

Mi dici con parole tue il lavoro che fai?

Modelli di text mining per l'analisi delle risposte aperte

Furio Camillo, Silvia Ghiselli

Consorzio Interuniversitario ALMALAUREA - Università di Bologna

Riassunto. Nell'indagine 2004 di ALMALAUREA sulla condizione occupazionale è stata inserita, per il momento in via sperimentale e su un campione di rispondenti, una domanda aperta relativa alla descrizione minuziosa dell'attività lavorativa svolta. Il testo è stato trattato con alcune moderne tecniche di text mining, che consentono di decifrare in modo automatico il contenuto semantico delle espressioni raccolte. Nel presente lavoro sono messe in evidenza, mediante un approccio di "tipo kernel" di analisi dei dati, le principali relazioni fra il percorso dei laureati (in termini di caratteristiche personali, curricolari e di aspettative all'uscita) e le risposte aperte fornite.

Parole chiave: Descrizione, Attività lavorativa, Text mining, Cluster, Aspettative, Curriculum, Database

1. Introduzione

Recenti studi, diversamente correlati al mondo del lavoro giovanile, hanno fatto ricorso a tecniche di natura qualitativa d'indagine, al fine di analizzare gli aspetti meno tangibili, di natura quasi immaginaria, relativi al lavoro svolto o desiderato¹.

Si tratta di studi che trovano numerose occasioni di diffusione (anche sugli strumenti di comunicazione di massa) e che per questo, molte volte, fanno parte di percorsi di indagine più ampi, sulla condizione giovanile o sulle dinamiche del mercato del lavoro, a volte con uno spiccato orientamento alla comunicazione, piuttosto che alle modalità tipicamente scientifiche. Il bisogno di approfondimenti qualitativi deriva dalla necessità di "decodificare" i segnali apparentemente deboli provenienti dalla relazione tra mondo giovanile e mondo del lavoro, al fine di migliorarne la gestione da parte di tutti gli operatori coinvolti nei processi che trasformano le competenze acquisite in professionalità. In questi ambiti l'analisi è sviluppata utilizzando il più possibile l'informazione "grezza", rilevata direttamente sul soggetto.

¹ In via esemplificativa si vedano: <http://www.futureconceptlab.com> e <http://www.iard.it>

L'indagine 2004 di ALMALAUREA sulla condizione occupazionale dei laureati ha approfondito, per la prima volta ed in via sperimentale, questo particolare filone di ricerca. Come è noto, l'indagine sull'occupazione di ALMALAUREA è condotta annualmente con l'obiettivo di indagare il percorso formativo e lavorativo compiuto dai laureati dopo il conseguimento del titolo universitario; alla classica struttura di rilevazione (con domande che, conservando nel tempo la medesima formulazione, permettono di rilevare l'evoluzione dei principali fenomeni legati al mercato del lavoro) si affiancano di anno in anno alcune domande che hanno come obiettivo quello di approfondire particolari temi di analisi (generalmente di attualità, di interesse scientifico o di ricerca). Nell'ultima indagine, quella condotta nel 2004, è stata inserita una specifica domanda relativa alla descrizione testuale dell'attività lavorativa svolta.

L'obiettivo generale è quello di integrare le informazioni rilevate (con domande chiuse) circa la posizione professionale e il contratto di lavoro, con una descrizione puntuale delle mansioni e delle funzioni di cui i laureati si avvalgono nel proprio lavoro. Inoltre, attraverso la raccolta di questo tipo di informazione, ovviamente sintetizzata in categorie semanticamente omogenee di lavori, si potrebbe integrare ed arricchire il curriculum vitae dei laureati, a corollario descrittivo del lavoro svolto e delle modalità soggettive della sua "narrazione".

2. Le caratteristiche della rilevazione ALMALAUREA

La documentazione utilizzata nel presente paper fa riferimento all'indagine del 2004², che ha coinvolto tutti i laureati pre-riforma della sessione estiva degli anni 2003, 2001 e 1999, intervistati rispettivamente a uno, tre e cinque anni dal conseguimento del titolo. La rilevazione, condotta con metodologia CATI, si è svolta tra settembre e novembre 2004³ e ha coinvolto quasi 56.000 laureati di 27 Atenei italiani: la partecipazione dei laureati è stata particolarmente ampia, tanto che il tasso di risposta ha raggiunto nel complesso ben l'82% (in particolare, 86% ad un anno, 81% a tre anni, 76 a cinque anni).

Nel questionario somministrato al collettivo del 2003 (composto da 23.459 laureati), in particolare, è stata introdotta una domanda aperta, con l'obiettivo di rilevare, in modo preciso e particolareggiato, l'attività lavorativa svolta dagli occupati, così da qualificarne ulteriormente la condizione professionale. È noto infatti che le domande chiuse, se da un lato agevolano la risposta dell'intervistato, facilitandone la riflessione e il ricordo, dall'altro ne limitano inevitabilmente la libertà di espressione.

In particolare questo è vero per le domande chiuse relative alla condizione professionale degli intervistati che, non potendo racchiudere tra le alternative di ri-

² Tutta la documentazione è disponibile all'indirizzo <http://www.almalaurea.it/universita/occupazione>.

sposta tutte le molteplici sfaccettature dell'attuale mercato del lavoro, finiscono inevitabilmente per assimilare posizioni lavorative in parte differenti. Ad esempio, all'interno della categoria "impiegato ad alta/media qualificazione" rientrano tutti coloro che formalmente ricoprono questo ruolo, indipendentemente dal livello di autonomia e dalle competenze richieste.

Queste considerazioni hanno portato all'inserimento, in via sperimentale, del quesito aperto "Quali sono i compiti principali che svolge nel suo lavoro? Dovrebbe descrivere le mansioni effettivamente svolte, cercando di non utilizzare termini generici come impiegato di concetto o collaboratore tecnico". La domanda è stata sottoposta ad un campione di occupati, selezionati attraverso un campionamento stratificato proporzionale per genere, area geografica di residenza, area geografica dell'ateneo e gruppo disciplinare³. Il processo di campionamento ha portato all'estrazione di un campione di 6.441 laureati; di questi hanno risposto alla domanda aperta, in quanto occupati ad un anno, 2.984 persone.

3. La metodologia utilizzata

Le risposte rilevate sono state trattate attraverso alcune tecniche di text-mining al fine di ottenere, dall'ampia mole di informazioni testuali a disposizione, una classificazione automatica rappresentativa dei principali concetti legati alla posizione lavorativa. Il primo obiettivo della ricerca testuale è stato infatti quello di classificare per gruppi tematici le informazioni rilevate e le relative correlazioni, a volte non prevedibili, ma di elevato contenuto informativo. Gli strumenti di analisi testuale adottati sono stati utilizzati proprio in un contesto oramai divenuto classico per la cosiddetta *business intelligence* analitica, dove la rilevazione di informazioni "qualitative" o testuali su un campione rappresentativo di soggetti permette la "colorazione" di un intero data base, ovvero l'estensione a tutto il data base di alcuni risultati sintetici (ad esempio l'appartenenza ad un Cluster) elaborati sulla base di informazioni raccolte sul campione.

In particolare, si è fatto uso di una delle più moderne metodologie di text mining che, mediante specifici algoritmi, consente di decifrare in modo automatico il contenuto semantico delle espressioni raccolte. Il software usato, denominato "TWID EXPert", è progettato e realizzato da Synthema: esso rappresenta una soluzione in-

³ Nella popolazione di riferimento sono stati individuati 475 strati, ognuno dei quali ha subito un processo di estrazione casuale di una frazione fissa di casi, pari al 25% circa, ipotizzando una varianza teorica interna agli strati costante, dal momento che la variabile da rilevare (descrizione della professione svolta) è una variabile espressa in forma testuale, libera. Gli strati con ridotta numerosità sono stati campionati per intero.

novativa per estrarre le informazioni più rilevanti da un insieme di documenti testuali e che permette, tra l'altro, di navigare su tali documenti sulla base dei concetti estratti, ovvero delle parole chiave e dei relativi contesti individuati.

Il motore linguistico utilizzato combina l'analisi morfologica a quella statistica: da un documento vengono estratti come significativi solo quei termini - parole singole o espressioni composte - che rispettano sia le regole morfologiche definite, sia le soglie di significatività e frequenza. Ogni termine viene ricondotto alla sua forma base, ovvero l'infinito per i verbi e la forma singolare per i nomi e gli aggettivi, registrandone statisticamente l'occorrenza. Ad esempio, definendo la regola morfologica "nome+base(nome)", entrambe le espressioni "capi stazione" e "capo stazione" vengono ricondotte alla stessa forma base "capo stazione". TWID EXpert permette inoltre di navigare sulle fonti documentali in modo semplice ed intelligente, di esportare i risultati dell'analisi su database, su file XML o su file Excel compatibili. In questo modo è possibile creare grafici e applicare sui concetti estratti modelli statistici complessi finalizzati alla spiegazione, mediante variabili esogene, delle relazioni semantiche rinvenute nei testi (Elia, Vietri, 1993; Raffaelli, 2000).

4. L'elaborazione dei testi e la classificazione morfologica delle risposte⁴

I testi raccolti erano costituiti per lo più da frasi ellittiche del verbo, o costrutti sintattici difficilmente riconducibili ad espressioni comuni. L'approccio più efficace è sembrato dunque quello morfologico-statistico, con riconoscimento di pattern morfologici più o meno complessi: "nome+aggettivo", "aggettivo+nome", "nome+preposizione+nome", ecc..

Questa scelta è infatti apparsa più adeguata rispetto alla costruzione di una grammatica ad hoc, che magari sarebbe dovuta transitare attraverso una lemmatizzazione complessa del materiale testuale raccolto.

Il sistema di riconoscimento morfologico ha definito, mediante confronti con i lessici tematici maggiormente legati al mondo del lavoro, circa 740 lemmi attinenti, dei quali in Tabella 1 ne viene riportato uno stralcio relativo ai primi 35, ordinati per frequenza di occorrenza nel corpo testuale complessivo.

Successivamente è stata ottenuta una classificazione delle risposte aperte grazie all'applicazione di un algoritmo di clustering non-gerarchico delle nuvole dinamiche aggiustato e gerarchizzato sulle forme stabili, applicato a una matrice lessicale *risposte*×*forme grafiche*. Il clustering adottato in effetti ha utilizzato la definizione

⁴ L'elaborazione dei testi si è avvalsa della collaborazione di Federico Neri, di Synthema srl.

Tabella 1 – Occorrenze nel corpo testuale: i primi 35 lemmi attinenti (a fianco di ciascun lemma è riportata la posizione e la relativa occorrenza)

posiz.	lemma	occorrenza	posiz.	lemma	occorrenza	posiz.	lemma	occorrenza
1	progettazione	1,93	13	consulenza	0,95	25	farmacista	0,59
2	gestione	1,92	14	vendita	0,85	26	progettista	0,58
3	lavoro	1,83	15	ricerca	0,83	27	organizzazione	0,58
4	insegnante	1,69	16	servizio	0,75	28	relazione	0,53
5	azienda	1,64	17	studio	0,74	29	consulente	0,53
6	scuola	1,57	18	tecnico	0,73	30	assistenza	0,53
7	responsabile	1,38	19	contabilità	0,73	31	addetto	0,53
8	ufficio	1,29	20	controllo	0,70	32	sviluppo	0,51
9	educatore	1,21	21	banca	0,64	33	prodotto	0,51
10	attività	1,20	22	contatto	0,62	34	assistente	0,51
11	cliente	1,13	23	analisi	0,62	35	operatore	0,50
12	progetto	1,09	24	personale	0,59			

semantica di “lemma attinente” in quanto ciascuno dei 100 lemmi più ricorrenti è stato utilizzato come seme iniziale del processo di aggregazione non gerarchico. L’incrocio di 5 partizioni diverse ha infine dato luogo alle forme stabili sulle quali è stato lanciato un clustering gerarchico a riduzione massima di varianza interna di tipo Ward e che ha prodotto la classificazione gerarchica finale (Bolasco, 1999).

La decisione del taglio della gerarchia a 20 Cluster è stata ottenuta sia studiando l’andamento del dendrogramma, sia osservando la sequenza degli pseudo-F associati a ogni passo dell’aggregazione. In realtà il taglio a 20 Cluster è un “ottimo locale”, in quanto il taglio ottimale avrebbe condotto all’individuazione di tre Cluster di testi. Tuttavia tale taglio è sembrato immediatamente troppo grossolano e comunque privo sia di capacità interpretativa che di rappresentatività delle risposte fornite, risultando pertanto carente del livello di analiticità richiesto (si ricordi che uno degli obiettivi di questo lavoro è quello di arricchire il data-base di ALMALAUREA grazie alla classificazione delle descrizioni della professione). La valutazione del χ^2 sulla matrice di frequenze collassata ha inoltre permesso di valutare ex-post la robustezza dei risultati ottenuti e quindi di procedere alla descrizione il contenuto semantico di ogni Cluster di risposte aperte fornite.

5. I Cluster individuati

Ciascuno dei 20 Cluster semantici individuati è descritto da una serie di parole chiave. L’ordine di presentazione delle parole nell’output non è casuale, bensì rispecchia il proprio grado di capacità descrittivo-caratterizzante del Cluster, in accordo a preci-

se misure di specificità lessicale legate alle metriche del χ^2 , che sono state applicate alla tavola *cluster*×*forme grafiche*. La descrizione completa di ciascun Cluster è riportata in Appendice, nella Tabella I, in corrispondenza della colonna “Descrizione lavoro”.

In questa sede pare opportuno, invece, fare qualche approfondimento sulla capacità descrittiva di alcuni Cluster realizzati: il Cluster 6, ad esempio, pare essere fortemente connotato da mansioni legate al mondo medico e paramedico, inerente però l’ambito farmacologico (comprese le attività commerciali).

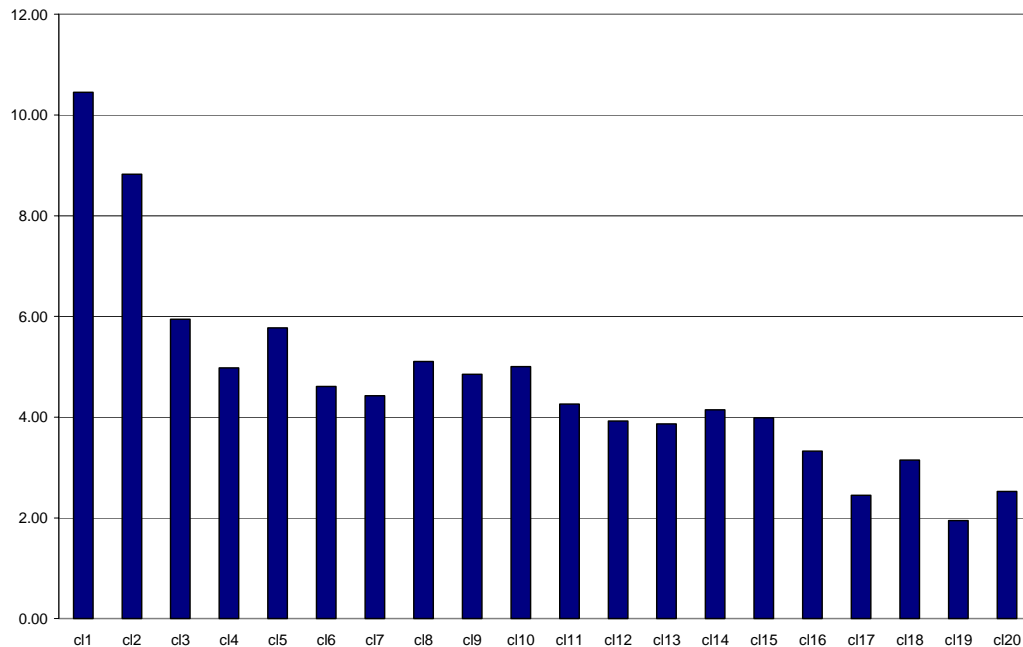
A prima vista, nell’insieme di parole connotanti il Cluster 6, sembrano dissonanti le forme “commessa” o “interprete”; in realtà, il sistema ha perfettamente riconosciuto due aspetti della commercializzazione dei prodotti farmaceutici, oggetto delle transazioni seguite e curate dai laureati. In particolare, la parola “commessa” fa riferimento alle mansioni di venditrice (quindi l’attività in farmacia, al banco). Inoltre, dal momento che numerose aziende farmaceutiche sono in contatto anche con società straniere, ai laureati del Cluster 6 è richiesta con una certa ricorrenza anche una competenza linguistica (di interprete), necessaria nei contatti diretti con clienti o fornitori all’estero.

Il Cluster 6, inoltre, presenta alcuni punti in comune con il successivo Cluster 7, poiché in entrambi i Cluster sono presenti termini che si riferiscono alla professione in ambito farmaceutico. Però, mentre il primo Cluster è connotato in particolare dalla parola “vendita” (con alcune peculiarità che rimandano alla figura dell’informatore scientifico del farmaco), il secondo è caratterizzato soprattutto dal concetto di progettazione. È probabile che all’interno di quest’ultimo Cluster si ritrovino anche gli addetti al banco della farmacia, ma è diversa la percezione che i laureati hanno della propria posizione, caratterizzata da aspetti diversi dalla pura “vendita”, che in questo Cluster non viene assolutamente menzionata.

La Figura 1 riporta i pesi dei 20 Cluster di risposte aperte individuati, misurati semplicemente mediante la percentuale di intervistati che, in base alla propria risposta, afferisce al Cluster. Appare evidente dal colpo d’occhio sul grafico che, fatta eccezione per i primi due Cluster, esiste una sostanziale uniformità nella distribuzione dei pesi, compresi generalmente fra il 3% e il 6%.

La classificazione morfologica delle risposte sembra dunque aver individuato Cluster molto specifici, i quali sono stati successivamente descritti grazie alle numerose variabili rilevate da ALMALAUREA e mediante le usuali tecniche di descrizione delle caratteristiche esogene connotanti i gruppi.

La descrizione, riportata in Appendice nella Tabella I, è stata effettuata grazie a un’applicazione contenuta nel software SPAD che permette di individuare le caratteristiche distintive di una data variabile nominale (Bolasco, 1999). Gli elementi caratterizzanti (intesi come modalità di un insieme di variabili) sono classificati secondo un test statistico che si distribuisce come una normale standardizzata: più il valore

Figura 1 – Composizione percentuale dei Cluster morfologico-semantiche individuati (in termini di rispondenti)

del test è elevato (come è noto, se $\alpha=5\%$ deve essere superiore a +2) più la probabilità ad esso associata è significativa e più, conseguentemente, l'elemento che si sta analizzando è in grado di caratterizzare il Cluster. Se il valore del test associato all'elemento che si sta valutando è inferiore a -2 (sempre per $\alpha=5\%$) allora risulta altrettanto significativo, ma nel "non" caratterizzare il Cluster di riferimento.

La descrizione dei Cluster è avvenuta utilizzando sia variabili nominali che variabili continue. Data una variabile nominale per la quale si voglia appurare se risulti caratterizzante un gruppo k di individui, si supponga di osservarne una qualunque modalità j . Nell'ipotesi di indipendenza, gli n individui del gruppo k , ovvero n_k , saranno stati estratti casualmente fra le n osservazioni e pertanto:

$$\frac{n_{kj}}{n_k} = \frac{n_j}{n}$$

dove n_{kj} indica il numero di individui del gruppo k per i quali si osserva la modalità j ; n_k indica il complesso gli individui del gruppo k ; n_j indica il numero di individui della popolazione in corrispondenza dei quali si è osservata la modalità j .

Nel caso in cui la variabile per la quale si vuole appurare la caratterizzazione del gruppo sia continua, allora si confronta la media del gruppo k con la media generale.

La descrizione dei 20 Cluster (riportata in modo esaustivo in Appendice nella colonna “Descrizione caratteristiche dei soggetti del Cluster” della Tabella 1) è risultata utile per completare la descrizione dell’attività lavorativa con l’insieme delle caratteristiche dei laureati (individuali, curricolari e relative alle aspettative). La capacità descrittiva di queste variabili è sottolineata (nelle descrizioni riportate in tabella) da avverbi come “probabilmente”, “verosimilmente”, che approssimano il livello di significatività, calcolato come detto sopra e associato alla modalità della variabile che si sta descrivendo. Una lettura attenta di queste schede non può esimersi dal considerare anche questi particolari elementi.

Dall’analisi di questi gruppi possono essere dedotte significative analisi comparative sulle caratteristiche dei Cluster semanticamente simili in base alla descrizione del lavoro svolto. In tal senso ci sembra utile soffermare l’attenzione sulle differenze fra il Cluster 2 e il Cluster 4. Tali differenze sono sottili ma interessanti, e percepibili con evidenza dalla lettura delle forme connotanti; in particolare, mentre il Cluster 2 è caratterizzato soprattutto dalla parola “insegnante” (primo elemento descrittivo), il Cluster 4 preferisce definirsi “educatore”. Nel Cluster 4 rientrano dunque mansioni che non sono propriamente legate all’insegnamento, le quali vanno a connotare il lavoro inteso anche come impegno di natura sociale o pseudo-sociale.

I due Cluster morfologico-semanticanti risultano tra l’altro diversi anche per le variabili di natura socio-demografica e legate agli atteggiamenti, probabilmente perché sono differenti le modalità con le quali i laureati dei due Cluster si collocano “psicologicamente” sul mercato del lavoro: i rispondenti del Cluster 2 ritengono di non fare un lavoro adeguato al proprio titolo di studio, sembra che insegnino, precariamente, e comunque non cercano un nuovo lavoro. I laureati del Cluster 4 invece cercano un nuovo lavoro, pur svolgendo mansioni che sembrano dinamiche ed interessanti, ma scalpitano e probabilmente vivono la condizione di precariato (anche reddituale) come una sorta di spinta propulsiva per proporsi al mercato del lavoro in modo forte e competente.

6. Professioni raccontate, caratteristiche dell’individuo e aspettative alla laurea

L’analisi delle caratteristiche dei soggetti, segmentati nei diversi Cluster semantici, mostra come i testi descrittivi della professione svolta dipendano da una molteplicità di elementi. Più in generale, la descrizione dipende da questioni sia soggettive che oggettive, ossia sia dalla percezione che i soggetti hanno della propria collocazione nello specifico posto di lavoro oltre che, ovviamente, da una serie di condizioni oggettive che si trovano a vivere.

Abbiamo quindi sviluppato una serie di studi delle relazioni della classificazione testuale con le caratteristiche dell'individuo, considerandole non tanto a livello di singolo Cluster (operazione di fatto contenuta in termini analitici già nelle schede proposte in appendice), quanto mediante strumenti che consentano una visione più globale delle relazioni incrociate. Di seguito, si farà cenno a tre categorie di studi eseguiti a questo proposito.

Occorre comunque sottolineare che l'importanza operativa di questo sforzo è legato alla caratteristica stessa della banca dati di *ALMALAUREA* che, costituendo di fatto un vero e proprio panel (i laureati sono contattati fino a 5 anni dal conseguimento del titolo), consente di misurare una serie di aspetti e di fenomeni legati in generale al mondo dei giovani e al rapporto fra il loro passato studentesco e il presente lavorativo. In particolare, questo rapporto determina, insieme al processo formativo intrapreso (ovvero al trattamento cui i giovani sono stati sottoposti), performance individuali diverse le quali, entrando a far parte comunque degli elementi di valutazione del successo di un corso di laurea, di una facoltà o di un ateneo, possono risultare strategici per gli addetti ai lavori nei processi di erogazione del servizio formativo e per la valutazione dell'efficacia esterna del sistema universitario.

Una prima relazione è individuabile fra gruppo disciplinare e Cluster di testi morfologico-semantici. Nelle figure di seguito proposte (Figure 2 e 3, lasciate distinte solo per chiarezza grafica) è stata rappresentata una mappa fattoriale delle corrispondenze binarie, sintesi della matrice di frequenza 20 righe (i Cluster di lavori descritti) per 15 colonne (i gruppi disciplinari). La forma delle nuvole e i diagnostici collegati ai grafici fattoriali mostrano una dipendenza elevata dei due insiemi. In particolare, è interessante notare come il linguaggio e la costruzione logica usati per la descrizione delle mansioni siano messi in relazione, sull'asse orizzontale, ad una sorta di dicotomia servizio-materia (da sinistra a destra), relativamente al "tipo di oggetto" che viene prodotto mediante la propria attività lavorativa.

In termini di percorsi di studio, invece, si ritrovano verso sinistra le lauree dei gruppi psicologico, insegnamento, letterario e linguistico, mentre verso destra quelli chimico-farmaceutico, architettura e ingegneria. Si ricordi che le mappe devono essere lette in termini di distanza. Più due punti sono vicini e maggiore sarà la corrispondenza fra di essi: se si tratta di due righe, le due righe hanno un profilo molto vicino, e lo stesso vale per le colonne.

Il fattore verticale sembra invece più difficile da battezzare, se non mediante la raffigurazione delle corrispondenze righe-colonne, ossia testi-gruppi disciplinari.

Un'altra serie di elementi che indicano quanto l'espressione testuale sia legata alla provenienza sociale degli individui può essere dedotta dalla Tabella 2, nella quale è sintetizzata la relazione fra i Cluster di descrizione del lavoro e la classe sociale dei genitori. Risulta evidente un certo grado di connessione, confermato da alcune

misure della connessione come il χ^2 e il Lambda di Goodman-Kruskal usualmente adoperati su tavole di questo tipo.

Figura 2 – Analisi delle corrispondenze: l'insieme delle righe

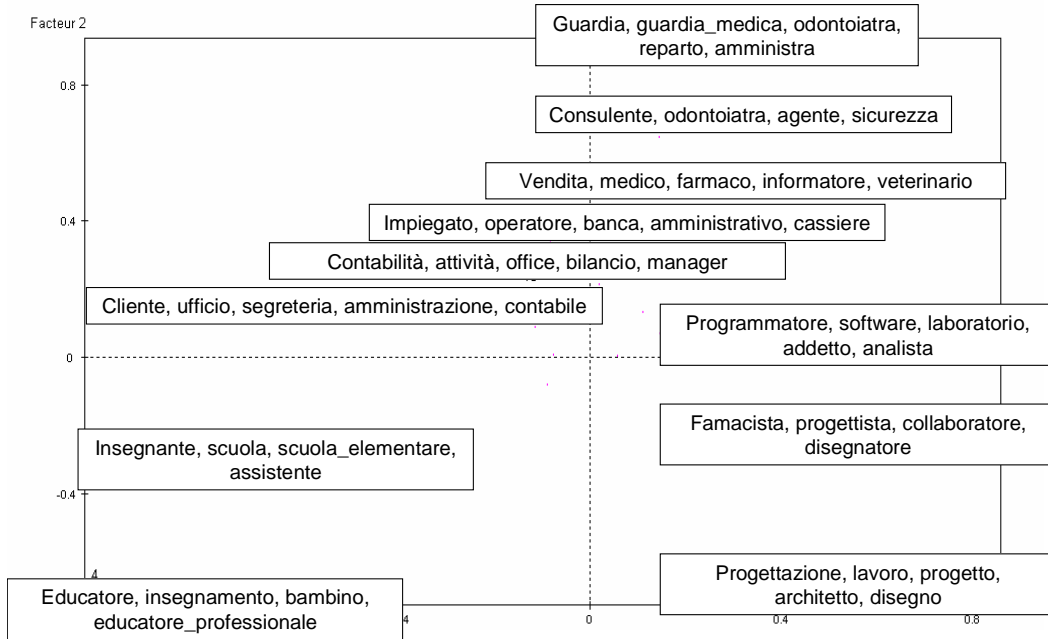


Figura 3 – Analisi delle corrispondenze: l'insieme delle colonne

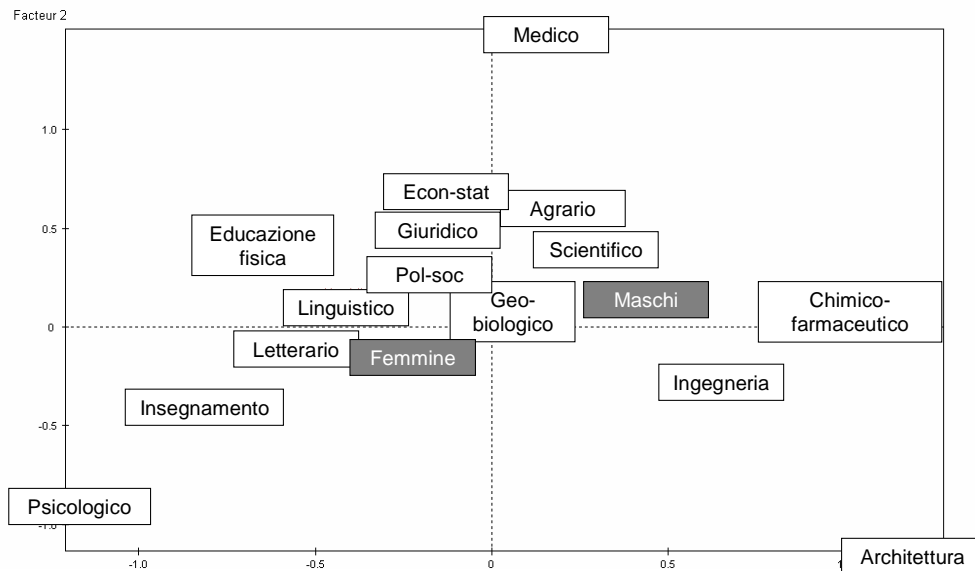


Tabella 2 – Cluster morfologico-semantiche e classe sociale del genitore
(sono evidenziate le celle superiori al profilo medio)

	Borghesia	Classe media impiegatizia	Piccola borghesia	Classe operaia	Non disponibile	Totale
Cluster 1	34,2	21,7	12,5	12,9	18,8	100,0
Cluster 2	20,5	25,6	12,0	11,6	30,2	100,0
Cluster 3	27,0	26,4	16,9	15,2	14,6	100,0
Cluster 4	19,9	31,3	13,1	17,0	18,8	100,0
Cluster 5	28,9	27,1	14,5	15,1	14,5	100,0
Cluster 6	34,6	26,9	7,7	10,9	19,9	100,0
Cluster 7	40,5	21,6	12,4	9,8	15,7	100,0
Cluster 8	34,9	21,7	15,1	13,8	14,5	100,0
Cluster 9	29,0	26,2	13,1	10,3	21,4	100,0
Cluster 10	29,0	28,3	11,0	20,0	11,7	100,0
Cluster 11	38,5	25,2	12,6	10,4	13,3	100,0
Cluster 12	27,4	23,4	12,9	12,9	23,4	100,0
Cluster 13	32,5	27,4	15,4	10,3	14,5	100,0
Cluster 14	35,9	30,8	8,5	8,5	16,2	100,0
Cluster 15	39,8	32,7	9,7	8,8	8,8	100,0
Cluster 16	37,3	25,5	8,8	5,9	22,5	100,0
Cluster 17	34,1	20,7	14,6	14,6	15,9	100,0
Cluster 18	26,8	35,1	9,3	16,5	12,4	100,0
Cluster 19	38,2	20,6	7,4	10,3	23,5	100,0
Cluster 20	23,9	31,3	14,9	13,4	16,4	100,0
TOTALE	30,9	25,8	12,5	12,7	18,1	100,0

Un curioso esempio di lettura di questa tabella è relativo ai Cluster 7 e 16. Questi due Cluster, soprattutto il Cluster 16, hanno in comune dal punto di vista morfologico-semantiche di descrizione del lavoro (cfr. Tabella 1, in Appendice) un'ampia componente libero-professionale (“free-lance” o “consulente”), in alcuni casi legata all'appartenenza ad un albo. E proprio questi due Cluster sono quelli dove la presenza di laureati di origine “operaia” è decisamente sotto-rappresentata: 9,8% per il Cluster 7 e addirittura 5,9% per il Cluster 16. Altrettanto interessante è la presenza più consistente di genitori di classe operaia in corrispondenza del Cluster 10, dove la descrizione morfologica del lavoro svolto ruota soprattutto attorno alle parole “responsabile” e “produzione”, ossia a parole che inducono a pensare che, seppure con mansioni di responsabilità, l’“officina” e la “fabbrica” restino sempre appannaggio delle “solite” classi sociali.

Scendendo maggiormente nel dettaglio, l'utilità dell'analisi condotta è orientata dalla necessità di esplorare e diminuire la complessità delle relazioni che esistono fra le variabili che concorrono a determinare la valutazione dell'efficacia di un

percorso di studi universitario. Nell'ambito della valutazione, un modello concettuale che di recente è stato oggetto di studio e di applicazioni quantitative nell'approccio della misura del capitale umano è, ad esempio, il seguente (Crisp, 2005; Vittadini, Lovaglio, 2006):

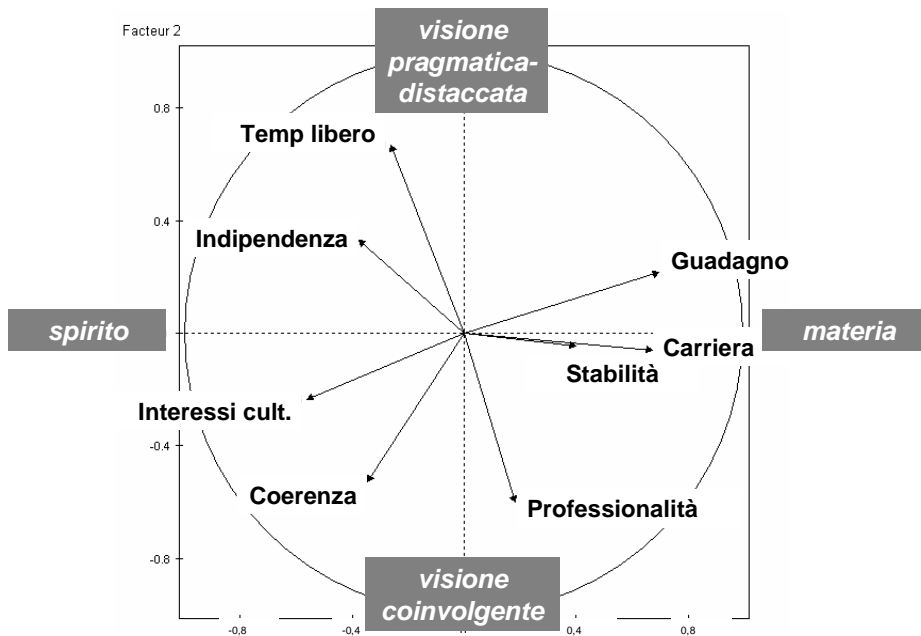
$$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$$

Componente Ateneo-Gruppo	Componente individuale	Componente comportamentale	Componente aspettative all'uscita
-----------------------------	---------------------------	-------------------------------	--------------------------------------

dove y è l'*outcome*, valutato in termini di reddito percepito dai laureati dopo cinque anni dal conseguimento del titolo. Nel seguito verrà mostrato come una delle componenti del pacchetto esogeno di variabili, in particolare la componente relativa alle aspettative espresse all'uscita dal laureato, risulti importante non solo per trovare o meno lavoro, come mostrato in altri studi (Camillo, Mignoli, Nardi, in corso di pubblicazione), ma addirittura, a parità di altre condizioni, anche per il tipo di lavoro trovato e per la sua descrizione in forma testuale libera.

La mappa riportata nella Figura 4 rappresenta la struttura delle correlazioni fra i diversi items (guadagno, carriera, stabilità e sicurezza dell'impiego, coerenza con gli studi compiuti, acquisizione di professionalità, rispondenza ai propri interessi culturali, indipendenza e autonomia sul lavoro, tempo libero) che costituiscono l'immaginario valutato dal laureando circa il lavoro desiderato e che, presumibilmente, dopo poco inizierà a cercare. Tale informazione è integrata nella banca dati ALMALAUREA ed è quindi direttamente analizzabile anche in questo contesto.

Figura 4 – Mappa delle aspettative: cerchio delle correlazioni



Più in dettaglio, la mappa è stata ricavata mediante una scomposizione in componenti principali dei giudizi depurati dall'effetto taglia e relativa ai soli laureandi cui è stata sottoposta la domanda a risposta aperta oggetto del presente lavoro.

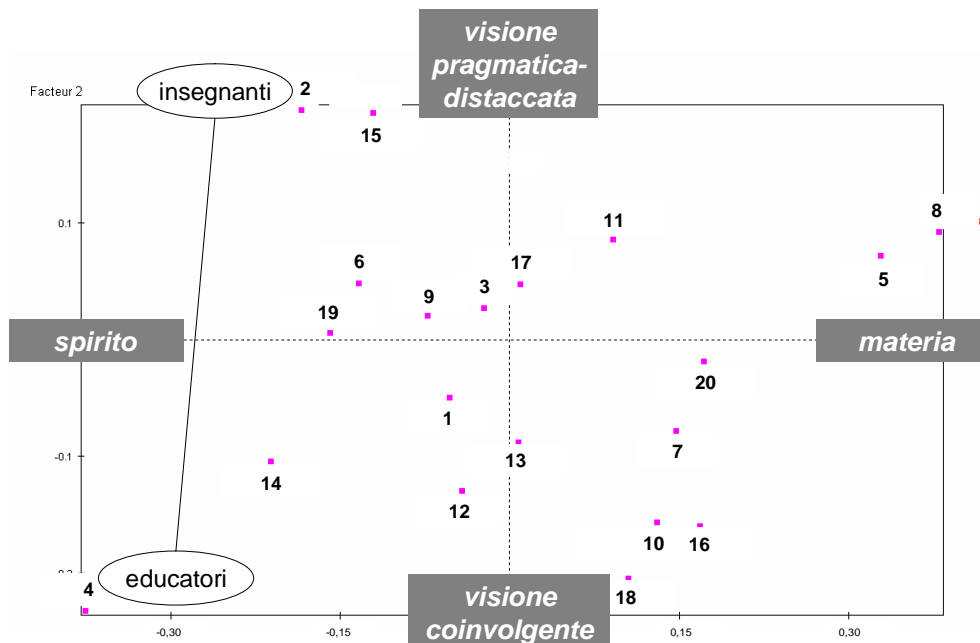
La struttura semantica degli assi fattoriali che si compongono non sembra differire molto da quella più generale usualmente riscontrata in altri studi nei quali sono stati considerati collettivi più ampi (Camillo, Ghiselli, 2005; Camillo, Mignoli, Nardi, in corso di pubblicazione).

Riprendendo come esempio la dicotomia fra gruppo semantico degli "educatori" e quello degli "insegnanti" sembra evidente, soprattutto dalla Figura 5, la relazione esistente fra l'immaginario valutato all'uscita dall'università e il tipo di lavoro svolto successivamente. In particolare nel nostro caso la dicotomia educatori-insegnanti si mostra associata a una diversa visione del lavoro ideale (asse verticale nella mappa) dimostrata all'uscita dall'università.

La relazione fra le risposte aperte e le aspettative che avevano dichiarato all'uscita della laurea sono state studiate anche da un punto di vista più analitico. A tale scopo si è partiti da una segmentazione a cinque tipologie del collettivo ALMA-LAUREA, che si basa sui giudizi d'aspettativa all'uscita e che è stata sviluppata in precedenti analisi (Camillo, Mignoli, Nardi, in corso di pubblicazione).

Riprendendo il modello della valutazione di un outcome y , sopra proposto, e considerando in questo caso come variabile dipendente l'appartenenza o meno ad

Figura 5 - Mappa delle aspettative: proiezione dei Cluster testuali



uno specifico Cluster semantico ($y = 1/0$, variabile dicotomica) è stata stimata una funzione discriminante di tipo kernel per ciascuno dei 20 Cluster:

$$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$$

Mediante processi di simulazione è stato poi possibile isolare sulla singola funzione discriminante l'influenza della sola appartenenza a ciascuno dei cinque gruppi d'aspettative sul lavoro. La Tabella 3 riporta il risultato, tradotto in termini di punteggio (procedura di scoring), di una funzione discriminante ricavata da un processo di Kernel Discriminant Analysis, aggiustata con un criterio di selezione del *best model* mediante l'indice ICOMP di Bozdogan⁵.

Tabella 3 – Modello di scoring a parità delle altre condizioni (simulazioni totali)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	
Aspetti caratterizzanti	Tempo libero	Coerenza, stabilità, interessi culturali	Carriera, stabilità, guadagno	Carriera, guadagno, professionalità	Tempo libero, guadagno, stabilità	Totale
p(4)	2	10	14	68	6	100
p(2)	12	16	21	4	47	100

Il contenuto della Tabella 3 deve essere letto nel senso delle righe, dove ogni cella rappresenta un punteggio che varia da 0 a 100 e che esprime il peso dell'appartenenza al singolo gruppo di aspettative circa il lavoro ideale. Ad esempio la simulazione mostra che aver dichiarato al momento della laurea di appartenere al Cluster 4 (caratterizzato dal ritenere più importanti della media carriera, guadagno, professionalità), ha implicato una probabilità molto alta che, successivamente, lo stesso intervistato finisse in Cluster semantico p(4), ossia quello degli "educatori", piuttosto che in quello p(2), degli "insegnanti".

Dalla tabella e da quanto detto finora risulta dunque evidente che esiste una serie di relazioni non semplici da individuare e formalizzare fra le variabili che determinano il risultato sul mercato del lavoro. E queste interrelazioni sembrano molto

⁵ La procedura di scoring permette di esplicitare in modo elementare i risultati di un'analisi di discriminazione. Lo scoring calcola di fatto una funzione di score che è una modificazione della funzione discriminante. In particolare vengono introdotte delle zone di decisione (rossa, verde e arancio o centrale) a partire da una tolleranza dell'errore di classificazione delle unità statistiche in questione. La procedura di scoring qui usata trasforma i coefficienti usando due regole (Bouroche e Saporta, 1988). La prima è che il coefficiente minimo di ogni variabile sia posto a zero. In questo modo lo score minimo possibile di un individuo è nullo. Tale score sarà ottenuto da un individuo che, per ogni variabile, possieda la modalità assegnata dal coefficiente trasformato nullo. La seconda regola è che il massimo possibile della funzione di scoring si ottiene fissandolo arbitrariamente, ad esempio come nel nostro caso, a 100. La funzione discriminante nel nostro caso è stata trovata con una strategia di data mining del tipo KDA (kernel discriminant analysis), con scelta delle opzioni di non linearità mediante l'approccio ICOMP di gestione della complessità del modello adottato (Bozdogan, 2004).

forti anche quando si mettono in gioco *outcome* apparentemente più “sfocati”, come la probabilità di descrivere, usando una risposta aperta, un certo tipo lavoro. Come in altra sede è stato messo a fuoco (Camillo, Mignoli, Nardi, in corso di pubblicazione), le condizioni socio-economiche familiari, oltre a influire sulla scelta del corso di laurea e perfino sulle scelte post-laurea, si riflettono anche sulla tipologia delle aspettative. Ma, al di là di queste valutazioni, anche l’analisi più dettagliata conferma il ruolo assolutamente significativo, non sempre tenuto sufficientemente in considerazione, che le aspettative dei laureati rivestono nella ricerca del lavoro e nella qualità effettiva e/o autopercepita del lavoro svolto dopo un anno dalla laurea.

7. Qualche conclusione

I risultati proposti individuano una stretta correlazione tra caratteristiche personali (famiglia di origine), percorso formativo intrapreso (gruppo di corso di laurea), aspettative all’uscita dal sistema universitario (caratteristiche del lavoro ideale cercato) e attività lavorativa svolta.

Dalle analisi effettuate emerge con evidenza che è possibile classificare con chiarezza gruppi di descrizione delle professioni, e che tali descrizioni risultano molto spesso correlate non solo alle effettive mansioni svolte dai laureati, ma anche alla percezione che essi hanno del proprio ruolo nel processo produttivo dell’azienda in cui sono inseriti. Tale percezione deriva in larga parte dalle esperienze e dal bagaglio formativo e culturale maturato negli anni.

Queste analisi, che come già detto sono state introdotte per la prima volta nel sistema di rilevazione di ALMALAUREA, ci spingono a ritenere che l’argomento sia attuale e di grande interesse. Questo interesse è legato non tanto all’arricchimento delle informazioni relative alla posizione nella professione e al contratto di lavoro (obiettivo che, date le finalità dell’indagine occupazionale, per noi è solo marginale), quanto all’inserimento della risposta aperta, opportunamente classificata, in un database più ampio, la cui funzione deve essere semplicemente quella di qualificare meglio i curricula dei laureati, specialmente quando questi sono messi a disposizione delle aziende con finalità di selezione del personale.

APPENDICE

Tabella I - Cluster morfologico-semantiche rinvenuti nel corpo testuale, e relativa descrizione

Cluster	Descrizione lavoro	Descrizione caratteristiche dei soggetti del Cluster ⁶
1	progettazione; lavoro; progetto; architetto; disegno; cantiere; direzione; edificio; edilizia; rilievo; studio; architettura; progettazione_architettonica; campo; ristrutturazione; esecutivo; pratica; documento; giornale; industriale; collaborazione; sicurezza; struttura; cad; civile; realizzazione; grafica; progettazione_edilizia; direzione_dei_lavori; tecnico; software; costruzione; impianto; ingegnere; progettazione_strutturale; coordinamento; interno; sviluppo; arredamento; libero; persona; progettazione_di_strutture; redazione; fase; studio_di_architettura; piano; gruppo; ambito; orientamento; computer	Gli intervistati sono con grande probabilità architetti, o al limite ingegneri; uomini, lavorano nell'industria a tempo pieno e hanno un contratto di lavoro autonomo, anche se potrebbero lavorare anche senza contratto. Probabilmente per questo motivo, non percepiscono, in genere, benefit aggiuntivi come tredicesima, buoni pasto, assicurazione sanitaria integrativa, premio di produttività. Hanno studiato al Politecnico di Torino; in misura minore, negli Atenei di Firenze, di Chieti-Pescara o in un altro ateneo del Centro Italia. È probabile che provengano dall'estero, e la loro sede di lavoro sia nel Centro Italia, forse in Toscana. Considerano la laurea molto efficace, non cercano un nuovo lavoro anche se ritengono che il guadagno percepito non sia adeguato alla posizione ricoperta. Hanno un indice di qualità per il lavoro svolto e un voto di laurea superiori alla media; sono soddisfatti del loro lavoro soprattutto per la coerenza con gli studi fatti e la rispondenza agli interessi culturali.
2	Insegnante; scuola; scuola_elementare; assistente; sostegno; insegnante_di_scuola; scuola_materna; insegnante_di_sostegno; superiore; lingua; assistente_sociale; inglese; scuola_superiore; insegna; italiano; media; insegnante_di_scuola_materna; insegnante_elementare; insegnante_di_scuola_elementare; matematica; maestra; scuola_media; musica; infanzia; docente; storia; lezione; spagnolo; ripetizione; scuola_privata; studente; liceo; insegnante_di_italiano; insegnante_di_musica; supplente; insegnante_di_matematica; bambino; materia; straniero; informatica; educazione; scuola_primaria; handicap; istituto; insegnante_di_scuola_superiore; fisica; lingua_italiana; lingua_inglese; scuola_d_infanzia; insegnamento	Sono con grande probabilità donne, occupate nel settore dei servizi, con lauree dei gruppi insegnamento e letterario, oppure con minor probabilità linguistico, conseguite forse nell'Ateneo di Cassino; ritengono inoltre il titolo molto efficace. Hanno un contratto a tempo determinato o anche atipico, a tempo parziale. Al momento dell'intervista sono impegnati in una scuola di specializzazione post-laurea. Non percepiscono benefit (come premio di produttività, computer portatile, cellulare aziendale, auto aziendale, rimborso delle spese di trasporto), tranne che la tredicesima, peraltro solo in alcuni casi. Non cercano un nuovo lavoro, anche se ritengono che il proprio guadagno non sia adeguato al titolo di studio e al lavoro che svolgono; infatti sono complessivamente più soddisfatti della media per il proprio lavoro, soprattutto per quanto riguarda la rispondenza agli interessi culturali e il tempo libero. C'è qualche possibilità che siano coniugate e residenti all'estero.
3	cliente; ufficio; segreteria; amministrazione; contabile; mansione; contatto; segretario; rapporto; estero; albergo; ditta; ufficio_commerciale; ordine; lavoro; viaggio; fornitore; personale; pubblica_amministrazione; prodotto; fattura; lavoro_di_segreteria; prenotazione; mansione_di_segreteria; archivio; contatto_con_i_clienti; agenzia; contatto_con_clienti; registrazione; clientela; archiviazione; bolla; consegna; marketing; pubblico; commerciale_estero; traduzione; informazione; ufficio_personale; ufficio_estero; contratto; inserimento; progetto; organizzazione; contatto_con_	I laureati sono per lo più donne, laureate nel settore linguistico, o anche politico-sociale ed economico-statistico, che considerano il titolo poco efficace, o al limite in alcuni casi abbastanza efficace, e il cui guadagno non risulta adeguato. Cercano un nuovo lavoro, pur percependo tredicesima e quattordicesima, premio di produttività e buoni pasto. Forse provengono dalla classe media impiegatizia, hanno studiato nell'Ateneo di Torino, o comunque in un altro Ateneo del Nord Ovest. Con qualche probabilità, hanno un contratto a tempo indeterminato e sono più soddisfatti della media degli intervistati per l'indipendenza, il coinvolgimento nei processi decisionale dell'azienda.

⁶ La descrizione dei Cluster è avvenuta con la collaborazione di Serena Cesetti e Matteo Gallerani del Consorzio Interuniversitario ALMALAUREA.

Cluster	Descrizione lavoro	Descrizione caratteristiche dei soggetti del Cluster ⁶
	il pubblico; export; ricevimento; cameriera; hostess; associazione	
4	educatore; insegnamento; bambino; educatore_professionale; ragazzo; nido; handicap; asilo; casa; centro; cooperativa; minore; sostegno; attività; famiglia; comunità; portatore_di_handicap; portatore; ripetizione; scuola; cooperativa_sociale; educatore_di_asilo; educatore_di_sostegno; scolastica; scuola_materna; educazione; progetto; adolescente; gruppo; persona; doposcuola; lavoro; riabilitazione; infanzia; animazione; disagio; integrazione; anziano; educatore_di_bambini; centro_diurno; istituto; domicilio; ambito; matematica; musica; compito; giovane; attività_di_sostegno; laboratorio; recupero	Sono soprattutto donne, forse coniugate, laureate nel settore insegnamento, o psicologico, occupate nei servizi, a tempo parziale. Non percepiscono premio di produttività, computer portatile, buoni pasto; potrebbero essere occupate con un contratto atipico oppure senza contratto, e guadagnare fra 250 e 500 euro al mese o anche meno di 250. Si sono laureati negli Atenei di Padova, Bari, Firenze, con un voto compreso tra 106 e 110, che risulta maggiore della media. Hanno poi portato a termine un praticantato e attualmente cercano un nuovo lavoro; infatti non ritengono che il guadagno sia adeguato al titolo e alla posizione, pur essendo molto soddisfatti del rapporto con i colleghi e dell'utilità sociale del lavoro svolto. Ritengono comunque la laurea molto efficace per il lavoro svolto.
5	gestione; controllo; produzione; controllo_di_gestione; cliente; personale; organizzazione; ufficio; contratto; punto; credito; gestione_personale; ordine; affare; ente; libreria; ricetta; vendita; finanziaria; lavoro; gestione_del_personale; addetto; portafoglio; amministrazione; rete; rapporto; relazione; recupero; impresa; data; costo; procuratore; prestito; gestione_commerciale; appalto; polizza; negozio; attività; formazione; impianto; ditta; magazzino; estero; contabilità; cantiere; qualità; programmazione; aspetto; risorsa; ufficio_personale	Con grande probabilità celibi, appartengono perlopiù al gruppo Economico-statistico e si sono laureati nell'Ateneo di Firenze o al limite in altri Atenei del Centro, con voto 91-100. Sono residenti al Centro e lavorano nella medesima area geografica, in particolare in Toscana. Lavorano a tempo pieno e con una certa probabilità hanno un contratto di lavoro a tempo indeterminato; il guadagno è fra i 1000 e i 1250 euro al mese (maggiore della media), che ritengono essere adeguato al proprio titolo di studio. Inoltre percepiscono a volte tredicesima e premio di produttività. Non hanno frequentato nessuna scuola di specializzazione, valutano la laurea abbastanza efficace, ed infine provengono da famiglie con entrambi i genitori laureati. Meno soddisfatti della media per tempo libero, rapporto con i colleghi.
6	vendita; medico; farmaco; informatore; veterinario; informatore_scientifico; commessa; traduttore; cameriera; visita; ambulatorio; funzione; base; medico_veterinario; pubblicitario; informazione; negozio; interprete; medico_di_base; informatore_scientifico_del_farmaco; abbigliamento; addetto; esercizio; informatore_medico; rappresentante; negozio_di_abbigliamento; tavolo; vendita_di_prodotti; addetto_alle_vendite; prodotto; medico_scientifico; informatore_medico_scientifico; esercizio_commerciale; sostituzione; settore; malattia; trattamento; classe; supplenza; reparto; farmaco_ai_medici; funzione_amministrativa; campo; amministrazione; libreria; ditta; controllo; lavoro; ordinazione; animale	Con grande probabilità hanno concluso un tirocinio finalizzato all'iscrizione ad un albo e perlopiù appartengono al gruppo Chimico-farmaceutico, Medico, Agrario o Linguistico. Considerano il titolo poco o per nulla efficace e tendenzialmente percepiscono benefit aggiuntivi (computer portatile, cellulare e auto aziendale, rimborso delle spese di trasporto, premio di produttività). Potrebbero non avere un contratto di lavoro o forse essere autonomi e probabilmente sono celibi. Si sono laureati con un voto compreso tra 91 e 100 e forse lavorano con un contratto a tempo parziale. C'è qualche probabilità che abbiano in corso una scuola di specializzazione. Ritengono il guadagno non adeguato al proprio titolo di studio (guadagnano meno di 250 euro al mese), si sono laureati in Atenei del Nord-Est, forse in quello di Parma. La sede di lavoro potrebbe trovarsi in Trentino Alto Adige o Lombardia e tendenzialmente provengono da famiglie con un solo genitore laureato. Molto soddisfatti per vari aspetti del lavoro svolto.
7	farmacista; progettista; collaboratore; disegnatore; farmacia; banco; impianto; lavoro_al_banco; progetto; impianto_elettrico; elettrice; meccanico; tecnico; farmaco; lavoro; software; studio; collaboratore_in_farmacia; farmacista_al_banco; autoveicolo; circuito; banco_della_farmacia; sviluppo;	Appartengono al gruppo Chimico-farmaceutico, forse a Ingegneria o Architettura. Considerano il titolo molto efficace, guadagnano 1001-1250 euro al mese. Lavorano nell'industria, stabilmente (autonomi effettivi o a tempo indeterminato) e tendenzialmente percepiscono, come benefit aggiuntivo, solo la tredicesima (non dispongono di auto aziendale, buoni pasto, rimborso spese per gli spostamenti di lavoro, premio di produttività e assicurazione sanitaria). È probabile che sia-

Cluster	Descrizione lavoro	Descrizione caratteristiche dei soggetti del Cluster ⁶
	architettura; analista; civile; ambito; ingegnere; infrastruttura; studio_privato; prodotto; titolare; ingegnere_civile; test; carico; cad; depurazione; aiuto; ricetta; laboratorio; calcolo_strutturale; servizio; spedizione; industriale; magazzino; bonifica; ufficio_tecnico; vendita; studio_di_ingegneria; calcolo	no laureati del Politecnico di Torino, di Padova o di altri Atenei del Nord-Ovest. Di provenienza borghese, lavorano a tempo pieno e il titolo di studio dei genitori potrebbe essere il diploma secondario superiore. Hanno concluso un tirocinio per l'iscrizione all'albo, ma non una scuola di specializzazione. Non cercano lavoro e probabilmente si sono laureati con un voto tra 66-90 o, con minore probabilità, tra 106-110. Forse lavorano e risiedono al Nord, in particolare al Nord-Ovest. Più soddisfatti della media per utilità sociale del lavoro e coerenza con gli studi compiuti.
8	impiegato; operatore; banca; amministrativo; cassiere; bancario; sportello; cassa; pubblico; cassiere_in_banca; ufficio; impiegato_amministrativo; sportello_bancario; istruttore; impiegato_bancario; conto; conto_corrente; supermercato; operatore_di_sportello; operazione; vice; contabile; filiale; versamento; tecnico; centro; ente; servizio; consiglio; manutenzione; retro; cassa_in_banca; grafico; amministratore; lavoro; settore; legale; atto_amministrativo; atto; comunità; apertura; assegno; settore_amministrativo; agenzia; cliente; addetto; titolo; gestore; area; bonifico	Probabilmente percepiscono i buoni pasto, forse la tredicesima, il premio di produttività o l'assicurazione sanitaria (non hanno computer portatile, auto e cellulare aziendale o alloggio). Guadagnano tendenzialmente 1251-1500 euro mensili, appartengono al gruppo Economico-statistico o al Giuridico. Lavorano nei servizi, potrebbero avere un contratto a tempo indeterminato, di formazione lavoro o, con minore probabilità, a tempo determinato. Considerano il titolo abbastanza efficace o, con qualche probabilità, poco o per nulla efficace. Ritengono che il guadagno sia adeguato alla posizione ricoperta, lavorano a tempo pieno e probabilmente sono donne. Potrebbero aver in corso un praticantato (non dovrebbero aver svolto una collaborazione volontaria o un dottorato). La sede di lavoro potrebbe trovarsi in Veneto o in Toscana; tendenzialmente cercano lavoro. Il titolo di studio dei genitori è un diploma di scuola secondaria superiore (o di licenza media). Potrebbero aver conseguito la laurea con voto tra 91 e 100 o 66-90, con qualche probabilità presso l'Università di Torino. Forse risiedono nel Nord-Est; potrebbero appartenere alla piccola borghesia. Più soddisfatti della media per la stabilità, sicurezza del lavoro.
9	consulenza; formazione; studio; pratica; formazione_professionale; agenzia; legale; impresa; commercialista; cliente; studio_dentistico; azienda; materia; dichiarazione; progettazione; auto; normativa; selezione; cura; prevenzione; attività; ente; sicurezza; intermediazione; collaborazione; banca; personale; orientamento; legge; ambito; consulenza_fiscale; lavoro; docente; consulenza_finanziaria; comunicazione; finanziaria; stesura; x; formatore; consulenza_legale; politica; centro; amministrazione; rapporto; familiare; docenza; tecnico; ingegneria; relazione_pubblica; applicazione	Generalmente hanno in corso un tirocinio finalizzato all'iscrizione ad un albo o una collaborazione volontaria non retribuita. Lavorano nei servizi e potrebbero provenire dai gruppi Giuridico o Economico-statistico, forse dal Politico-sociale. Per quel che riguarda i benefit non percepiscono tredicesima o buoni pasto ma, con minore probabilità, dispongono del computer portatile, del cellulare, dell'auto aziendale e del rimborso delle spese di trasporto. Probabilmente si sono laureati con un voto tra 91 e 100 e tendenzialmente ritengono che il guadagno non sia adeguato al titolo. Si tratta perlopiù di collaboratori, forse di lavoratori autonomi, o di occupati con contratto di formazione lavoro; potrebbero guadagnare tra 250 e 500 euro mensili. Tendenzialmente provengono dalla piccola borghesia, convivono, si sono laureati nell'Ateneo di Siena e probabilmente la sede di lavoro si trova in Toscana. Ci sono alcune probabilità che risiedano nel Nord-Ovest, che lavorino a tempo parziale e che abbiano genitori di cui uno solo in possesso della laurea. Più soddisfatti della media per la flessibilità dell'orario e dei tempi di lavoro.
10	responsabile; qualità; insegna; ufficio; docente; prodotto; responsabile_ufficio; acquisto; logistica; responsabile_commerciale; gestione; azienda; servizio; progetto; comune; sistema; tecnico; gara; sviluppo; respon-	Hanno a disposizione: cellulare aziendale, premio di produttività, auto aziendale o rimborso spese di trasporto; non ricevono la tredicesima e l'assicurazione sanitaria. Lavorano con contratto a tempo indeterminato, forse nell'industria e c'è qualche probabilità che siano donne. Potrebbero essere ingegneri (anche eco-

Cluster	Descrizione lavoro	Descrizione caratteristiche dei soggetti del Cluster ⁶
	sabile_di_produzione; produzione; responsabile_amministrativo; cantiere; appalto; agenzia; negozio; responsabile_di_cantiere; ufficio_tecnico; gara_d_appalto; banca; responsabile_di_un_ufficio; vendita; lavoro; sicurezza; relazione; controllo; rapporto; informatica; coordinamento; magazzino; amministrativo; ginnastica; organizzazione; marketing; centro; coordinatore; anziano; monitoraggio; amministrazione; volontariato	nomisti e statistici), guadagnano fra 1501 e 1750 euro al mese o oltre 2000 euro. Si sono laureati in atenei del Sud (Chieti-Pescara o Salerno) e probabilmente provengono da una famiglia di origine operaia. Ritengono che il guadagno sia adeguato al titolo, potrebbero lavorare stabilmente, con qualche probabilità in Lombardia o anche al Sud. Considerano il titolo abbastanza efficace, si sono laureati con voto tra 101 e 105 e, attualmente, non cercano un nuovo lavoro. Con qualche probabilità il titolo di studio dei genitori è al più la licenza elementare. Non dovrebbero aver svolto alcuna scuola di specializzazione post-laurea. Più soddisfatti della media per la stabilità e la sicurezza del lavoro.
11	azienda; marketing; direttore; cliente; grafica; interno; azienda_privata; vendita; contatto; mobile; azienda_agricola; bilancio; controllo; prodotto; gestione; assistente; dirigente; servizio; commercio; tecnica; comunicazione; informatica; produzione; personale; assistente_del_direttore; costo; ufficio; procedura; liquidatore; telecomunicazione; chimica; materiale; analisi; multinazionale; segretario; titolare; coordinamento; stampa; fatturazione; estero; telefonia; funzionario; progettista; ambito; lavoro; direttore_di_cantiere; attività; rapporto; relazione; processo	Ricevono benefit aggiuntivi (computer portatile o assicurazione sanitaria; forse anche premio produttività, buoni pasto, tredicesima o cellulare aziendale; con minor probabilità auto aziendale o rimborso spese). Lavorano a tempo indeterminato, forse con contratto di formazione, a tempo pieno e probabilmente nell'industria. Ritengono che il guadagno sia adeguato al titolo; non hanno svolto alcuna collaborazione volontaria né un dottorato di ricerca; con qualche probabilità stanno svolgendo un tirocinio per l'iscrizione all'albo. Laureati in un Ateneo del Nord-Ovest (Politecnico di Torino), il guadagno è tendenzialmente tra 1501-1750 euro al mese. Perlopiù appartengono al gruppo Economico-statistico o Linguistico; meno probabilmente ad Ingegneria. Potrebbero lavorare in Lombardia, essere donne e risiedere al Nord-Ovest. Con qualche probabilità hanno conseguito la laurea con voto tra 66-90, provengono da una famiglia di origine borghese e forse convivono. Più soddisfatti della media per la stabilità, sicurezza del lavoro.
12	assistenza; servizio; tecnico; comune; tecnica; informatica; traduzione; testo; esame; tecnico_commerciale; catalogazione; docenza; orientamento; studente; redazione; assistenza_sociale; assistenza_tecnica; documentazione; lavoro; anagrafe; offerta; promozione; accompagnatore; vigilanza; clinico; ufficio_tecnico; utente; aspetto; ufficio; indagine; ente; comunicazione; cliente; settore; giornalista; alunno; informatico; bancario; associazione; direttore; centro; sicurezza; servizio_sociale; rivista; atto; operatore; supporto; x; civile; funzionario	Risiedono al Sud, sono lavoratori atipici, per lo più con contratto di collaborazione. Come benefit, dispongono tendenzialmente solo dell'auto aziendale, mentre non dovrebbero avere l'assicurazione sanitaria, buoni pasto o tredicesima. Sono alla ricerca di un nuovo lavoro, potrebbero essere ingegneri che lavorano in Sicilia o in Piemonte. Probabilmente si sono laureati presso l'Ateneo di Catania, al Politecnico di Torino ma anche all'Università di Messina; con minori probabilità a Trieste. Tendenzialmente hanno ottenuto il massimo dei voti (110 e lode) o un voto tra 66 e 90. Probabilmente sono occupati a tempo parziale, non hanno seguito alcuna scuola di specializzazione e forse guadagnano meno di 250 euro o tra 250 e 500 euro al mese. C'è qualche possibilità che lavorino al Sud (o al Nord Italia), e che abbiano entrambi i genitori con la laurea. Più soddisfatti della media per: acquisizione di professionalità, utilità sociale del lavoro, tempo libero.
13	contabilità; attività; office; bilancio; manager; pagamento; amministrazione; revisione; compito; libro; compagnia; coordinamento; società; nota; finanziaria; azienda; assicurazione; export; database; cliente; addeito; gestione; contabile; impiegato; centro; segreteria; rapporto; monitoraggio; la-	Appartengono perlopiù al gruppo Economico-statistico o Letterario. Potrebbero star svolgendo un tirocinio per l'iscrizione all'albo o aver svolto un master universitario, mentre non dovrebbero aver svolto una scuola di specializzazione. Probabilmente percepiscono la tredicesima, i buoni pasto (non l'auto aziendale o il computer portatile). Lavorano tendenzialmente con contratto di formazione lavoro/Apprendistato o a tempo deter-

Cluster	Descrizione lavoro	Descrizione caratteristiche dei soggetti del Cluster ⁶
	voro; opera; fatturazione; fornitore; studio; preparazione; aspirante; cantiere; analisi; paga; documento; recupero; settore; direzione; sportello; credito; fotografo; università; tenuta; dirigente; veterinario; filiale	minato, convivono e ci sono possibilità che si siano laureati all'Università di Genova, a Salerno o, meno probabilmente, a Messina. La sede di lavoro potrebbe trovarsi in Liguria o Campania; considerano la laurea abbastanza efficace, probabilmente sono maschi e provengono da famiglie in cui solo un genitore è laureato. Ci sono alcune possibilità che siano alla ricerca di un nuovo lavoro e che risiedano nel Nord-Ovest. Non dichiarano il proprio guadagno. Più soddisfatti della media per la stabilità, sicurezza del lavoro e per il luogo di lavoro.
14	ricerca; sviluppo; campo; personale; settore; elaborazione; selezione; programmazione; attività di ricerca; selezione del personale; cinema; mercato; attività; ricerca di mercato; processo; risorsa; progetto; sistema; economia; statistica; ricerca nel settore; strumento; lavoro; risorsa umana; cartografia; pubblicazione; paga; indagine; studio; sistema informativo; intervento; redazione; comunicazione; docenza; analisi; territorio; centro; società; intervista; ente; gestione; coordinamento; laboratorio; marittimo; promozione; assunzione; marketing; turismo; problematica; supporto	Hanno con grande probabilità un contratto di collaborazione oppure di formazione lavoro/Apprendistato, per lo più lavorano in Trentino Alto Adige, potrebbero essersi laureati all'Università di Siena o anche in un Ateneo del Nord-Ovest. Forse sono occupati nel settore dei servizi, non percepiscono la tredicesima, tendenzialmente provengono dal gruppo Geo-Biologico oppure Psicologico o, con una probabilità minore, dai gruppi Letterario e Ingegneria. Ci sono alcune possibilità che il guadagno mensile sia compreso tra 250 e 500 euro o tra 751 e 1000, che siano femmine e che ritengano il guadagno non adeguato alla posizione ricoperta. Potrebbero aver conseguito la laurea con il massimo dei voti (110 e lode). Più soddisfatti della media per la coerenza con gli studi compiuti (meno per stabilità del lavoro e il rapporto con i colleghi).
15	analisi; evento; relazione; stampa; organizzazione; comunicazione; pubblica relazione; chimica; ufficio; organizzazione di eventi; pianificazione; motore; produzione; addetto; giornalista; mercato; organizzatore; quotidiano; laboratorio; agenzia; soggetto; figura; pubblico; promozione; rassegna; biologo; traduzione; ente; organizzatore di eventi; lavoro; promotore; chimico; didattica; associazione; finanziaria; attività; indagine; pubblicità; componente; media; congresso; clinica; rapporto; casa; marketing; ambito; analisi di laboratorio; analisi di mercato; contatto; vendita	Molto probabilmente appartengono al gruppo Politico-sociale, oppure, con probabilità del tutto minori, a quello Geo-Biologico o Letterario. Potrebbero essersi laureati presso l'Ateneo di Trento ma anche in un altro del Nord-est dell'Italia. Tendenzialmente sono lavoratori atipici e ci sono possibilità che abbiano un contratto di collaborazione e che non percepiscano benefit aggiuntivi come tredicesima, premio di produttività, assicurazione sanitaria. Non hanno svolto un tirocinio finalizzato all'iscrizione ad un albo, potrebbero guadagnare oltre 2000 euro al mese e lavorare all'estero o, con qualche probabilità, in Trentino Alto Adige. Considerano il titolo poco o abbastanza efficace, probabilmente risiedono all'estero ma anche in zone del Nord-Ovest dell'Italia e forse lavorano a tempo parziale. Più soddisfatti della media per prospettive future di carriera, rapporto coi colleghi, prospettive future di guadagno.
16	consulente; odontoiatra; agente; sicurezza; consulente finanziario; finanziamento; professionista; consulente informatico; area; informatico; commercio; agente di commercio; valutazione; ambito; ambiente; azienda; protesi; consulente aziendale; banca; studio; animale; lavoro; polizia; impresa; progetto; bancario; studio privato; prodotto; qualità; finanziaria; programma; agenzia; assicurazione; agricoltura; test; ditta; sviluppo; software; cura; arte; clientela; sinistro; norma; europeo; mutuo; ambulatorio; campo; coordinatore; mobile; commerciale estero	Con buone probabilità sono lavoratori autonomi, occupati stabilmente. Potrebbero essere medici e percepire alcuni benefit: computer portatile, premio di produttività, forse il rimborso delle spese di trasporto o l'alloggio (non dovrebbero ricevere, invece, assicurazione sanitaria o buoni pasto). Per lo più lavorano nei servizi e probabilmente guadagnano oltre 2000 euro al mese, ma anche (con probabilità minori) 1251-1500 euro. Tendenzialmente lavorano a tempo pieno, forse sono donne e potrebbero aver studiato nell'Ateneo di Catania o a Trento. Ritengono che il guadagno sia adeguato al titolo, attualmente non cercano lavoro e ci sono alcune possibilità che risiedano in una delle nostre maggiori Isole. Probabilmente si sono laureati con voto tra 101 e 105, la sede di lavoro potrebbe trovarsi in Lombardia o in Sicilia, ma anche in Trentino A.A. È probabile che provengano da famiglie in cui un solo genitore è laureato, non esprimono giudizi sull'efficacia del titolo e tendenzialmente non hanno svolto una scuola di specializzazione. Più soddisfatti della media per il prestigio che ricevono dal lavoro.

Cluster	Descrizione lavoro	Descrizione caratteristiche dei soggetti del Cluster ⁶
17	cameriere; barista; dentista; ingegnere; centralinista; ricercatore; teatro; giornalista; tutor; hostess; guida; baby; baby-sitter; bibliotecario; atto; marittimo; congresso; assicuratore; biologia; service; giardiniere; idraulica; fotografo; video; contributo; baby_sitter; sitter; stipulazione; commesso; museo; domanda; automazione; scienza; guida_turistica; processo; erogazione; carico; apprendista; trasporto; nuovo_prodotto; impianto; informazione; cinema; vino; provincia; tavolo; mansione; fiera; percorso; ricevimento	Ritengono che il titolo non sia efficace. Probabilmente svolgono un lavoro a tempo parziale o senza contratto; non dovrebbero percepire alcun benefit (tredicesima, cellulare aziendale, computer portatile, auto aziendale, buoni pasto, assicurazione sanitaria). Tuttavia, ci sono alcune possibilità che dispongano di un alloggio. Tendenzialmente sono alla ricerca di un lavoro, per lo più appartengono al gruppo Giuridico, Geo-biologico o Letterario. Potrebbero lavorare nei servizi e guadagnare tra 250 e 500 euro al mese, ma anche meno di 250 euro; ritengono che la retribuzione non sia adeguata al titolo. Ci sono possibilità che si siano laureati con il massimo dei voti o con voto compreso tra 101 e 105, che stiano svolgendo un praticantato, una scuola di specializzazione o, con minori probabilità, un master universitario. Forse la sede di lavoro si trova in Friuli V.G.; potrebbero essere uomini laureati nell'Ateneo di Bari, conviventi.
18	programmatore; software; laboratorio; addetto; analista; materiale; sviluppatore; funzionario; chimico; laboratorio_chimico; hardware; manovale; sviluppo; collaudo; livello; fascia; operaio; sistemista; magazzino; industria; fatturazione; controllo; prova; cucina; addetto_controllo; programmazione; riparazione; sviluppo_software; elettronica; tecnico_di_laboratorio; preparazione; informatico; credito; prodotto; analisi; tecnica; mansione; impianto; tecnico; cliente; uscita; sezione; appuntamento; progettista; prodotto_finito; agricoltura; assistenza; progetto; investimento; ambito	Sono perlopiù Ingegneri, ma si rilevano anche laureati del Geo-biologico e, con minori probabilità, del Chimico-farmaceutico. Potrebbero essere donne, occupate a tempo pieno, che percepiscono buoni pasto, tredicesima, forse alloggio e premio di produttività; non dovrebbero ricevere cellulare o auto aziendale. Probabilmente lavorano al Nord, guadagnano mensilmente 1001-1250 o 1251-1500 euro; ci sono possibilità che la sede di lavoro sia in Piemonte o in Campania. Dopo la laurea non hanno svolto attività di formazione (praticantati, collaborazioni volontarie, master universitari). Potrebbero risiedere al Nord (in particolare Nord-Ovest); perlopiù hanno studiato al Politecnico di Torino, a Salerno o, meno probabilmente, a Padova. Di origine sociale medio-impiegatizia o operaia, ritengono che la laurea sia abbastanza efficace; ci sono possibilità che abbiano un contratto a tempo indeterminato o determinato e che lavorino nell'industria. Forse si sono laureati con un voto compreso tra 101 e 105; il titolo di studio dei genitori è al più la licenza media inferiore. Meno soddisfatti della media per: utilità sociale del lavoro, rapporto coi colleghi, luogo di lavoro, prestigio che ricevono dal lavoro.
19	guardia; guardia_medica; odontoiatria; reparto; amministratore; visita; dirigente; finanza; capo; ristorante; istruttrice; paziente; nuoto; centro_commerciale; allievo; sala; professione; ufficiale; diagnosi; chirurgia; cuoca; dipendente; gestore; ispettore; polizia; centro; medico; meccanica; servizio; amministratore_di_sistema; medico_generico; società; grande_distribuzione; titolare; distribuzione; sostituzione; sistema; vaccinazione; gara; vigilanza; igiene; cambio; settimana; militare; azienda_sanitaria; organo; preventivo; sezione; ambulatorio; macchina_automatica	Sono perlopiù medici, ma anche laureati appartenenti al gruppo Giuridico e, meno probabilmente, a quello Agrario. Dichiarano di disporre di: auto aziendale e buoni pasto (non della tredicesima); tendenzialmente si tratta di lavoratori autonomi (comunque occupati stabilmente), che potrebbero aver in corso una scuola di specializzazione o avere concluso un tirocinio per l'iscrizione ad un albo. Inoltre, ci sono alcune possibilità che stiano seguendo un master universitario o che abbiano portato a termine una collaborazione volontaria. Guadagnano tra 1501 e 1750 euro o 1251-1500 euro, probabilmente sono coniugati e considerano la laurea molto efficace. Potrebbero essere donne, occupate nei servizi, laureate con voto tra 66 e 90 che lavorano nel Lazio, forse in Liguria. Ci sono possibilità che abbiano un contratto part-time, che risiedano al Sud e che si siano laureati nell'Ateneo di Parma. Ritengono che il guadagno non sia adeguato alla posizione ricoperta, e probabilmente lavorano al Sud. Più soddisfatti della media per: prestigio che ricevono dal lavoro, luogo di lavoro, rapporto coi colleghi, coinvolgimento e partecipazione all'attività lavorativa e ai processi decisionali.

<i>Cluster</i>	<i>Descrizione lavoro</i>	<i>Descrizione caratteristiche dei soggetti del Cluster⁶</i>
20	call; center; computer; call_center; telefono; calcolo; inserimento; opera; struttura; posta; scavo; hotel; scavo_archeologico; informazione; certificazione; accoglienza; telemarketing; fiera; aeroporto; grafica; arte; lettera; calcolo_strutturale; università; inserimento_lavorativo; cantiere; pubblicazione; utenza; disegno; assistente; fatturazione; studio; rilevazione; ordinazione; coordinamento; iscrizione; pubblico; progetto; capo; tramite; negozio; ambiente; catalogazione; elaborato_grafico; urbanistica; acquisto; prenotazione; straniero; casa; apparecchiatura	Ritengono che la laurea sia poco o per nulla efficace e probabilmente sono alla ricerca di un lavoro. Potrebbero aver conseguito la laurea al Politecnico di Torino o presso l'Ateneo di Chieti-Pescara o anche, con probabilità minori, a Trento. Ci sono possibilità che guadagnino tra 750 e 1000 euro o tra 250-500 euro mensili e che provengano dal gruppo Letterario. Per quel che riguarda i benefit dovrebbero percepire solo l'assicurazione sanitaria (non il computer portatile, il cellulare o l'auto aziendale). Perlopiù si sono laureati con il massimo dei voti (110 e lode), ma esistono alcune probabilità che abbiano conseguito il titolo con un voto tra 66 e 90. Forse lavorano in Veneto, hanno portato a termine un tirocinio finalizzato all'iscrizione ad un albo. Ci sono alcune probabilità che siano uomini, che lavorino con un contratto part-time e che risiedano al Sud. Poco soddisfatti per vari aspetti del lavoro svolto.

Riferimenti bibliografici

- BOLASCO S. (1999) *Analisi multidimensionale dei dati*, Carocci Editore, Roma.
- BOUROCHE, J.M., SAPORTA G. (1988) Les méthodes et les applications du scoring, *Atti 34° Riunione Scientifica della Società Italiana di Statistica*: 19-26.
- BOZDOGAN H. (2004) *Intelligent Statistical Data Mining with Information Complexity and Genetic Algorithms*, in *Statistical Data Mining and Knowledge Discovery*, Chapman and Hall/CRC, Boca Raton-Florida.
- CAMILLO F., GHISELLI S. (2005) Dall'aspirazione alla soddisfazione: tipologia di transizione e percorsi laurea-lavoro definiti su basi soggettive. In: CAMELLI A. (a cura di), *La transizione dall'università al lavoro in Europa e in Italia*, Il Mulino, Bologna: 75-103.
- CAMILLO F., MIGNOLI G.P., NARDI E. (2006) Lavoro ideale e tipologia delle aspettative: un modello esplorativo sui laureati 2004 e una simulazione della condizione occupazionale. In: CONSORZIO INTERUNIVERSITARIO ALMALAUREA (a cura di), *L'università in transizione: laureati vecchi e nuovi alla luce della riforma*, Il Mulino, Bologna: 87-100.
- CAMELLI A., LA ROSA M. (2004) *I laureati in Italia. Le indagini di ALMALAUREA su scelte formative, orientamento al lavoro e occupabilità*, FrancoAngeli, Milano.
- CONSORZIO INTERUNIVERSITARIO ALMALAUREA (2005) *Lavorare dopo la laurea. Caratteristiche e percorsi occupazionali*, Il Mulino, Bologna.
- CRISP (2005) *Simulazione dell'applicazione dell'indicatore "capitale umano" per la valutazione dell'efficacia dei corsi di laurea*, Rapporto finale
- ELIA A., VIETRI S. (2000) *Electronic dictionaries and linguistic analysis of Italian large corpora*, JADT 2000, 5th International Conference on the Statistical Analysis of Textual Data, École Polytechnique Fédérale de Lausanne, Lausanne-Switzerland, 2-4
- FRANCHI M. (2005) *Mobili alla meta. Giovani tra Università e lavoro*, Donzelli, Roma.
- ISTAT (2005) *Inserimento professionale dei laureati. Indagine 2004*, Roma.
- LEBART L., MORINEAU A., PIRON M. (1985) *Statistique exploratoire multidimensionnelle*, Dunod, Paris.
- LEBART L., SALEM A., BERRY A. (1998) *Exploring Textual Data*, Kluwer Academic Publishers, Dordrecht/Boston/London.
- RAFFAELLI R. (1992), *An inverse parallel parser using multi-layered grammars*, IBM Technical Disclosure Bulletin, 2Q.
- VITTADINI G., LOVAGLIO P.G. (2006) *Il capitale umano in Italia*. In CROCETTA C. (a cura di) *Metodi e modelli per la valutazione del sistema universitario*, CLEUP, Padova: 1-24.

***Could you tell me with your words the job you're doing?
Text mining models for the analysis of the open questions***

Summary. *In the 2004 AlmaLaurea Survey on the occupational condition of graduates it was experimented, on a small sample of working respondents, an open question demanding to describe in detail the carried out job. The collected texts have been processed through modern text mining techniques, that allow to unscramble semantic content of the expressions. In this paper, we have conducted a “kernel type” data analysis, and the main relations between graduates paths (in terms of personal characteristics, curriculum and relevant aspects of further career) and given responses have been pointed out.*

Keywords. *Description, working activity, text mining, Cluster, relevant aspects of further career, curriculum, data-base.*