Il lavoro e i suoi contenuti. Un'applicazione di Text Mining per categorizzare le attività dettagliate di lavoro nell'indagine campionaria sulle professioni Istat [1]

Francesca della Ratta-Rinaldi, Barbara Lorè

Istat - Direzione Centrale sulle Condizioni di Vita, Via Ravà 150, 00142 Roma - Italy

Riassunto

In questo lavoro si presenta la strategia di trattamento delle risposte aperte utilizzata per codificare le attività specifiche di lavoro dell'indagine campionaria sulle professioni (1ª edizione 2006-07), finalizzata alla costruzione di un sistema descrittivo delle professioni presenti nel mercato del lavoro italiano. Le attività specifiche di lavoro sono state rilevate mediante una domanda aperta in cui veniva chiesto agli intervistati di indicare i compiti e le attività richiesti dalla propria professione. Il testo inserito dai rilevatori è stato sottoposto ad alcune tecniche di analisi testuale per standardizzare le risposte e produrre una lista delle attività dettagliate per ciascuna delle circa 800 unità professionali (UP) cui era rivolta l'indagine. Per categorizzare le attività è stata utilizzata la funzione di Ricerca Entità presente nel software TALTAC 2.0, che considera come unità di analisi non soltanto la singola forma testuale, ma anche l'intero frammento, all'interno del quale è possibile ricercare alcune combinazioni di parole cui associare un'etichetta, poi aggiunta in una nuova variabile nel dataset di partenza. La lista definitiva delle attività include 7.618 voci diverse.

Abstract

The aim of this work is to show the open answers analysis technique used for categorizing detailed work activities coming from Istat survey on occupations (1st edition, 2006-07), which gives a full description of jobs that shape the Italian labour market. Detailed work activities were collected by asking respondents to list the tasks they were expected to perform. The text entered by interviewers was processed using text analysis techniques in order to produce standardized lists of activities for all the studied occupations. For this purpose the Taltac2.0 "Entity Search" utility has been used because it allowed us to have as unit of analysis not single words but the entire fragments coming from respondents descriptions. This utility searches a distinctive word sequence inside fragments, labels it and adds it to the dataset as a new variable. This technique has brought improvements to data production process, both in terms of shortening working times and enhancing quality of data. The whole list of activities includes 7.618 different labels.

Keywords: text mining, automatic categorization, open questions, professions

1. Introduzione

In questo lavoro viene presentata la strategia utilizzata nell'ambito dell'indagine campionaria sulle professioni (1ª edizione, 2006-07) per trattare e codificare le risposte al quesito aperto sulle attività specifiche di lavoro. L'indagine, condotta dall'Istat d'intesa con il Ministero del

[1] Il lavoro è frutto di un lavoro comune. I paragrafi 1 e 6 sono stati redatti da entrambe le autrici, i paragrafi 2, 3 e 4 da Francesca della Ratta, il 5 da Barbara Lorè.

lavoro e l'Isfol, ha avuto come obiettivo la costruzione di un sistema informativo sulle professioni presenti nel mercato del lavoro italiano, utile ai diversi operatori del mercato del lavoro (soggetti che gestiscono l'incontro tra domanda e offerta, esperti di formazione, decisori politici ecc.) [2].

L'indagine ha riguardato circa 16.000 lavoratori, rappresentativi delle circa 800 Unità professionali (UP) [3] oggetto dell'indagine, cui è stato sottoposto un questionario composto da oltre 150 quesiti, mutuato dal sistema americano O*net (Occupational Information Network; www.online.onetcenter.org). L'indagine ha consentito di rilevare numerose informazioni sia sulle caratteristiche dei lavoratori che svolgono determinate professioni (conoscenze, competenze, attitudini, caratteristiche di personalità ecc.), sia sulle caratteristiche intrinseche alla professione stessa (condizioni entro cui si svolge la professione, percorsi di accesso ecc.).

Riguardo i compiti e le mansioni svolte, l'interesse era rivolto sia alle attività lavorative trasversali a più professioni che alle attività specifiche che caratterizzano ogni singola occupazione. L'intervista iniziava con la richiesta di indicare le principali attività svolte nell'ambito della propria professione, da un minimo di cinque a un massimo di dieci. Queste informazioni testuali sono state trattate con l'obiettivo di produrre una lista delle attività dettagliate per ciascuna Unità Professionale, standardizzando le risposte registrate dai rilevatori, ma conservando per quanto possibile la specificità e variabilità del testo originario.

2. La descrizione delle attività

Le tecniche di analisi testuale consentono di descrivere il linguaggio utilizzato dagli intervistati per raccontare le loro attività di lavoro. Il testo in analisi conta complessivamente oltre 430.000 occorrenze e circa 33.500 forme grafiche. L'analisi delle corrispondenze realizzata sulla distribuzione del vocabolario per Grandi Gruppi [4] professionali mostra la sostanziale omogeneità interna dei raggruppamenti professionali anche dal punto di vista del linguaggio utilizzato per descrivere il lavoro, confermando la validità della classificazione (Fig. 1).

Per evidenziare le caratteristiche più salienti del testo l'analisi delle corrispondenze è stata condotta su una selezione del vocabolario (linguaggio peculiare), composta dai termini sovrarappresentati rispetto a un modello di riferimento dell'italiano standard. Per ridurre l'impatto delle basse frequenze sono state considerate attive soltanto le parole con frequenza superiore a 10. Nel grafico, infine, sono visualizzate soltanto le forme che hanno maggiormente contribuito alla determinazione degli assi, in modo da semplificarne la lettura.

Il primo fattore mostra l'opposizione tra il linguaggio delle professioni intellettuali e quello delle professioni manuali. La distribuzione a ferro di cavallo suggerisce una parziale sovrapposizione tra i primi due fattori (effetto Guttman); peraltro, il secondo fattore sembra contrapporre

[2] Se si esclude il sistema informativo Orfeo dell'Isfol su formazione professionale, orientamento e istruzione, prima del 2006 nulla o poco era stato fatto in Italia per descrivere i contenuti delle professioni; a partire dalla seconda metà del 2010 avrà luogo una seconda edizione dell'indagine sulle professioni.

[3] Per unità professionale si intende un insieme di professioni omogenee sia rispetto alle attività che il lavoro richiede, sia rispetto alle caratteristiche soggettive, cognitive e di personalità di chi le svolge.

[4] I Grandi Gruppi, il primo livello classificatorio della Nomenclatura delle Unità Professionali, sono nove: I - Legislatori, dirigenti e imprenditori, II - Professioni intellettuali, scientifiche e di elevata specializzazione, III - Professioni tecniche, IV – Impiegati, V - Professioni qualificate nelle attività commerciali e nei servizi, VI - Artigiani, operai specializzati e agricoltori, VII - Conduttori di impianti e operai semiqualificati addetti a macchinari fissi e mobili, VIII - Professioni non qualificate, IX - Forze armate. L'indagine non ha riguardato nessuna delle professioni delle forze armate. http://www.istat.it/strumenti/definizioni/professioni/nup/.

tra due diversi modi di descrivere le attività di lavoro: da un lato il lavoro è descritto in termini di obiettivi, oggetti e processi generali, dall'altro vi è una descrizione più circoscritta dei contenuti e dei risultati del lavoro, tipica soprattutto del VI e del II Grande gruppo.

Tuttavia, per ottenere un elenco di attività standardizzate per ciascuna Unità professionale è stato necessario utilizzare le tecniche di analisi testuale in modo innovativo, riconducendo la risposta di ciascun record a un sistema di categorie definito. La strategia messa a punto è basata sulla funzione di "Ricerca Entità" (Entity Research by Regular Expressions) presente in TALTAC2, che ha permesso di ricercare ed etichettare nell'intero frammento gli insiemi di parole necessari a definire le attività dettagliate.

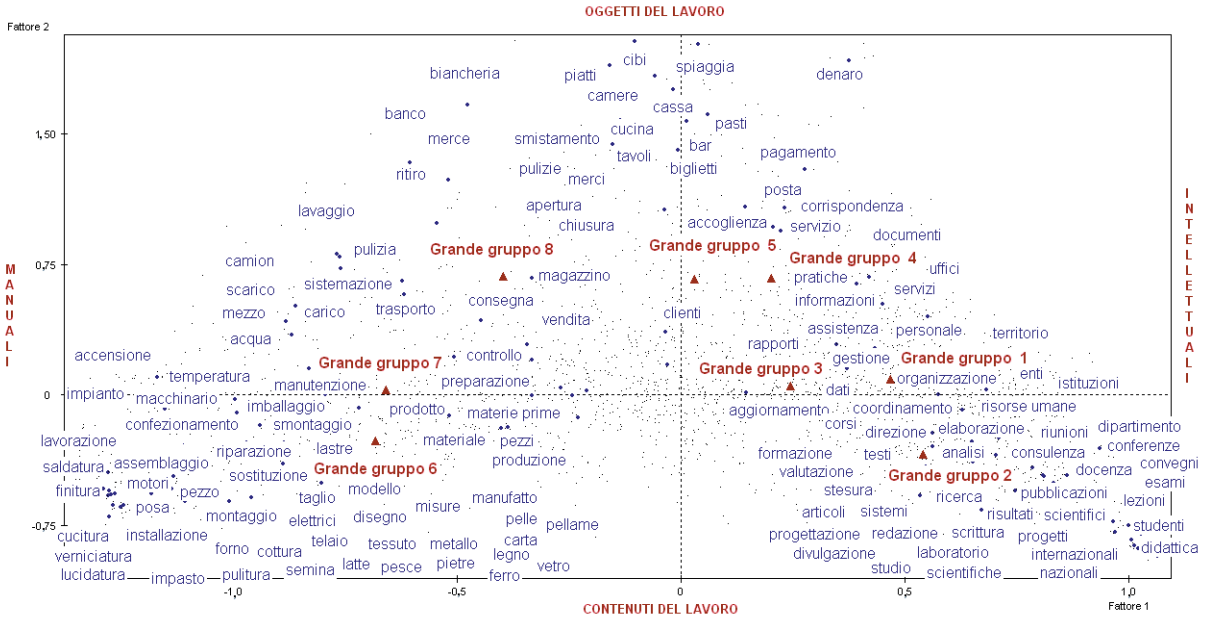

Figura 1: Piano fattoriale: distribuzione del vocabolario per Grandi gruppi professionali

3. La categorizzazione automatica delle attività

Successivamente alla predisposizione del testo si è proceduto alla categorizzazione delle attività mediante la funzione di "Ricerca Entità". L'utilità di questa funzione consiste nel considerare come unità di analisi non più la singola forma testuale ma l'intero frammento, all'interno del quale è possibile cercare alcune combinazioni di parole cui associare automaticamente un'etichetta da inserire in una nuova variabile aggiunta al file di partenza.

La ricerca sul testo delle informazioni utili ad alimentare questa nuova variabile avviene ricorrendo a query complesse definite da "espressioni regolari" (Bolasco, 2007). Queste fanno uso di operatori booleani (and, or, and not), di riduzioni lessematiche (caratteri jolly "*" e "?" come ad esempio contatt* e client?), di distanze [LAGgxx] fra parole in sequenza fra loro, che consentono di individuare nel testo le diverse espressioni che possono essere ricondotte ad un unico concetto. Ad esempio, la query "contatt* LAG3 client?" è in grado di individuare diverse risposte come "contattare il cliente", "contatti con i clienti", "contatto i miei principali clienti" e il valore da immettere nella nuova variabile potrebbe essere per ciascuna di queste risposte "curare i rapporti con i clienti".

Il primo passo di questo processo è la definizione di che cosa cercare, vale a dire del sistema di categorie. Per raggiungere questo obiettivo è necessario effettuare un'analisi del contenuto del

testo che consenta di individuare quali attività vi sono menzionate, stabilendo poi come queste devono essere etichettate.

Una prima panoramica sul tipo di attività citate dagli intervistati può essere ottenuta grazie all'insieme dei "segmenti ripetuti" presenti nel testo. Dalla lista dei segmenti sono stati selezionati soltanto quelli riconducibili ad una attività, come "gestione amministrativa" o "contatti con i clienti" che, al contrario di altri segmenti significativi (ad esempio "scuola primaria", "pentola a pressione", "risonanza magnetica" ecc.), consentono già in prima battuta di individuare una specifica attività. Questa lista è stata il punto di partenza per procedere alla categorizzazione delle attività utilizzando la funzione di "Ricerca Entità".

Poiché il lavoro di categorizzazione è iniziato durante la fase di rilevazione, non è stato possibile sottomettere alle query di codifica il testo nella sua interezza. Si è proceduto quindi per blocchi di interviste, man mano che venivano completate le interviste per ciascuna UP.

Nel primo blocco di testo esaminato (1.849 frammenti) le query costruite sui segmenti ripetuti con frequenza maggiore di 8 hanno consentito di etichettare l'11% del totale dei frammenti. Con tutti i segmenti con frequenza maggiore di 4 si arriva a 2.651 frammenti, pari al 17% del totale (circa 200 query). In totale, le operazioni di categorizzazione condotte sulla lista dei segmenti hanno portato alla classificazione automatica del 35% dei frammenti.

La funzione può essere utilizzata anche in modalità esplorativa: è possibile infatti verificare a quali frammenti corrisponde una determinata query per valutare la correttezza dell'etichetta che si è stabilito di apporre: a questa esplorazione può corrispondere anche una precisazione della query stessa in modo da evitare il problema di categorizzazioni non pertinenti.

Inoltre, la sequenza con cui vengono sottoposte le query è molto importante, in quanto il risultato di una query viene sovrascritto su quello eventualmente ottenuto in precedenza. Per questo motivo si deve lavorare con una logica "a imbuto", in cui le query successive specificano il risultato ottenuto con quelle più generiche. Ad esempio, una prima query generica relativa al segmento "attività didattica" ha consentito di classificare circa 100 frammenti. Di questi però alcuni possono essere meglio specificati con query più specifiche, relative ad esempio all'attività di coordinare o programmare la didattica.

Si tratta di una caratteristica che presenta aspetti sia positivi sia negativi. Da un lato la possibilità di specificare meglio le query consente una categorizzazione più accurata, che probabilmente non dovrà subire modifiche nella fase finale di revisione, dall'altro non sempre è possibile determinare la sequenza più appropriata (cioè quella che produce il risultato più vicino alla codifica stabilita in fase di validazione). Questa stessa logica a imbuto può essere utile anche per la categorizzazione di quei frammenti in cui le attività sono espresse con una sola parola (didattica, manutenzione, gestione, amministrazione, riparazione ecc.). Tali frammenti sono problematici sia perché non consentono di lavorare con la combinazione di parole, sia perché la loro categorizzazione è più dipendente dal contesto della UP cui sono riferiti. Pertanto, la categorizzazione automatica dei frammenti composti da una sola parola deve essere condotta con cautela in quanto si rischia di apporre un'etichetta troppo generica che probabilmente sarà modificata nella successiva fase di validazione.

Inoltre, nel caso di frammenti lunghi nei quali siano presenti più di una delle combinazioni di parole, la sovrascrittura del risultato rende problematica la categorizzazione multipla. Si è deciso di duplicare i frammenti troppo lunghi nella successiva fase di validazione, inserendo in campi diversi tutte le attività menzionate nel frammento.

Un'ulteriore tecnica di analisi del testo utile per individuare le attività su cui costruire le query di codifica è l'analisi delle concordanze delle parole chiave del testo. La possibilità di con-

trollare con l'analisi delle concordanze il contesto d'uso delle parole significative (vale a dire l'intorno sinistro e l'intorno destro di determinate parole pivot) consente di esplorare il testo originario individuando ulteriori attività da ricercare ed etichettare nel testo.

Questa modalità di analisi, un po' più onerosa dal punto di vista dei tempi di realizzazione, ha consentito di arrivare alla categorizzazione del 50 per cento dei frammenti.

Una volta raggiunta una soglia prefissata di categorizzazione del testo si può procedere alla ricostruzione del testo, operazione che consente di riscrivere il file originario con l'aggiunta della nuova variabile che riporta le etichette definite in fase di categorizzazione.

4. La validazione del risultato

La categorizzazione ottenuta in automatico è stata successivamente validata mediante un'analisi puntuale del risultato, finalizzata sia a completare il lavoro per i frammenti non classificati, sia a controllare l'aderenza dell'etichetta apposta in automatico al testo originario e al contesto dell'Up.

Il lavoro di validazione è stato svolto in due fasi: in una prima fase le Up sono state divise per Grandi gruppi e ciascun componente il gruppo di ricerca ha controllato le Up relative a uno o più Grandi gruppi, modificando e integrando le etichette poste in automatico: in questo modo ciascun codificatore è diventato esperto delle attività realizzate da professioni relativamente simili, rendendo progressivamente più rapido il lavoro di codifica.

In questa fase la presenza di una etichetta automatica per una parte delle attività si è rivelata molto utile perché ha orientato il lavoro di codifica. La fase conclusiva della validazione è stata invece realizzata da un'unica persona, che ha riunificato le singole Up, controllando soprattutto l'uniformità delle etichette nel caso di attività comuni a più di una Up, in modo da assicurare la possibilità di confronto e individuazione di attività trasversali.

La fase di validazione della codifica effettuata sul primo insieme di frammenti ha inoltre permesso al gruppo di lavoro di raggiungere una certa omogeneità nelle modalità di codifica delle attività. Le query realizzate sono state salvate in modo da renderle operative per le successive tornate di codifica. La possibilità di scrivere in un file .txt le query da sottoporre al testo ha consentito di velocizzare significativamente le operazioni. Le query sono state salvate in diversi file distinti per tipologia di argomenti (contatti-relazioni con clienti o fornitori; attività amministrative; risorse umane, didattica, ricerca, manutenzione ecc.), in modo da poter operare distinzioni tra i file da mandare in esecuzione. La sequenza delle query registrate segue la logica a imbuto, per cui sono state inserite prima le query più generiche e successivamente quelle più dettagliate che specificano il risultato.

Poiché si è ritenuto indispensabile procedere comunque a un'accurata validazione e integrazione manuale delle risposte si è ritenuto di ottimizzare i tempi della categorizzazione automatica non andando oltre il 50 per cento dei record categorizzati. A volte, infatti, il tempo richiesto per l'analisi del testo e la formulazione della query poteva mostrarsi eccessivo rispetto al risultato raggiunto.

Tuttavia, man mano che si è proceduto con le successive tornate di Up da categorizzare la quota di record classificati utilizzando le query già mandate in esecuzione è andata via via aumentando. Inoltre, la disponibilità di un dizionario di attività sempre più dettagliato ha reso più agevole la fase di classificazione manuale, limitando sempre più la necessità di creare ex novo le etichette.

In sintesi la strategia di categorizzazione adottata presenta i seguenti passi:

1) normalizzazione di base del testo;
2) individuazione e selezione segmenti ripetuti riferibili alle attività;
3) analisi delle concordanze delle parole chiave per individuare ulteriori gruppi di attività;
4) sottomissione delle query per etichettare la porzione più ampia possibile di testo (ottimale fino al 50 per cento);
5) registrazione delle query per le successive porzioni di testo;
6) ricostruzione del testo aggiungendo al file originario una nuova variabile in cui è riportata l'etichetta attribuita nel corso della Ricerca Entità;
7) individuazione dei record brevi da eliminare (xxx, aaa ecc.);
8) controllo della lista di attività per ciascuna Up per validare, completare o integrare la categorizzazione effettuata e sdoppiare i frammenti lunghi con due attività.

5. L'omogeneizzazione e la validazione del dizionario delle attività

Una volta etichettate tutte le attività si è provveduto omogeneizzare il dizionario, che nella sua versione finale presenta 7.618 voci, per oltre 70 mila occorrenze, con una frequenza media di circa nove occorrenze per ogni voce. Per ridurre la variabilità introdotta nella fase di revisione manuale è stato necessario un ulteriore intervento di validazione, finalizzato a uniformare le etichette simili nel contenuto che presentavano variazioni dovute al singolare/plurale, congiunzioni/disgiunzioni, uso di sinonimi, specifiche superflue, parole invertite ecc. Si è cercato, dunque, di rendere il più possibile uniformi quelle espressioni che pur facendo riferimento ad attività comuni a più di una Up, presentavano un certo grado di variabilità dovuta ai diversi stili lessicali e/o cognitivi utilizzati dai revisori nella fase di revisione manuale delle risposte.

Dopo una prima correzione ortografica, si è proceduto ad eliminare, quando possibile, la variabilità legata al singolare/plurale, alle congiunzioni/disgiunzioni o alla presenza/assenza di articoli. Nei casi in cui la scelta tra l'una o l'altra etichetta non comportava alcuna alterazione di significato, la scelta è ricaduta su quella che presentava un numero di occorrenze maggiore in termini di Up a cui era stata attribuita. Negli altri casi è stato necessario, invece, entrare nel merito della descrizione della professione e delle altre attività per accertare che le modifiche non danneggiassero l'omogeneità interna di ciascuna Up. Interventi di tipo sostanziale, invece, sono stati adottati nei casi in cui la stessa attività era espressa da sinonimi o da espressioni equivalenti nel contenuto ma differenti per l'ordine o il numero delle parole. Nel primo caso, dopo aver accertato che si trattasse realmente di sinonimi e che non ci fossero differenze semantiche riconducibili a sottili sfumature, ci si è basati unicamente sul criterio della frequenza. Nel caso, invece, di etichette che differivano per la presenza o assenza di specifiche si è entrati nel merito delle specifiche stesse per valutarne l'effettiva necessità. La stessa attenzione è stata necessaria nel caso di etichette molto simili e che presentavano differenze sia rispetto al numero delle parole sia al loro ordine. Molto spesso si sono presentate congiuntamente più situazioni tra quelle descritte, rendendo necessario un attento lavoro di analisi delle Up coinvolte e di valutazione delle modifiche più opportune da apportare.

L'inevitabile variabilità introdotta dai codificatori, nonostante la disponibilità di un dizionario integrato da cui attingere le etichette, dimostra l'utilità di un approccio semiautomatico, che ha il merito di ridurre o almeno controllare la variabilità delle decisioni che possono essere prese in fase di codifica.

A titolo esemplificativo si riportano in Tab. 1 le attività menzionate con maggiore frequenza dagli intervistati, ripartite per Grandi gruppi professionali.

Nonostante la fase di validazione manuale del risultato sia risultata piuttosto onerosa, la possibilità di categorizzare in automatico una parte dei frammenti costituisce un risultato molto

interessante, in quanto ha ridotto i tempi di lavoro, ha garantito un buon livello di standardizzazione del risultato, ha permesso di individuare le attività trasversali a diverse Up e ha costituito un criterio guida per la fase di codifica manuale.

ATTIVITÀ DETTAGLIATE	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*Totale*
Controllare la qualità del processo o del prodotto	39	69	122	6	9	313	852		1.410
Curare i rapporti con i clienti	209	84	242	46	150	346	93		1.170
Organizzare il lavoro o le attività	101	113	160	70	93	179	333	8	1.057
Gestire e/o coordinare le risorse umane	508	156	227	23	72	59	46		1.091
Studiare e aggiornarsi	26	437	162	10	30	46	5		716
Curare i rapporti con i fornitori	153	26	123	32	111	164	37	11	657
Svolgere attività di manutenzione ordinaria o straordinaria su attrezzature o impianti		10	40		8	207	323	10	598
Controllare i macchinari o le attrezzature		9	42		3	77	464	1	596
Coordinare il lavoro o le attività	142	77	83	7	10	33	7		359
Gestire gli ordinativi (acquisire ed evadere)	9	15	58	6	44	70	14	15	231
Partecipare a corsi di formazione e aggiornamento	10	55	33	11	24	29	16	2	180
Gestire il magazzino	8	7	7	6	32	35	37	5	137
Totale attività dettagliate	*6.078*	*15.988*	*14.481*	*3.022*	*5.911*	*16.111*	*10.073*	*1.588*	*73.252*

Tabella 1: Attività dettagliate citate con maggiore frequenza dagli intervistati per Grande gruppo professionale (valori assoluti)

Secondo l'esperienza di chi ha condotto il lavoro di validazione e integrazione puntuale delle liste, la possibilità di disporre di una parte delle attività già classificate ha semplificato notevolmente il lavoro, orientandolo verso binari già predefiniti. Inoltre, man mano che il lavoro di categorizzazione è andato avanti, la disponibilità del dizionario delle attività via via implementato ha facilitato e velocizzato il lavoro, limitando la soggettività della codifica da parte di differenti codificatori. Peraltro, le liste verranno sottoposte a validazione empirica nel corso della seconda edizione dell'indagine che si svolgerà nel periodo 2010-2011.

6. Conclusioni

Ci sembra di poter affermare che la strategia adottata ha consentito di migliorare il risultato finale, sia dal punto di vista dei tempi di esecuzione sia da quello della standardizzazione del dizionario di attività.

Resta da considerare se la percentuale di frammenti categorizzati in automatico possa considerarsi un buon risultato, tale da suggerire di adottare gli strumenti dell'analisi testuale per la codifica di testo non strutturato. Benché in genere le applicazioni di classificazione automatica dei testi consentano di classificare porzioni più ampie di testo, in questo caso sono stati classificati soltanto la metà dei frammenti a causa della eterogeneità del testo e per la necessità di preservare e restituire la specificità delle singole Up. In applicazioni analoghe ma condotte su gruppi più omogenei di rispondenti è stato infatti possibile etichettare correttamente oltre il 70% dei frammenti, proprio grazie alla maggiore omogeneità delle risposte.

Si può pertanto ipotizzare che la particolarità del testo analizzato costituisca un "caso critico" dal punto di vista del grado di automazione raggiungibile con questo tipo di applicazioni, i cui risultati migliorano decisamente a fronte di una maggiore omogeneità del testo trattato.

Riferimenti bibliografici

Bolasco S. (2007). Analisi dei diari giornalieri con strumenti di statistica testuale e text mining. In Istat, *I tempi della vita quotidiana*, Argomenti, n. 32. Roma: Istat.

Bolasco S. (1999). *Analisi multidimensionale dei dati.* Roma: Carocci.

della Ratta-Rinaldi F. (2007). L'analisi testuale computerizzata. In Cannavò, L. and Frudà, L., editors, *Manuale di ricerca sociale applicata. Tecniche speciali di rilevazione, trattamento e analisi*, Roma: Carocci, cap. V.

della Ratta-Rinaldi F., Lorè B. and La Rocca G. (2007). Textual analysis perspectives on categorisation of activities in Istat survey on occupations, In *Classification and Data Analysis 2007, Book of short papers*, Macerata: EUM, pp. 263-266.

Gallo F., Scalisi P. and Scarnera C. (editors) (2009). *L'indagine sulle professioni. Anno 2007. Contenuti, metodologia e organizzazione.* Roma: Istat, *Metodi e Norme. n. 42*.

Gallo F. and Lorè B. (2006) Descrivere le professioni: il modello adottato nell'indagine Isfol/Istat. In Crocetta, C., editor, *Metodi e modelli per la valutazione del sistema universitario*, Padova: CLEUP, pp. 367-379.

Istat (2008). *Competenze, attività e condizioni lavorative delle professioni in Italia – Anno 2007*, Statistica in breve. Roma.

Peterson N.G., Munford M.D., Borman W.C., Jeanneret P.R. and Fleishman E.A. (1999). A occupational information system for the 21st century: the development of O*net. *American Psychological Association*, Washington, DC 20002.

Scarnera, A. (editor) (2003). Il Dizionario delle professioni tecniche: uno studio di fattibilità. Roma: Istat.