

SESSIONE II

PREVENZIONE, VALUTAZIONE E TRATTAMENTO DEGLI ERRORI NON CAMPIONARI

Sperimentazione di un sistema di controllo
e correzione per la codifica dell'attività
economica nell'indagine sulle Forze di
lavoro

Francesca Della Ratta e Mauro Tibaldi

Sperimentazione di un sistema di controllo e correzione per la codifica dell'attività economica nell'indagine sulle Forze di lavoro

Francesca della Ratta e Mauro Tibaldi

Istat, DISA - DCSE

dellarat@istat.it – tibaldi@istat.it

Sommario

Per migliorare la qualità dei dati il Servizio Istruzione, formazione e lavoro dell'Istat ha messo in campo una procedura di Text Mining per il controllo e la correzione della codifica dell'attività economica effettuata dai rilevatori nell'indagine sulle Forze di lavoro. La strategia è resa possibile grazie alla presenza nel file dati di un campo testuale in cui i rilevatori riportano le caratteristiche dell'attività economica svolta, così come descritta dagli intervistati, accompagnato dalla codifica relativa all'attività economica. La procedura, attiva dal quarto trimestre 2011, si basa sul confronto tra il vocabolario adottato dai rispondenti e il vocabolario specifico di ciascuna divisione, desunto dalla classificazione ufficiale (classificazione Ateco 2007, Nace rev. 2). I record nelle cui descrizioni compaiono parole non presenti nel vocabolario ufficiale della divisione specifica sono automaticamente estratti e controllati. In caso di errori ricorrenti, le istruzioni di ricerca sono eseguite automaticamente nelle future sessioni di correzione. Il processo, che ha permesso di migliorare la qualità dei dati prodotti, si completa attraverso una verifica puntuale della coerenza degli stessi che, in ogni sessione, prevede sempre un intervento manuale di esperti per la validazione delle correzioni effettuate e la definitiva attribuzione del codice corretto.

Parole chiave: analisi statistica dei dati testuali; qualità dei dati; controllo dei dati; classificazioni statistiche; Indagine sulle forze di lavoro.

Abstract

In order to improve data quality, Istat Division "Education, Training and Labour" experienced a text mining procedure for editing and imputation of the economic activity coding carried out by interviewers in the Labour Force Survey. This strategy is made possible by the presence of a text field in the data file in which the interviewers report economic activity features and the occupation as described by respondents. The procedure has been operating since the fourth quarter of 2011 and is based on a comparison between respondent vocabulary and the specific dictionary of each official classification division (Ateco 2007 classification, originated from Nace rev. 2). Records containing words used by the interviewers and not reported in the official dictionary are automatically extracted and verified. In the case of recurring errors, queries can be saved to be used in future editing and imputation sessions. The process is completed by a thorough examination of data consistency in each session, to validate the editing and imputation work done and for the final assignment of proper codes.

Key words: text mining; data quality; data checking; statistical classifications; Labour Force Survey.

1. Premessa

Con l'introduzione della nuova classificazione delle attività economiche Ateco 2007 nella rilevazione sulle forze di lavoro, a fronte di risultati poco coerenti tra congiuntura economica e andamenti occupazionali riscontrati in alcuni settori di attività economica nel corso del 2011, sono stati compiuti alcuni controlli a campione finalizzati a verificare la congruenza tra la codifica Ateco inserita dal rilevatore e la descrizione libera di attività economica e professione. L'esito di tali controlli ha evidenziato la presenza di consistenti incongruenze soprattutto in alcune sezioni (agricoltura, costruzioni, istruzione, sanità e assistenza sociale e servizi generali della PA), sulle quali si è deciso di intervenire in sede di correzione del dato.

A partire dal IV trimestre 2011 è stata quindi messa a punto una procedura di controllo, in parte automatizzata, finalizzata all'individuazione e alla correzione dei record mal classificati. Tale strategia non si limita soltanto alla preliminare correzione dei record, ma ha consentito la definizione progressiva di un sistema integrato di attività mirate alla prevenzione degli errori, quali ritorni formativi presso i rilevatori, esercitazioni svolte a distanza, segnalazioni automatiche in seguito a codifiche imprecise e controlli a campione sull'insieme dei record.

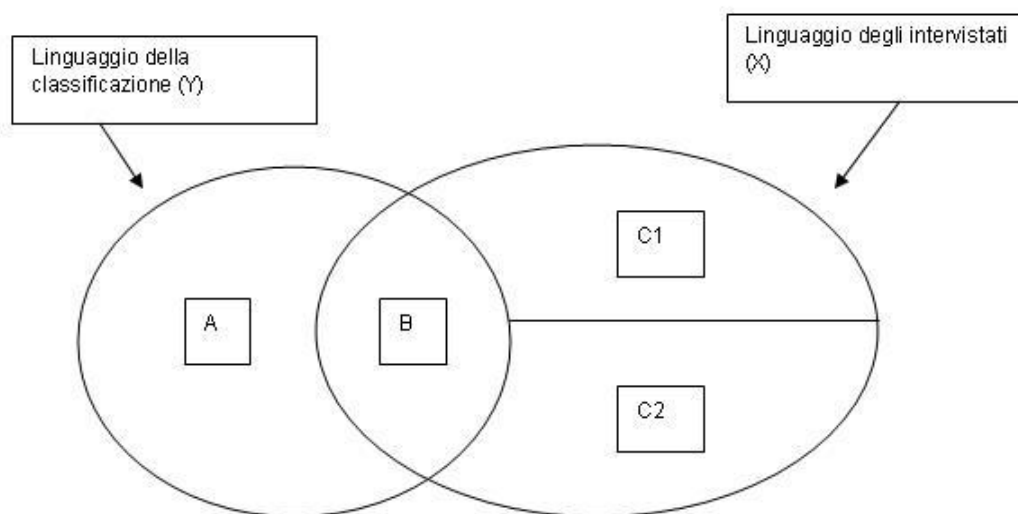
2. La procedura di controllo "Dizionari"

Per la fase di correzione si considerano congiuntamente i testi costituiti dalle risposte alle due domande aperte presenti nel questionario, relative a professione e attività economica. Anche se la variabile professione non è oggetto dell'attività di correzione, soprattutto in alcuni comparti la descrizione della professione consente di individuare meglio l'unità locale presso cui viene svolta l'attività economica. Ad esempio le stringhe contenenti le informazioni "raccolta rifiuti - comune", oppure "cuoco di cooperativa ristorazione - scuola dell'infanzia" ci consentono di individuare eventuali errori che non sarebbe possibile evidenziare analizzando la sola informazione contenuta nella risposta sull'Ateco. In sostanza, la lettura integrata di queste informazioni consente di specificare meglio il contesto di svolgimento dell'attività, rendendo di fatto più agevole l'individuazione della codifica più pertinente.

Per la fase di controllo, le interviste codificate dai rilevatori in ciascuna delle sezioni critiche sono state analizzate distintamente. Per valutare la congruenza tra le parole impiegate per descrivere l'attività di un settore e le caratteristiche del settore stesso è stato effettuato un confronto tra il vocabolario (inteso come insieme di parole utilizzate) adottato dai rispondenti e un vocabolario specifico, riferito nel nostro caso ad una particolare branca di attività economica, facilmente ottenibile a partire dall'insieme delle descrizioni dei codici afferenti le singole voci della classificazione Ateco 2007.

L'intersezione tra i due insiemi determina quattro blocchi logici di parole (Figura 1). Il sottoinsieme di parole comuni ai due vocabolari (B) è quello meno problematico, in quanto è costituito sia da parole specifiche che caratterizzano con elevate probabilità il settore di riferimento, sia da parole strumentali della lingua (articoli, preposizioni ecc.) che non sono di interesse per la presente analisi. Al contempo anche il sottoinsieme di parole presenti esclusivamente nel vocabolario Ateco (A) è meno problematico per l'individuazione degli errori, contenendo sia le parole riferite a settori non presenti nel campione in esame sia la terminologia specialistica non utilizzata dai rispondenti. Di maggiore interesse è invece il sottoinsieme composto dalle parole riportate dai rilevatori e non presenti nel dizionario Ateco della divisione specifica (C).

Figura 1 – Intersezione tra il linguaggio della classificazione e il linguaggio degli intervistati



Questo sottoinsieme può essere a sua volta scomposto in due blocchi: da un lato la terminologia “pertinente” (C1) che gli intervistati utilizzano per descrivere la propria attività attraverso vocaboli non presenti nella classificazione ufficiale (ad esempio laddove la classificazione parla di ‘coltivazione di agrumi’ gli intervistati potrebbero parlare direttamente di ‘limoni’ o di ‘arance’); dall’altro invece la terminologia “non pertinente” (C2) che potrebbe nascondere la presenza di errori di codifica (ad esempio termini quali ‘imbottigliamento’, ‘confezionamento’, ‘commercio’ nelle interviste classificate nel settore dell’agricoltura).

Proprio questi due sottoinsiemi costituiscono il punto di partenza della procedura di individuazione degli errori: da un lato i termini da noi attribuiti al sottoinsieme C1 vengono aggiunti di volta in volta al vocabolario di partenza dell’Ateco, incrementandolo e permettendo quindi di ridurre i tempi dell’analisi nei trimestri successivi; dall’altro i termini del sottoinsieme C2 sono utilizzati per le procedure di controllo.

Il software di analisi testuale utilizzato (Taltac2) consente di rintracciare tutti i record che contengono una determinata parola o una combinazione di parole. Una volta appurato che la presenza di uno specifico termine o combinazione di termini dà luogo a un errore di classificazione, è possibile aggiungere un *flag* al record, inserendo il codice Ateco corretto in un apposito campo che viene aggiunto al file di partenza. Il processo di correzione, tuttavia, si completa necessariamente con un intervento manuale da parte di esperti per la validazione delle correzioni effettuate e la corretta attribuzione del codice Ateco al massimo dettaglio possibile.

Peraltro, le sessioni di correzione successive alla prima “apprendono” dalle operazioni effettuate in precedenza: dopo oltre due anni e mezzo dalla messa a punto della strategia di controllo è stato possibile semplificare ulteriormente - con una conseguente riduzione dei tempi - la fase di controllo testuale. Difatti, da un lato le *query* che girano in automatico consentono di individuare circa il 60% dei record da correggere, dall’altro la progressiva riduzione del sottoinsieme di parole da controllare (insieme C2) rende piuttosto agevole

etichettare tutti i record che contengono una parola anomala. Nel corso della successiva fase di verifica del risultato finale è possibile abbinare il codice Ateco corretto ai record che risultano effettivamente classificati in maniera erranea.

Nel complesso, tra il IV trimestre 2011 e il II trimestre 2014, sono stati corretti 3.484 record, pari all'1,7% degli oltre 210 mila controllati. Le divisioni più critiche risultano i servizi generali della PA (con il 3,9% di errori), l'agricoltura (1,3%), le costruzioni e l'istruzione (entrambe con 1,1% di errori).

L'analisi puntuale degli errori più ricorrenti per ciascuna divisione ha inoltre consentito di individuare le principali lacune cognitive da colmare in occasione dei successivi *debriefing* presso i rilevatori. Ad esempio, le codifiche erroneamente attribuite all'agricoltura rivelano una confusione tra "prodotto" e "processo", come nel caso delle attività di trasformazione industriale di prodotti agricoli (industria alimentare, macellazione carni, trasformazione) o di commercializzazione dei prodotti (vendita di frutta, piante, mercato ortofrutticolo, ecc.). Allo stesso modo, tra i record erroneamente attribuiti alle costruzioni vi sono alcune attività da ricondurre invece all'industria manifatturiera poiché riferite alla produzione dei materiali utilizzati in edilizia (ad esempio pavimenti, prefabbricati, tettoie).

La divisione nella quale si riscontra il maggior tasso di errore è quella dei Servizi generali della Pubblica amministrazione, che presenta una casistica di errore molto articolata. L'errore più frequente riguarda il personale sanitario (medici, infermieri) o amministrativo (impiegati, operatori) che lavora in strutture della sanità o dell'assistenza, erroneamente classificato nella Pubblica Amministrazione anziché nella sanità. Un altro errore frequente riguarda la codifica delle attività connesse alla gestione dei rifiuti, delle reti idriche o al trasporto pubblico locale, che secondo le regole della classificazione bisogna codificare tra le attività dell'industria o dei trasporti in presenza di unità locali che svolgono tali attività in via principale.

3. La prevenzione dell'errore e la qualità delle stringhe descrittive

L'analisi delle stringhe testuali ha consentito di far luce su un problema a monte, vale a dire l'insufficiente accuratezza delle stringhe descrittive inserite dai rilevatori, che se eccessivamente generiche non consentono di codificare l'attività economica con il dettaglio richiesto dall'indagine sulle Forze di Lavoro (sesto digit per l'Ateco e quinto per le professioni). Per comprendere la portata di questo problema è necessario considerare le caratteristiche delle classificazioni in uso, che utilizzano variabili di tipo categoriale e sottostanno a criteri di ordinamento astratti che rendono piuttosto complessa l'attività di codifica (Vicari, Ferrillo, Valery, 2009). Mentre la classificazione delle professioni, tra le due, appare relativamente meno problematica poiché richiede di descrivere un insieme di attività lavorative concretamente svolte da ciascun individuo, la classificazione Ateco è di più complesso utilizzo. Quest'ultima, infatti, è una classificazione piatta, non ordinabile e molto analitica (prevede in tutto 1.224 sottocategorie) che rimanda a un livello più astratto, ovvero l'attività svolta dall'ente/impresa presso cui il soggetto presta la propria opera, attività che non sempre gli intervistati conoscono in maniera sufficiente. Un contributo fondamentale al processo di codifica è pertanto fornito dal rilevatore, che attraverso la propria esperienza e capacità relazionali dovrebbe tendere a ottenere il maggior numero di informazioni riguardo le variabili indagate, in modo da rendere il più possibile pertinente ed esaustiva la codifica.

Nel corso dell'attività di correzione sono invece stati individuati numerosi casi in cui nelle stringhe testuali inserite mancava la descrizione del tipo di attività, oppure non era evidenziato il bene o il servizio oggetto di tale attività, o ancora era assente il materiale del bene prodotto. Per tali ragioni spesso le descrizioni fornite non consentono né di costruire una

codifica pertinente né di controllare l'esattezza del codice Ateco inserito dal rilevatore. Per limitare questo fenomeno, dal settembre 2012 è stata avviata un'operazione di monitoraggio continuo, che consiste nell'invio quotidiano di un file contenente proprio i casi con descrizioni troppo generiche. Mentre nel caso delle professioni si è attinto a una lista costituita da 120 voci generiche costruita nel corso dei monitoraggi finora effettuati, nel caso dell'Ateco la lista di descrizioni generiche è stata individuata a partire da una selezione ragionata di tutte le stringhe composte da meno di 13 caratteri inserite dai rilevatori nei primi due trimestri del 2012. La lista di descrittivi generici si compone di circa 800 termini, cui è stata aggiunta l'indicazione dei motivi per cui la descrizione è da ritenersi inadeguata e il suggerimento sul modo più appropriato di inserire le informazioni. Per incrementare l'efficienza di questo tipo di controllo e arricchirlo altresì di un contenuto formativo, da marzo 2013 l'invio è stato corredato da brevi spiegazioni ed esempi in merito agli errori che si verificano con maggior frequenza. Anche se una descrizione generica non comporta necessariamente un errore di codifica, si ritiene che richiamare l'attenzione dei rilevatori sull'importanza di una descrizione esaustiva possa migliorare l'accuratezza della codifica stessa.

Nei primi 9 mesi del 2013 sono state inviate circa 3.500 segnalazioni riferite alla descrizione dell'attività economica e 2.800 alla professione. Analizzando l'insieme delle codifiche generiche riferite all'Ateco è interessante osservare la loro concentrazione in alcuni settori, quali la sanità, i servizi generali della PA, le costruzioni, le attività dei servizi per la persona, l'agricoltura, la ristorazione, l'istruzione e il commercio. Nel corso dei mesi, tuttavia, la numerosità delle segnalazioni si è più che dimezzata, passando da 1.965 del primo trimestre a 949 del terzo trimestre 2013. La riduzione del volume delle segnalazioni costituisce un risultato molto positivo che mostra l'efficacia della strategia fin qui messa in atto e che verrà quindi rafforzata nei successivi incontri formativi.

Un elemento critico che permane nell'insieme delle operazioni di descrizione e codifica, è la mancata individuazione/esplicitazione dell'unità locale presso la quale l'intervistato lavora, fattore che genera spesso una descrizione dell'attività economica non pertinente e quindi un errore di codifica. Le regole di attribuzione dell'attività economica, infatti, prevedono che se un ente/azienda dispone di più sedi/unità locali, deve essere rilevata l'attività principale della sede presso cui l'intervistato lavora, che può essere diversa da quella dell'impresa o ente madre. Il caso più frequente, come sottolineato, riguarda gli enti locali e in particolare i comuni, che erogano una pluralità di servizi ai cittadini cui sottostanno attività economiche differenti. Nei comuni di una certa dimensione tali attività vengono esercitate in unità locali diverse, per cui il rilevatore dovrebbe raccogliere questo tipo di informazione in modo da descrivere l'attività economica dell'unità locale e codificarla correttamente.

L'analisi degli errori ricorrenti e l'introduzione delle azioni di monitoraggio sulla qualità delle stringhe descrittive costituiscono uno dei tasselli di un processo più generale rivolto a migliorare la qualità dei dati, la cui architettura poggia su periodici ritorni formativi presso i rilevatori sia CATI sia CAPI. La formazione, infatti, svolge un ruolo fondamentale perché consente di migliorare ex-ante la qualità dei dati prodotti attraverso un'opera di prevenzione, finalizzata a ridurre in maniera significativa l'insorgere di errori. Per tale ragione sono stati programmati dei ritorni formativi presso i rilevatori, con l'obiettivo di consolidare e incrementare le conoscenze utili a supportare una corretta codifica dell'attività economica. In particolare, nel dicembre 2013 è stato realizzato un intervento formativo frontale con tutti i rilevatori della rete CAPI, a tre anni di distanza dall'ultimo intervento di formazione diretto dedicato all'introduzione della nuova classificazione. Questo evento ha consentito, attraverso un'interazione diretta, di comprendere meglio le dinamiche reali del processo di

codifica, i problemi sottostanti e le difficoltà incontrate in particolari contesti territoriali o economici.

L'intervento si è sviluppato da una parte sull'approfondimento dei contenuti e della logica della classificazione dell'attività economica e sulle regole che presiedono all'individuazione, descrizione e codifica dell'Ateco, dall'altra su simulazioni di interviste ed esercitazioni pratiche di codifica accompagnate da sessioni di discussione e verifica dei risultati. Vista l'indubbia efficacia emersa da questo tipo d'intervento, si sta progettando di replicare la formazione frontale con tutti i rilevatori almeno a cadenza biennale. Nel frattempo, è stata avviata la somministrazione di esercitazioni di codifica per via telematica a livello trimestrale, proprio per consolidare le conoscenze acquisite nell'ultimo incontro formativo. Accertata l'utilità di tale strumento le esercitazioni, peraltro, sono state estese a tutta la rete di rilevazione.

Tra gli altri, è emerso da parte della maggioranza dei rilevatori un atteggiamento di quasi totale delega nei confronti del navigatore della classificazione inserito nel questionario elettronico. I rilevatori sembrano spesso considerare il navigatore come l'unico strumento deputato a una corretta codifica attraverso la mera digitazione dei descrittivi rilevati, e più raramente ricorrono alle regole e alle conoscenze che presiedono al processo stesso di codifica.

Un altro aspetto critico sul quale si è intervenuti è costituito dall'interazione fra rilevatori e supervisori, che probabilmente rappresentava l'anello debole della catena formativa. Ciò è avvenuto da un lato coinvolgendo attivamente i supervisori della rete di rilevazione come destinatari del processo formativo, grazie a un incontro svoltosi a marzo 2013, dall'altro incentivando il flusso comunicazionale tra i rilevatori e questi ultimi che, con la preparazione acquisita, possono fungere da "filtro" per ridurre le impurità contenute nelle codifiche.

4. Linguaggio degli intervistati e linguaggio della classificazione

L'attività di correzione e l'interazione con i rilevatori ha evidenziato un limite dello strumento di codifica utilizzato (navigatore), che utilizza esclusivamente il linguaggio ufficiale della classificazione Ateco 2007, caratterizzato da un vocabolario piuttosto tecnico e lontano dal modo di esprimersi degli intervistati (della Ratta, Gallo, Loré, 2011). Tale navigatore consente, attraverso l'inserimento di stringhe di testo, di individuare la codifica più appropriata nella classificazione Ateco. Tuttavia, se la parola menzionata dall'intervistato non è presente nel dizionario ufficiale, il rilevatore dovrà ricorrere esclusivamente alle sue nozioni sulla classificazione per individuare il codice giusto, esponendosi a maggiori rischi di errore. Al contrario un dizionario arricchito dalla terminologia impiegata dai rispondenti potrebbe facilitare il lavoro di codifica riducendo gli errori.

Un confronto puntuale tra le parole utilizzate dai rispondenti per descrivere l'attività economica e il vocabolario ufficiale ha consentito di accertare una certa distanza tra i due linguaggi, individuando alcuni termini correttamente utilizzati per descrivere l'attività non presenti nel dizionario (della Ratta, Tibaldi, Pontecorvo, 2014). Ad esempio, in agricoltura si incontrano forme grafiche che fanno riferimento a specifici oggetti dell'attività agricola (vigneto/i, campo, viti, oliveto ecc.) mentre nelle costruzioni spiccano parole inerenti il tipo di sede (impresa, ditta, azienda), e vocaboli comuni che connotano l'attività edilizia (ristrutturazione, abitazioni, case, interni). Nel settore dell'istruzione gli intervistati utilizzano molto gli aggettivi che specificano il tipo di scuola utilizzando il linguaggio comune al posto di quello istituzionale (materna, tecnico, alberghiero, industriale, agrario ecc), mentre nella PA le parole più diffuse specificano soprattutto il tipo di amministrazione (comunale, provinciale ecc.). Infine, in sanità le parole non presenti nel dizionario Ateco riguardano in particolar modo le Asl, come si evince sia dalla specifica forma grafica, sia dai termini

“azienda” e “locale”.

Questo tipo di risultato suggerisce l'opportunità di arricchire lo strumento di codifica attualmente in uso dai rilevatori: una maggiore aderenza del linguaggio utilizzato dal navigatore non potrà che facilitare tutto il processo, contribuendo a ridurre l'insorgenza di errori. Attualmente infatti il navigatore in uso risente di una maggiore distanza dal linguaggio reale perché non riesce a render conto della forte variabilità, genericità e indeterminazione del linguaggio comune degli intervistati, che si arricchisce quotidianamente con una terminologia molto variegata. L'enorme base testuale accumulata nel campo nelle indagini sulle famiglie e in particolare nell'indagine sulle forze di lavoro, potrebbe costituire il punto di avvio per un processo più ampio, in cui le descrizioni (corrette) fornite dagli intervistati possono contribuire ad arricchire il dizionario del navigatore utilizzato per la codifica con esempi tratti dal modo concreto di esprimersi degli intervistati. Tale operazione, che dovrebbe coinvolgere diverse professionalità all'interno dell'Istituto, completerebbe il percorso avviato fornendo ai rilevatori uno strumento più utile nella fase della codifica.

La fase di correzione, quindi, ha permesso di costruire progressivamente un sistema organico e strutturato orientato a prevenire l'errore di codifica attraverso una pluralità di strumenti. Formazione generale sull'Ateco e sulle regole di codifica a tutti i soggetti coinvolti (rilevatori Capi, Cati, Astat, responsabili uffici regionali, supervisor Capi), ritorni formativi specifici sull'Ateco a tutti i rilevatori, invio di un warning automatico quotidiano a tutti i rilevatori che inseriscono un descrittivo generico, inserimento nel questionario di un promemoria sulle informazioni da rilevare per una corretta descrizione dell'Ateco, invio trimestrale personalizzato a ciascun rilevatore delle codifiche errate con relative spiegazioni, istituzione di un indirizzo email specifico attraverso cui degli esperti possono chiarire i dubbi sulle codifiche poste dai rilevatori, esercitazioni periodiche di codifica estesa a tutta la rete di rilevazione. I risultati sono stati apprezzabili, tanto che il tasso di errore totale si è quasi dimezzato, passando dal 2,1% nel secondo trimestre 2012 all'1,1% due anni dopo.

In definitiva, è stato implementato un processo produttivo innovativo che ha consentito di elevare il livello di qualità statistica dei dati dell'indagine, un'attività a regime e ormai irrinunciabile per i risultati conseguiti. Per la sua prosecuzione e consolidamento, tuttavia, è indispensabile individuare risorse aggiuntive con l'obiettivo di continuare ad assicurare e, se possibile, migliorare la qualità dei dati prodotti.

Bibliografia

Bolasco S. *L'analisi automatica dei testi. Fare ricerca con il text mining*. Roma: Carocci, 2013.

De Angelis R., Macchia S., Mazza L. [Applicazioni sperimentali della codifica automatica: analisi di qualità e confronto con la codifica manuale](#). Istat, *Quaderni di ricerca - Rivista di statistica Ufficiale*, n.1/2000, pp.29-54.

della Ratta-Rinaldi F., Loré B. “Il lavoro e i suoi contenuti. Un'applicazione di Text Mining per categorizzare le attività dettagliate di lavoro nell'indagine campionaria sulle professioni Istat”. In: S. Bolasco, I. Chiari, L. Giuliano (eds). *Statistical Analysis of Textual Data. Proceedings of 10th International Conference 9-10 June 2010*; pp. 195-202, 917-928 e 929-937. Roma: Led, 2010.

della Ratta-Rinaldi F., Gallo F., Loré B. “How do you name your occupation? A text mining application on the language used by workers and by the standard occupational classification”. In: *Cladag 2011, 8th Scientific Meeting of the Data Analysis Group of the Italian Statistical Society*. Pavia: Pavia University press, settembre 2011.

della Ratta-Rinaldi F., Tibaldi M., Pontecorvo M. E. “Strategie di Text Mining per il controllo e la correzione della codifica dell'attività economica nell'indagine Istat sulle forze di lavoro”. In: E. Née,

- J.M. Daube, M. Valette, S. Fleury, (eds). *Actes des 12es Journées internationales d'Analyse statistique des Données Textuelles*, Paris : Sorbonne Nouvelle – Inalco, 2014.
- della Ratta-Rinaldi F., Tibaldi M. “Sperimentazione di un sistema di controllo e correzione per la codifica dell’attività economica”, Roma: *Istat working papers*, n. 4, 2014, in via di pubblicazione.
- Ferrillo A., Macchia S., Mazza L., Valery A., Vicari P. *La funzione su web per l’individuazione del codice ATECO sulla base di una descrizione sintetica e monitoraggio delle performance*. Roma: Istat working papers, n. 4, 2012.
- Gallo F., Loré B. *Training on the new occupational classification: the Italian experience*. Roma: Istat Working paper n. 12, 2012.
- Istat (Aa. Vv.). [Linee guida per la qualità dei processi statistici](#). Roma, 2011.
- Vicari P., Ferrillo A., Valery A. (eds). *Classificazione delle attività economiche - Ateco 2007*. Metodi e norme n. 40. Roma: Istat, 2009.