

La rappresentazione delle competenze nel linguaggio delle università e in quello degli operatori del terzo settore: uno studio sulle relazioni tra *corpora*

Simona Balbi, Maria Gabriella Grassia, Giorgio Infante¹

Dipartimento Matematica e Statistica, Università di Napoli “Federico II”

Riassunto: In questo lavoro si affronta il problema della comunicazione fra domanda e offerta di competenze e di professionalità in un nel mercato del lavoro particolare, quello del terzo settore. L’attenzione si è concentrata sull’erogazione di servizi legati all’immigrazione. Si è confrontato il linguaggio utilizzato dagli operatori per descrivere, nel corso di interviste in profondità, le competenze richieste ai laureati, con il linguaggio utilizzato dalle università, per descrivere, nelle schede ministeriali, i contenuti e gli sbocchi professionali dei propri corsi di laurea. Da un punto di vista metodologico il problema affrontato è quello del confronto fra corpora. Nel lavoro vengono presentati i principali risultati, in termini di misure di dissimilarità e rappresentazioni grafiche, ottenute ricorrendo a strumenti propri dell’analisi di dati simbolici e dell’analisi statistica multivariata classica.

Parole chiave: dati testuali, oggetti simbolici, misure di dissimilarità, analisi delle correlazioni canoniche, rotazioni procustiane

1. Introduzione

Obiettivo del presente lavoro è valutare la rispondenza fra la domanda di competenze e di professionalità proveniente dal mondo lavorativo del terzo settore e il tipo di formazione che il sistema universitario italiano offre. L’attenzione è rivolta, in particolare, alle attività legate all’immigrazione.

¹ Il presente lavoro è stato finanziato nell’ambito del progetto “Modelli e metodi per abbinare profili formativi e bisogni di professionalità di comparti del terziario avanzato”, cofinanziato dal MIUR. Coordinatore nazionale è L. Fabbris, coordinatore dell’Unità di Napoli è S. Balbi. La nota, opera congiunta dei tre autori, è stata redatta da S. Balbi per i paragrafi 1e 5, da M. G. Grassia per i paragrafi 2 e 4 e da G. Infante per i paragrafi. 3 e 6.

Per raggiungere questo obiettivo si è fatto ricorso ai metodi statistici per l'analisi dei dati testuali, confrontando alcune interviste semi-strutturate a testimoni privilegiati, con le presentazioni che le università rivolgono ai potenziali studenti dei propri corsi di laurea, ricavate dal sito del Ministero dell'università.

Lo studio delle relazioni tra i due *corpora* ha comportato un problema di natura metodologica: le due tabelle lessicali corrispondenti non presentano né le stesse righe (documenti), né le stesse colonne (termini), pur avendo come oggetto gli stessi temi.

Nella soluzione di questo problema si è fatto riferimento alla letteratura specifica in tema di *competenze*, così da poter introdurre, nell'analisi statistica, una informazione esterna di natura contestuale. Si sono così costruite unità lessicali di ordine superiore, definite come gruppi di parole che condividono lo stesso ambito di caratteristiche lavorative (ad esempio, *sapere, saper fare, ...*). Questa operazione è stata effettuata ricorrendo a conoscenza esperta. E' stato così possibile misurare la dissimilarità nell'uso dei termini riferiti a sei diverse tematiche (*attitudini personali, capacità, conoscenze, termini generali, professionalità, servizi*), da parte dagli operatori e dalle università, ricorrendo ad un indice precedentemente proposto per oggetti simbolici (Balbi, Grassia, 2007).

In una fase successiva, maggiormente *data oriented*, si è fatto ricorso a strumenti di analisi statistica multivariata, proposti per il confronto di strutture concettuali di riferimento (Balbi, Misuraca, 2006), ottenendo una rappresentazione grafica, relativa alle distanze fra i singoli termini utilizzati. Questo ha reso possibile una sintesi ed una possibile interpretazione non soltanto delle criticità sostanziali legate a ciò che si insegna e ciò che si applica nel mondo del lavoro, ma anche di una discrasia dovuta, in parte, a difetti di comunicazioni, legati alla non condivisione di un linguaggio.

Nel lavoro vengono introdotte le nozioni di base dei metodi utilizzati (par. 2), la struttura dei dati (par. 3), la strategia di analisi adottata (par. 4) e i principali risultati ottenuti (par. 5). Le conclusioni sono tratte nel par. 6.

2. Alcuni richiami metodologici

I principali riferimenti metodologici di questo lavoro sono l'Analisi dei dati simbolici (Diday, 1987), per il trattamento di unità statistiche complesse non rappresentabili in una classica matrice unità *per* variabili, e l'Analisi ruotata delle correlazioni canoniche (Lafosse, 1989).

2.1 Oggetti simbolici ed oggetti testuali

Un oggetto simbolico (Bock, et al., 2000) è definito da una terna $s = (a, R, d)$, dove $d = (d_1, \dots, d_j, \dots, d_p)$ è la descrizione dell'oggetto (*intensione*), come insieme dei valori assunti da p descrittori $(Y_1, \dots, Y_j, \dots, Y_p)$, a è la funzione di riconoscimento booleana $\{vero, falso\}$ che consente di individuare gli elementi che appartengono all'insieme di descrizione d (*estensione* dell'oggetto) e $R = (R_1, \dots, R_j, \dots, R_p)$ è la relazione su cui si basa il riconoscimento delle singole osservazioni come realizzazioni individuali della descrizione fornita a livello concettuale da d . I descrittori di un oggetto simbolico possono essere di tipo nominale, continuo o discreto, e presentare più modalità o valori per ciascun oggetto.

Balbi, Bolasco e Verde (Balbi, et al., 2002; Bolasco, et al., 2002) propongono un'analisi su dati testuali derivati dalla definizione formale di concetti e dalla loro composizione in termini di unità elementari (parole), introducendo nozioni proprie dell'analisi di dati simbolici per lo studio di particolari strutture linguistiche. Si tratta di strutture lessicali complesse, caratterizzate dalla presenza di particolari prefissi (*anti-*, *filo-*, ecc.) e particolari suffissi (*-ismo*, *-ista*, ecc.) al fine di approfondire alcuni aspetti del linguaggio giornalistico italiano nel corso degli anni Novanta.

2.2 Analisi ruotata delle correlazioni canoniche

L'analisi delle correlazioni canoniche (Hotelling, 1936) si pone l'obiettivo di identificare combinazioni lineari (fattori canonici) di variabili, appartenenti a due gruppi distinti, che siano fra loro massimamente correlate. Questa tecnica, pur avendo una grandissima rilevanza metodologica, presenta notevoli difficoltà applicative, dal punto di vista dell'interpretazione e rappresentazione grafica dei risultati, poiché le coppie di fattori canonici individuate appartengono a spazi differenti, ciascuno generato dall'insieme di variabili che lo costituisce. Una interessante soluzione a questo problema (Lafosse, 1989, ripresa da Balbi, Esposito, 2000) è quella di ruotare preliminarmente le due matrici in senso procustiano (Gower, 1975), così da ricondurre i fattori canonici ad uno stesso spazio.

Formalmente, date 2 matrici \mathbf{X} (n, p) e \mathbf{Y} (n, q), riferite a due gruppi di variabili osservate sugli stessi individui, centrate e standardizzate, la miglior rotazione (nel senso dei minimi quadrati) di \mathbf{Y} verso \mathbf{X} è data da \mathbf{YR} :

$$\mathbf{R} = \mathbf{V}_{\mathbf{YX}}(\mathbf{V}_{\mathbf{XY}}\mathbf{V}_{\mathbf{YX}})^{-1/2} \quad (1)$$

dove $\mathbf{V}_{\mathbf{XY}} = \mathbf{X}'\mathbf{Y}$ è la matrice di cross-varianza fra \mathbf{X} e \mathbf{Y} e $\mathbf{V}_{\mathbf{YX}}$ la sua trasposta.

Poiché l'operazione base di un'analisi delle correlazioni canoniche è l'identificazione della struttura caratteristica della matrice $\mathbf{V}_{XX}^{-1}\mathbf{V}_{XY}\mathbf{V}_{YY}^{-1}\mathbf{V}_{YX}$, con \mathbf{V}_{XX} e \mathbf{V}_{YY} matrici di varianza e covarianza, rispettivamente di \mathbf{X} e \mathbf{Y} , una volta ruotato \mathbf{Y} verso \mathbf{X} , la matrice da decomporre è, dopo pochi passaggi:

$$\mathbf{V}_{XY}\mathbf{R} = (\mathbf{V}_{XY}\mathbf{V}_{YX})^{1/2} \quad (2)$$

Tale matrice è simmetrica e, quindi, le combinazioni lineari identificate sia per le variabili in \mathbf{X} che in \mathbf{Y} sono associate allo stesso sistema di vettori \mathbf{u} , appartenenti così allo stesso spazio di riferimento:

$$\mathbf{V}_{XY}\mathbf{R} = \sum_{j=1}^p \zeta_j \mathbf{u}\mathbf{u}' \quad (3)$$

Nell'ambito dell'analisi dei dati testuali, Balbi, Misuraca (2006) hanno proposto l'utilizzo dell'analisi ruotata delle correlazioni canoniche al fine di misurare la bontà di una traduzione e di rappresentare graficamente le criticità, come distanza fra termini in lingua originale e in traduzione.

3. La struttura dei dati da analizzare

Come anticipato, obiettivo del presente lavoro è valutare, per il terzo settore ed in particolare per i servizi all'immigrazione, la rispondenza fra la domanda di competenze e di professionalità proveniente dal mondo lavorativo e il tipo di formazione offerto dal sistema universitario italiano.

Si è così fatto ricorso a interviste in profondità rivolte a venti testimoni privilegiati (presidenti di associazioni di volontariato o cooperative, attive nella regione Campania) al fine di delineare la domanda di competenze e di professionalità proveniente dai principali attori istituzionali dei servizi offerti agli immigrati. Le interviste sono state successivamente sbobinate ed hanno costituito il *corpus della domanda di competenze*, aggregando le diverse interviste in tre gruppi principali, definiti in base alla dimensione dell'organizzazione (*piccola, media, grande*, secondo parametri specifici del settore²).

Dal lato dell'offerta sono state considerate le schede di presentazione dei 135 corsi di laurea triennale che nell'anno accademico 2005-2006 facevano esplicito riferimento, nella descrizione dei propri sbocchi occupazionali, ad attività del terzo

² La dimensione è stabilita in funzione del numero di operatori e volontari: se l'organizzazione ha un numero di operatori e volontari *inferiore a 10* è classificata come *piccola*, tra *11 e 30* è classificata come *media*, *oltre 30* è classificata come *grande*

settore o al *non profit*, estratte dalla banca dati *OFF.F* del Ministero dell'Università (disponibili sul sito www.cercauniversita.it). Data l'eterogeneità presente in questo tipo di attività, si tratta di corsi di laurea appartenenti a classi diverse: scienze economiche; scienze del servizio sociale; scienze sociali per la cooperazione, lo sviluppo e la pace; scienze sociologiche; scienze politiche e delle relazioni internazionali. Nel seguito si è proceduto aggregando i corsi di laurea in 16 classi definite sulla base della classe di appartenenza e della denominazione prescelta.

Della scheda si sono considerati i due campi scritti in forma libera, relativi agli obiettivi formativi caratterizzanti ed agli ambiti occupazionali previsti. Questi due campi hanno costituito il *corpus* dell'*offerta di competenze*.

Tabella 1. *I corpora della domanda e dell'offerta di competenze*

Domanda di competenze

ID	classe dimensionale	Testo
1	2	Le principali competenze richieste agli operatori laureati di questa associazione riguardano la conoscenza delle lingue, la conoscenza delle tradizioni e della cultura degli immigrati
2	1	Sicuramente quello che manca agli operatori quando si avvicinano agli immigrati è la conoscenza della loro cultura. E' necessario adeguarsi a loro capire le loro usanze e mediare rispetto alle nostre.
...		
20	1	Innanzitutto, rispetto come dire alle attività o all'aspetto sociale e culturale in questo ambito, ci vuole una forte motivazione una forte capacità relazionale.

Offerta di competenze

ID	classe di laurea	Testo
1	6	il corso di laurea prevede una formazione interdisciplinare nell'ambito delle discipline sociali e sociologiche, giuridiche con particolare riferimento al diritto di famiglia e al penale,
2	6	il corso di laurea in servizio sociale è orientato alla preparazione alla professione di assistente sociale e quindi a fornire ai laureati: una buona conoscenza delle discipline di base per il servizio sociale, una buona padronanza dei metodi e delle tecniche del servizio sociale, competenze pratiche ed operative, relative al rilevamento ed al
...		
135	2	il corso di laurea ha l'obiettivo di fornire agli studenti conoscenze adeguate di metodi e contenuti culturali e scientifici volte al conseguimento del livello formativo richiesto dall'area professionale del servizio sociale.

Da una prima lettura, emerge un elemento non trascurabile nell'analisi successiva: lo stile, legato, nel caso delle interviste, al linguaggio parlato, in quello delle schede a documenti ufficiali. Questa considerazione ha rappresentato un ulteriore elemento a favore di un'analisi che operasse su aggregazioni di termini, piuttosto che sui singoli termini utilizzati.

Entrambi i *corpora* (*domanda ed offerta di competenze*) sono stati normalizzati e lessicalizzati con procedure automatiche (grazie al *software* TALTAC), evidenziando *poliformi* e *polirematiche* di interesse per l'analisi. Dalle forme grafiche, sono state, quindi, individuate le forme testuali (definite come componenti significative minime del discorso, non ulteriormente decomponibili, siano semplici, composte o complesse, Bolasco, 1994), quali unità di analisi.

Si sono così ottenute due tabelle lessicali, di termine generico la frequenza di una forma in un documento: T_1 di dimensione (20, 87) relativa alla domanda di competenze (20 testimoni privilegiati e 87 forme testuali) e T_2 di dimensione (135 corsi di laurea, 73 forme testuali), relativa all'offerta. Le tabelle lessicali aggregate corrispondenti sono, quindi, T^*_1 (3,87), avendo considerato i testi generati dalle interviste secondo le 3 classi dimensionali delle organizzazioni, e T^*_2 (16, 72) avendo aggregato i corsi di laurea per classe e denominazione. Si è, inoltre, proceduto ad una ricodifica dei dati, considerando la sola presenza /assenza di una forma testuale in un documento (codifica booleana).

4. La strategia di analisi

L'obiettivo di confrontare, dal punto di vista della comunicazione, offerta e domanda di competenze nel terzo settore, si traduce, da un punto di vista metodologico, nella necessità di analizzare congiuntamente tabelle lessicali costruite a partire da due *corpora* differenti, che si ritiene però condividano tematiche comuni (la natura e le caratteristiche delle competenze necessarie per operare in attività connesse al terzo settore). Da un punto di vista algebrico si tratta di operare su matrici con un numero di righe (*documenti*) e di colonne (*vocabolario*) differenti.

La soluzione proposta si base sulla costruzione *a priori* di *oggetti testuali* relativi alle tematiche affrontate nei documenti prodotti, vuoi dall'università, vuoi dagli operatori di settore. L'interesse di costruire *oggetti "testuali"* consiste, infatti, nella possibilità di introdurre nell'analisi informazioni esterne (assimilabili a *metadati*), relative al contenuto del testo, derivanti vuoi da conoscenza esperta, vuoi dalla elaborazione di informazioni elementari rilevate nella fase di raccolta dei dati (*metadati interni*).

Da un punto di vista computazionale, l'utilizzo degli oggetti nel campo dell'analisi dei dati testuali evita i problemi connessi al trattamento di matrici sparse.

Se, infatti, l'aggregazione dei documenti, quando necessaria, avviene attraverso una variabile classificatrice (nel nostro caso, da un lato la dimensione dell'organizzazione, dall'altro una variabile composita che tiene conto della classe di laurea e della denominazione), l'aggregazione dei termini pertiene alle scelte effettuate nella fase di pretrattamento del testo, quali l'eliminazione di termini poco frequenti, la definizione di una *stopping list*, oppure la lemmatizzazione o lo *stemming*. A questi ultimi due casi può assimilarsi la costruzione di oggetti testuali, come riduzione delle forme attraverso una loro organizzazione in unità di ordine superiore, descritte da variabili modali, le cui modalità sono le forme stesse, descritte dalle corrispondenti frequenze osservate.

Questo modo di operare consente, inoltre, di confrontare *corpora* provenienti da fonti diverse e relativi a vocabolari e lessici differenti. Nel caso specifico la forma parlata delle interviste e quella scritta delle schede predisposte dalle università.

La costruzione di oggetti testuali sui due *corpora* ha consentito di calcolare una misura di dissimilarità (Balbi, Grassia, 2007) fra coppie di oggetti.

Si considerino due generici oggetti simbolici s_k e $s_{k'}$ descritti dalle stesse H variabili categoriche Z_h , ciascuna con m_h modalità, con $h = 1, \dots, H$:

$$s_k = \bigwedge_{h=1}^H [Z_{kh} = \{z_{km_{j_h}}, f_{km_{j_h}}\}] \quad s_{k'} = \bigwedge_{h=1}^H [Z_{k'h} = \{z_{k'm_{j_h}}, f_{k'm_{j_h}}\}] \quad (4)$$

dove $f_{km_{j_h}}$ è la frequenza relativa della m_j -esima modalità della variabile Z_h per l'oggetto s_k (analogamente per $s_{k'}$). La dissimilarità è calcolata come media della differenza della distribuzione relativa delle singole modalità per ogni singola variabile Z_h :

$$d(s_k, s_{k'}) = \sum_{h=1}^H \frac{1}{m_{j_h}} \sum_{j=1}^{m_{j_h}} \left| f_{km_{j_h}} - f_{k'm_{j_h}} \right| / \max [f_{km_{j_h}} - f_{k'm_{j_h}}] \quad (5)$$

L'indice varia tra 0 ed 1. Quando assume valore 1 c'è massima dissimilarità; quando è uguale a 0 le due distribuzioni sono identiche.

Attraverso la conoscenza esperta, le forme dei due vocabolari sono state suddivise nelle seguenti sei categorie (le variabili delle competenze):

1. Attitudini personali (Saper Essere)
2. Capacità (Saper Fare)
3. Conoscenze (Sapere)
4. Termini generali
5. Professionalità
6. Servizi

Dalla fusione delle forme testuali dei due *corpora* si sono ottenute 122 forme

distinte. Una prima considerazione: le forme testuali totali sono 160 (87 della domanda + 73 dell'offerta); l'intersezione fra questi due insiemi, riferiti allo stesso tema, è appena di 38 termini (160 - 122). Pur volendo tener conto della profonda diversità delle due forme di espressione (interviste orali/documenti ufficiali), un sospetto di diverse modalità di espressione sorge.

Tornando all'analisi, si sono costruiti 19 oggetti testuali (3 tipi di organizzazione + 16 tipi di corso di laurea), descritti da 6 variabili modali (le competenze), le cui categorie sono i 122 termini, ciascuno assegnato da un esperto esclusivamente ad una delle 6 variabili. Si sono così calcolate le distanze fra oggetti, utilizzando la misura precedentemente richiamata (Balbi, Grassia, 2007), ponendo:

$$f_{kmj} = \frac{\sum_{i=1}^{n_j} \tilde{t}_{m_i}}{n_j} \quad (6)$$

dove f_{kmj} è la frequenza della m -esima modalità appartenente alla j -esima variabile del k -esimo oggetto, come somma delle frequenze degli n_j termini \tilde{t}_{m_i} che la compongono, diviso il totale n_j .

Per la rappresentazione grafica congiunta delle competenze offerte e domandate, si è ritornati alle tabelle lessicali originarie: \mathbf{T}_1 e \mathbf{T}_2 . Per poter applicare l'analisi ruotata delle correlazioni canoniche lessicale occorre che esista una corrispondenza fra i termini confrontati (Balbi, Misuraca, 2006). Si è così tenuto conto ancora una volta dell'unione dei due vocabolari, a 122 termini.

Si chiamino $\tilde{\mathbf{T}}'_1$ (122, 20) e $\tilde{\mathbf{T}}'_2$ (122, 135) le matrici aventi in riga il vocabolario comune e in colonna, rispettivamente, le 20 interviste e le 135 descrizioni dei corsi di laurea. Si considerino, inoltre, le matrici $\tilde{\mathbf{O}}'_1$ (6, 20) e $\tilde{\mathbf{O}}'_2$ (6, 135) in cui le righe sono state raggruppate nelle sei categorie delle competenze, precedentemente individuate, con termine generico pari alla sommatoria delle frequenze delle rispettive forme testuali appartenenti alle categorie. La giustapposizione, rispettivamente, delle matrici $\tilde{\mathbf{T}}'_1$ e $\tilde{\mathbf{O}}'_1$, $\tilde{\mathbf{T}}'_2$ e $\tilde{\mathbf{O}}'_2$ ha generato le matrici \mathbf{X} ed \mathbf{Y} su cui è stata effettuata l'analisi ruotata delle correlazioni canoniche.

5. I principali risultati

Nella Fig. 1 sono riportate le dissimilarità tra gli oggetti della domanda e dell'offerta di competenze, da cui è possibile comprendere dove si incontrano e dove, invece, appare minore la sintonia fra università e mondo del lavoro. Non sembra emergere un vantaggio particolare di alcune classi sulle altre, né la chiave dimensionale sembra essere rilevante sul valore globale dell'indice. Complessivamente, la dissimilarità maggiore riguarda l'utilizzo dei termini generali relativi alle competenze (Tab. 2), ma appare comunque piuttosto alta.

Figura 1. La matrice di dissimilarità tra oggetti

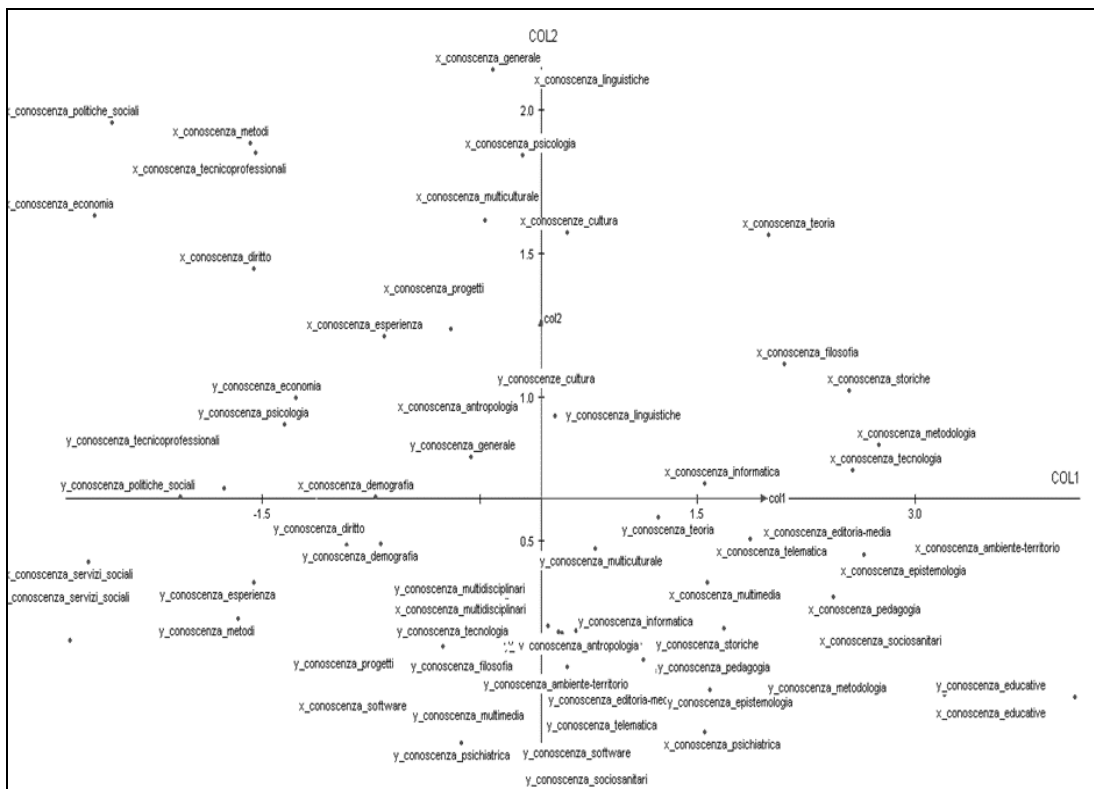
CLASSI	servizio sociale	servizio internazionale	scienze sociali e cooperazione e sviluppo	scienze sociali e cooperazione e sviluppo	scienze sociali e cooperazione e sviluppo	economia nonprofit	discipline economiche e aziendali	scienze per la pace	scienze dell'educazione	infanzia	educatore sociale, culturale e territoriale	professioni educative	educatore interculturale	promozione e sviluppo risorse umane	scienze dell'educazione e della formazione	scienze dei processi formativi
	44	2	12	2	4	4	1	2	16	9	8	12	2	5	9	3
Numerosità interna	0,43188	0,14583	0,19896	0,23958	0,27188	0,22396	0,27083	0,26458	0,33475	0,09766	0,36458	0,34422	0,33333	0,14583	0,41059	0,39583
attitudini_1	0,41053	0,23526	0,43333	0,34263	0,3774	0,38574	0,23397	0,36346	0,47989	0,28405	0,36409	0,42998	0,34744	0,36406	0,48275	0,51058
capacità_1	0,45334	0,32339	0,42986	0,31102	0,38324	0,37254	0,35484	0,47715	0,49634	0,30956	0,37788	0,43434	0,39866	0,3833	0,50862	0,4578
conoscenza_1	0,5919	0,35119	0,56245	0,56548	0,56746	0,50099	0,69048	0,57738	0,58624	0,51171	0,49235	0,5	0,39762	0,45655	0,56002	0,44762
generale_1	0,54962	0,50476	0,5131	0,40476	0,48651	0,46429	0,4881	0,54167	0,41916	0,51736	0,4951	0,54589	0,48095	0,48782	0,54498	0,43571
professione_1	0,50094	0,3451	0,41085	0,41373	0,39951	0,34926	0,43137	0,44559	0,40396	0,38268	0,49272	0,47304	0,38186	0,32497	0,4254	0,45245
servizi_1	0,52656	0,27083	0,44792	0,36458	0,52083	0,34896	0,39583	0,48958	0,49579	0,24479	0,38988	0,52312	0,45833	0,27083	0,4974	0,39583
attitudini_2	0,31968	0,35577	0,36394	0,42115	0,41341	0,44396	0,38141	0,41955	0,43151	0,3237	0,3553	0,37999	0,41442	0,42185	0,41613	0,54359
capacità_2	0,5207	0,32473	0,42472	0,20806	0,36398	0,34149	0,24731	0,39624	0,57402	0,30009	0,40296	0,50842	0,38548	0,40022	0,57849	0,4707
conoscenza_2	0,58708	0,46429	0,59828	0,49405	0,55992	0,44643	0,83333	0,74405	0,59351	0,50802	0,56719	0,70476	0,57024	0,54476	0,68542	0,60833
generale_2	0,58845	0,63095	0,41483	0,44643	0,48909	0,48909	0,5	0,52976	0,39156	0,35863	0,36074	0,50795	0,3619	0,29214	0,57184	0,30476
professione_2	0,55852	0,26912	0,29724	0,25	0,29412	0,19771	0,22059	0,40686	0,55609	0,45343	0,60011	0,55508	0,22794	0,44549	0,56878	0,48088
servizi_2	0,40093	0,24107	0,32708	0,24033	0,39658	0,20424	0,28571	0,37946	0,46801	0,31052	0,25122	0,42634	0,33408	0,24107	0,44647	0,23532
attitudini_3	0,35771	0,39267	0,40023	0,45746	0,4476	0,46696	0,40659	0,46346	0,42938	0,30547	0,33103	0,38878	0,45888	0,42668	0,42033	0,55458
capacità_3	0,46191	0,312	0,36014	0,24979	0,28961	0,30679	0,3341	0,40486	0,52194	0,33094	0,34763	0,48756	0,32295	0,32248	0,54351	0,49409
conoscenza_3	0,58518	0,49082	0,52833	0,55816	0,37466	0,48963	0,70408	0,55485	0,45528	0,52612	0,44625	0,62132	0,50298	0,30969	0,63278	0,54082
generale_3	0,5831	0,60306	0,56347	0,54974	0,625	0,6014	0,53061	0,53444	0,47076	0,56035	0,50055	0,4991	0,50485	0,51017	0,53496	0,48135
professione_3	0,4192	0,43309	0,46512	0,47234	0,50683	0,4475	0,48739	0,49247	0,4284	0,48319	0,45574	0,46075	0,47847	0,34819	0,48567	0,42248
servizi_3																

Tabella 2. Indice di dissimilarità aggregato tra domanda ed offerta

Variabili	Indice di Dissimilarità
Attitudini	0,48
Capacità	0,48
Conoscenza	0,44
Generale	0,56
Professione	0,43
Servizi	0,45

Un approfondimento è fornito dall'analisi ruotata delle correlazioni canoniche, che consente un confronto più puntuale. In Fig. 2 sono riportate le *conoscenze*: emergono rare coincidenze nella loro rappresentazione concettuale così come sono disegnate dal sistema universitario e come sono descritte dagli operatori del settore indagato (si noti la prossimità di *conoscenza servizi sociali*, *conoscenze educative*, *conoscenze multidisciplinari*, ma anche *conoscenze psichiatriche* o *epistemologiche*). Prevale, invece, la rappresentazione di mondi differenti. Questa distanza nella rappresentazione sembra nascere non soltanto da esigenze di contenuti differenti, ma anche da un linguaggio non condiviso.

Figura 2. *La rappresentazione congiunta delle principali forme testuali*



6. Conclusioni

Con il presente lavoro si è tentato di portare un ulteriore contributo al processo che il sistema universitario sta mettendo in atto, ossia di rendere la propria offerta formativa coerente con gli sbocchi occupazionali previsti per i suoi laureati.

Questo si traduce in un dialogo più stretto fra docenti e operatori. L'idea che si è voluta qui sviluppare è che non basta parlare di più, se si parlano linguaggi differenti. Per arrivare ad un accordo sui contenuti formativi dei corsi di laurea è necessario comprendersi. E fondamentale è che questa comprensione nasca da uno sforzo congiunto dei diversi attori sulla scena. Da una parte il mondo produttivo desidera degli strumenti funzionali ai propri obiettivi, dall'altro l'università non può abbandonare il suo ruolo culturale e innovatore.

L'esperimento che si è qui condotto tocca un settore particolarmente delicato sul piano umano, sia perché riguarda servizi offerti a soggetti particolarmente deboli, sia perché in queste attività è molto forte la presenza del volontariato (e, in qualche misura, la *non professionalità*).

I risultati cui si è pervenuti sembrano non contraddire le ipotesi iniziali: il ricorso a oggetti testuali ha consentito di riferirsi a contenuti comuni (le competenze così come classificate dalla letteratura). L'analisi ruotata delle correlazioni canoniche ha consentito di concentrare l'attenzione sulle differenze terminologiche.

Riferimenti bibliografici

- BALBI S., BOLASCO S., VERDE R. (2002) Text Mining on Elementary Forms in Complex Lexical Structures, in: MORIN A., SEBILLOT P. (eds) *JADT 2002*, St Malo I: 89-100.
- BALBI S., ESPOSITO V. (2000) Rotated Canonical Analysis onto a Reference Subspace, *Computational Statistics & Data Analysis*, **32**, 3-4: 395-410.
- BALBI S., GRASSIA M. G. (2007) Profiling and labour market accessibility for the graduates in economics at Naples university. In: FABBRIS L. (eds) *Effectiveness of University Education in Italy*, Physica Verlag, New York: 345-356.
- BALBI S., MISURACA M (2006) Rotated Canonical Correlation Analysis for Multilingual Corpora, In: VIPREY J.M. (eds) *JADT 2006*, Besançon: 99-106.
- BOLASCO, S. (1994) L'individuazione di forme testuali per lo studio statistico dei testi con tecniche di analisi multidimensionale. In: *Atti della XXXVII Riunione Scientifica della SIS*, Sanremo: 95-103.

- BOLASCO S., VERDE R., BALBI S. (2002) Outils de Text Mining pour l'analyse de structures lexicales à éléments variables in: MORIN A., SEBILLOT P. (eds.) *JADT 2002*, St Malo, **1**:197-208.
- BOCK H. H., DIDAY E. (2000). *Analysis of Symbolic Data*, Springer
- DIDAY, E. (1987): The symbolic approach in clustering and related methods of data analysis, in: BOCK H. (eds): *Classification and Related Methods of Data Analysis*, North-Holland, Amsterdam: 673-684.
- GOWER, J. C. (1975). Generalized Procrustes analysis, *Psychometrika*, **40**: 33-51
- HOTELLING H. (1936) Relations between two sets of variables, *Biometrika*, **28**: 321-377
- LAFOSSE R. (1989). Ressemblance et différence entre deux tableaux totalement appariés, *Statistique et analyse des données*, **14**: 1-24

***The representation of the competencies in the language
of universities and third sector's operators:
a study on the relationships among corpora***

Summary. *In this paper we propose a strategy for comparing supply and demand of competences on the "third sector" labour market, focusing attention on activities connected with immigration. The adopted viewpoint is communication: if competences offered by university courses unfit competences required by organisations active in the third sector, the reason is that actors on the stage do not correctly communicate. And this can be partially due to the use of different languages. Therefore, we jointly analyse documents derived from in-depth interviews to chief officers of organisations and from official descriptions of university courses. From a methodological viewpoint, we have to compare two corpora, by textual data analysis. We propose the use of a symbolic data analysis distance measure together with classic multivariate statistical methods, namely canonical correlation analysis together with Procrustean rotations, in order to understanding and representing communication gaps.*

Keywords. *Textual Data, Symbolic Objects, Dissimilarity Measures, Canonical Correlation Analysis, Procrustian Rotation.*