

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall’art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale, possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali (www.clearedi.org; e-mail autorizzazioni@clearedi.org).

Stampa: Global Print s.r.l., Via degli Abeti n. 17/1, 20064 Gorgonzola (MI)

Indice

Introduzione , di Roberto Cipriani, Simonetta Montemagni	pag.	7
1. L'approccio qualitativo computer-assistito	»	7
2. Le indagini sulla religiosità in Italia	»	9
Riferimenti bibliografici	»	13
1. I programmi per l'analisi qualitativa: un auspicabile lavoro congiunto , di Domenico Schiattone, Martina Lippolis	»	15
1. Introduzione	»	15
2. Il <i>software NVivo</i> : una presentazione	»	16
3. Verso l'analisi di quattro concetti sensibilizzanti	»	19
4. Chiesa come Istituzione	»	21
5. Papa Francesco	»	36
6. Dio	»	49
7. Preghiera	»	58
8. Conclusioni	»	67
Riferimenti bibliografici	»	67
2. Trattamento automatico della lingua ed estrazione di informazione: materiali e metodi , di Simonetta Montemagni, Giulia Venturi	»	69
1. Introduzione	»	69
2. Oltre il testo come sequenza di caratteri	»	71
3. Il <i>corpus</i> di interviste	»	79
4. Conclusioni	»	86
Riferimenti bibliografici	»	86
3. Dal testo alla conoscenza , di Andrea Cimino, Felice Dell'Orletta, Giulia Venturi	»	88
1. Introduzione	»	88
2. La piattaforma Text-to-Knowledge (T2K)	»	89

3. Indicizzazione e analisi aggregata	pag.	124
4. Conclusioni	»	135
Riferimenti bibliografici	»	137
4. Analisi della polarità delle interviste: <i>Sentiment Analysis</i>, di Andrea Cimino, Felice Dell’Orletta, Giulia Venturi	»	139
1. Introduzione	»	139
2. La metodologia di <i>Sentiment Analysis</i>	»	140
3. Discussione dei risultati	»	146
4. Conclusioni	»	173
Riferimenti bibliografici	»	174
Conclusioni, di Andrea Cimino, Felice Dell’Orletta, Giulia Venturi	»	177

Introduzione

di Roberto Cipriani, Simonetta Montemagni*

1. L'approccio qualitativo computer-assistito

Per lungo tempo la sociologia ha visto schierati su fronti contrapposti i sostenitori dell'approccio quantitativo da una parte e i fautori dell'analisi qualitativa dall'altra. I primi hanno avuto la meglio per oltre un secolo, i secondi hanno riguadagnato spazio e credibilità negli ultimi cinquant'anni. Occorre parlare di un riguadagno perché nella seconda metà degli anni Trenta del Novecento era convinzione dei sociologi, statunitensi nella fattispecie, che una delle opere più importanti nella storia della sociologia fosse quella, a carattere qualitativo, su *Il contadino polacco in Europa e America*, di cui erano autori Thomas e Znaniecki (1918-1920). La richiesta di valutazione per individuare quali fossero le pubblicazioni più influenti era stata avanzata ad un gruppo di sociologi esperti dal *Committee on Appraisal of Research* del *Social Science Research Council* con sede a New York. A Herbert Blumer (1939) era stato poi chiesto un ulteriore parere, modificato in parte in prosieguo di tempo (Blumer, 1979). Nel replicare alle prime critiche di Blumer, Thomas (1939: 87) affermava: «è evidente che gli studi statistici del comportamento delle popolazioni avrà un significato limitato fino a quando i dati statistici non saranno corroborati da storie individuali di caso».

Successivamente, però, si sono sviluppati soprattutto gli studi quantitativi, relegando in second'ordine le ricerche di tipo qualitativo, aiutati in questo dall'avvento di programmi informatici in grado di elaborare i dati numerici in modo sempre più rapido e sicuro.

* Roberto Cipriani è Emerito di Sociologia dell'Università Roma Tre; Simonetta Montemagni è Dirigente di ricerca presso l'Istituto di Linguistica Computazionale "A. Zampolli" del CNR di Pisa.

Solamente a partire dalla prima metà degli anni Ottanta del secolo scorso si è avuta una svolta, con lo sviluppo e la messa a disposizione di strumenti informatici peculiarmente dedicati alla ricerca qualitativa. Si tratta di strumenti, oggi collettivamente denominati *Computer-Assisted Qualitative Data Analysis Software* (CAQDAS), le cui funzionalità spaziano dalla raccolta, digitalizzazione ed organizzazione dei dati, alla loro trascrizione, per arrivare all'analisi e presentazione dei risultati.

La prima esigenza da soddisfare era quella di ordinare e categorizzare le note sul campo raccolte dagli etnologi e dagli antropologi in primo luogo, ma anche dai sociologi. Non a caso, il primo *software* efficace e di largo successo fra gli addetti ai lavori fu *The Ethnograph*, lanciato nel 1984 (giunto adesso alla sua sesta versione) e subito messo alla prova per le ricerche svolte presso l'allora Istituto di Sociologia dell'Università "La Sapienza" di Roma. I risultati acquisiti furono apprezzabili.

Nel frattempo, su iniziativa dei coniugi Richards (1987), un altro strumento cominciava a farsi strada, dapprima con il nome di *NUD*IST* (*Non numerical Unstructured Data Indexing Searching and Theorizing*), dal 1981 al 1997 con quattro versioni e poi come *N4* nel 1997, *N5* nel 2000 e *N6* nel 2002, e successivamente come *NVivo* (dal 2002 in poi, con dodici versioni, a partire da *NVivo 2*, che riprendeva la serie interrotta con *N6*). *ATLAS.ti* nasceva invece nel 1993, con caratteristiche affini a quelle di *NVivo* ed una maggiore facilità a livello operativo ma con limitazioni nel caso di indagini complesse.

A parte le differenze tra i diversi sistemi, la categoria dei CAQDAS raccoglie una famiglia di *software* che offre strumenti di interrogazione, ricerca, organizzazione e annotazione di materiali testuali (o anche multimediali) finalizzati all'analisi qualitativa del contenuto. Utilizzati prevalentemente per analisi in ambito sociologico, psicologico ed etnografico, questi prodotti si presentano come strumenti di ausilio per il ricercatore nella fase di annotazione del testo con categorie concettuali e, essenzialmente, cercano di rendere più rapide le operazioni di ricerca di informazioni nei testi e di ricodifica automatica, nonché di visualizzazione dei dati in forma aggregata (e non).

All'interno di questa categoria di *software*, l'analisi del testo è tipicamente circoscritta alle stringhe di caratteri che lo compongono e al computo delle relative frequenze. Nonostante il progresso e i considerevoli successi raggiunti con i CAQDAS, l'accesso ai contenuti testuali rimaneva ancora fortemente limitato dall'insufficiente capacità di penetrare la superficie linguistica dei testi, per accedere al loro contenuto di informazione. Affinare questa capacità significava dotare i sistemi di analisi di un'adeguata "intelligenza linguistica", che rendeva possibile l'estrazione e la classificazione dei nuclei di informazione implicitamente codificati nelle strutture linguistiche sottostanti

al testo. Un contributo importante in tal senso è offerto da metodi e strumenti di trattamento automatico della lingua e del testo, essenziali per aumentare la quantità e la qualità dell'informazione recuperata e rielaborata.

Nel tentativo di rendere possibili interrogazioni che astraessero dalla sequenza di caratteri nelle parole del testo, facendo riferimento al lemma di riferimento oppure a categorie concettuali maggiormente astratte, i sistemi CAQDAS sono stati nel tempo affiancati da strumenti di analisi testuale che ne hanno potenziato le capacità di analisi del testo e, più recentemente, anche di estrazione di conoscenza e *Text Mining*. Già negli anni '90, insieme a *The Ethnograph*, il software adoperato per l'analisi qualitativa computer-assistita (Schiattone, 1993), è stato usato il *DBT (Data Base Testuale)*, messo a disposizione dall'Istituto di Linguistica Computazionale per l'analisi testuale automatica (Picchi, 1989). Grazie al *DBT*, è stato possibile analizzare i testi alla base delle ricerche, ottenendo sofisticate elaborazioni lessico-statistiche ancorate al testo (con produzione di concordanze). In pratica, il *DBT* ha permesso analisi incentrate sul lessico e soprattutto ha velocizzato significativamente il lavoro d'indagine.

Più recentemente, grazie alla maturità degli strumenti di analisi linguistica automatica e di *Text Mining* la sinergia con le comunità che fanno uso di sistemi CAQDAS si è rafforzata. Ne sono testimonianza la serie di *workshops* su *Natural Language Processing and Computational Social Science (NLP+CSS)*¹ organizzati a partire dal 2016, che ha fornito l'avvio di un dialogo tra le comunità dei sociologi e dei linguisti computazionali sull'apporto di strumenti di *Natural Language Processing (NLP)* in ricerche di tipo sociologico. Particolarmente rilevante ai fini del presente volume è il contributo di Karlgren *et al.* (2020) che discute le sfide connesse all'elaborazione di testi di interviste, dove l'obiettivo non è circoscritto all'identificazione dei contenuti, ma anche del modo in cui essi sono presentati, inclusi gli atteggiamenti e i sentimenti mostrati dagli intervistati.

2. Le indagini sulla religiosità in Italia

Nel 1995 vennero pubblicati i risultati di quella che si può considerare la prima ricerca quantitativa nazionale su *La religiosità in Italia* (Cesareo, Cipriani, Garelli, Lanzetti, Rovati, 1995); il questionario era stato somministrato ad un ampio campione di 7.500 intervistati, fra i quali vennero poi selezionati, con criteri scientifici, 4.500 casi. In precedenza, furono intervi-

¹ <https://aclanthology.org/venues/nlpcss/>

state 84 persone, di cui si analizzarono (mediante una griglia costruita con 8 categorie e 201 sottocategorie), le storie di vita raccolte da 14 coppie di ricercatori e ricercatrici, a ciascuna delle quali erano state assegnate 6 interviste da registrare, trascrivere e analizzare (Cipriani, Corradi, Costa, Schiattone, 1993). Il campionamento prevedeva la suddivisione per sesso (uomo, donna), età (18-35, 36-60, 61 e +), stato civile (coniugate/i, non coniugate/i), condizione professionale (occupate/i, non occupate/i), zona (centro/medio centro/periferia), reddito annuo in milioni di lire (fino a 5, da 6 a 10, da 11 a 20, da 21 ed oltre); in relazione alla condizione professionale si distinguevano anche il rapporto di lavoro (dipendenti, autonome/i) ed il settore di attività (industria, commercio, servizi). Le analisi condotte in occasione di questa prima indagine qualitativa si sono avvalse del *software DBT* che, affiancato al programma *The Ethnograph* in un primo momento e a *NUD*IST* e *NVivo* successivamente, ha permesso elaborazioni raffinate dal punto di vista lessico-concettuale.

Nel 2017 si è svolta un'altra inchiesta sul fenomeno religioso in Italia, utilizzando due metodologie congiunte: 1) un'indagine quantitativa con questionario somministrato a 3.238 persone (campione nazionale statisticamente rappresentativo), di cui sono già stati pubblicati i risultati (Garelli, 2020); 2) uno studio con approccio qualitativo riguardante 164 soggetti, debitamente individuati sul territorio italiano, seguendo soprattutto il criterio del titolo di studio conseguito (licenza media inferiore, diploma, laurea e oltre). Questa seconda modalità d'investigazione è stata implementata con due diversi metodi: 78 interviste sono state raccolte senza porre domande predefinite, lasciando gli intervistati liberi di parlare della loro vita, dei loro sentimenti e delle loro abitudini; altri 86 soggetti sono stati interpellati per una prima parte in forma aperta, non guidata, e per una seconda parte sono stati invitati a parlare anche di vita quotidiana e festiva, felicità e dolore, vita e morte, Dio, preghiera, Chiesa come istituzione e papa Francesco.

I dati qualitativi raccolti sono stati analizzati con diversi metodi, applicati all'insieme delle 164 interviste debitamente trascritte, in forma letterale, parola per parola audio-registrata. In seguito, vari gruppi di ricerca hanno utilizzato procedimenti specifici, finalizzati all'individuazione degli aspetti più significativi e suscettibili di spiegazione e interpretazione. In particolare si sono seguite le seguenti procedure:

- analisi del contenuto come inchiesta, analisi delle corrispondenze lessicali (*SPAD*), *VoSpec (Vocabulaire Spécifique des Groupes d'individus)* (*SPAD*), *cluster analysis*, analisi delle concatenazioni probabilistiche di nodi semantici (*T-LAB*), creazione di micro-mappe e macro-mappa della dinamica discorsiva (Cipriani, Faggiano, Piccini, 2020);

- linguaggio peculiare (*T-LAB*), analisi delle co-occorrenze, analisi

delle corrispondenze lessicali (*SPAD-T*), *cluster analysis* (*SPAD-T*), *Social Network Text Analysis* (Punziano, 2020);

- *grounded theory* (codifica riga per riga, sintesi delle codifiche più ricorrenti, codifica di primo livello per induzione, confronto tra gli analisti, codifica di secondo livello focalizzata per abduzione, codifica di terzo livello teorica per abduzione, ipotesi interpretative) (Quagliata, 2020);

- analisi bimatriciale e multidimensionale (incremento dell'universo d'indagine da 164 a 177 casi, individuazione dei 155 casi corrispondenti a cattolici battezzati, analisi qualitativa dei 155 cattolici battezzati, individuazione dei 122 casi corrispondenti a credenti anti-clero ed a cattolici incoerenti, analisi quali-quantitativa dei 122 casi, individuazione di categorie attitudinali e comportamentali) (Delli Poggi, Bolla, Fiorentini, Valletta, 2020).

A questa serie di procedimenti analitici si sono aggiunte le elaborazioni condotte con strumenti di estrazione della conoscenza e di *Sentiment Analysis* sviluppati dall'Istituto di Linguistica Computazionale "Antonio Zampolli" del Consiglio Nazionale delle Ricerche (ILC-CNR).

Come illustrato nei capitoli che seguono, per l'estrazione di conoscenza è stata utilizzata la piattaforma *software* originariamente progettata e sviluppata dall'ILC-CNR e dal Dipartimento di Linguistica dell'Università di Pisa, denominata *Text-to-Knowledge*, in breve T2K (Dell'Orletta *et al.*, 2008). In particolare, nel progetto è stata utilizzata la versione corrente di T2K, specializzata ed estesa dall'*Italian Natural Language Processing Lab* dell'ILC-CNR (Dell'Orletta *et al.*, 2014), che è stata sperimentata con successo all'interno di diversi domini applicativi (tra cui, quelli legislativo, educativo, sanitario). T2K offre una batteria integrata di strumenti avanzati di analisi linguistica del testo, analisi statistica e apprendimento automatico del linguaggio, destinati a offrire una rappresentazione accurata del contenuto di una base documentale non strutturata, per scopi di indicizzazione avanzata e navigazione intelligente. T2K trasforma le conoscenze implicitamente codificate all'interno di un corpus di testi in conoscenza esplicitamente strutturata: il risultato finale di questo processo interpretativo spazia dall'acquisizione di conoscenze lessicali e terminologiche complesse alla loro organizzazione in strutture proto-concettuali.

Tuttavia, un sistema come T2K non era sufficiente, da solo, per affrontare le sfide connesse all'elaborazione di testi di interviste, dove l'obiettivo è duplice, in quanto non riguarda soltanto i contenuti oggettivi dell'intervista ma anche il modo in cui l'intervistato si rapporta ad essi, attraverso emozioni, attitudini, valutazioni e sentimenti. Per esplorare la sfera delle valutazioni soggettive si è fatto ricorso a un algoritmo di *Sentiment Analysis* sviluppato da Cimino e Dell'Orletta (2016), finalizzato alla classificazione

della polarità (positiva, negativa o neutra) del contesto che descrive un dato contenuto.

Questo volume si propone di indagare l'apporto di tecnologie basate sul *Natural Language Processing* (NLP) o *Trattamento Automatico della Lingua* (TAL) all'analisi qualitativa di un corpus testuale come le trascrizioni delle interviste alla base di un'indagine sociologico-religiosa. Partendo dai risultati di uno studio condotto con *NVivo*, uno dei sistemi CAQDAS più diffusi oggi, il lettore viene condotto gradualmente a scoprire quale sia il valore aggiunto delle analisi basate su NLP e quali nuovi orizzonti di ricerca possano essere aperti da queste analisi. Lasciamo al lettore scoprire passo passo i significativi vantaggi che derivano dalle elaborazioni basate su NLP in modo da poter prospettare, in un futuro non troppo lontano, un'integrazione dei sistemi CAQDAS con moduli di estrazione di conoscenza e *Sentiment Analysis*.

Il volume è organizzato come segue. Il capitolo 1 contiene un contributo invitato, curato da Domenico Schiattone e Martina Lippolis, che è finalizzato a illustrare i risultati dell'analisi del corpus di interviste alla base della ricerca con uno strumento appartenente alla categoria dei sistemi CAQDAS, *NVivo*. Il capitolo mostra come *NVivo* consenta di condurre uno studio focalizzato su alcuni concetti selezionati, considerati particolarmente rilevanti sulla base delle categorie teoriche definite da sociologi esperti. In particolare, attraverso questo capitolo il lettore comincia a familiarizzare con il *corpus* delle interviste alla base della ricerca sulla religiosità presentata in questo volume e con la tipologia di analisi qualitative rese possibili da *NVivo*.

Il capitolo 2, redatto da Giulia Venturi con il contributo invitato di Simonetta Montemagni, introduce il lettore alla domanda se e in che misura il ricorso a tecnologie di *Trattamento Automatico della Lingua* e di *Estrazione di Conoscenza* possa contribuire in modo significativo all'analisi qualitativa di un corpus testuale come le trascrizioni delle interviste alla base di questa ricerca, fornendo evidenza complementare rispetto a quella ottenuta con sistemi CAQDAS. Sono inoltre illustrate le sfide poste per l'analisi automatica dalla specifica varietà d'uso della lingua rappresentata nelle trascrizioni di interviste orali.

I due capitoli che seguono, redatti dagli autori di questo volume, conducono il lettore per mano offrendo un esempio concreto di come metodi e tecniche di *Trattamento Automatico della Lingua* e di *Estrazione di Conoscenza* possano, da un lato, offrire ai ricercatori esperti evidenza a supporto di ipotesi già formulate in materia di studi socio-religiosi in Italia, dall'altro portare alla luce elementi in grado di aprire nuovi orizzonti di ricerca. In particolare, il capitolo 3 illustra la piattaforma T2K all'opera nel complesso

processo di trasformazione dei testi non strutturati delle interviste in conoscenza organizzata. Il percorso che porta dal testo alla conoscenza è caratterizzato da un costante alternarsi di fasi di estrazione di informazione dai testi in analisi e di fasi di impiego di informazione di dominio già nota *a priori*. Il capitolo mostra i risultati conseguiti passo passo, dall'identificazione dei termini rilevanti per giungere a un'analisi aggregata dei risultati sulla base delle variabili socio-demografiche considerate e delle dimensioni macro-concettuali nelle quali sono stati raggruppati i concetti sensibilizzanti selezionati. Il capitolo 4 si addentra invece nello studio della polarità positiva e negativa dei racconti degli intervistati, permettendo di qualificare la conoscenza estratta da T2K.

Le Conclusioni, infine, ripercorrono quanto mostrato gradualmente nel volume per arrivare a delineare quali possano essere auspicabili sviluppi metodologici nell'ambito di ricerche sociologiche, resi possibili da metodi e tecniche di Trattamento Automatico della Lingua e di Estrazione di Conoscenza.

Riferimenti bibliografici

- Blumer H. (1939), "Critiques of Research in the Social Sciences: I. An Appraisal of Thomas and Znaniecki's *The Polish Peasant in Europe and America*. With statements by William I. Thomas and Florian Znaniecki, a panel discussion, and summary and analysis by Read Bain", *Social Science Research Council, Bulletin*, 44.
- Blumer H. (1979), *Critiques of Research in the Social Sciences: An Appraisal of Thomas and Znaniecki's The Polish Peasant in Europe and America. With a new introduction*, Transaction, New Brunswick, N. J.
- Cesareo V., Cipriani R., Garelli F., Lanzetti C., Rovati G. (1995), *La religiosità in Italia*, Mondadori, Milano.
- Cimino A. and Dell'Orletta F. (2016), "Tandem LSTM-SVM Approach for Sentiment Analysis", *Proceedings of Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA)*, Napoli, 5-6 dicembre, pp. 1-6.
- Cipriani R. (2020), *L'incerta fede. Un'indagine quanti-qualitativa in Italia*, FrancoAngeli, Milano.
- Cipriani R., Corradi C., Costa C., Schiattone D. (1993), *Sentieri della religiosità. Un'indagine a Roma*, a cura di R. Cipriani, Morcelliana, Brescia.
- Cipriani R., Faggiano M. P., Piccini M. P. (2020), *La religione dei valori diffusi. Intervista qualitativa e approccio misto di analisi*, FrancoAngeli, Milano.
- Costa C., Morsello B. (a cura di) (2020), *L'incerta religiosità. Forme molteplici del credere*, FrancoAngeli, Milano.
- Delli Poggi S., Bolla F., Fiorentini S., Valletta G. L. (2020), *Italiani: scomunicati o*

- liberi pensatori? Eresia, Apostasia, Scisma o libertà di credere. Profili di cattolici in Italia*, pubblicazione indipendente, 2019-2020, https://www.amazon.it/dp/1693714280?ref_=pe_3052080_397514860
- Dell'Orletta F., Lenci A., Marchi S., Montemagni S., Pirrelli V., Venturi G. (2008), "Dal testo alla conoscenza e ritorno: estrazione terminologica e annotazione semantica di basi documentali di dominio", *Aida Informazioni*, n. 1-2, a. 26, gennaio-giugno 2008, pp. 185-206.
- Dell'Orletta F., Venturi G., Cimino A. and Montemagni S. (2014), "T2K². A System for Automatically Extracting and Organizing Knowledge from Texts", *Proceedings of 9th Edition of International Conference on Language Resources and Evaluation (LREC 2014)*, European Language Resources Association (ELRA), Reykjavik, 26-31 May 2014, pp. 2062-2070.
- Garelli F. (2020), *Gente di poca fede. Il sentimento religioso nell'Italia incerta di Dio*, il Mulino, Bologna.
- Karlgren J., Li R., Meyersson Milgrom E.M. (2020), "Text Mining for Processing Interview Data in Computational Social Science", *arXiv:2011.14037 [cs.CL]*.
- Picchi E. (1989), *DBT. Data Base Testuale*, CNR, Pisa.
- Punziano G. (a cura di) (2020), *Parole, dimensioni e forme della religiosità in Italia. Espressioni multiple tra pratiche e sentire*, FrancoAngeli, Milano, 2020.
- Quagliata A. (a cura di) (2020), *Il dogma inconsapevole. Analisi del fenomeno religioso in Italia. Il contributo qualitativo della Grounded Theory costruttivista*, FrancoAngeli, Milano, 2020.
- Richards L., Richards, T. J. (1987), "Qualitative Data Analysis: can computers do it?", *Australian and New Zealand Journal of Sociology*, 23(1), 23-35.
- Schiattone D. (1993), "La religione diffusa: un'analisi computerassistita di una storia di vita". In Cipriani R., Corradi C., Costa C., Schiattone D. (1993), *Sentieri della religiosità. Un'indagine a Roma*, a cura di R. Cipriani, Morcelliana, Brescia, 211-262.
- Thomas W. I., Znaniecki F. (1918-1920), *The Polish Peasant in Europe and America*, University of Chicago Press, Chicago, 5 voll.; ed. it., *Il contadino polacco in Europa e America*, Comunità, Milano, 1968.
- Thomas W. I. (1939), "Comment by W. I. Thomas". In "Critiques of Research in the Social Sciences: An Appraisal of Thomas and Znaniecki's *The Polish Peasant in Europe and America*. With statements by William I. Thomas and Florian Znaniecki, a panel discussion, and summary and analysis by Read Bain", *Social Science Research Council, Bulletin*, 44: 82-87.

1. I programmi per l'analisi qualitativa: un auspicabile lavoro congiunto

di *Domenico Schiattone, Martina Lippolis**

1. Introduzione

Grazie al coinvolgimento in quest'ultima indagine nazionale sulla religiosità, ho potuto ripercorrere un'esperienza vissuta 30 anni fa quando incontrai il professor Eugenio Picchi presso l'Istituto di Linguistica Computazionale (ILC) del CNR di Pisa. Era già qualche anno che ci occupavamo di analisi qualitativa *computer*-assistita e cercavamo di esplorare i pochi *software* disponibili del tempo. Tra questi il Data Base Testuale (D.B.T.) un programma di indicizzazione dei testi, che abbiamo fin da subito utilizzato per una ricerca sulla religiosità (Cipriani, Corradi e Costa, 1993).

Il lavoro che oggi affrontiamo non è molto dissimile da quello compiuto allora: procedere alla comparazione tra un programma di analisi qualitativa e un *software* di analisi linguistica.

Ciò è stato possibile grazie alla collaborazione con i ricercatori dell'Istituto di Linguistica Computazionale di Pisa che, continuando la tradizione scientifica dell'ILC, hanno sviluppato nuovi applicativi per gli studi linguistici.

A differenza del passato oggi esistono molte applicazioni informatiche utilizzabili e disponibili sia sul mercato che negli ambienti accademici, quindi i ricercatori qualitativi hanno ora la possibilità di scegliere fra molti programmi. In questa indagine ne sono stati utilizzati almeno 6.

Il distanziamento dovuto alla pandemia in corso da più di un anno ha

* Domenico Schiattone, sociologo perfezionato in metodologia della ricerca, Direttore dell'Ufficio del Garante nazionale delle persone private della libertà, è autore dei par. 1, 3, 4, 5 e 8; Martina Lippolis, dottoranda di ricerca in Teoria e ricerca educativa e sociale, curriculum in Ricerca sociale teorica e applicata, nell'Università Roma Tre: è autrice dei par. 2, 3, 6, 7 e 8.

impedito ai diversi Gruppi di lavoro di incontrarsi in presenza. Non ci si è quindi potuto avvalere di strumenti di ricerca che già nella fase iniziale dell'indagine avrebbero potuto ottenere “il massimo rendimento” dalle applicazioni informatiche attualmente disponibili. Allo stesso modo non si è potuto utilizzare pienamente il programma *NVivo*, nonostante l'approfondita esperienza di utilizzo del software a Roma Tre, applicato anche in ambito non accademico (Schiattone 2015, 2020).

In questa ricerca sono stati coinvolti 28 tra studenti e studiosi che hanno partecipato a questi corsi. Gli stessi che hanno svolto il ruolo di intervistatori prima e di analisti dopo la raccolta delle interviste hanno contribuito alla identificazione dei concetti sensibilizzanti. Il loro coinvolgimento nella segmentazione dei testi delle interviste non ha potuto fare riferimento ad un vocabolario comune. Ciò ha comportato che i nodi (così nella nomenclatura di *NVivo* vengono definiti) fossero associati a porzioni di testo non sempre coerenti. Spesso alcuni concetti fondanti i temi della ricerca (quelli riferiti al fenomeno religioso) non sono stati codificati a pieno o sono stati codificati solo in parte.

L'impianto della ricerca è stato saldamente condotto dal punto di vista metodologico da chi ha coordinato l'indagine, ma dalla lettura dei rapporti si evince come gli applicativi non siano stati utilizzati pienamente e soprattutto non sia stato possibile attuare un lavoro congiunto.

Come già scritto, ci ritroviamo ora a collaborare con l'Istituto di Linguistica Computazionale di Pisa. In questa sede presentiamo – seguendo lo stesso filo rosso degli anni '90 – il *software* di analisi qualitativa *NVivo* con lo scopo di metterne in luce le potenzialità di analisi quando impiegato nello studio del contenuto del *corpus* di interviste raccolte. La nostra intenzione è quella di mostrare come i risultati raggiunti possano essere considerati complementari, se opportunamente indagati, rispetto a quelli ottenuti impiegando *Text-to-Knowledge* (T2K), la piattaforma di estrazione e organizzazione della conoscenza da *corpora* testuali di dominio basata su metodi e strumenti di Trattamento Automatico della Lingua, oggetto di questo libro.

Procediamo pertanto a descrivere ed esemplificare le funzioni di base ed avanzate di *NVivo*, mentre rimandiamo il lettore alla lettura del capitolo 2 per una comparazione guidata dei risultati ottenuti con *NVivo* e T2K.

2. Il *software NVivo*: una presentazione

Per presentare il *software NVivo* è necessario sciogliere il suo acronimo, Nud*IstVivo: *Non-numerical Unstructured Data* Indexing, Searching and Theorizing Vivo*. Si tratta di un programma realizzato intorno agli anni '90

da Tom e Lyn Richards e prodotto dalla *Qualitative Software Research*, QSR International, di Melbourne. Il *software* si colloca tra gli strumenti con approccio semi-automatico, i cosiddetti CAQDAS (*Computer Assisted Qualitative Data Analysis*) (Coppola, 2011) per coadiuvare i ricercatori nella lettura, organizzazione e analisi dei materiali di ricerca semi-strutturati e non strutturati, che si caratterizzano per la loro natura eterogenea.

Il valore aggiunto del programma *NVivo* è la sua flessibilità; è stato pensato inizialmente per l'applicazione della metodologia qualitativa della *Grounded Theory* ma può essere utilizzato come semplice strumento, in base alla scelta metodologica e alle necessità del gruppo di ricerca che lo utilizza.

Sebbene vi sia una struttura che “guida” i ricercatori, con una sorta di processo di fasi di lavoro apparentemente ordinate, *NVivo* consente di ricorrere continuamente (riportando sempre il contesto e le informazioni utili) ai materiali di ricerca importati.

La ricerca qualitativa e in particolare l'approccio *grounded* sono caratterizzati dalla circolarità tra teoria e dati e con tale programma risulta semplice “tornare indietro”, inserire nuovi dati e nuove categorizzazioni, fare interrogazioni prima di procedere con l'analisi: i ricercatori sono dunque facilitati nel poter lavorare all'interno di un processo iterativo.

Tra le principali caratteristiche e funzioni del programma *NVivo* si segnalano:

- la gestione e archiviazione dei dati prodotti e raccolti;
- la suddivisione in segmenti o regioni testuali (o multimediali) mediante la creazione dei *nodi*;
- la creazione dei *memos*, importanti annotazioni prodotte durante il processo di ricerca;
- il raggruppamento o il collegamento tra i diversi materiali e contenuti codificati;
- l'elaborazione di schemi concettuali, grafici e *reports* sui risultati della ricerca.

Il *software NVivo* presenta le seguenti caratteristiche e funzionalità, espresse in un lessico specifico, che di seguito riportiamo:

- i *Nodes* (nodi) sono le categorie, i concetti sensibilizzanti che guidano il processo di interpretazione e analisi dei dati. Troviamo, nelle ultime versioni di *NVivo*¹ una distinzione tra *Theme nodes* e *Case nodes*;

¹ La versione *NVivo 11* è disponibile dal 2015 e la versione 12 dal 2018: www.qsrinternational.com

- i *Theme nodes* sono contenitori logici di tipo descrittivo o più analitico;
- i *Case nodes* (casi) sono le “unità di osservazione” e potrebbero includere persone, luoghi, siti o organizzazioni;
- gli *Attributes* (attributi) sono utilizzati per valorizzare documenti (fonti o *sources*), nodi e casi, ad es. il genere, lo stato civile, il titolo di studio e la nazionalità. Risulta interessante elaborare delle interrogazioni (*Queries*) sulla base delle classificazioni individuate (assegnazione di attributi nei documenti, nei nodi e nei casi);
- i nodi, documenti e casi possono essere raggruppati in *Sets*, contenitori di elementi con specifiche caratteristiche;
- le *Queries* sono le operazioni di ricerca, corrispondono al modo di “interrogare” i dati, esplorandone i contenuti per aiutare i ricercatori a delineare le interpretazioni. Con *NVivo* è possibile elaborare le seguenti *queries*:
 - *Text Search* (Ricerca di testo): consente di trovare una parola o una combinazione di parole contenute nei documenti o nei nodi del progetto;
 - *Word Frequency* (Creazione del vocabolario): estrae dai documenti, dai nodi o casi le parole usate più frequentemente;
 - *Coding* (Ricerca nella codifica): consente di effettuare ricerche utilizzando nodi, casi e/o attributi;
 - *Matrix coding* (Costruzione di matrici di codifica): consente di effettuare ricerche incrociando nodi, casi e/o attributi, costruendo matrici di dati;
 - *Compound* (Combinazione di ricerche di testo e di codifica): consente di combinare ricerche di testo e di nodi o casi;
 - *Coding comparison* (Ricerche comparative): consente di confrontare la codifica fatta da due utenti o gruppi di utenti;
 - *Group* (Ricerche di raggruppamento): consente di trovare oggetti che sono associati in modo particolare con altri elementi del progetto (ad esempio, i nodi utilizzati per codificare due diverse interviste);
 - i grafici, i *Models* e le *Maps* permettono di visualizzare le informazioni relative al progetto e ai risultati, per rappresentare documenti, nodi e attributi;
 - i *Reports* consentono di visualizzare dettagliatamente, e in qualsiasi fase del lavoro, tutti i materiali presenti e archiviati nel programma.

Nell’ambito della ricerca sulla religiosità presentata in questo volume, *NVivo* è stato scelto come programma di analisi per il contributo qualitativo della *Grounded Theory* costruttivista, per le 164 interviste svolte, nel periodo giugno-dicembre 2017, dai 28 ricercatori coinvolti, che, ricordiamo, avevano partecipato ai corsi sull’utilizzo del *software*.

I ricercatori, intervistatori e analisti, hanno utilizzato tale programma per l’assegnazione dei 219 nodi (concetti sensibilizzanti), definiti e condi-

visi nel corso della riunione nazionale, tenutasi il 17 maggio 2018 (Lippolis, 2020). In particolare, ciascun analista ha assegnato i nodi, predisposti in un progetto *NVivo* denominato “Religiosità in Italia”, utilizzato come base per l’analisi delle proprie interviste. Una volta raccolti i 28 progetti, con tutte le interviste analizzate e codificate, sono stati riuniti in un unico progetto, che permette di interrogare e presentare i dati nella loro totalità.

3. Verso l’analisi di quattro concetti sensibilizzanti

Al fine di procedere all’illustrazione del supporto del programma *NVivo* alle tematiche della ricerca abbiamo fatto riferimento alle pubblicazioni relative all’indagine che sono state pubblicate o che sono in via di pubblicazione, sviluppate con metodologie sia quantitative che qualitative. Tra queste i riferimenti più rilevanti sono stati i due testi principali della ricerca *L’incerta fede* di Roberto Cipriani (Cipriani, 2020) e *Gente di poca fede* di Franco Garelli (Garelli, 2020). Nello stesso tempo ci siamo basati sulle elaborazioni ottenute dai ricercatori dell’ILC di Pisa attraverso l’utilizzo del programma T2K, ampiamente illustrate nei capitoli successivi.

In particolare, abbiamo preso come riferimento i primi 30 termini più rilevanti estratti in modo automatico da T2K a partire dall’intero *corpus* di interviste² e ne abbiamo estrapolato inizialmente i 13 più significativi per le tematiche di ambito religioso. Il risultato è la seguente lista: ‘ateismo’, ‘catechismo’, ‘cattolicesimo’, ‘Chiesa’ (istituzione), ‘chiesa’ (tempio), ‘cristianesimo’, ‘domenica’, ‘istituzioni religiose’, ‘papa Francesco’, ‘papa Giovanni Paolo II’, ‘pratiche’, ‘preghiera’, ‘vita’.

L’intento iniziale di lavorare alla comparazione tra nodi/termini, dopo la lettura approfondita dei testi dell’indagine, ci ha portato a selezionare ulteriormente i concetti sensibilizzanti oggetto dell’analisi di questo capitolo.

La selezione dei concetti si è quindi orientata seguendo le risultanze da cui è scaturita la categoria teorica de *L’incerta fede* costruita da Cipriani nel suo omonimo testo (Cipriani, 2020, p. 429).

Cipriani procede nel suo lavoro operando la distinzione tra diverse classi di concetti, emersi nel corso dell’indagine:

- primari generali (Fede-Incertezza);
- primari particolari (Chiesa-religione-Religiosità-Spiritualità-Valori);
- secondari particolari (Dio - Preghiera - papa Francesco).

² Cfr. Tabella 1, capitolo 3 di questo libro.