

Se pensa al suo futuro, di cosa ha più paura? ¹

Francesca della Ratta-Rinaldi

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

Riassunto

A distanza di cinque anni l'IRES ha replicato nel 2009 un'indagine campionaria sui bisogni e le aspettative delle donne nei confronti del lavoro. All'interno di un questionario centrato sulle tematiche della partecipazione al mercato del lavoro e della conciliazione lavoro-famiglia è stata inserita una domanda volutamente generica che chiedeva appunto "Se pensa al suo futuro, di cosa ha più paura?". L'obiettivo era quello di valutare l'entità delle paure legate alla vita lavorativa rispetto a quelle più generali, quali ad esempio le malattie, le guerre, la situazione internazionale, la criminalità e così via. Oltre ai risultati descrittivi dell'analisi, che consentono tra l'altro di evidenziare l'impatto della crisi economica attuale sull'immaginario delle intervistate, saranno utilizzate tecniche di Text Mining finalizzate a conteggiare in modo puntuale i riferimenti all'attività lavorativa rispetto alle altre tematiche. La possibilità di categorizzare le risposte sulla base di una griglia costruita in seguito all'analisi descrittiva del testo consentirà inoltre di associare i temi delle risposte aperte alle caratteristiche delle intervistate.

Abstract

Five years later, IRES repeated in 2009 a sample survey on women's needs and expectations concerning work. The questionnaire was focused on labor market participation and work-family reconciliation and contained an open question deliberately unspecific on "Thinking to your future, what are you most afraid of?". The aim was to evaluate the job-related fears compared to more general ones, such as illness, wars, international situation, crime and so on. Beside analysis descriptive results, which point out among others economic crisis impact on interviewee imaginary, Text Mining techniques will be used to count job-related references by comparison with other fears. The opportunity to obtain an answer classification based on a grid built starting from text descriptive analysis, will allow to link open answers themes to interviewee characteristics.

Keywords: open questions, text mining, female work, economic crisis

1. Introduzione

Dal punto di vista della ricerca sociale, una delle applicazioni più interessanti dell'analisi testuale è la possibilità di analizzare in profondità le risposte a domande aperte. Queste rappresentano una risorsa molto utile poiché consentono all'intervistato di esprimere liberamente il proprio pensiero, eludendo sia il rischio di costringere l'intervistato entro una risposta eccessivamente rigida, sia di suggerirgli una risposta nel caso in cui sia privo di opinione (Pitrone, 1984). Nonostante questi vantaggi la codifica delle risposte a domande aperte è considerata troppo onerosa, tanto da limitarne il ricorso ai casi in cui non è possibile fare ricorso a domande predefinite. Inoltre, le tradizionali operazioni di post-codifica sono spesso "riduttive" rispetto alla messe di informazioni contenute nelle risposte libere (Pitrone, 1984: 64).

¹ Il presente lavoro fa riferimento a un più ampio lavoro condotto dell'IRES e coordinato da Giovanna Altieri.

Le tecniche di analisi testuale, invece, riducono l'onere dell'analisi conservando più informazione rispetto alla semplice identificazione di modalità realizzata con la post-codifica: già Lebart e Salem ricordano come la post-codifica manuale costituisca una irrimediabile perdita di informazione a causa della mediazione del codificatore, in seguito alla quale «l'informazione è mutilata nella sua forma, e spesso *impoverita* nel suo contenuto» (1994: 29, corsivo mio).

Invece, la possibilità di esplorare il testo e di individuare i profili lessicali di determinate tipologie di rispondenti, consente non solo di individuare dimensioni inattese, ma offre la possibilità di lavorare sul linguaggio dei soggetti. Inoltre, come si cerca di mostrare in questo lavoro, l'integrazione del tradizionale approccio "lessicale" con un approccio di tipo "testuale" (Bolasco and Pavone, 2010), finalizzato alla categorizzazione automatica delle singole risposte, consente di recuperare i vantaggi e la logica della post-codifica, riconducendo ciascuna risposta a una modalità di risposta definita sulla base dei risultati dell'analisi "lessicale" del testo. L'interesse di questo approccio è la possibilità di identificare e costruire le modalità della post-codifica in modo più sistematico e meno dipendente dalla sensibilità del ricercatore, sfruttando le potenzialità descrittive delle tecniche statistiche di analisi del testo.

2. Il testo in analisi: risultati dell'analisi lessicale

A distanza di cinque anni l'IRES (Istituto Ricerche Economiche e Sociali) ha replicato nel 2009 un'indagine campionaria sui bisogni e le aspettative delle donne nei confronti del lavoro. In ciascuna delle due tornate di indagine sono state intervistate con tecnica CATI circa 1.000 donne, nella condizione di occupate, in cerca di occupazione e inattive, residenti su tutto il territorio nazionale. Nel questionario, centrato sulle tematiche della partecipazione al mercato del lavoro e della conciliazione lavoro-famiglia, è stata inserita una domanda che chiedeva "Se pensa al suo futuro, di cosa ha più paura?". L'inserimento di una domanda piuttosto generica e volutamente non circoscritta era finalizzato a stabilire il peso delle paure legate alla vita lavorativa rispetto a quelle più universali, come la malattia, la guerra, la criminalità e così via.

I testi raccolti nelle due indagini presentano dimensioni simili: poco più di 4.000 parole nell'indagine del 2004 e circa 3.000 nel 2009 (considerando le mancate risposte si tratta in media di circa 4 parole per intervistata). Nonostante si tratti di un testo di piccole dimensioni (anche per via della tecnica telefonica utilizzata), esso presenta alcuni elementi di interesse che contribuiscono ad arricchire l'analisi svolta e a fornire una risposta sull'importanza della dimensione lavorativa nell'immaginario delle intervistate.

Nelle risposte del 2009 le parole piene con frequenza più elevata sono *lavoro*², *figli*, *futuro*, *salute*, *malattie*, che con un ordine leggermente diverso comparivano anche nel 2004. Nel 2009 si affacciano anche termini come *crisi*, *economica*, *mancanza*, *disoccupazione*, che possono essere collegati alla particolare situazione congiunturale in cui sono state realizzate le interviste (tra marzo e aprile 2009).

Tra i segmenti ripetuti (vale a dire le combinazioni di parole che si presentano con la stessa sequenza nel testo), troviamo numerosi riferimenti alla crisi economica in atto, soprattutto la paura di non trovare o perdere il lavoro o non riuscire a sostenere le difficoltà economiche (*crisi economica*, *mancanza di lavoro*, *instabilità del lavoro*, *rimanere senza lavoro*). Vi sono poi i timori relativi al futuro dei figli o dei giovani in generale (*futuro dei miei figli*, *instabilità lavorativa future generazioni*). Un peso non irrilevante è occupato dal timore della precarietà,

² Da questo punto in poi le parole in corsivo sono quelle effettivamente presenti nel testo.

associato sia alla situazione contingente (*non mi venga rinnovato il contratto*) sia al proprio futuro (*non trovare un lavoro adeguato o non percepire la pensione*).

Oltre ai segmenti ripetuti è poi possibile estrarre dal vocabolario le parole chiave: queste si ottengono confrontando la frequenza relativa delle parole presenti nel testo in analisi con quella con cui le stesse parole compaiono in un lessico di frequenza – assunto come rappresentativo del “linguaggio comune”³. A ciascuna parola viene quindi associata una misura di significatività (scarto standardizzato): le parole chiave sono pertanto quelle che risultano sovra-rappresentate nel testo in analisi (e a cui corrispondono i valori più elevati dello scarto).

Per comodità di lettura le parole chiave sono state ricondotte ad alcune tematiche generali: crisi economica, precarietà, giovani e figli e paure universali: anche in questo caso sono prevalenti i riferimenti alla crisi economica, alle incertezze per il futuro dei figli e alla crescente precarietà del lavoro (Tab. 1).

<i>Forma grafica</i>	<i>Occ. totali</i>	<i>Scarto</i>	<i>Forma grafica</i>	<i>Occ. totali</i>	<i>Scarto</i>
CRISI ECONOMICA			PAURE UNIVERSALI		
disoccupazione	45	620,2	malattie/malattia	44	1.084,7
crisi	102	612,0	futuro	97	727,0
lavoro	148	451,2	salute	65	535,1
mancanza	42	450,9	instabilità	18	504,2
economica/ economia	87	439,5	vecchiaia	18	495,8
sicurezza	16	132,2	criminalità	22	374,3
famiglia	13	75,4	solitudine	16	343,5
difficoltà	12	66,9	violenza	27	283,4
PRECARIETÀ			degrado	9	235,9
precarietà/precariato	45	1.561,1	paura	26	219,5
incertezza	24	461,2	menefreghismo	1	124,3
pensione	24	328,3	donne	12	86,5
insicurezza	4	215,1	delinquenza	3	79,0
precaria/precario	8	171,8	valori	8	64,6
GIOVANI E FIGLI			ambiente	6	63,4
figli/ figlio/ figlia	124	739,5	inquinamento	2	62,5
giovani	24	177,9	povertà	4	53,3
generazioni	8	148,4	immigrati	3	48,1

Tabella 1: Selezione di parole chiave ordinate per valori decrescenti dello scarto standardizzato. Corpus 2009

Se si raggruppano in un unico testo le risposte ottenute nelle due indagini del 2004 e del 2009 è possibile confrontare il vocabolario utilizzato nelle due tornate d'indagine per ottenere indicazioni su come sono cambiate le aspettative e i timori verso il futuro negli ultimi cinque anni. Nel 2004 risultavano caratteristici i riferimenti alle paure universali, come la salute o la situazione internazionale, con termini come *guerra, malattie, terrorismo, morte*. Tra i temi legati al lavoro emergeva anche il tema della conciliazione vita-lavoro. Nel 2009 invece sono prevalenti soprattutto parole come *disoccupazione, precarietà e crisi economica*.

³ Le parole chiave sono state estratte utilizzando il programma TALTAC, che usa come modello di riferimento il lessico di frequenza dell'italiano standard espresso in forme grafiche (POLIF), composto da oltre quattro milioni di occorrenze (Cfr. Bolasco, 1999: 223).

Una sintesi grafica delle differenze tra il vocabolario utilizzato nelle due indagini è presentata in Fig. 1, ottenuta grazie all'analisi delle corrispondenze applicata all'insieme delle risposte registrate nelle due indagini. La matrice analizzata è una tabella forme*testi, che presenta in riga le parole chiave e in colonna la frequenza con cui queste compaiono nei due anni di indagine e nei tre gruppi di intervistate corrispondenti alle occupate, alle disoccupate e alle inattive.

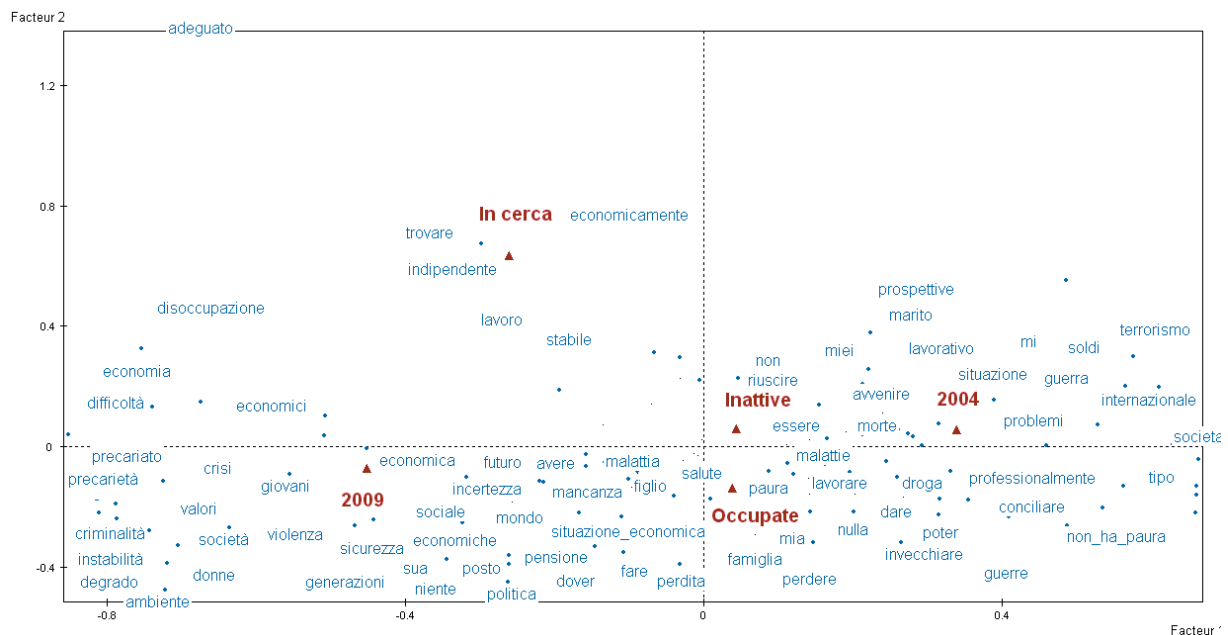


Figura 1: Piano fattoriale, confronto 2004-2009 e caratteristiche intervistate

Il primo fattore è chiaramente riconducibile alla dimensione cronologica e mostra l'evoluzione della terminologia utilizzata nelle due indagini: se in corrispondenza del 2004 troviamo una pluralità di tematiche, le risposte del 2009 sono più centrate su quella lavorativa. Nel 2004, peraltro, le paure legate al lavoro erano riferite soprattutto alle prospettive lavorative, alla conciliazione lavoro-famiglia o alla soddisfazione professionale.

Il secondo fattore invece riproduce le differenze tra le occupate da un lato e le inattive e disoccupate dall'altro. A caratterizzare le disoccupate sono soprattutto i riferimenti al timore di non trovare un lavoro stabile o di non essere indipendente economicamente; le inattive si caratterizzano per i riferimenti alla salute e alle paure per i figli o i familiari: sono soprattutto le occupate a parlare della situazione economica, delle pensioni, dell'incertezza dovuta alla precarietà e del timore di perdere il lavoro.

L'analisi sin qui svolta fornisce un quadro dettagliato del tipo di paure citate dalle intervistate nelle loro risposte e consente di individuare le principali differenze tra l'indagine del 2004 e quella del 2009. L'elevata frequenza dei termini riferiti al lavoro mette in evidenza come questo sia sicuramente un tema cruciale nell'immaginario delle intervistate, almeno al pari di altri temi più generali come la paura delle malattie o degli eventi tragici.

Tuttavia, il tradizionale approccio lessicale non consente di quantificare il peso di una tematica rispetto a un altro né di mettere in relazione in modo sistematico le caratteristiche degli intervistati con un determinato tipo di risposta.

3. Analisi testuale e codifica automatica delle risposte

3.1. La strategia di codifica automatica adottata

Per quantificare in modo più puntuale la frequenza con cui ciascuna tematica è stata citata dalle intervistate è stata utilizzata una tecnica di analisi utile per ricondurre le risposte a un insieme di modalità definite. Si tratta della funzione di “Ricerca Entità” (Entity Research by Regular Expressions) presente in TALTAC2, che, a partire dalla somministrazione di query complesse basate sulla combinazione di parole all’interno di ciascuna risposta, consente di codificarne automaticamente il contenuto.

Questa funzione presenta un elemento di grande interesse: l’unità di analisi non è più la singola parola o forma grafica contenuta nel testo ma l’intera risposta (o frammento testuale), all’interno della quale è possibile cercare determinate combinazioni di parole, con il fine di individuare le diverse espressioni riconducibili a un medesimo concetto. Alle risposte che contengono combinazioni di parole considerate simili è possibile associare automaticamente una etichetta o modalità, che viene inserita 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 et al., 2007). Queste fanno uso di operatori booleani (and, or, e not), di riduzioni lessematiche (caratteri jolly: * e ?), di distanze [LAGgxx] fra parole in sequenza fra loro. Ad esempio la query “perd* LAG4 lavor? OR perd* LAG4 occupazion?” è in grado di individuare diverse risposte come “perdere il mio lavoro”, “perdita del lavoro”, “che perdo il lavoro”, “perdere l’occupazione e non trovarne un’altra”. Se si considerano simili tutte queste risposte sarà possibile associare a ciascuna di esse la stessa etichetta “perdere il lavoro” nella nuova variabile “risposta codificata” aggiunta al file di partenza. In questo modo a risposte espresse con terminologia differente corrisponderà un’unica modalità di risposta assegnata in automatico.

Naturalmente, per poter codificare automaticamente il contenuto delle risposte è indispensabile aver svolto una fase preliminare di analisi: solo così infatti è possibile definire che cosa cercare nel testo. Le query infatti vengono definite sulla base dei principali risultati dell’analisi lessicale (segmenti ripetuti, parole chiave, analisi delle concordanze ecc.).

In particolare sono molto utili i segmenti ripetuti, soprattutto nel caso di risposte brevi in cui spesso il segmento coincide con il contenuto espresso nella risposta. Segmenti come *perdere il lavoro*, *manca di lavoro*, *crisi economica* ecc. sono stati il punto di partenza per la definizione delle query. Peraltro, la possibilità di utilizzare le riduzioni lessematiche e gli operatori di distanza consente di “pescare” nel testo anche combinazioni di parole che non direttamente coincidenti con il segmento in questione. Riprendendo l’esempio precedente, se si parte dal segmento “perdere il lavoro” la query “perd* LAG4 lavor?” consente di individuare non solo tutte le risposte in cui è effettivamente contenuta l’espressione “perdere il lavoro” ma anche altre combinazioni, come ad esempio “che perdo il mio lavoro” o “la perdita del mio lavoro” o “perdita del lavoro”, accrescendo il numero di risposte che possono essere codificate con un’unica query.

Contando sia le risposte del 2004 sia quelle del 2009 le risposte valide da codificare erano in tutto 1.848 (899 risposte nell’indagine del 2004 e 949 nel 2009). Le circa 240 query sottoposte al corpus hanno consentito di codificare automaticamente 1.549 risposte, pari all’84% del totale (il tempo necessario a formulare ed eseguire le query è stato pari a circa una giornata lavorativa).

Le oltre 40 modalità ottenute nella fase di interrogazione e codifica del testo sono state ulteriormente aggregate in 10 macro categorie. Il risultato è stato poi controllato manualmente, sia per completare il lavoro di codifica sui restanti 300 record, sia per modificare eventuali codifiche non rispondenti al contenuto effettivo della risposta. Delle 1.549 risposte codificate in automatico è stato necessario modificare la modalità assegnata in automatico 152 volte, vale a dire in meno del 10% dei casi.

3.2. Risultati della codifica automatica

L'interesse della tecnica utilizzata consiste nella possibilità di fornire un'indicazione quantitativa della frequenza di ciascuna modalità o categoria e in particolare dell'insieme dei riferimenti al lavoro, offrendo informazioni molto interessanti su come sono cambiate le paure di una parte dell'universo femminile (Fig. 2).

Innanzitutto è importante sottolineare la rilevanza delle macro categorie riferite al lavoro e all'economia in generale (si ricorda che la domanda era volutamente generica e soprattutto non era riferita nello specifico alla vita lavorativa): queste rappresentano il 35% del totale delle intervistate nel 2004 e quasi la metà nel 2009. Nel 2009 la rilevanza delle questioni lavorative ed economiche fa inoltre diminuire il peso di paure più universali come i problemi di salute o gli eventi tragici. Tra le paure più rilevanti registrate nel 2004 spiccano invece i riferimenti alle questioni politiche, alla situazione internazionale, al terrorismo e ai conflitti in corso.

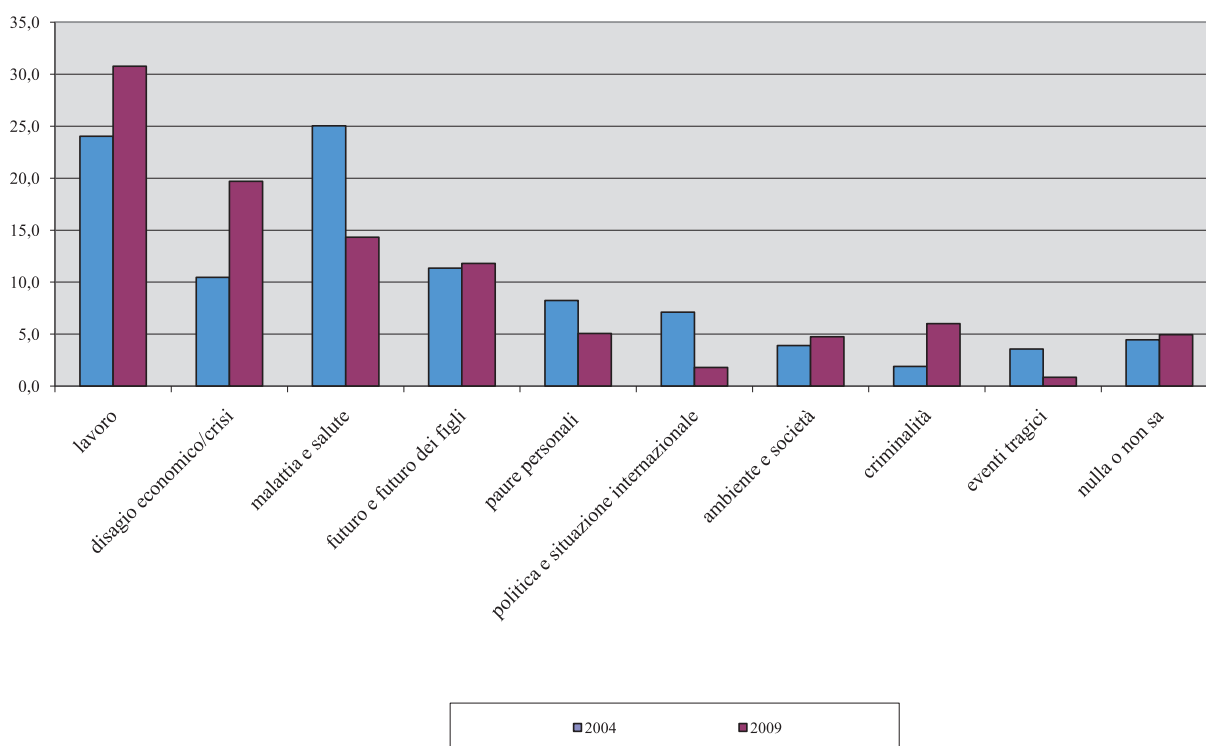


Figura 2: Macro categorie utilizzate per la codifica automatica delle paure. Confronto 2004-2009 (valori percentuali sul totale delle risposte valide in ciascuna tornata di indagine)

Nel 2009 si registra anche un incremento delle paure legate alla criminalità e alla violenza sulle donne, probabilmente per effetto dei drammatici fatti di cronaca verificatesi nelle settimane in cui è stata condotta l'indagine.

In Tab. 2 si può osservare il dettaglio delle modalità e delle macro categorie ottenute in seguito all'operazione di codifica, distinte per anno di indagine. Tra le voci riferite al lavoro spiccano il timore per la disoccupazione e per il lavoro precario, diffuse in misura più rilevante tra le intervistate del 2009. Seguono le preoccupazioni per il futuro lavorativo dei figli e delle giovani generazioni. Un'altra paura ricorrente è quella di non riuscire a realizzarsi nel lavoro, a causa della difficoltà a trovare un impiego coerente con gli studi fatti o con le proprie aspirazioni. In totale, un quarto delle risposte del 2004 e quasi un terzo di quelle del 2009 sono riferite al lavoro.

Macro categoria	Modalità	Valori assoluti			Valori percentuali sul totale risposte valide		
		2004	2009	Totale	2004	2009	Totale
LAVORO	non trovare lavoro/disoccupazione	45	86	131	5,0	9,1	7,1
	lavoro precario	18	54	72	2,0	5,7	3,9
	futuro lavorativo dei figli	34	31	65	3,8	3,3	3,5
	perdere lavoro	29	30	59	3,2	3,2	3,2
	non realizzarmi nel lavoro	28	20	48	3,1	2,1	2,6
	futuro lavorativo dei giovani	5	23	28	0,6	2,4	1,5
	non riuscire a conciliare lavoro e famiglia	15	1	16	1,7	0,1	0,9
	non trovare lavoro in Italia	1	2	3	0,1	0,2	0,2
	dover lavorare tutta la vita		1	1	0,0	0,1	0,1
	<i>non codificata</i>	41	44	85	4,6	4,6	4,6
Tot. categoria		216	292	508	24,0	30,8	27,5
DISAGIO ECONOMICO/CRISI	crisi economica	24	117	141	2,7	12,3	7,6
	non avere la pensione	23	21	44	2,6	2,2	2,4
	riduzione potere d'acquisto	16	3	19	1,8	0,3	1,0
	dipendere dalla famiglia	7	6	13	0,8	0,6	0,7
	difficoltà economiche	2	9	11	0,2	0,9	0,6
	non farcela	1	5	6	0,1	0,5	0,3
	povertà	1	3	4	0,1	0,3	0,2
	<i>non codificata</i>	20	23	43	2,2	2,4	2,3
Tot. categoria		94	187	281	10,5	19,7	15,2
MALATTIA SALUTE	malattie	215	127	342	23,9	13,4	18,5
	malattia invalidante	3	5	8	0,3	0,5	0,4
	<i>non codificata</i>	7	4	11	0,8	0,4	0,6
Tot. categoria		225	136	361	25,0	14,3	19,5
FUTURO	futuro dei figli	79	76	155	8,8	8,0	8,4
	incertezza futuro	2	9	11	0,2	0,9	0,6
	<i>non codificata</i>	21	27	48	2,3	2,8	2,6
Tot. categoria		102	112	214	11,3	11,8	11,6
Totale risposte valide		899	949	1.848	100,0	100,0	100,0

Tabella 2 (prima parte): Paure per il futuro per anno di indagine. Valori assoluti e percentuali sul totale risposte valide in ciascun anno

Strettamente collegate al lavoro, soprattutto nel 2009, sono le paure riferite alla situazione economica e alle difficoltà cui può andare incontro la propria famiglia. Le risposte riferibili a questa categoria passano dal 10,5% nel 2004 al 19,7% nel 2009. Oltre ai timori per la crisi economica, le intervistate citano il rischio di percepire una pensione insufficiente, le difficoltà

determinate dalla riduzione del proprio potere d'acquisto, la preoccupazione di dover continuare a dipendere economicamente dalla famiglia o più semplicemente il timore di “non farcela”.

Macro categoria	Modalità	Valori assoluti			Valori percentuali sul totale risposte valide		
		2004	2009	Totale	2004	2009	Totale
PAURE PERSONALI	solitudine	29	14	43	3,2	1,5	2,3
	vecchiaia	21	17	38	2,3	1,8	2,1
	rimandare la famiglia	1	8	9	0,1	0,8	0,5
	non essere serena	2	1	3	0,2	0,1	0,2
	<i>non codificata</i>	21	8	29	2,3	0,8	1,6
Tot. categoria		74	48	122	8,2	5,1	6,6
POLITICA	guerre	45	4	49	5,0	0,4	2,7
	situazione internazionale	7		7	0,8	0,0	0,4
	situazione politica	3	2	5	0,3	0,2	0,3
	instabilità politica		3	3	0,0	0,3	0,2
	multietnicità	1		1	0,1	0,0	0,1
	<i>non codificata</i>	8	8	16	0,9	0,8	0,9
Tot. categoria		64	17	81	7,1	1,8	4,4
AMBIENTE	degrado socio-ambientale	10	23	33	1,1	2,4	1,8
	terremoto	3		3	0,3	0,0	0,2
	<i>non codificata</i>	22	22	44	2,4	2,3	2,4
Tot. categoria		35	45	80	3,9	4,7	4,3
CRIMINALITÀ	criminalità	9	23	32	1,0	2,4	1,7
	società violenta	7	17	24	0,8	1,8	1,3
	violenza sulle donne		11	11	0,0	1,2	0,6
	<i>non codificata</i>	1	6	7	0,1	0,6	0,4
Tot. categoria		17	57	74	1,9	6,0	4,0
EVENTI TRAGICI	morte	20	5	25	2,2	0,5	1,4
	incidenti stradali	4	1	5	0,4	0,1	0,3
	perdere persone care	4	1	5	0,4	0,1	0,3
	<i>non codificata</i>	4	1	5	0,4	0,1	0,3
Tot. categoria		32	8	40	3,6	0,8	2,2
NULLA O TUTTO	di nulla in particolare	22	28	50	2,4	3,0	2,7
	non so	10	10	20	1,1	1,1	1,1
	di tutto	4	3	7	0,4	0,3	0,4
	<i>non codificata</i>	4	6	10	0,4	0,6	0,5
Tot. categoria		40	47	87	4,4	5,0	4,7
Totale risposte valide		899	949	1.848	100,0	100,0	100,0

Tabella 2 (segue): Paure per il futuro per anno di indagine. Valori assoluti e percentuali sul totale risposte valide in ciascun anno

Alle paure legate alla vita lavorativa e alla situazione economica si affiancano timori più universali, soprattutto riferiti alla salute. Questo tipo di risposte sono citate più spesso nel 2004, anno in cui rappresentavano il 25% del totale rispetto al 2009, quando raggiungono il 14,3%.

Rilevanti anche le paure legate al futuro, soprattutto quello dei figli, che come abbiamo visto sono menzionate soprattutto dalle donne che non lavorano. Vi sono poi le paure legate alla vita personale, tra cui la paura della solitudine o della vecchiaia, o di dover rimandare troppo a lungo per farsi una famiglia. Anche in questo caso si tratta di paure citate con maggior frequenza nel 2004 (8,2% del totale delle risposte rispetto al 5,1% del 2009). Decisamente più frequenti nel 2004 rispetto al 2009 i timori nei confronti della situazione politica internazionale, che passano dal 7,1% all'1,8% del 2009. Alcune intervistate hanno citato infine paure riferite alla società e al suo degrado culturale e ambientale: i riferimenti a questo tipo di paure sono leggermente più frequenti nel 2009 rispetto al 2004 (rispettivamente dal 3,9% al 4,7% del 2009).

Un altro timore decisamente in crescita nel 2009 è quello per la criminalità (dall'1,9% del 2004 al 6% del 2009) in particolare per la violenza sulle donne, paura questa probabilmente alimentata dai drammatici episodi di cronaca che hanno caratterizzato l'inizio del 2009.

Infine, le intervistate citano la paura di trovarsi davanti a eventi tragici come la morte o la perdita di persone care: come per molte delle paure più universali anche in questo caso questa categoria è citata con maggior frequenza nel 2004 (con rispettivamente il 3,6% delle risposte nel 2004 e lo 0,8% nel 2009).

Vi è poi un piccolo gruppo di intervistate che sia nel 2004 sia nel 2009 non ha saputo individuare nessuna paura in particolare (il 4,4% nel 2004 e il 5% nel 2009).

Nel 2009 sono soprattutto le intervistate più giovani a fornire una risposta legata alla dimensione lavorativa: quasi la metà delle intervistate fino a 30 anni fa riferimento a timori per il futuro, soprattutto in considerazione della propria esperienza di lavoratrice precaria (Tab. 3). Le donne meno giovani invece citano più spesso timori legati alla contingente situazione di crisi economica. Sono soprattutto le donne con più di 40 anni a temere i problemi di salute o a mostrare preoccupazioni per il futuro dei propri figli.

<i>Macro categoria</i>	<i>fino a 30 anni</i>	<i>da 31 a 40 anni</i>	<i>da 41 a 60 anni</i>	<i>Totale</i>
Lavoro	48,3	25,2	15,0	29,6
Economia	16,6	19,5	20,9	19,0
Malattie	5,4	13,1	23,0	13,8
Futuro	5,1	15,8	13,2	11,4
Criminalità	6,9	4,6	5,8	5,8
Personalì	3,9	3,6	7,1	4,9
Ambiente e società	3,0	5,8	4,9	4,6
Politica	1,5	1,5	2,1	1,7
Eventi tragici	0,6	0,9	0,9	0,8
Nulla in particolare	4,5	7,0	2,8	4,8
Nessuna risposta	3,9	3,0	4,3	3,8
<i>Totale</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>

*Tabella 3: Macro categorie sul tipo di paure per il futuro per classe di età (valori percentuali).
Indagine 2009*

Se si guarda alla condizione occupazionale si nota che sono soprattutto le donne in cerca di occupazione o le inattive a citare i timori legati alla sfera lavorativa. Le disoccupate sono anche le più preoccupate per la crisi economica e le difficoltà economiche che ne possono conseguire. La paura per la criminalità è diffusa soprattutto tra le donne che non lavorano, mentre le paure legate alla sfera personale o a quella politica sono leggermente più frequenti tra le occupate (Tab. 4).

Macro categoria	occupate	disoccupate	Inattive	Totale
Lavoro	25,7	47,0	30,1	29,6
Economia	18,9	23,8	14,4	19,0
Malattie	14,9	6,6	15,8	13,8
Futuro	11,8	7,9	13,0	11,4
Criminalità	5,5	2,0	11,0	5,8
Personalì	5,4	4,6	2,7	4,9
Ambiente e società	5,1	3,3	3,4	4,6
Politica	2,2	0,0	1,4	1,7
Eventi tragici	0,9	0,0	1,4	0,8
Nulla in particolare	5,7	1,3	4,1	4,8
Nessuna risposta	4,1	3,3	2,7	3,8
<i>Totale</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>

Tabella 4: Macro categorie sul tipo di paure per il futuro per condizione occupazionale (valori percentuali). Indagine 2009

3.3. Un confronto tra domanda chiusa e domanda aperta

Infine, un'interessante verifica del risultato ottenuto con l'analisi testuale si può ottenere confrontando per il 2009 le risposte ottenute con la domanda aperta sulle paure per il futuro con quelle ottenute e a una successive domanda sempre relativa al tema delle paure, ma in cui le modalità di risposta erano predefinite. Anche nel caso della domanda chiusa la disoccupazione, il timore di non riuscire a mantenere il tenore di vita attuale e non riuscire a far fronte ai debiti sono paure dichiarate da una intervistata su due (Fig. 3).

Tuttavia vi sono altri item che nella risposta aperta sembravano più marginali e che invece sono indicati da un numero più elevato di intervistate quando si tratta di effettuare una scelta tra un numero limitato di modalità. È il caso ad esempio della paura per la diffusione della criminalità, menzionata dal 31% delle intervistate che hanno risposto alla domanda chiusa. Nella precedente domanda aperta soltanto il 6% aveva fornito una risposta riconducibile al tema della criminalità e della violenza sulle donne.

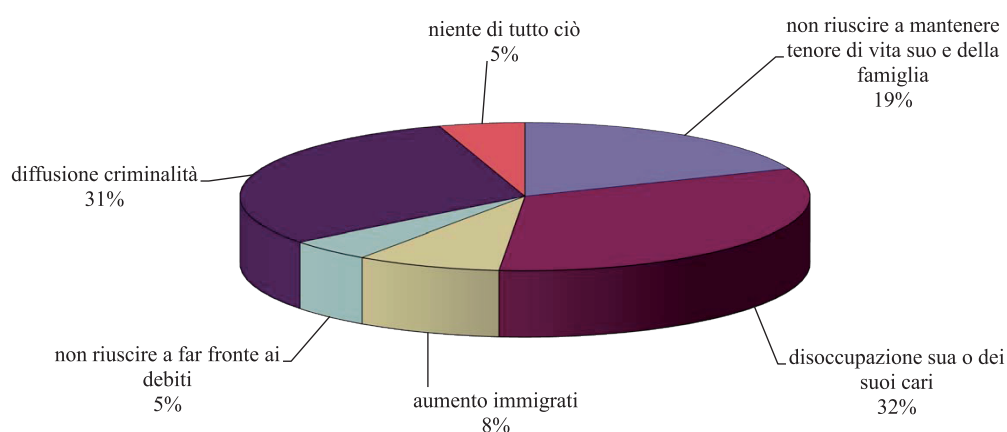


Figura 3: Paure per il futuro: risposte alla domanda chiusa - 2009 (valori percentuali)

Si tratta di un risultato soltanto apparentemente sorprendente e che può essere ricondotto al differente compito cognitivo richiesto all'intervistato quando risponde a una domanda risposta chiusa o aperta: infatti, mentre le domande chiuse richiedono un "recognition task" (Schwarz

and Hippler, 1991: 44), le domande aperte comportano per l'intervistato un "fre-recall task", con una differenza analoga a quella che intercorre tra "riconoscere e rievocare" (Gobo, 1997: 93). Questa differenza risulta in genere evidente quando si confrontano le risposte a domande che richiedono all'intervistato di scegliere all'interno di una lista predefinita di *items* con quelle ottenute con la stessa domanda lasciata libera (Schuman and Presser, 1979), come avvenuto nel nostro caso per il tema della criminalità, significativamente più "attraente" quando viene proposto in una lista di item rispetto a quanto non lo sia nella "rievocazione" spontanea delle intervistate.

4. Conclusioni

L'esperimento di codifica automatica qui condotto ha consentito di confermare uno dei risultati principali emerso nel corso dell'indagine sui bisogni e le aspettative delle donne nei confronti del lavoro: il lavoro assume una centralità sempre maggiore nella vita delle donne, tanto da occupare una posizione di primo piano anche in relazione alle preoccupazioni verso il futuro. Perdere il lavoro o restare a lungo disoccupata, la crisi economica in atto e le sue ricadute sulla situazione familiare insieme alla maggiore diffusione della violenza sulle donne sono le paure che agitavano le intervistate nella prima parte del 2009.

Oltre all'interesse del risultato in sé è utile soffermarsi sull'enorme potenzialità dalle tecniche di text mining per il trattamento semiautomatico di domande a risposta aperta. La possibilità di ottenere una nuova variabile che consente di effettuare conteggi sul contenuto delle risposte libere costituisce sicuramente un'opportunità che prefigura prospettive di analisi innovative.

La tradizione statistica considera le domande a risposta aperta uno strumento sconsigliabile e poco applicabile nel caso di survey di grandi dimensioni («Le domande aperte sono facili da porre, difficili da rispondere, e ancor più difficili da esaminare», Oppenheim, 1966: 113). Le tecniche di text mining costituiscono invece un interessante passo in avanti che può indurre a *ripensare* il ruolo stesso che le domande aperte occupano nell'ambito della ricerca sociale. Queste tecniche consentono allo stesso tempo di ridurre l'onere dell'analisi di un elevato numero di risposte, e di conservare molta più informazione rispetto alla semplice identificazione di categorie realizzata nelle fasi di post-codifica manuale.

Riferimenti bibliografici

- Altieri G., della Ratta-Rinaldi F., Dota F. and Ferrucci G. (2009). *Donne e lavoro. Rapporto di ricerca 2009*. Ires. Mimeo.
- Bolasco S. (1999). *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*. Roma: Carocci.
- Bolasco S. and Pavone P. (2010). Automatic dictionary and rule-based systems for extracting information from text. In Palumbo, F., Lauro, C.L. and Greenacre, M.J., editors, *Data Analysis and Classification*, London: Springer, pp. 189-198.
- Bolasco S., D'Avino E. and Pavone P. (2007). Analisi dei diari giornalieri con strumenti di statistica testuale e text mining. In ISTAT, *I tempi della vita quotidiana. Un approccio multidisciplinare all'analisi dell'uso del tempo*. Collana Argomenti n. 32. Roma.
- 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, pp.133-152.

- Gobo G. (1997). *Le risposte e il loro contesto. Processi cognitivi e comunicativi nelle interviste standardizzate*. Milano: Franco Angeli.
- Lebart L. and Salem A. (1994). *Statistique textuelle*. Paris: Dunod.
- Oppenheim A.N. (1966). *Questionnaire Design and Attitude Measurement*. New York: Basic Books.
- Pitrone M.C. (1984). *Il sondaggio*. Milano: Franco Angeli.
- Schwarz N. and Hippler H.J. (1991). *Response Alternative: The Impact of Their Choice and Presentation Order*. In Biemer, P., Groves, R., Lyberg L., Mathiowetz, N. and Sudman, S., editors, *Measurement Errors in Surveys*, New York: J. Wiley & Sons, pp. 41-56.
- Schuman H. and Presser S. (1979). The Open and Closed Question. *American Sociological Review*, XLIV, No. 5.