

# La stampa italiana e l'informazione demografica: un'applicazione dell'analisi testuale

*Maura Misiti*

Istituto Ricerche sulla Popolazione (IRP)

[misiti@irp.rm.cnr.it](mailto:misiti@irp.rm.cnr.it)

*Stella Iezzi*

[s.iezzi@ismae.it](mailto:s.iezzi@ismae.it)

## Abstract

*Population issues are increasingly reported by media. This interest has been driven not only by important events such as international conferences or news concerning population but also by an increasing demographers attention in dissemination of demographic issues. Certainly the Italian media have become quite interested in demographic issues, particularly on highly visible questions like the arrival of numbers of Albanians in Italy, the impending pension crisis, and the progressive decline in fertility. It is therefore important to see whether this increase in coverage is synonymous with an increase in effective communication).*

*A textual analysis on the demographic information reported on Italian press is presented in the paper based on a press collection on main Italian daily newspapers and magazines. The collection is related to two press conferences organised by IRP to present the results of his surveys on "Italian's Ideals" and "Stay at Home Children".*

## Riassunto

*L'informazione demografica in Italia è progressivamente trattata dai media. Questo interesse è connesso non soltanto ad importanti eventi come conferenze internazionali o fatti legati alla popolazione, ma anche da una crescente attenzione dei demografi nella diffusione dell'informazione demografica. Senza altro i media italiani stanno riservando una progressiva attenzione alle problematiche demografiche come l'arrivo degli albanesi o la questione delle pensioni o il calo della fecondità. A questo punto è importante valutare se ad un aumento dell'interesse corrisponda una maggiore efficacia comunicativa.*

*In questo lavoro è presentata un'analisi testuale dell'informazione demografica riportata dalla stampa italiana in occasione della presentazione alla stampa dei risultati di due indagini dell'IRP su "gli ideali degli italiani" e "Giovani che non lasciano il nido".*

*Parole chiave: informazione demografica, stampa, segmenti ripetuti, analisi delle corrispondenze, cluster analysis.*

## 1. Introduzione

Il rapporto fra l'informazione demografica e la stampa in Italia rappresenta un terreno estremamente delicato, poiché la prima si presta a forti manipolazioni, affonda le radici nella recente storia del paese quando la popolazione era considerata un obiettivo "politico" e non sembra poggiarsi su una base conoscitiva solida. Per gli italiani l'unica fonte di conoscenza/informazione nel campo dei fenomeni e delle dinamiche demografiche sono i media. L'informazione veicolata da giornali, radio e tv è divenuta il principale canale di conoscenza su questi temi, ma i giornalisti non sono chiamati a dare informazioni bensì a fornire notizie che, anche se corrette, devono comunque essere interessanti, catturare

l'attenzione di chi legge o ascolta e spesso devono essere legate ad un fatto di cronaca che ne giustifichi la diffusione. Non c'è dubbio che la questione demografica ha conquistato spazi crescenti sui media italiani, soprattutto in funzione della visibilità offerta dalle cosiddette emergenze demografiche come i ripetuti sbarchi di albanesi in Italia, la crisi del sistema pensionistico a ripartizione o il calo progressivo della fecondità. E' importante perciò valutare se all'aumento di attenzione corrisponda un analogo incremento dell'efficienza della comunicazione. Per questo motivo abbiamo analizzato con metodi di analisi testuale gli articoli comparsi sulle testate di quotidiani nazionali e locali e di magazines in occasione di due conferenze stampa dell'Istituto di ricerche sulla popolazione, la prima relativa alla presentazione dei risultati della 4° Indagine sugli atteggiamenti e le opinioni degli italiani verso le tendenze demografiche, e la seconda relativa alla presentazione dei risultati dell'Indagine sui giovani e la loro permanenza in famiglia. La scelta di questi due eventi è legata al fatto che in entrambi i casi si conosceva sia il contenuto del comunicato stampa che le modalità di contatto con i giornalisti e dunque si poteva essere certi che l'informazione originale fosse ben confezionata e correttamente predisposta alla comunicazione divulgativa.

## 2. La metodologia

L'analisi del *corpus* si snoda in 5 fasi progressive:

1. Pre-processing, con la numerizzazione delle forme grafiche e calcolo delle frequenze associate, sia nell'intero corpus che nei singoli testi;
2. ricerca automatica dei segmenti;
3. analisi delle corrispondenze e classificazione delle unità statistiche sulla base dei segmenti "grezzi";
4. eliminazione delle sequenze-funzione;
5. calcolo di indici volti ad individuare l'impatto delle tematiche emergenti nei gruppi selezionati nella fase 3.

La prima fase è di tipo esplorativo e consente di valutare la consistenza del corpus ed individuare le eventuali discrasie formali tra i singoli testi. In questa fase si determinano i valori minimi, massimi, il range e lo scarto quadratico medio delle frequenze delle occorrenze, forme grafiche e ricchezza del vocabolario all'interno dei singoli testi.

Nella seconda fase, si ricercano i segmenti con una soglia ragionevolmente bassa (composta da 2 a 5 parole, con un frequenza minima pari a 2), in modo tale da poter successivamente lavorare su un numero elevato di frasi.

Nella terza fase, i segmenti ripetuti, ottenuti automaticamente, costituiscono l'input per un'analisi delle corrispondenze e per l'applicazione di una classificazione gerarchica, attraverso il metodo di Ward. Il criterio prescelto per l'aggregazione è particolarmente idoneo, in questo caso, poiché, come è noto, i dati di input sono coordinate fattoriali. Il numero di gruppi è stato individuato sia attraverso la visualizzazione grafica delle proiezioni delle prime coordinate fattoriali sia attraverso il taglio del dendrogramma (stopping rules).

Nella quarta fase sono eliminate le sequenze non significative (poiché costituite soltanto da forme grammaticali, incomplete o banali) ossia si eliminano le sequenze funzione (De Beugrande e Dressler, 1981) e si ricercano le strutture tematiche.

Nell'ultima fase, si valuta l'impatto dei diversi temi trattati nelle due indagini dell'IRP nei articoli considerati nell'analisi attraverso alcuni indicatori: il peso percentuale della somma delle occorrenze nel gruppo rispetto all'intero corpus (1) e l'indice di variazione tra le tematiche affrontate nei vari articoli (corpus) e la dimensione del gruppo, intesa sempre come somma di occorrenze (2).

$$(1) P_i = \frac{f_{ik}}{f_c} \times 100; \text{ dove } f_c \text{ è la dimensione nel corpus e } \sum f_{ik} \text{ è somma delle occorrenze di testi}$$

che concorrono alla formazione di un gruppo; questo indice varia tra zero e 100, tanto più si avvicina al cento tanto maggiore è l'incidenza del gruppo (quindi dei temi che lo caratterizzano) sull'intero corpus.

$$(2) \Delta_i = \frac{(f_c - \sum f_{ik})}{f_c}; \text{ questo indice varia tra zero e 1, tanto più si approssima all'unità}$$

tanto minore è l'incidenza della tematica sull'intero corpus.

## 2. Applicazione

I documenti analizzati presentano una alta variabilità interna, segno distintivo di una eterogeneità formale dei testi. Un rapido sguardo alla tabella 1 consente di verificare l'ampio raggio di oscillazione tra i valori massimi e minimi.

**Tabella 1**

Rass. Stampa	<i>ideali (39 testi)</i>			<i>giovani (36 testi)</i>		
	occorrenze	Forme grafiche	segmenti	occorrenze	forme	segmenti
freq. ass.	17.683	2.378	1.049	21.390	3.115	2.170
Minimo	86	73	gruppo 1 (160)	73	69	gruppo 1 (280)
Massimo	1.523	621	gruppo 2 (551)	1.034	689	gruppo 2 (1.141)
Range	1.437	548	391	961	620	861

L'analisi delle corrispondenze ha dato luogo all'individuazione della struttura portante della ricezione mediatica dei risultati delle due indagini. Per quanto riguarda la prima (Gli ideali degli italiani) sono stati identificati due fattori principali: il primo rappresenta "le relazioni tra individuo e società", mentre il secondo "l'organizzazione della vita in comune". Il grafico sui segmenti caratteristici della rassegna stampa dell'indagine sui giovani definisce due figure ben distinte: una a forma di ellissoide, posizionata tra il I e IV quadrante, che rappresenta "l'uscita dal nido" e l'altra a ferro di cavallo, posizionata tra il II e III quadrante, che rappresenta "il comportamento dei giovani tra le mura domestiche". Al fine di validare i risultati ottenuti con l'analisi delle corrispondenze si è successivamente applicata un'analisi di classificazione gerarchica (metodo di Ward) sulle coordinate fattoriali. L'output dell'analisi ha mostrato che le testate giornalistiche hanno trattato in modo molto eterogeneo i temi proposti dai risultati delle indagini Irp nelle conferenza stampa. Nel caso dell'indagine sugli ideali degli italiani, sono emerse 4 tematiche distinte: l'immigrazione e la presenza straniera (gruppo 1), ruoli familiari ,politiche a sostegno e reti (gruppo 2), il matrimonio (gruppo 3), moglie ideale e lavoro part-time (gruppo 4). Nell'indagine sui giovani, si formano sempre 4 classi, che si caratterizzano sui seguenti temi: il conflitto generazionale(gruppo 1), i problemi economici

connessi all' uscita da casa (gruppo 2), i giovani in famiglia (gruppo 3) e le condizioni per l' uscita dalla famiglia (gruppo 4).

Dall'osservazione della tabella 2 emerge come alcune tematiche siano privilegiate dalla stampa: il tema dell'immigrazione nell' indagine sugli Ideali degli italiani è trattato in 17 articoli e risulta preponderante (43,6% delle testate), mentre per quanto riguarda l' indagine sui giovani il tema che ha catalizzato l' attenzione dei giornali è stato quello delle modalità della vita in famiglia dei giovani, trattato da 21 testate (58,3%).

**Tabella 2**

<i>Tematiche: Indagine sugli ideali</i>	<b>Somma n. occorrenze</b>	<b>Peso % occorrenze</b>	<b>Δc</b>
<i>Gruppo 1 l'immigrazione</i>	9.139	51,7	0,48
<i>Gruppo 2 ruoli familiari</i>	1.968	11,1	0,89
<i>Gruppo 3 matrimonio</i>	5.073	28,7	0,71
<i>Gruppo 4 moglie ideale</i>	1.502	8,5	0,91
	<b>17.682</b>	<b>100</b>	
<i>Tematiche: Indagine sui giovani</i>	<b>Somma n. occorrenze</b>	<b>Peso %</b>	<b>Δc</b>
<i>Gruppo 1 conflitto generazionale</i>	3.413	16,0	0,84
<i>Gruppo 2 problemi economici</i>	1.908	8,9	0,99
<i>Gruppo 3 giovani in famiglia</i>	13.066	61,1	0,39
<i>Gruppo 4 condizioni per l'uscita</i>	3.003	14,0	0,86
	<b>21.390</b>	<b>100</b>	

## 4. Conclusioni

L'analisi testuale ci ha consentito di misurare e valutare attraverso l'analisi delle frequenze e la strumentazione specifica della sua metodologia, le caratteristiche di base del linguaggio giornalistico applicato all'informazione demografica, evidenziando, attraverso la cluster analysis dei segmenti ripetuti, la struttura portante delle rassegne stampa esaminate.

## Bibliografia

- Bolasco S.. (1999) *Analisi multidimensionale dei dati* . Carocci.
- Bolasco S.and Morrone A.. (1998b) A frequency dictionary of polyforms as linguistic database for text disambiguation . In Proc. of *TALTAC.(VI Conference of IFCS)*..
- De Beugrande R, and Dressler W. (1984). *Einführung in die Textlinguistik. Introduzione alla linguistica testuale*. Il Mulino.
- Lebart. L. and Salem. A. (1994). *Statistique textuelle*. Dunod.
- Lebart. L. and Salem. A. Berry J. (1998). *Exploring textual data*. Kluwer Academic Publ.
- Misiti M. (1999). Demographic Information in the Italian Press. Paper presented at European Population conference 1999.
- Muller. Ch. (1977). *Principes et méthodes de statistique lexicale*. Hachette.