



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

**CORSO DI LAUREA IN SCIENZE STATISTICHE, DEMOGRAFICHE E
SOCIALI**

Tesi di Laurea

**L'IMMAGINE DELL'UNIVERSITÀ
FORNITA DAI QUOTIDIANI**

Relatore: Prof. Lorenzo Bernardi

Laureando: Paolo Nadalutti

Anno Accademico 2008/2009

Indice

Introduzione	6
1 Università e giornali di inizio millennio	10
1.1 Storia di una riforma	10
1.1.1 Motivi di una riforma	12
1.2 Alcune note sul mondo giornalistico	14
1.2.1 Gli Uffici Stampa delle università	16
2 Le ipotesi	18
2.1 La prospettiva temporale	20
2.1.1 Quali anni prendere in considerazione?	20
2.1.2 Il dettaglio temporale	20
2.2 I quotidiani analizzati	22
2.3 La selezione degli articoli	23
2.4 Alcuni numeri sul materiale raccolto	24
3 Analisi esplorative	26
3.1 Il vocabolario per forme grafiche	26
3.1.1 Misure lessicometriche	28
3.1.2 Una prima normalizzazione	28
3.1.3 Le specificità	30
3.2 Il vocabolario per segmenti	33
3.2.1 Una seconda normalizzazione	36
3.2.2 Il riconoscimento dei segmenti e la lessicalizzazione	38
3.2.3 Le specificità	40
4 Diamo tempo al tempo	44
4.1 Il ciclo di vita delle parole, un'analisi descrittiva	45
4.2 Ulteriori spunti sul trend	47

5	Quali differenze?	52
5.1	Una breve analisi dei titoli	52
5.1.1	L'interpretazione della soluzione fattoriale	55
5.2	Analisi fattoriale sul corpus esteso: alcune considerazioni pre- liminari	59
5.3	Un primo piano fattoriale	60
5.4	Un secondo piano fattoriale	62
5.4.1	La soluzione fattoriale: i primi e tre fattori	64
5.5	Conclusioni e spunti	66
6	E le opinioni?	72
6.1	I diritti ed i doveri dell'Università	74
6.2	Il vil denaro	84
6.3	La ricerca e la cultura	86
6.4	E il 2001?	87
6.5	Alcune considerazioni	94
	Conclusioni	96
A	Analisi del contenuto: cenni di base	99
A.1	Corpus, vocabolario e unità del discorso	99
A.2	Variabili e specificità	100
A.3	L'indice IS	101
A.4	Il ciclo di vita delle parole	103
	Il rapporto tra indice IT e IT_0	104
	Il modello lineare stimato	104
A.5	L'indice TFIDF	104
B	Analisi delle corrispondenze, cenni di base	107
B.1	Le tabelle di contingenza	107
B.2	Il calcolo della soluzione fattoriale	108
	Variabili attive e passive	108
	L'interpretazione dei fattori	108
	Bibliografia	109

Elenco delle figure

4.1	Parole con significatività del parametro quadratico < 0.01 . . .	48
4.2	Parole con significatività del parametro quadratico < 0.05 . . .	49
4.3	Parole con significatività del parametro quadratico < 0.1 . . .	49
5.1	Piano fattoriale costruito dal vocabolario dei titoli	56
5.2	Piano fattoriale costruito dal vocabolario dei titoli, con Sole-24Ore variabile supplementare	57
5.3	Screeplot	61
5.4	Piano fattoriale I° e II° fattore, con Sole-24Ore	68
5.5	Piano fattoriale I° e II° fattore, Sole-24Ore variabile supplementare . . .	69
5.6	Piano fattoriale I° e III° fattore, Sole-24Ore variabile supplementare . . .	70
5.7	Piano fattoriale II° e III° fattore, Sole-24Ore variabile supplementare . . .	71
6.1	Distribuzione dell'indice TFIDF	75
6.2	Distribuzione dell'indice TFIDF	84
6.3	Distribuzione dell'indice TFIDF	87
6.4	Distribuzione dell'indice TFIDF	89

Elenco delle tabelle

3.1	Forme grafiche di fascia alta	27
3.2	Parole specifiche esclusive per la variabile ANNO	32
3.3	Parole con specificità positiva per l'anno 2001	34
3.4	Parole con specificità positiva per l'anno 2002	35
3.5	Parole specifiche esclusive per la variabile TESTATA	36
3.6	Parole specifiche per i diversi quotidiani, con i relativi p-values	37
3.7	Alcuni segmenti selezionati dal software, con indice IS assoluto e relativo	39
3.8	Segmenti esclusivi per la variabile ANNO	41
3.9	Segmenti esclusivi per la variabile TESTATA	42
3.10	Parole esclusive per la variabile TESTATA, ad esclusione del Sole-24Ore	43
4.1	Parole con trend significativo individuate dall'indice IT	46
4.2	Parole con indice $IT = 0$ con le relative stime dei parametri di regressione e della significatività degli stessi	50
5.1	Alcune tra le forme grafiche piene a maggiore frequenza	54
A.1	Liste di parole utilizzate per il calcolo dell'indice TFIDF	106

Introduzione

Possiamo dire che l'occasione di sfogliare vecchi quotidiani offre, oltre al fascino di notare come in pochi anni cambino lo stile, l'impaginazione, come la carta cambi colore ingiallendo, l'occasione di notare parecchie similarità con il mondo attuale.

Letizia Moratti non è più un pugnace ministro ma è un pugnace sindaco, ma molti altri protagonisti dell'epoca della riforma universitaria relativa al decreto 509 del 1999 sono ancora molto attivi e presenti sulla scena. Per protagonisti non si intende solamente i personaggi politici, ovviamente, ma anche i giornalisti e gli imprenditori; non si parla solo di persone, ma anche di situazioni, di avvenimenti, di proteste e di applausi.

Fino a pochi mesi fa (si sta scrivendo nel giugno del 2009) le università ribollivano e protestavano contro i tagli alla ricerca e alla didattica, esattamente come 8 anni prima. E se stessimo scrivendo nel 2001, anno di avvio della riforma universitaria, potremmo benissimo riferirci alla decade precedente negli stessi termini, se qualche lettore ricorda l'esperienza della Pantera. Questi movimenti sono un esempio di fenomeno comunicativo, prendono forma anche e soprattutto a livello mediatico e una parte del loro potenziale è espresso dalla comunicazione. Oggi l'Onda e la Gelmini, l'altro ieri la Pantera. E ieri? Berlinguer, la Moratti e la Riforma.

Dal 2001, anno di avvio della riforma universitaria sono cambiate molte cose nel mondo accademico. Accanto ai mutamenti strategici, come le finalità dei vari corsi di studio e dei due nuovi livelli di laurea, accanto anche all'architettura generale degli atenei nella loro riorganizzazione amministrativa e gestionale, si è modificata anche l'immagine esterna dell'Università. Denominata anche immagine sociale, è una delle dimensioni sulle quali è andata

maggiormente ad influire la riforma (Bernardi, 2004).

L'immagine dell'università è venuta a delinearsi più chiaramente agli occhi dei media negli ultimi anni (Boldrini, 2005). Prima della legge 150/2000 le università non erano obbligate a mantenere un ufficio stampa. Ma da 15 anni circa a questa parte l'università è diventata sempre di più un centro di comunicazione, in grado di produrre informazioni su se stessa; si è dovuta confrontare anche con quello che è il concetto di notiziabilità che sicuramente porta i fatti negativi, come inchieste e problemi vari del sistema, a godere di tanto spazio mediatico quanto è quello dedicato alle innovazioni e ai contributi portati alla scienza.

Questo spazio mediatico, anche se ultimamente si è diversificato parecchio rispetto agli anni passati, è il campo di battaglia principale sul quale combatte ogni giorno l'immagine di un'istituzione. Nel 2001, qualche passo indietro rispetto ad oggi sul fronte delle tecnologie e dell'accessibilità al web, una buona rappresentazione di questo campo di battaglia era la carta stampata, intesa nel senso dei quotidiani.

L'approcciarsi ai quotidiani utilizzando sistemi prettamente qualitativi (lettura comparata delle testate, dopo aver stabilito criteri adeguati a garantire l'estrazione omogenea delle informazioni) può rivelarsi un'impresa ardua: la grande mole di testo rende difficile il mantenere stabili nel tempo i criteri di analisi, dato che col procedere della lettura possono disvelarsi continuamente nuovi punti di vista, costringendo l'analista a riprendere l'analisi da principio. I metodi dell'analisi del contenuto cercano di mettere ordine nella massa di parole che i quotidiani riversano sulle loro pagine, sfruttando un approccio, possiamo così dire, olistico, ovvero considerando il testo nel suo insieme e bypassando (non del tutto, una lettura del corpus non può essere del tutto nociva) la lunga fase di acquisizione dei dati tramite lettura, anche ripetuta, degli articoli.

L'analisi del contenuto ha assunto negli anni recenti una stabilità sia a livello di tecnica che a livello di metodo, abbastanza elevata da consentire di poter effettuare un'analisi seguendo un percorso ormai "classico", che garantisca un'estrazione di informazione di efficacia ed efficienza comprovati.

Le tecniche dell'ambito dell'analisi del contenuto utilizzate sono state ap-

plicate utilizzando il software Taltac², acronimo di Trattamento Automatico Lessicale e Testuale per l'Analisi del Contenuto di un Corpus, trattandoli, selezionandoli ed esplorandoli con l'utilizzo del DBMS MS Access, analizzando i dati grazie al software statistico R, e aiutandosi spesso, a livello soprattutto di visualizzazione dei dati, con il foglio di calcolo elettronico MS Excel.

Lo scopo di questa tesi è di capire e analizzare quella gamma di elementi che caratterizzano e distinguono fra di loro alcuni dei principali quotidiani italiani ed estrarre tutta la massa di informazioni che è stata pubblicata a riguardo della riforma 509/99, mettendo ordine nelle diverse opinioni. Pertanto non si andrà a verificare se la riforma abbia o meno raggiunto i suoi obiettivi, ma si andrà a testare se l'Università abbia o meno appeal sui quotidiani.

Questa considerazione ci guida immediatamente attraverso tutta una serie di implicazioni: i quotidiani percepiscono allo stesso modo la comunicazione emessa dalle università? c'è intenzionalità comunicativa da parte dell'Università? e se sì, con quale efficacia? coincidono l'immagine che ha di se stessa e quella che percepiscono e diffondono i quotidiani? questa immagine è fedele a se stessa nel tempo?

I capitoli 1 e 2 gettano le basi per la presente tesi. Viene descritta per brevi cenni la situazione dell'Università e della stampa all'inizio del decennio, presentando il cammino legislativo e operativo della riforma universitaria. Non si entrerà nel merito delle valutazioni effettuate da fonti interne o esterne al mondo accademico, per evitare di influenzare a priori le analisi dei capitoli successivi. Si presentano inoltre le ipotesi operative e le basi per l'estrazione di informazione dai testi. Ipotesi come il metodo di costruzione del corpus (quattro quotidiani, scelti in base alla tiratura e alla diffusione) e si motiverà la scelta del periodo di tempo preso in esame (anni 2001 e 2002)

Il capitolo 3 è dedicato principalmente alle analisi esplorative effettuate sul corpus mediante analisi delle specificità. A questo livello i risultati ottenuti non sono facilmente generalizzabili, ma si possono già intuire alcune differenze tra i quotidiani, soprattutto a livello di conferma delle ipotesi di partenza.

Al fine di considerare al meglio il fattore temporale e di capire se determinate parole vengano utilizzate in maniera sempre più intensa o sempre più

raramente, il capitolo 4 presenta un'ulteriore analisi descrittiva del corpus basata sulla serie temporale formata dalle frequenze di ogni parola.

Mediante l'applicazione di un'analisi delle corrispondenze è possibile dare una rappresentazione complessiva del contenuto di un corpus in una mappa fattoriale (Bolasco, 2005). L'analisi fattoriale consente di individuare eventuali similarità tra profili lessicali: ciò consente di identificare e di indagare al meglio le differenze che intercorrono tra i quotidiani e tra le annate. La tecnica è applicata nel capitolo 5.

Inoltre il capitolo 5 scende maggiormente nel dettaglio nel cercare di indagare le diverse aree semantiche che possono essere presenti nel corpus. Se il capitolo 5, sfruttando l'analisi delle corrispondenze, analizza il corpus grazie ad una selezione trasversale del lessico, andando a pescare indifferenziatamente dal corpus, la tecnica del *ranking* TFIDF seleziona le parti da analizzare in modo verticale, consentendo di ottenere quegli articoli che rispondono maggiormente alle ipotesi operative effettuate. Il capitolo 6 ha come obiettivo per l'appunto questa analisi dei singoli articoli, selezionati grazie al *ranking*.

Attraverso lo studio analitico del contenuto dei quotidiani il rischio maggiore è quello di giungere alla scoperta di posizioni e punti di vista che si rivelano poi, alla luce della tipologia di quotidiano analizzato, scontati. Questo pericolo è presente in tutta questa tesi, ma la selezione di quotidiani che per natura sono molto simili fra loro (tre quotidiani "di opinione" a diffusione nazionale) e di un quotidiano maggiormente "tecnico" che possa fare da paragone, garantisce che le eventuali differenze che emergeranno siano reali, e non solamente frutto della natura, per così dire, congenita dei quotidiani.

Capitolo 1

Università e giornali di inizio millennio

1.1 Storia di una riforma

Per capire quali siano stati i motori che hanno spinto la riforma universitaria si deve per lo meno risalire alla fine degli anni ottanta. Seguiremo ora passo passo il cammino legislativo che ha percorso il mondo universitario¹.

La legge 168/89 istituiva il ministero dell'Università e della Ricerca scientifica e tecnologica(MURST), sottraendo alla Direzione per l'istruzione universitaria del ministero della Pubblica Istruzione le competenze che gli erano state proprie per un secolo. Questa legge è importante soprattutto per aver sancito il riconoscimento dell'autonomia statutaria, amministrativa, finanziaria e didattica dell'Università.

La legge 341/90 riguarda gli ordinamenti didattici. E' significativa soprattutto per l'introduzione della tripartizione degli studi universitari, che affianca alla laurea e al dottorato di ricerca il diploma universitario (conseguibile attraverso un ciclo di istruzione breve, biennale o triennale, che consente in linea di principio il trasferimento al corso di laurea).

Il 9 dicembre 1997 viene resa pubblica la cosiddetta bozza Martinotti,

¹Gran parte delle informazioni contenute in questo capitolo sono tratte da CRUI, Scheda sulla legislazione in materia universitaria, 2002

il rapporto finale della Commissione intitolato “Autonomia didattica e innovazione dei corsi di studio a livello universitario e post-universitario”: la terza parte della relazione contiene le proposte di intervento vere e proprie. I punti principali sono due: il sistema dei crediti e i curricula. I crediti vengono a sostituire la definizione dei corsi in annualità e ore per rispondere all’articolarsi delle autonomie dei singoli Atenei, alla maggiore complessità dei percorsi formativi, alla domanda di una loro maggiore flessibilità e fruibilità, alla necessità di una codificazione comune, all’esigenza di incrementare i rapporti tra le istituzioni universitarie sia italiane che europee. Il sistema di crediti è sviluppato coerentemente con quello europeo, l’ECTS, nel quale i crediti quegli ormai familiari numeri che vanno da 1 a 60, e sono associati alle unità di corso per descrivere il carico di lavoro totale che ciascuna unità di corso richiede.

Il 25 maggio 1998, alla Sorbona, il ministro Berlinguer firma coi ministri dell’istruzione francese, inglese e tedesco una dichiarazione congiunta su “L’armonizzazione dell’architettura dei sistemi di istruzione superiore in Europa”. Ecco un passaggio significativo:

“Il riconoscimento internazionale e il potenziale d’attrazione dei nostri sistemi sono direttamente connessi alla loro trasparenza esterna ed interna. Sembra emergere un sistema in cui due cicli universitari principali, uno di primo ed uno di secondo livello, saranno riconosciuti ai fini dell’equiparazione e dell’equivalenza in ambito internazionale. Gran parte dell’originalità e della flessibilità sarà ottenuta attraverso l’utilizzazione dei crediti e dei semestri”.

Tra giugno e ottobre 1998 il ministro Berlinguer invia ai rettori delle università due note di indirizzo, documenti di valore politico ma non normativo: la seconda di queste due note delinea l’architettura del sistema universitario come cornice vincolante per i “decreti di area”. Il punto principale è l’articolazione dei corsi di studio su due livelli. Il primo livello, di durata triennale,

“ha per obiettivo di fornire allo studente una formazione culturale e professionale compiuta, spendibile sul mercato del lavoro, tale da poter dare accesso, di norma, alle attività per le quali attualmente si richiede la laurea, salvo specifiche e motivate eccezioni”.

Il secondo livello durerà due anni, per un ulteriore carico didattico di 120 crediti; i corsi di secondo livello “hanno per obiettivo una formazione culturale e professionale comprensiva della specializzazione”.

Il 19 giugno 1999 viene firmata a Bologna, da tutti i ministri europei dell’istruzione superiore, una dichiarazione congiunta su “Lo spazio europeo dell’istruzione superiore”, che ribadisce la validità del cosiddetto 3+2 e del sistema dei crediti.

Il decreto ministeriale sull’autonomia didattica viene emanato dal ministro Zecchino il 3 novembre 1999 (DM 509/99).

I Decreti ministeriali 4 agosto e 28 novembre 2000 (durante il governo Amato): individuano rispettivamente le 42 classi delle lauree e le 104 classi delle lauree specialistiche;

Febbraio 2001: Testo Unico in materia di Università. Unificherà 350 leggi, cioè la normativa universitaria stratificata in sette decenni di leggi (il precedente Testo Unico sulla pubblica istruzione era del 1933).

1.1.1 Motivi di una riforma

I motivi che hanno spinto il governo a varare una riforma di ampio respiro per l’Università italiana sono noti, e riassumibili essenzialmente in quattro punti (Bernardi, 2002):

- un basso numero di laureati (punto che comprende in una certa qual maniera anche i due seguenti);
- un alto tasso di abbandono (quasi la metà degli studenti che si iscrivevano non conseguiva il titolo di studio);
- una durata eccessiva degli studi;

- un’offerta didattica spesso poco flessibile e distante, in termini di contenuti e di competenze trasmesse, rispetto alle necessità del mondo del lavoro.

Non è obiettivo di questa tesi stabilire se questi obiettivi siano stati o meno raggiunti, ma piuttosto di prendere spunto da quali fossero in realtà gli scopi prefissati in origine dagli organi istituzionali, per poter trarre qualche spunto nell’analizzare le differenze tra i media italiani. Per definire in maniera ancora più precisa quella che è la “missione” che tutto l’ambiente amministrativo, politico ed universitario si diede alla vigilia della riforma, riportiamo i 10 doveri dell’università, redatti dalla Conferenza dei Rettori delle università italiane nel 2001:

1. Adozione di politiche di formazione e di ricerca sempre più aperte alle esigenze della società;
2. Superamento di logiche corporative e individualistiche, a favore degli interessi della comunità universitaria;
3. Affermazione di una nuova deontologia della professione del docente, e crescita di un nuovo rapporto tra docenti e studenti;
4. Centralità della ricerca come investimento primario per la crescita della conoscenza e per lo sviluppo armonico della società;
5. Centralità dei diritti degli studenti nelle attività didattiche degli atenei;
6. Difesa della qualità, senza deroghe, nella didattica, nella ricerca e nel reclutamento del personale;
7. Valutazione costante e accurata del rispetto degli obiettivi prefissati in coerenza con la missione dell’università;
8. Efficienza, trasparenza e comportamenti innovativi nella gestione dell’università;
9. Attenzione al territorio ed in particolare al sistema delle imprese;

10. Rispetto della persona e delle diversità per essere luogo di confronto e di crescita.

In questi dieci punti vengono toccati tutti i settori strategici dell'università e tutte le categorie di persone coinvolte.

I doveri dell'università vengono integrati dai diritti:

1. Pieno riconoscimento e rafforzamento del ruolo dell'università come risorsa strategica nazionale;
2. immediati investimenti nelle università per sostenere la crescita economica e culturale del Paese e la competizione internazionale;
3. adeguati stanziamenti per un'efficace attuazione della riforma didattica universitaria;
4. riconoscimento della funzione innovativa della ricerca universitaria e del suo diritto a competere per l'accesso a tutte le fonti di finanziamento;
5. effettivo diritto allo studio con forme adeguate ed efficaci di supporto agli studenti e alle loro famiglie;
6. precise e trasparenti strategie nazionali di valutazione degli atenei dirette a riconoscere e garantire la loro qualità;
7. adeguati strumenti per un modello innovativo di organizzazione del lavoro e di gestione del personale;
8. Certezze legislative;

1.2 Alcune note sul mondo giornalistico

La stampa italiana nella sua storia ha avuto una particolare evoluzione. La frammentazione storica cui per anni è stata soggetta la penisola ha avuto un influsso anche sullo sviluppo dei quotidiani, favorendo la diffusione di un gran numero di testate locali, a scapito di un modello basato maggiormente su quotidiani di livello nazionale (Sorrentino, 1995). Ciò non toglie che anche

in Italia i ruoli delle persone all'interno del mondo della stampa siano riconducibili ad un modello, e che pertanto il mondo dell'informazione ubbidisca a delle regole diffuse.

Un modello per interpretare il mondo dei media (Rosten, 1937) presuppone che il compito del giornalista consista principalmente nel selezionare, gerarchizzare e presentare la realtà. Ciò ci fa intuire la portata della riduzione della complessità di un fatto reale, una volta che questo diviene notizia. Ma questo non spiega il motivo per il quale questo fatto, reale e complicatissimo nella sua realtà, diventi una notizia. Sicuramente questo processo dipende dal contesto sociale e dall'ambito temporale. Inoltre entrano in gioco anche altre variabili, o meglio attori. L'attuazione del processo di selezione, gerarchizzazione e presentazione è demandata ovviamente ai giornalisti e ai giornali: o più astrattamente alla cultura professionale dei giornalisti e alla cultura organizzativa delle redazioni (o dei luoghi ove si svolga l'attività giornalistica) (Robinson, Robinson 1981). Infatti le scelte lavorative compiute sono la conseguenza, da un lato, del modo in cui nasce e si costruisce la professionalità giornalistica; dall'altro lato dell'organizzazione del lavoro interno alla redazione.

Insomma, se il giornalista con la sua sensibilità è in grado di influenzare il lavoro che esegue, il giornalista stesso sarà influenzato da tutta quella serie di regole, scritte e non, alle quali ubbidisce la redazione. Rimane ancora da capire quale criterio faccia sì che un fatto venga sottoposto al processo di generazione della notizia. Golding e Elliot (1979) definiscono questo criterio, che viene comunemente denominato come criterio di *notiziabilità*: si tratta di una serie di regole pratiche comprendenti un corpus di conoscenze professionali che implicitamente e spesso esplicitamente spiegano e guidano le procedure lavorative redazionali [...]. Sono qualità degli eventi o della loro costruzione giornalistica, la cui relativa assenza o presenza li raccomanda per l'inclusione in un prodotto informativo. Tali qualità degli eventi devono essere applicabili facilmente e rapidamente, flessibili per potersi adattare alla varietà senza fine di eventi disponibili, relazionabili e comparabili, dato che l'opportunità di una notizia dipende sempre da quali altre sono disponibili (Golding, 1979).

Una notizia, una volta creata, segue un suo cammino proprio, come una creatura dotata di vita. Una volta pubblicata, la notizia influenzerà tutte le notizie simili o relative allo stesso argomento che nasceranno. Questo fa sì che le notizie relative alla riforma universitaria (un evento a lungo respiro temporale) siano spesso interrelate fra loro. O meglio, quando per un motivo di attualità (ad esempio la rivolta dei rettori di fine 2002) irrompe nella scena mediatica relativa alla riforma, lo fa portandosi appresso tutti i giudizi passati che sono stati formulati sulla riforma stessa dalla stampa.

1.2.1 Gli Uffici Stampa delle università

Tranne per alcune università che ne sono già dotate, gli Uffici Stampa nascono per legge a partire dall'anno 2000 (con la legge 150/2000). Per i primi anni si limitano ad una comunicazione di tipo prevalentemente istituzionale. Ad esempio vengono comunicate alla stampa novità didattiche e amministrative, l'indire di conferenze stampa o di eventi interni all'università. Si parla insomma soprattutto di eventi di carattere concreto e soprattutto ufficiale. Inoltre in questi anni i media hanno tempi di reazione leggermente più lenti di oggi: mancano ancora alcune delle tecnologie che oggi vengono usate estensivamente (internet) e perciò era molto più usato il contatto diretto e un rapporto personale tra i giornalisti e l'Ufficio Stampa, il che rallentava le relazioni tra le due istituzioni.

Con il passare del tempo (a partire circa dal 2003) la comunicazione si evolve, e i comunicati stampa cominciano a veicolare anche informazioni di carattere meno ufficiale. Ad esempio si cerca di rendere noti i successi della ricerca, o di sponsorizzare eventi a cui partecipa l'università in qualsiasi forma, sia interni che esterni. Lo scopo principale di questa strategia è "sponsorizzare" l'università, in modo da renderla più "attraente" agli occhi della comunità: lo scopo non dichiarato è di catturare fondi. Nel giro di pochi anni, grazie alla spinta data dalla sempre maggiore autonomia degli atenei l'atteggiamento dell'ufficio stampa cambia, passando da organo di supporto, sia per l'università che per i quotidiani, a soggetto attivo di comunicazione, il cui obiettivo diventa quello di selezionare le notizie: ovvero cercare

di cambiare il criterio di notiziabilità. Questo cambiamento si riflette anche sulle pagine dei quotidiani, che costituiscono una serie di organi preferenziali selezionati per comunicare con le università.

Gli anni che verranno considerati nella presente tesi fanno ancora riferimento ad un'epoca di "risveglio" della comunicazione mediatica di ateneo.

Capitolo 2

Le ipotesi

Il titolo di questa tesi richiama tre concetti di non semplice portata: i quotidiani, l'istituzione universitaria e il concetto di immagine.

La scelta di analizzare i quotidiani nazionali nasce da una selezione che potrebbe trarre i suoi elementi da un ventaglio di proposte diverse: esistono e sono analizzabili i anche periodici di opinione o specializzati, gli organi di comunicazione via *web*, la stampa internazionale oppure la stampa di settore o le pubblicazioni scientifiche che riguardino il mondo universitario o in particolare la riforma.

Tutti questi diversi soggetti del mondo dell'editoria potrebbero essere presi in esame per andare ad alimentare un *database* testuale che si rivelerebbe utilissimo nel tracciare in modo ampio e completo quella che si rivelerà essere l'“immagine” dell'Università. La scelta di analizzare proprio i quotidiani prende spunto soprattutto dalla loro diffusione e dalla loro capacità di fornire una quantità significativa di materiale, e non da ultimo dalla disponibilità immediata di materiale sotto forma di testi già digitalizzati¹. Se la scelta di considerare come base del campione proprio i quotidiani nazionali non è da mettersi in discussione, considerare anche altri tipi di materiale mediatico potrebbe dare consistenza alle analisi: confrontare le opinioni dei maggiori quotidiani con quelle di quotidiani minori o “di parte” consentirebbe di an-

¹i siti *web* www.repubblica.it, www.corriere.it, www.lastampa.it, www.sole24ore.it offrono la possibilità, gratuitamente, di usufruire delle proprie banche dati storiche degli articoli.

dare ad evidenziare al meglio le differenze che possono esserci nelle diverse opinioni. Non meno importante sarebbe confrontare il contenuto dei messaggi emessi dai *mass media* con quelli emessi dalla stampa scientifica o di settore, al fine di controllare al meglio se l'Università si rispecchi o meno nella visione che il mondo della stampa ha di essa. Anche la stampa internazionale (dopo aver tradotto gli articoli) potrebbe essere inserita nel monitoraggio effettuato su questo insieme di fonti.

Il secondo concetto coinvolto nel titolo è quello di Università. L'Università che verrà analizzata in questa tesi è intesa come Mondo Accademico, come Istituzione Universitaria, cioè come elemento della società e come istituzione nazionale. Non si prenderanno in considerazione elementi relativi alle realtà locali. L'Università è inoltre costituita da diverse aree che interagiscono fra di loro e che la completano. Elementi come la formazione, la ricerca, la produzione economica, l'amministrazione, l'organizzazione possono essere oggetto di comunicazione; si cercherà di identificare quale comunicazione sia generica e quale invece concorra a delineare singolarmente uno o più di questi profili. L'arco di tempo che prenderemo in considerazione è fondamentale per capire di che Università si stia parlando. Essendo soggetta, oltre che ad una crescita costante e frutto di una naturale evoluzione, ad una crescita "a scalini" frutto di cambiamenti legislativi, prendere in esame gli anni corretti si può rivelare fondamentale.

Il terzo concetto è l'immagine. Dei tre elementi che si sta discutendo è sicuramente quello per il quale sono state spese più parole nel tempo. Letteralmente è una "descrizione più o meno accurata" (definizione data dal dizionario Devoto-Oli). Nel contesto di questa tesi la si deve intendere come il messaggio di uno schema comunicativo. In pratica si va ad intercettare l'immagine dell'Università non per come è percepita dai lettori, ma per come è emessa o percepita dai giornali.

La possibilità affascinante di poter costruire uno schema² in cui la fonte di comunicazione (l'Università) invia un messaggio ai quotidiani i quali lo ritrasmettono, e la possibilità di poter andare ad intercettare questa "doppia

²facendo riferimento al modello comunicativo della teoria matematica dell'informazione elaborato da Shannon-Weaver

comunicazione” al livello di emissione da parte dell’Università e di ricezione e di ri-trasmissione del messaggio da parte dei quotidiani è, ai fatti, limitata in alcuni aspetti dal corpus di cui si dispone.

Non di meno si rinuncerà a provare a testare tutte le ipotesi qui sopra presentate, cercando di sfruttare al meglio il materiale a disposizione.

2.1 La prospettiva temporale

2.1.1 Quali anni prendere in considerazione?

Per considerare nell’analisi la maggior parte di informazione possibile a riguardo della riforma universitaria, la scelta non può che cadere sul biennio 2001-2002. Il 2001 è un anno cruciale per la vita universitaria e politica in Italia: se a Maggio hanno avuto luogo le elezioni politiche per la costituzione della XIV legislatura, a settembre veniva ufficialmente avviata la riforma universitaria; sembra poi naturale prolungare l’analisi fino alla fine dell’anno accademico 2001/2002, e proseguire nell’analizzare anche la parte finale dell’2002, che comprende l’inizio dell’Anno Accademico successivo. In definitiva si monitorano due anni solari interi, un anno accademico intero e due mezzi anni accademici.

La scelta alternativa di prendere in considerazione due anni accademici consecutivi (e non due anni solari) non avrebbe consentito di poter capire quali fossero le opinioni preliminari dei quotidiani all’inizio della riforma, nonché quelle che seguono il primo anno accademico “riformato”.

2.1.2 Il dettaglio temporale

Durante un anno accademico, ma anche in un anno solare, a diversi periodi di tempo corrispondono diversi criteri di notiziabilità. In altre parole, bisogna cercare di capire quale dettaglio temporale consente di estrarre la maggior informazione dai testi. Tale criterio non è semplice da individuare, dato che spesso i cicli di vita di una notizia sono molto variabili, e si possono avere dei cicli lunghi uno o più mesi oppure cicli lunghi pochi giorni.

Il trade-off che separa la scelta di una unità di tempo-base breve da una lunga è la quantità di informazione contenuta in ogni periodo: scegliendo un'unità base breve (il mese) si può andare incontro al rischio di avere qualche unità di tempo nella quale non è stato scritto alcun articolo, senza contare la possibilità di generare confusione nella grande quantità di dettagli che hanno origine da tale suddivisione; un'unità base ampia (semestrale, ad esempio) rischia di non riuscire a monitorare con la dovuta efficacia molti avvenimenti che verrebbero oscurati da avvenimenti molto duraturi nel tempo.

D'altronde le dimensioni del nostro corpus ci indicano che la scelta semestrale si rivela la più adatta. Le tabelle che seguono cercano di fornire un quadro della cadenza da un lato degli eventi reali del mondo accademico (o di ciò che lo circonda), dall'altro quello del ritmo di pubblicazione di articoli.

Anno	Mesi	Avvenimenti di interesse
2001	Gennaio-Maggio	Inaugurazioni anno accademico, elezioni politiche.
	Giugno-Agosto	Esami di stato, scelta delle facoltà da parte delle matricole.
	Settembre-Dicembre	Inizio dei corsi.
2002	Gennaio-Maggio	Inaugurazioni anno accademico.
	Giugno-Agosto	Esami di stato, scelta delle facoltà da parte delle matricole.
	Settembre-Ottobre	Inizio dei corsi, varo della prima manovra finanziaria del neo-governo.

	2001		2002		Totale Complessivo
	Primo Semestre	Secondo Semestre	Primo Semestre	Secondo Semestre	
	Corriere	47	15	7	
Sole-24Ore	66	32	46	41	185
La Stampa	15	22	15	19	71
Repubblica	17	11	18	39	85
Totale complessivo	145	80	86	120	431

Come possiamo vedere c'è un netto sbilanciamento del numero degli articoli: la maggior parte viene pubblicata nel primo e nell'ultimo dei semestri presi in considerazione, mentre il Sole-24Ore è, semestre per semestre il quotidiano che pubblica di più.

La punta di maggior ritmo di pubblicazione si ha nel marzo 2001, mese in cui il Sole-24Ore e il Corriere assieme pubblicano quasi 40 articoli.

Considerando una frequenza più alta (quadrimestri, trimestri) il rischio di ottenere sub-corpora dalle dimensioni limitate è molto alto.

2.2 I quotidiani analizzati

Al giorno d'oggi i quotidiani a maggior diffusione, con una tiratura superiore alle 300.000 copie non sono molti, e soprattutto alcuni rimangono legati alle loro origini locali, prediligendo nella diffusione delle notizie tutto ciò che riguarda la propria regione di appartenenza.

Il criterio di scelta dei quotidiani da analizzare è sostanzialmente il numero di copie vendute e il grado di diffusione sul territorio nazionale. Nonostante tutti i quotidiani che sono stati selezionati pubblichino regolarmente delle edizioni locali, la loro tiratura (e vendita) ha portata nazionale.

Premesso ciò, la nostra scelta nella selezione del materiale è ricaduta su:

- il Sole 24Ore: (tiratura 343.855 copie) un quotidiano che tratta per la maggior parte questioni legate all'economia, che pubblica articoli dal taglio molto tecnico;
- la Repubblica: (tiratura 630.080 copie) è il primo vero quotidiano concepito sin dalla nascita come "nazionale";
- il Corriere della Sera: (tiratura 681.854 copie) per anni il quotidiano più venduto sul suolo nazionale, è nato a Milano a fine '800, per espandersi poi nel tempo per numero di copie vendute e tematiche trattate;
- la Stampa: (tiratura 317.092 copie) storicamente è il quotidiano della città di Torino. Nel tempo, data la sua alta diffusione a livello nazionale

ha allargato anche il bacino da cui trarre le notizie, rimanendo però sempre legato alla regione del Piemonte.

I dati sulla tiratura sono aggiornati agli anni 2005 e 2006, valore medio annuo fonte www.adsnotizie.it. A ben vedere, questi quotidiani rappresentano una buona parte della stampa nazionale italiana.

2.3 La selezione degli articoli

Non si può prescindere da una selezione coerente del materiale all'interno dei quotidiani. Se l'orizzonte temporale ci costringe senza fallo a selezionare gli articoli a partire dal primo di gennaio 2001 al 31 dicembre 2002, non altrettanto semplice si presenta la ricerca e la selezione del materiale compreso tra queste due date.

I quotidiani che sono stati scelti per l'analisi mettono a disposizione dei lettori in formato digitale il loro archivio storico, con tanto di motore di ricerca e pertanto forniscono la possibilità di rintracciare i testi in base al loro contenuto. I criteri di ricerca scelti sono nella forma:

“rintracciare tutti gli articoli che contengono le parole *università* e *riforma*”.

C'è solamente una remota possibilità, in verità non sondata, che la ricerca non individui qualche articolo che, pur parlando della riforma universitaria, non contenga queste due parole: pertanto l'analisi sarà al limite tra campionamento e analisi censuaria dei dati. Certamente tra gli articoli rintracciati sono presenti molti articoli che, pur parlando di tutt'altro, contengono queste due parole. La “pulizia” del corpus è stata eseguita manualmente, eliminando tutti quegli articoli che non hanno come argomento principale la riforma universitaria.

A metà strada tra l'argomento della ricerca e altro si collocano due argomenti che in quegli anni sono stati molto d'attualità, e toccavano marginalmente il mondo universitario: la riforma scolastica e quella degli ordini professionali, legate entrambe al mondo universitario.

La prima di queste due riforme complementari alla nostra ha avuto am-

pia copertura mediatica, mentre la seconda ha animato il dibattito solamente presso circoli molto specifici: si è scelto di mantenere nel corpus gli articoli che riguardano questi argomenti solamente quando i riferimenti alla riforma universitaria presenti nel testo fossero palesi e rievanti. La creazione di un criterio rigido per la discriminazione di tali articoli non sarebbe stato affidabile, in quanto basarsi sulla presenza o meno di alcune parole (ad esempio “albi” oppure “scuola”) non avrebbe necessariamente stabilito il reale contenuto dell’articolo.

2.4 Alcuni numeri sul materiale raccolto

In totale sono stati raccolti 431 articoli, 225 scritti nel 2001 e 206 nel 2002, dei numeri abbastanza simili, che testimoniano una distribuzione piuttosto uniforme nel tempo.

La tabella seguente riporta il numero di articoli suddivisi per trimestre e per testata:

	Anno 2001	Anno 2002	Totale
Corriere	62	28	90
Sole-24Ore	98	87	185
Repubblica	37	34	71
Stampa	28	57	85
TOTALE	225	206	431

Il Sole-24Ore è il quotidiano che pubblica di più sull’argomento. Seguono Corriere e Stampa, e per ultima la Repubblica. Tra questi il Sole-24Ore è il quotidiano che ha più contatti con gli uffici stampa delle università in qualità di organo informatore sia di eventuali nuove tecnologie, sia di ogni possibile innovazione didattica o formativa. Notoriamente si tratta di un quotidiano dedicato a coloro che sono interessati a nuove figure professionali od a possibili risorse intellettuali.

La distribuzione degli articoli tra i due anni è essenzialmente omogenea (225 articoli nel 2001, 206 nel 2002), anche se con vistose differenze tra i

contributi che i singoli quotidiani apportano al totale. Colpiscono soprattutto le differenze interne di Corriere e Stampa: il primo che pubblica soprattutto nel 2001, il secondo nel 2002.

Capitolo 3

Analisi esplorative

Il programma scelto per l’elaborazione dei dati testuali è Taltac¹. Per rendere il file gestibile più facilmente da tale software si è proceduto ad eliminare molti dei refusi dovuti all’importazione dei caratteri di testo da pagine HTML. La codifica ASCII trasforma infatti alcuni simboli utilizzati nei testi HTML in simboli diversi. Così ad esempio sono stati sostituiti i “trattini lunghi” (codice ASCII 45) con dei “trattini brevi” (codice ASCII 173). Questa pulizia a priori del corpus sgrava l’elaborazione di TalTac da un lavoro molto time-consuming: la preparazione del testo ha interessato perciò anche:

- eliminazione di errori ortografici;
- trasformazione degli apostrofi in accenti;
- regolarizzazione dei nomi più celebri che non verrebbero riconosciuti da TalTac (Letizia Moratti diventa Letizia.Moratti);
- le sigle di associazioni o enti non riconosciute dal software sono state trasformate nella loro forma estesa.

3.1 Il vocabolario per forme grafiche

Il parsing ci restituisce un corpus costituito da 266.438 forme grafiche, una buona quantità per rispondere ad un trattamento statistico del testo (Giu-

¹www.taltac.it

liano, 2008): il vocabolario per forme grafiche ha una dimensione di 22.307 termini. La tabella 3.1 ci mostra un risultato preliminare che proviene direttamente dal caricamento dei dati. Come per quasi tutti i possibili corpus, le forme grafiche più utilizzate sono congiunzioni, articoli o preposizioni; ma se consideriamo solamente le parole piene (che hanno significato in sé) di fascia alta (vedi appendice A.1) possiamo individuare quelli che sono i possibili “titoli” che si potrebbero assegnare al corpus (Tuzzi, 2003). Infatti le prime due parole piene che incontriamo sono *riforma* e *università*. Certamente i due termini si possono riferire a concetti differenti (riforma dell’università, oppure riforma degli ordinamenti, oppure università di Roma - o Milano-etc.). Queste due parole in particolare presentano un numero di occorrenze molto alto in quanto sono le parole chiave utilizzate per estrarre gli articoli dai quotidiani.

Per ora limitiamoci a constatare come compaiano in questa lista solamente parole che si riferiscono ad un mondo “interno” all’università e la prima forma grafica riferita a qualcosa di altro è *ministro*: una parola che proviene dalla politica.

Forma grafica	Rango	Occorrenze totali
di	1	10598
...
da	23	1376
riforma	24	1349
...
università	30	1150
degli	31	1075
anche	32	1020
alla	33	1010
laurea	34	982
studenti	35	898
...
ministro	61	435

Tabella 3.1: Forme grafiche di fascia alta

3.1.1 Misure lessicometriche

Come abbiamo già detto il corpus è composto da più di 250 mila forme grafiche, delle quali 11.083 sono hapax, cioè parole che compaiono nel corpus una sola volta. Prendendo in considerazione anche l'estensione del corpus possiamo ricavare alcune delle statistiche di riferimento (Tuzzi, 2003) per valutare la possibilità di effettuare un'analisi statistica del corpus. Dato N il totale delle occorrenze all'interno del corpus (in forme grafiche differenti), V l'ampiezza del vocabolario (in numero di forme grafiche), V_1 il numero di hapax, abbiamo che:

- l'indice di ricchezza lessicale $\frac{V}{N} = 8.37\%$
- la percentuale di hapax vale $\frac{V_1}{V} = 49.68\%$
- la frequenza media $\frac{N}{V} = 11.94$
- il coefficiente α approssimato della legge di Zipf $\frac{\log N}{\log V} = 1.25$

Dato che si dispone di un corpus sufficientemente esteso non ci sarebbe potuto aspettare che questi indicatori potessero assumere valori anomali. Il loro calcolo comunque mostra come non ci siano evidenti anomalie nel testo.

3.1.2 Una prima normalizzazione

Il vocabolario appena costruito è composto da una serie di forme grafiche che differiscono molto fra di loro per la forma in cui sono salvate. Se in sede preliminare sono stati eliminati molti difetti dei testi dovuti all'importazione dei dati, permangono tuttavia molte forme non univoche dei dati. Taltac riconosce come differenti le lettere maiuscole e minuscole: per questo corpus ciò significa per il software attribuisce valori diversi alle forme grafiche *RI-FORMA* scritta in maiuscolo in quanto compare nei titoli, e *rimforma* scritta in minuscolo in quanto compresa nei testi. Questa irregolarità è presente per molte forme grafiche del vocabolario, ma può essere risolta procedendo ad una normalizzazione del testo. I parametri scelti per la normalizzazione sono:

- riconoscimento dei nomi propri (che potrà distinguere *Letizia* nome proprio del ministro dalla *letizia* sostantivo);
- riconoscimento delle celebrità (compaiono molti nomi di ministri, di uomini di cultura e di spettacolo nel corpus);
- riconoscimento dei giornali (ovviamente);
- riconoscimento dei toponimi, titoli, sigle e società;
- riduzione sia delle minuscole ad inizio parola, sia delle parole con tutti i caratteri in maiuscolo.

Per tutte le forme grafiche che non sono state riconosciute come appartenenti alle categorie sopraelencate si forza sia la riduzione delle maiuscole iniziali, sia la riduzione delle parole scritte tutte in maiuscolo. Nonostante TalTac sia in grado di riconoscere le parole con iniziale maiuscola precedute da punteggiatura indicante l'inizio di un periodo, forziamo la riduzione delle maiuscole in quanto la punteggiatura (soprattutto nei titoli) non è sempre ben riportata. Insomma, si preferisce perdere l'informazione relativa ad alcuni nomi propri (*Rossi* nome o *rossi* aggettivo) piuttosto che avere informazione scorretta, derivante da una cattiva importazione della punteggiatura (ricordiamo che i testi sono stati importati da web, e pertanto possono essere possibili alcuni refusi). Un esempio di questa situazione è esposto nel periodo seguente:

La recente riforma dei concorsi, poi, non ha funzionato. “Va recuperato il sistema del vincitore unico in luogo del vigente sistema dei due candidati idonei”, ha proposto la Moratti. I temi chiave del progetto di Moratti STATI GENERALI DELLA SCUOLA Vi faranno parte genitori, insegnanti, studenti.

Il secondo periodo comincia dopo un punto, ma l'inizio del secondo periodo non è altro che un titoletto all'interno dell'articolo e pertanto privo di punteggiatura. Il riconoscimento da parte del lettore della diversa parte dell'articolo è affidato all'interpretazione del testo tutto in maiuscolo, o forse ad un formato differente del carattere (al quale non possiamo risalire).

Perciò il riconoscimento automatico delle maiuscole potrebbe confondere una fra le parole *STATI GENERALI DELLA SCUOLA* come un nome proprio, e mantenere la maiuscola iniziale.

La riduzione totale dei caratteri maiuscoli ad eccezione di quelli che vengono riconosciuti come facenti parte di parole che compaiono nel database di Taltac comporta la perdita di alcuni nomi propri e di quei riferimenti a categorie astratte come *Università* o *Scuola*, che, sebbene importanti, non vengono utilizzate in maniera uniforme dai giornalisti, che, interpretando liberamente quella che è una mancanza di regole della nostra lingua, non sempre utilizzano la maiuscola iniziale.

3.1.3 Le specificità

Taltac è in grado di riconoscere quali unità lessicali siano peculiari per una data partizione del corpus, riconoscendo tali parole in base ad un loro sopra o sotto utilizzo per quella data partizione. L'identificazione è basata sul risultato di un test con distribuzione ipergeometrica applicato alle suboccorrenze di ogni parola. Il corpus che stiamo elaborando dispone di una serie di informazioni che sono collegate ad ognuno degli articoli, informazioni riferite sia al periodo temporale della pubblicazione, sia alla testata, sia all'autore. Per orientarci meglio in queste analisi preliminari e non eccedere nel dettaglio di informazioni, prendiamo in esame due delle possibili partizioni: l'anno di pubblicazione che è una divisione temporale ad ampio respiro, e la testata di appartenenza dell'articolo.

(vedi appendice A.2)

La soglia presa in considerazione per l'individuazione delle forme specifiche corrisponde ad un p-value di 0.025 (vedi Appendice A.2), che consente di selezionare dal vocabolario una lista di forme grafiche abbastanza ristretta da essere facilmente analizzata. Le forme grafiche che compaiono nelle tabelle di questo capitolo, oltre che in base alla probabilità calcolata da TalTac, sono state selezionate in base a criteri personali: si è cercato di fornire un esempio per ogni tipologia semantica delle parole.

La variabile anno

La tabella 3.2 mostra un esempio di evoluzione temporale degli articoli, e grazie alle forme grafiche esclusive rispetto all'annata di pubblicazione, notiamo come alcuni personaggi e alcune tematiche scompaiano. Si tratta di forme grafiche riferite a personaggi politici come membri di commissioni o comitati (*Serao* -presidente del Consiglio Nazionale dei Dottori Commercialisti-, *Validitara* -responsabile nazionale per la Scuola-, *Giampaolo Rossi* -presidente della commissione per il raccordo dei vecchi e nuovi titoli di laurea-), oppure di rettori universitari (*d'Ascenzo*). Insomma si tratta di termini molto tecnici, come tutti quelli che riguardano particolari figure professionali: *commercialisti*, *revisori*, *geometra*.

Altri termini (*libro bianco*, *meeting*, *Rimini*, *iunior* -sic-) sono riferiti ad eventi di attualità, e pertanto, al pari dei personaggi politici, soggetti ad una decadenza della loro notiziabilità.

Tutte queste forme grafiche occupano la fascia bassa del vocabolario, anche se più della metà si trovano nel primo decile delle basse frequenze (soglia considerata utile per le analisi sul vocabolario (Giuliano, 2008)).

Volendo approfondire maggiormente l'evoluzione temporale del lessico, possiamo fare riferimento alle tabelle 3.3 e 3.4, che contengono alcune delle forme grafiche scelte a discrezione del ricercatore con specificità positiva rispetto alla variabile anno.

Possiamo notare come i gruppi semantici di appartenenza delle forme grafiche relative ai due anni siano piuttosto differenti: i termini che compaiono negli articoli nel 2001 sono molto più specifici del mondo universitario rispetto a quelli del 2002, anno durante il quale prendono il sopravvento riferimenti di ampio respiro come *cultura* e *scienza*. Ovviamente molte forme grafiche risentono di eventi contingenti (*albo* e tutte le professioni legate alla riforma degli albi come *geometri*, *ingegneri* per il 2001; *risorse* e *tagli* per il 2002).

Alcuni termini per i quali ci si potrebbe aspettare una distribuzione piuttosto uniforme nei due anni sono quelli legati più strettamente al mondo universitario. Emblematica ad esempio è la posizione della parola *ricerca*: compare 129 volte in più nel 2002, a fronte di una dimensione del corpus

Forma Grafica	Occorrenze anno 2001	Occorrenze anno 2002	Fascia
commercialisti	94	0	Bassa
ascenzo	35	0	Bassa
commercialista	30	0	Bassa
revisori	24	0	Bassa
serao	21	0	Bassa
Giampaolo	21	0	Bassa
concerto	20	0	Bassa
iunior	16	0	Bassa
transitorio	15	0	Bassa
perito	15	0	Bassa
regolamentate	14	0	Bassa
4/99	13	0	Bassa
Palumbo	12	0	Bassa
agronomi	12	0	Bassa
elenchi	11	0	Bassa
definisce	11	0	Bassa
geometra	11	0	Bassa
santorelli	10	0	Bassa
Sessantotto	10	0	Bassa
crusca	10	0	Bassa
conferenze	10	0	Bassa
supplemento	10	0	Bassa
agronomo	10	0	Bassa
valditara	0	21	Bassa
meeting	0	19	Bassa
Rimini	0	17	Bassa
bollino	0	16	Bassa
stanziamento	0	14	Bassa
occasioni	0	12	Bassa
restauro	0	12	Bassa
campusone	0	11	Bassa
anticipo	0	11	Bassa
giustamente	0	11	Bassa
cesana	0	10	Bassa
citati	0	10	Bassa
stl	0	10	Bassa
Dpef	0	10	Bassa
Buttiglione	0	10	Bassa

Tabella 3.2: Parole specifiche esclusive per la variabile ANNO

inferiore.

Il tipo di giornale

Le differenze tra gli utilizzi di parole originali tra differenti quotidiani è visibile in tabella 3.5, che mostra come l'unica testata a differenziarsi sostanzialmente dalle altre sia il Sole-24Ore. Ovviamente ciò è dovuto a due fattori piuttosto evidenti:

- la maggiore dimensione del sub-corpus del Sole-24Ore, che tende a “schiacciare” il lessico delle altre testate;
- la mission particolare del Sole-24Ore, che si permette un linguaggio composto anche da termini tecnici.

Per meglio apprezzare le differenze tra i vari quotidiani possiamo fare riferimento alla tabella 3.6.

Ancora una volta emerge come il lessico utilizzato dal Sole-24Ore (le prime 18 righe della tabella 3.6) sia differente rispetto a quello relativo alle altre testate: le forme grafiche che presentano specificità positiva sia per il Sole-24Ore che per un altro quotidiano qualsiasi sono rarissime. Inoltre le parole sottoutilizzate dal Sole-24Ore hanno in molti casi specificità positiva per più di un quotidiano fra i restanti.

3.2 Il vocabolario per segmenti

Il vocabolario per forme grafiche può non cogliere tutta l'informazione presente nel corpus: soprattutto il livello di ambiguità di molte parole rimane alto. Abbiamo ad esempio già accennato a come il termine *riforma* si riferisca contemporaneamente a tre eventi relativi al biennio 2001-2002, cioè la riforma universitaria, quella scolastica e quella degli ordini professionali. Per aumentare la portata informativa delle singole parole verranno prese in analisi delle sequenze di parole presenti nel testo. Ad esempio considereremo come una singola parola la sequenza *riforma dell'università*, che, nella sua

Forma Grafica	Frequenza 2001	Frequenza 2002	p-value
università	940	783	0.02
riforma	846	573	0.00
laurea	601	417	0.00
studenti	518	411	0.01
laureati	264	96	0.00
lavoro	240	174	0.01
universitaria	238	145	0.00
ordini	176	16	0.00
autonomia	149	88	0.00
triennale	143	95	0.01
Roma	141	108	0.08
crediti	138	78	0.00
professioni	134	31	0.00
ministero	133	79	0.00
ateneo	125	84	0.01
specialistica	124	85	0.02
dottori	114	7	0.00
rettore	109	74	0.02
albo	101	5	0.00
iscritti	99	65	0.02
guerzoni	88	2	0.00
studente	83	35	0.00
ragazzi	76	51	0.05
ingegneri	74	22	0.00
ragionieri	71	1	0.00
ordine	69	31	0.00
Politecnico	68	36	0.00
tasse	63	26	0.00
architettura	57	24	0.00
Berlinguer	56	21	0.00
Cnr	46	21	0.00
CUN	44	19	0.00
umanistiche	42	26	0.07
Cattolica	38	12	0.00
lire	35	20	0.06
Zecchino	31	10	0.00
Giustizia	31	4	0.00
de Mauro	28	6	0.00
Modica	25	13	0.06
panebianco	23	3	0.00
psicologia	22	6	0.00
geometri	22	2	0.00
sociologia	21	10	0.06
corteo	19	1	0.00
medici	18	3	0.00
chimici	18	2	0.00
Medicina	17	6	0.03
biotecnologie	17	5	0.01
chimica	16	6	0.04
Accademia	16	5	0.02
elezioni	16	2	0.00
contestazione	14	6	0.08
rivoluzione	14	5	0.05
Polo	14	3	0.01
Ortensio	14	3	0.01

Tabella 3.3: Parole con specificità positiva per l'anno 2001

Forma Grafica	Frequenza 2001	Frequenza 2002	p-value
ricerca	231	362	0.00
ministro	182	266	0.00
scuola	216	260	0.00
Moratti	99	239	0.00
formazione	188	210	0.02
docenti	153	198	0.00
Italia	124	167	0.00
istruzione	123	174	0.00
risorse	66	162	0.00
governo	94	159	0.00
3+_2	99	150	0.00
didattica	102	119	0.05
euro	25	92	0.00
milioni	23	92	0.00
fondi	43	91	0.00
rettori	44	90	0.00
valutazione	76	89	0.07
cultura	50	82	0.00
Paese	61	75	0.06
Finanziaria	11	71	0.00
politica	40	58	0.02
Tremonti	2	58	0.00
scienza	15	56	0.00
senato	30	53	0.00
filosofia	36	51	0.03
tagli	16	47	0.00
maggioranza	32	44	0.05
Ciampi	4	44	0.00
aziende	30	41	0.06
giurisprudenza	26	40	0.03
europeo	21	39	0.01
internazionale	21	32	0.05
Berlusconi	13	32	0.00
Bologna	17	27	0.05
Bocconi	11	27	0.00
sperimentazione	8	26	0.00
Siena	5	26	0.00
dimissioni	3	26	0.00
de_maio	5	25	0.00
an	5	25	0.00
tosì	2	25	0.00
Ocse	9	23	0.01
letteratura	4	22	0.00
Francia	11	18	0.09
banca	9	16	0.08
mobilitazione	3	16	0.00
facoltà_di_giurisprudenza	2	16	0.00
Pavia	8	15	0.07
Sud	5	14	0.02
Pil	7	13	0.10
Nobel	3	13	0.01
Ds	3	12	0.01
facoltà di economia	4	11	0.04

Tabella 3.4: Parole con specificità positiva per l'anno 2002

Forma Grafica	Occorrenze Sole-24Ore	Occorrenze Repubblica	Occorrenze Stampa	Occorrenze Corriere
abilitati	27	0	0	0
contabili	16	0	0	0
operatori	14	0	0	0
regolamentate	14	0	0	0
direttive	13	0	0	0
4/99	13	0	0	0
consulenti	13	0	0	0
campusone	11	0	0	0
santorelli	10	0	0	0
stl	10	0	0	0
disposizioni	10	0	0	0
esaurimento	10	0	0	0
restauro	0	12	0	0
citati	0	10	0	0
Pacini	0	0	10	0
meneghello	0	0	10	0

Tabella 3.5: Parole specifiche esclusive per la variabile TESTATA

unitarietà riduce l'ambiguità che è insita nel significato delle tre parole che la formano prese singolarmente.

3.2.1 Una seconda normalizzazione

Il software Taltac è in grado di individuare automaticamente nel testo molte espressioni di uso comune composte da più parole, che vengono dette poliformi di uso comune. Queste espressioni possono essere individuate a priori durante la fase di normalizzazione, sfruttando una banca dati interna al software in grado di riconoscere questi poliformi come non ambigui e di marcarli. I parametri di normalizzazione utilizzati nella costruzione del vocabolario per forme grafiche vengono ripetuti anche per la costruzione del vocabolario per segmenti, ma chiederemo al software di ricercare anche alcune forme lessicali note come le polirematiche. Per lo scopo delle nostre analisi non siamo interessati alle locuzioni grammaticali, utilizzate generalmente nelle analisi di tipo lessicale.

Forma Grafica	Corriere	Sole 24Ore	Stampa	Repub -blica	Corriere	Sole 24Ore	Stampa	Repub -blica
riforma	pos	pos	neg	neg	0,01	0,03	0,04	0,00
corsi	pos	pos	neg	neg	0,00	0,00	0,00	0,00
lauree	pos	pos	neg	neg	0,00	0,00	0,03	0,00
triennale	pos	pos	neg		0,10	0,05	0,00	0,19
lavoro	neg	pos		neg	0,05	0,00	0,21	0,01
formazione	neg	pos	neg	neg	0,00	0,00	0,06	0,00
studi	neg	pos		neg	0,03	0,00		0,04
laureati	neg	pos	neg	neg	0,00	0,00	0,00	0,00
3_+_2	neg	pos		neg	0,01	0,00	0,36	0,00
specialistica	neg	pos	neg	neg	0,02	0,00	0,00	0,02
risorse	neg	pos			0,00	0,01	0,22	0,23
didattica	neg	pos			0,00	0,00	0,18	0,21
atenei		pos	neg	pos	0,26	0,00	0,00	0,02
laurea		pos	neg	neg	0,32	0,00	0,00	0,00
universitaria		pos	neg	neg	0,25	0,00	0,00	0,02
autonomia		pos	neg	neg	0,47	0,00	0,00	0,00
crediti		pos	neg	neg	0,16	0,00	0,00	0,00
ministero		pos	neg	neg	0,13	0,00	0,01	0,00
università	pos	neg	neg	pos	0,00	0,00	0,00	0,00
studenti	pos	neg		pos	0,00	0,00	0,49	0,10
governo	pos	neg	pos	pos	0,00	0,00	0,02	0,00
scuola	pos	neg	pos		0,01	0,00	0,00	0,34
ministro	pos	neg	pos		0,00	0,00	0,00	0,48
Moratti	pos	neg	pos		0,01	0,00	0,00	0,11
ateneo	pos	neg	pos		0,00	0,00	0,02	0,39
facoltà	neg	neg	pos	pos	0,06	0,00	0,00	0,00
ricerca		neg	pos	pos	0,22	0,00	0,00	0,00
docenti		neg		pos	0,20	0,03	0,12	0,00
giovani		neg		pos	0,49	0,07		0,04
istruzione	pos		pos	neg	0,10	0,26	0,02	0,00
Italia	neg		pos	pos	0,00	0,34	0,03	0,04
presidente	neg				0,07	0,14		0,44
Roma			neg	pos		0,15	0,07	0,00
scienze			pos	neg	0,11	0,44	0,00	0,00
corso			neg		0,24	0,17	0,01	0,48

Tabella 3.6: Parole specifiche per i diversi quotidiani, con i relativi p-values

3.2.2 Il riconoscimento dei segmenti e la lessicalizzazione

I segmenti sono unità di testo che si ripetono nel corpus. Taltac è in grado di individuare i segmenti, agendo in funzione di alcuni parametri:

- soglia di frequenza minima delle parole appartenenti al segmento: se di default vengono considerati anche gli hapax, per la nostra analisi possiamo selezionare come parametro il valore di frequenza superiore o uguale a 2;
- numero massimo di parole nel segmento: effettueremo la ricerca su segmenti costituiti al massimo da 10 forme grafiche.

La ricerca dei segmenti può essere integrata da un indice che possa misurare quanto le singole parole che costituiscono il segmento siano utilizzate specificatamente all'interno dello stesso. Questo indice prende il nome di indice *IS* che viene calcolato in base alla formula:

$$IS = \left[\sum_{i=1}^L \frac{f_{segm}}{f_{fg_i}} \right] \times P$$

dove L è il numero di forme grafiche del segmento, f_{segm} è la frequenza del segmento del quale calcolare l'indice, f_{fg_i} è la frequenza della forma grafica i che fa parte del segmento, e P è il numero di parole piene che costituiscono il segmento. Si rimanda all'appendice A.3 per ulteriori informazioni.

Applicando alla ricerca dei segmenti i parametri menzionati il software seleziona 19.842 segmenti, alcuni dei quali - elencati in un ordine tale da mettere in evidenza le ambiguità insite in alcuni dei termini che li compongono - sono illustrati in tabella 3.7 assieme al corrispondente indice *IS* assoluto e relativo. Insomma, se la parola *rimforma* venisse sempre usata assieme alle parole *dell'università*, l'indice *IS* per il segmento *rimforma dell'Università* avrebbe un valore molto alto.

Il gruppo di segmenti ritenuti utili alla ricerca consiste di 231 elementi (4.472 occorrenze), la cui somma delle frequenze ammonta a 4.531 . Come possiamo notare dalla tabella la selezione dei segmenti non viene effettuata

basandosi solamente sul calcolo dell'indice IS (alcuni segmenti presentano un valore molto basso dell'indice sia assoluto che relativo), ma anche cercando di cogliere tutti quei segmenti che possono aumentare la quantità di informazione riassunta dal gruppo dei segmenti stessi.

Segmento	Occorrenze	Indice IS	Indice IS
	Totali	Assoluto	Relativo
laurea specialistica	162	1.92	0.48
laurea triennale	141	1.50	0.38
corsi di laurea	151	0.84	0.09
fuori corso	41	0.59	0.15
corso di studi	22	0.31	0.03
ordini professionali	40	0.86	0.22
sbocchi professionali	30	1.59	0.40
riforma Moratti	19	0.14	0.03
riforma della scuola	33	0.23	0.03
riforma dei cicli	31	1.05	0.12
riforma Berlinguer	19	0.52	0.13
contro la riforma	13	0.09	0.01
riforma universitaria	206	1.38	0.35
autonomia universitaria	25	0.35	0.09
autonomia didattica	36	0.65	0.16
ordinamenti didattici	24	1.27	0.32
formazione universitaria	24	0.26	0.07
scienze della formazione	14	0.25	0.03
mondo della scuola	12	0.17	0.02
mondo del lavoro	34	0.45	0.11
ricerca di base	16	0.20	0.02
enti di ricerca	11	0.89	0.10

Tabella 3.7: Alcuni segmenti selezionati dal software, con indice IS assoluto e relativo

Ciò che otteniamo dopo l'inserimento nel vocabolario dei segmenti individuati è un nuovo vocabolario (quello che possiamo definire come definitivo) composto da 20673 unità ($\frac{V}{N} = 0.08$, $\frac{V_1}{V} = 0.47$).

3.2.3 Le specificità

L'aver introdotto nel nostro vocabolario una serie di forme flesse dovrebbe averne ridotto di molto le ambiguità. Ciò può portare ad una maggiore caratterizzazione dei vari sub-corpora, ovvero ad un numero maggiore di forme esclusive per ognuna delle annate o delle testate.

La variabile anno

Facendo riferimento anche alla tabella 3.2, integriamo la lista delle forme esclusive con i segmenti esclusivi (tabella 3.8). Il risultato, come si è già avuto modo di appurare, risente pesantemente del lessico tecnico-giuridico-economico tipico del Sole-24Ore, relativo soprattutto alla questione del riordino degli albi professionali. Cionondimeno notiamo che i segmenti esclusivi nell'anno 2001 siano più numerosi di quelli del 2002 (*schema di regolamento, nuovi titoli universitari, ministro dell'università, periodo transitorio*).

Facendo comunque un piccolo salto logico, confrontiamo questa tabella con la 3.9, e notiamo come non tutti i segmenti esclusivi dell'anno 2001 siano imputabili al Sole-24Ore, e perciò a quell'ormai già nominata notizia che è stata il rinnovo degli albi.

La variabile testata

Come è illustrato in tabella 3.9, l'aspettativa di riscontrare delle esclusività più marcate grazie all'introduzione nel vocabolario di alcuni segmenti è esattamente il primo risultato ad essere confutato. Come veniva anticipato dal vocabolario per forme grafiche, i tre quotidiani "non economici" formano un blocco comune, mentre il solo a distinguersi dagli altri è il Sole-24Ore. Repubblica, Corriere e Stampa differiscono almeno per qualche caratteristica? corrono una gara esclusiva fra loro tre, letteralmente "rubandosi le parole di bocca"?

Dato che le dimensioni dei loro sub-corpora sono simili, proviamo a ricavare le forme esclusive per questi tre quotidiani (tabella 3.10). Per poter meglio differenziarli, la soglia in occorrenze dei termini esclusivi caratteristici è stata abbassata a 4. Notiamo immediatamente come per il Corriere e per

Forma grafica	ANNO 01	ANNO 02
dottori commercialisti	53	0
d'Ascenzo	35	0
schema di regolamento	33	0
sottosegretario all'università	25	0
ordinamenti professionali	22	0
dottori commercialisti e ragionieri	17	0
nuovi titoli universitari	16	0
ministro dell'università	16	0
assistenti sociali	14	0
periodo transitorio	13	0
dottore commercialista	13	0
commissione Rossi	13	0
albo unico	12	0
accesso agli ordini	12	0
bozza di regolamento	11	0
ministero della Giustizia	11	0
riforma dei cicli scolastici	11	0
revisori contabili	10	0
nuovi titoli accademici	10	0
assistente sociale	10	0
requisiti minimi	0	28
tassa sul fumo	0	16
ministro dell'economia	0	16

Tabella 3.8: Segmenti esclusivi per la variabile ANNO

la Stampa i termini le parole più frequenti e usate in modo esclusivo siano *Censis*, *citati*, *Pacini* e *Meneghello*, cioè quattro nomi propri, oggetto, come abbiamo già anticipato, di trattazioni molto legate a realtà geografiche o temporali limitate.

E' leggermente differente la situazione della Repubblica, che si concede qualche "intromissione" più o meno importante in quello che è il lessico tecnico del Sole-24Ore, utilizzando termini come *commercialisti*, oppure, anche se con una frequenza bassa, *accesso alle professioni*.

Segmento	Occ. Sole-24Ore	Occ. Repubblica	Occ. Stampa	Occ. Corriere
schema di regolamento	33	0	0	0
dottori commercialisti e ragionieri	17	0	0	0
assistenti sociali	14	0	0	0
riforma dell'autonomia didattica	13	0	0	0
albo unico	12	0	0	0
bozza di regolamento	11	0	0	0
ministero della Giustizia	11	0	0	0
riforma dell' autonomia	10	0	0	0
d'Amato	10	0	0	0
titolo accademico	10	0	0	0
revisori contabili	10	0	0	0
nuovi titoli accademici	10	0	0	0

Tabella 3.9: Segmenti esclusivi per la variabile TESTATA

Parola	Occ. Totali	Occ. Repubblica	Occ. Corriere	Occ. Stampa
Decleva	19	15	0	0
bicocca	16	15	0	0
Palumbo	12	11	0	0
commercialisti	24	10	0	0
Rossi	56	9	0	0
iulm	14	9	0	0
Tor Vergata	18	7	0	0
periti	10	7	0	0
accesso alle professioni	17	6	0	0
psicologi	17	6	0	0
dopo la riforma	11	6	0	0
professionalizzanti	19	5	0	0
corsi di laurea triennali	10	5	0	0
geologi	10	5	0	0
Censis	23	0	19	0
citati	10	0	10	0
nuovi corsi di studio	11	0	4	0
ddl	11	0	4	0
Pubblica Amministrazione	10	0	4	0
modificata	10	0	4	0
Pacini	10	0	0	10
Meneghello	10	0	0	10
esami di Stato	12	0	0	2

Tabella 3.10: Parole esclusive per la variabile TESTATA, ad esclusione del Sole-24Ore

Capitolo 4

Diamo tempo al tempo

Per approfondire i risultati ottenuti nelle analisi esplorative dei dati procederemo a testare un'ulteriore ipotesi, che potrà fornire interessanti spunti per il prosieguo del presente lavoro: si procede nel testare una possibile relazione del contenuto degli articoli con il tempo. Il legame tra notiziabilità e tempo è sempre verificabile [Sorrentino, 1995], tanto più che il concetto di *scoop* riguarda notizie temporalmente limitate. A volte però una notizia genera strascichi che possono replicare nel tempo alcuni contenuti della stessa notizia, o addirittura alcune notizie si ripropongono con una certa ciclicità nel tempo.

La pubblicazione degli articoli potrebbe seguire infatti delle logiche predefinite (una breve introduzione che riguarda la possibile relazione tra notiziabilità e tempo è stata presentata nel par. 2.1.2), che possono necessitare di un'analisi più approfondita rispetto all'analisi delle specificità del paragrafo 3.1.3. In quel paragrafo l'unità di tempo base analizzata è l'anno, e l'arco di tempo coperto dalla tesi verrebbe diviso in due sole parti, troppo poche per poter affermare che le occorrenze delle parole seguono un trend.

In questo capitolo invece l'unità tempo di base utilizzata sarà il semestre, in modo da poter rispettare dei punti fermi:

- disporre di una serie di unità di tempo che possano essere allineate e confrontate al di là della differenza dicotomica tra anni;
- garantire una dimensione sufficiente dei sub-corpus, dato che nel ritmo

di pubblicazione degli articoli ci sono alcuni “punti morti” nei quali la pubblicazione veniva rallentata (e pertanto si potrebbe non disporre di testi sufficientemente ampi da essere significativi nelle analisi).

La differenza tra un’analisi delle specificità e l’analisi temporale sta nella possibilità di allineare i sub-corpus rispetto al tempo, oltre che nelle metodologie utilizzate.

4.1 Il ciclo di vita delle parole, un’analisi descrittiva

Bolasco nel 2003 propose un indice sintetico capace di identificare in un corpus diviso in subtesti riferibili al tempo quali parole seguissero un’evoluzione data da una variazione del numero delle occorrenze. Nel tempo il numero delle occorrenze può aumentare oppure diminuire, oppure può esserci un andamento misto. Insomma, la serie storica delle occorrenze riferite al vocabolario di un corpus diviso in subtesti ordinabili rispetto al tempo può essere trattata esattamente come tutte le altre serie temporali.

Considerando gli scarti tra le occorrenze normalizzate in ciascun semestre Occ_j e le occorrenze che si avrebbero nel caso di equidistribuzione Occ_M (scarti indicati nella seguente formula come $Scarto_Norm_j$) nonché i prodotti tra gli scarti normalizzati adiacenti, possiamo costruire l’indice IT_0 :

$$IT_0 = \left\{ \frac{\prod_{j=2}^n [(Scarto_Norm_j) \times (Scarto_Norm_{j-1})]}{\left| \prod_{j=2}^n [(Scarto_Norm_j) \times (Scarto_Norm_{j-1})] \right|} - 1 \right\} / 2 \quad (4.1)$$

Dove j rappresenta la frazione di tempo utilizzata, n il totale delle frazioni di tempo considerate. Moltiplicando poi il risultato di tale indice per la quantità $\frac{(Occ_1 - Occ_n)}{(Occ_1 + Occ_n)}$ si ottiene un indice che varia tra -1 e 1, e che grazie ai valori intermedi è in grado di identificare se il trend è in crescita, in calo, stazionario, o se la parole è obsolescente oppure un neologismo. Per indicazioni vedi appendice A.4.

La tabella 4.1 riporta le parole che sono risultate sensibili all'indice IT: oltre a confermare alcune osservazioni effettuate nei capitoli precedenti, offre anche indicazioni sul ciclo di vita di molte parole.

In estinzione $IT = -1$	Obsolescenti $-1 < IT \leq -0.6$	Crescenti $0.6 \leq IT < 1$	Neologismi $IT = 1$
architetti	ordini	Moratti	milioni di euro
ragionieri	Guerzoni	Finanziaria	Tremonti
sezione	regolamento	tagli	
dottori commercialisti	esame di Stato	finanziaria	
laureati specialisti	ingegnere	euro	
	dottori	Ciampi	
	professionisti	Letizia	
	sezioni	ricercatori	
	professioni	ministro dell'istruzione	
	professione	reclutamento	
	tecnici		
	dottori		
	laureati		
	riforma universitaria		
	ingegneri		
	lauree triennali		
	professionali		
	orientamento		
	CUN		
	Berlinguer		
	iscrizione		
	tirocinio		
	sapienza		
	classi		
	tecnologie		

Tabella 4.1: Parole con trend significativo individuate dall'indice IT

Innanzitutto le parole in estinzione *architetti*, *ragionieri* e *dottori commercialisti* sono parole con specificità molto alta esclusiva del Sole-24Ore del 2001. Infatti tra le parole obsolescenti troviamo molti altri termini che riportano alla riforma degli albi professionali, come *ingegnere*, *professione/i*, *professionisti*.

Non è una sorpresa trovare *Berlinguer* e *CUN* tra le parole obsolescenti, in quanto il primo è un personaggio dell'allora uscente governo di centrosinistra, mentre il secondo termine riguarda un organismo che poi lascia le scene della ribalta alla CRUI (che sia nella sua denominazione estesa che nella forma di sigla presenta indice IT positivo).

Tra le parole obsolescenti troviamo però il termine *riforma universitaria*: nei quattro semestri presenta rispettivamente 112, 32, 26, e 26 occorrenze. Da notare che anche *riforma dell'università* è segnalata come obsolescente ($IT =$

−0,438) con 23, 19, 6 e 9 occorrenze. Non è semplice spiegarsi come mai queste parole si presentino nel tempo con frequenze decrescenti: cercheremo di rispondere al problema nei paragrafi successivi, una volta raccolti maggiori indizi.

Tra le parole crescenti (e i neologismi) troviamo due dei personaggi politici di spicco dell'allora subentrante governo di centrodestra e soprattutto una serie di termini (*Finanziaria, milioni, tagli*) legati all'economia e soprattutto ai problemi finanziari che hanno tormentato l'Università nella seconda parte del 2002. Dobbiamo però segnalare come i termini *milioni* e *miliardi* (probabilmente di lire) non subiscono flessioni nel tempo, ma rimangono piuttosto costanti ($IT = 0$), anche se vedremo di seguito come possano comunque presentare un trend leggermente decrescente.

4.2 Ulteriori spunti sul trend

L'indice IT rappresenta una statistica descrittiva molto efficace nell'individuare delle tipologie di trend molto marcate, ma presenta una piccola limitazione, che consiste nel considerare nella valutazione dell'intensità dell'effetto-tempo solamente il numero di occorrenze di inizio e fine periodo. Se in questa tesi il numero di segmenti temporali è limitato (4), è comunque possibile che una parola presenti un picco di frequenza al secondo o terzo semestre che non può essere individuato dall'indice IT.

Assodato che se le frequenze aumentassero o diminuissero costantemente nel tempo queste verrebbero identificate dall'indice IT , al fine di cercare di dare visibilità anche a possibili trend non monotoni, procediamo con lo stimare un modello lineare che presenta variabile dipendente le occorrenze normalizzate rispetto al semestre di appartenenza (esattamente come fatto per il calcolo dell'indice IT) e come covariata il tempo:

$$OccNorm = \beta_0 + \beta_1 \times Tempo + \beta_2 \times Tempo^2 + \epsilon$$

In tabella 4.2 e in figura 4.1, 4.2 e 4.3 sono rispettivamente rappresentati le stime del modello, e i grafici relativi alle occorrenze normalizzate utilizzate

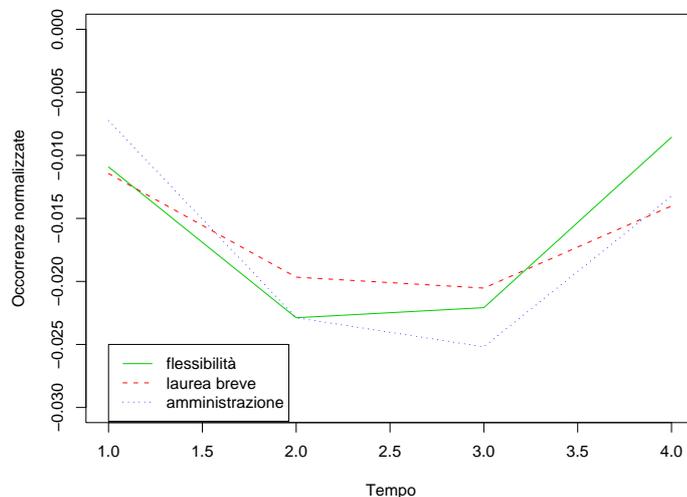


Figura 4.1: Parole con significatività del parametro quadratico < 0.01

per la costruzione dei modelli. I grafici contengono le parole raggruppate in base alla significatività del loro parametro quadratico.

Al fine di determinare la non monotonicità della serie delle occorrenze normalizzate possiamo valutare la significatività della stima del parametro di regressione relativo a $Tempo^2$: nel caso in cui questo sia significativo abbiamo che, laddove l'indice IT potrebbe non rilevare alcun andamento crescente/decescente, le occorrenze sono “gonfiate” o “sgonfiate” in corrispondenza di uno o due semestri centrali.

Il modello di regressione verrà applicato a tutte quelle parole per le quali l'indice IT vale 0, e che pertanto potrebbero non presentare un trend crescente/decescente.

L'indice *IT* aveva segnalato come alcune parole legate direttamente al lessico universitario fossero obsolescenti. Il modello lineare fa notare come anche i termini *laureato* e *laurea breve* (vedi tabella 4.2) subiscono una flessione nel loro utilizzo.

D'altra parte questa flessione è compensata dalla parola più generica *laurea*, che presenta un andamento fortemente parabolico con parametro quadratico negativo (figura 4.2), e in ogni caso un numero di occorrenze totali

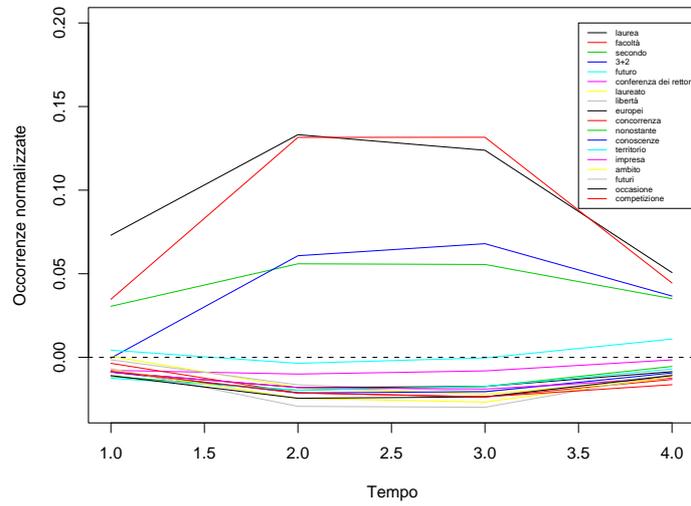


Figura 4.2: Parole con significatività del parametro quadratico < 0.05

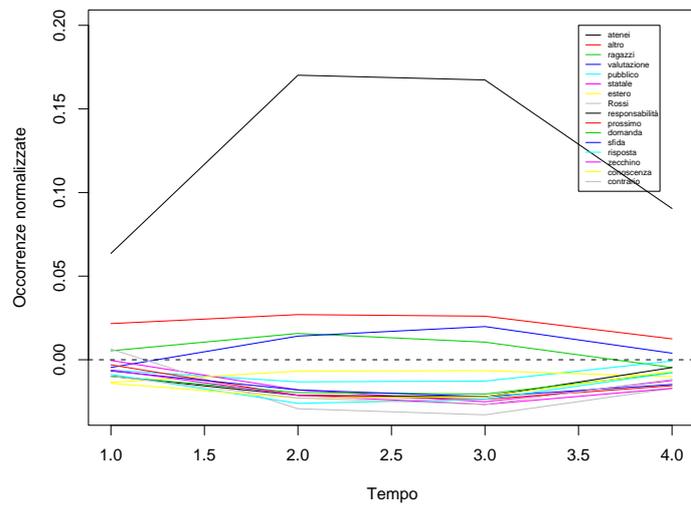


Figura 4.3: Parole con significatività del parametro quadratico < 0.1

Forma grafica	Intercetta	signific. β_0	Lineare	Signific. β_1	Quadratico	Signific. β_2
Rossi	0,065	0,535	-0,072	0,063	0,013	0,070
futuri	0,037	0,020	-0,054	0,012	0,011	0,012
risposta	0,023	0,107	-0,041	0,056	0,008	0,055
ambito	0,027	0,034	-0,042	0,020	0,008	0,020
statale	0,035	0,106	-0,042	0,080	0,008	0,087
responsabilità	0,018	0,166	-0,035	0,080	0,007	0,075
laureato	0,033	0,012	-0,040	0,009	0,007	0,010
concorrenza	0,028	0,037	-0,038	0,024	0,007	0,026
contrario	0,020	0,191	-0,036	0,097	0,007	0,098
amministrazione	0,022	0,016	-0,037	0,009	0,007	0,010
occasione	0,015	0,460	-0,033	0,028	0,007	0,028
prossimo	0,027	0,080	-0,037	0,053	0,007	0,059
flessibilità	0,014	0,001	-0,031	0,000	0,006	0,000
zecchino	0,023	0,095	-0,035	0,057	0,006	0,063
libertà	0,027	0,042	-0,034	0,030	0,006	0,033
conoscenza	0,008	0,225	-0,028	0,061	0,006	0,055
conoscenze	0,015	0,074	-0,030	0,034	0,006	0,033
domanda	0,012	0,147	-0,028	0,061	0,006	0,059
competizione	0,015	0,027	-0,028	0,013	0,005	0,014
nonostante	0,007	0,042	-0,023	0,011	0,005	0,010
futuro	0,021	0,400	-0,022	0,051	0,005	0,045
pubblico	0,011	0,139	-0,022	0,066	0,005	0,060
sfida	0,015	0,105	-0,026	0,055	0,005	0,061
europei	0,009	0,071	-0,022	0,026	0,005	0,026
territorio	0,003	0,151	-0,020	0,022	0,004	0,020
impresa	0,008	0,043	-0,021	0,015	0,004	0,016
laurea breve	0,004	0,002	-0,019	0,000	0,004	0,000
estero	-0,025	0,043	0,014	0,070	-0,003	0,075
altro	0,005	0,411	0,021	0,108	-0,005	0,094
ragazzi	-0,016	0,134	0,028	0,071	-0,006	0,062
valutazione	-0,043	0,079	0,046	0,067	-0,009	0,071
secondo	-0,016	0,138	0,059	0,036	-0,011	0,036
3+2	-0,104	0,409	0,127	0,044	-0,023	0,048
laurea	-0,052	0,042	0,159	0,013	-0,033	0,012
atenei	-0,126	0,110	0,237	0,054	-0,046	0,055
facoltà	-0,152	0,025	0,233	0,015	-0,046	0,015

Tabella 4.2: Parole con indice $IT = 0$ con le relative stime dei parametri di regressione e della significatività degli stessi

superiore. Anche il termine *3+2* presenta parametro parabolico negativo.

Dare una spiegazione a questi andamenti non è semplice: probabilmente dopo il primo periodo in cui i quotidiani si sono dilungati in spiegazioni riguardanti le novità della riforma, l'appeal del lessico tecnico si è ridotto.

Il lessico tecnico non è composto unicamente da queste parole, ma la presenza di *riforma universitaria* tra le parole obsolescenti è un importante indizio. La forma generica *riforma* presenta invece un andamento costante ($IT = 0$, trend lineare positivo, trend quadratico negativo ma entrambi non significativi). Inoltre la parola *riforma* è di gran lunga più utilizzata nel corpus rispetto ai segmenti che genera. Nonostante la parola *riforma* produca molti segmenti diversi la forma grafica grezza rimane quella a frequenza maggiore (l'indice IS dei segmenti prodotti da *riforma* non è mai particolarmente alto).

Abbiamo ottenuto solamente un indizio che riguarda la perdita di “tecnicismo” del lessico dei quotidiani. Le analisi che verranno effettuate da questo punto in poi dovranno in ogni caso tener conto di questa possibilità, che apre il campo ad uno scenario ad una visione più ampia: la riforma vista come notizia essa stessa, con una sua origine e una sua durata, posta all'interno di una cornice più ampia dall'andamento costante.

Capitolo 5

Quali differenze?

Per visualizzare al meglio le caratteristiche del corpus in analisi si effettuerà un'analisi delle corrispondenze su una tabella di contingenza “parole \times variabili”, dove le variabili di riga sono costituite dai 231 segmenti estratti grazie alla lessicalizzazione del corpus e da una serie di forme grafiche debitamente scelte, mentre le variabili colonna sono costituite dall'interazione tra la variabile quotidiano e la variabile anno. Lo scopo è quello di poter identificare al meglio le diverse posizioni relative che contraddistinguono i quotidiani fra loro e le differenze tra le annate 2001 e 2002.

5.1 Una breve analisi dei titoli

Il titolo di un articolo veicola, concentrata, moltissima dell'informazione contenuta nell'articolo stesso. Spesso contiene le parole chiave dell'articolo, sicuramente riporta i protagonisti principali della vicenda che deve essere esposta.

Ogni articolo solitamente è costituito da più sezioni, graficamente e concettualmente distinte. Il nostro corpus in particolare tiene conto di alcune di queste possibili sezioni, in particolare ogni testo è suddiviso in:

- occhiello: posizionato in alto, sopra il titolo principale, presenta la vicenda che verrà esposta;
- titolo: veicola al suo interno le parole chiave del testo dell'articolo;

- sottotitolo: posizionato sotto il titolo è a volte un periodo di lunghezza superiore al titolo;
- testo: il testo vero e proprio dell'articolo.

Quello che in questa analisi consideriamo come “titolo” è composto da occhiello, titolo e sottotitolo, cosa che ci consente anche di ottenere un corpus di dimensioni superiori a quello che otterremmo considerando solamente i titoli non in senso lato.

In seguito a questa operazione otteniamo un totale di 9.046 occorrenze, per 2.454 forme grafiche complessive ($\frac{V}{N} = 2,7\%$, $\frac{V_1}{V} = 63\%$) che sono cifre che non sono molto confortanti per quel che riguarda il trattamento statistico di questo testo; nondimeno possiamo estrarre dal corpus ottenuto alcune informazioni che potranno essere utili anche per l'interpretazione dei risultati delle analisi effettuate sul corpus completo.

Come possiamo notare dalla tabella 5.1 le parole piene a frequenza più alta sono molto simili a quelle rilevate per l'intero corpus: uno dei primi utilizzi del vocabolario “ridotto” è quello di fornire una traccia per la lettura dei risultati delle analisi sul vocabolario esteso. Tanto più che in alcune forme grafiche manca la dualità dovuta alla presenza di nomi coniugati in modi differenti. Le parole *studente - studenti* comparivano entrambe con una frequenza piuttosto rilevante nel vocabolario esteso (116 - 889), mentre nel vocabolario dei titoli la frequenza si abbassa fino quasi ad esaurire una di esse (3 - 43). Lo stesso vale per la coppia *ateneo - atenei, professori* e tutte le sue altre possibili forme (*professore-professor-professori universitari-etc...*). Il vocabolario dei titoli nella sua semplicità ci consente di concentrare l'attenzione solamente su una realizzazione grafica del concetto veicolato dalla parola: della terna *professore-professor-professori* presente nel corpus esteso, nel corpus dei titoli resta solamente la forma *professori*.

Forma grafica	ANNO 2001	ANNO 2002	Corriere	Sole 24Ore	La Stampa	Repubblica
università	106	62	47	71	21	29
riforma	72	42	34	48	20	12
atenei	31	30	20	30	5	6
Moratti	18	34	14	19	15	4
laurea	30	21	15	24	8	4
3+2	15	29	2	36	5	1
ministro	20	23	14	17	10	2
corsi	20	23	14	26	1	2
studenti	26	17	23	9	9	2
lauree	26	16	11	22	6	3
scuola	12	23	12	6	13	4
ricerca	11	19	10	8	8	4
istruzione	9	14	4	10	8	1
rettori	9	11	5	4	2	9
fondi	7	13	7	5	6	2
studio	10	8	2	14	1	1
intervento	7	10	2	15	0	0
governo	7	10	10	0	4	3
lavoro	8	9	3	11	3	0
giovani	12	5	3	10	3	1
formazione	9	7	0	16	0	0
futuro	10	6	6	7	0	3
riforme	14	2	4	12	0	0
ordini	15	1	4	12	0	0
albi	13	1	1	13	0	0

Tabella 5.1: Alcune tra le forme grafiche piene a maggiore frequenza

Due piani fattoriali

Considerando in totale 70 forme grafiche, proviamo a mappare con la tecnica dell'analisi delle corrispondenze¹ le frequenze con cui compaiono le parole all'interno dei sub-corpora a nostra disposizione. Al fine di non creare confusione nel proiettare 70 forme grafiche in una stessa mappa, senza poter evitare che i punti si sovrappongano, la figura 5.1 contiene due diversi grafici: il primo riporta un lessico (selezionato a discrezione del ricercatore) di natura economica e lavorativa (*commercialista, finanziaria, fondi, stage, aziende*) che dovrebbe fare riferimento maggiormente al Sole-24Ore, oppure un lessico di natura più generica (*ricerca, formazione, scuola*) che dovrebbe essere utilizzato maggiormente dagli altri tre quotidiani; il secondo riporta due tipologie di parole che si riferiscono rispettivamente al mondo universitario e a quello politico.

Il secondo piano fattoriale (figura 5.2) riporta invece la mappatura delle stesse parole, ma in questo caso il Sole-24Ore è stato trattato come variabile supplementare, al fine di limitare il suo impatto nella determinazione del primo fattore.

5.1.1 L'interpretazione della soluzione fattoriale

Nel primo piano fattoriale, l'asse orizzontale crea una separazione tra i diversi quotidiani (in particolare distingue tra Sole-24Ore e altri quotidiani); l'asse verticale divide le due principali annate.

I quotidiani non si dispongono in maniera regolare nel grafico: in particolare Repubblica e Corriere non si distanziano quasi per nulla l'uno dall'altro, lasciando intuire come i loro profili siano del tutto simili. La Stampa completa il terzetto di quotidiani non economici, mentre il Sole-24Ore viene a posizionarsi all'estrema destra dell'asse orizzontale.

Le variabili interagiscono fra di loro, e in particolar modo notiamo come Repubblica e Corriere siano leggermente spostati verso il 2001, mentre la

¹Per l'analisi si utilizzano i primi due fattori, che spiegano il 74% dell'inerzia totale, che ammonta a 0,22. Gli altri fattori spiegano rispettivamente il 14,98% e l'11,03% dell'inerzia residua

Stampa si nettamente più vicina al 2002. Il Sole-24Ore è equidistante tra le due annate, anche se possiamo pensare che sia leggermente influenzato dalle parole relative alla riforma degli ordini professionali, e pertanto possa essere leggermente avvicinato all'annata 2002.

Possiamo notare in prima istanza la presenza, nel cerchietto rosso del primo grafico, di una serie di termini riferiti alla riforma degli albi professionali; inoltre nel terzo quadrante (quello marcato da Corriere e Repubblica) la presenza di termini quali *Finanziaria*, *protesta*, e *contro*.

Nel secondo grafico di figura 5.1 possiamo immediatamente identificare: *Berlinguer* nel terzo quadrante (e molto spinto verso Corriere e Repubblica, *Moratti* nel quarto quadrante (che, anche se piuttosto debolmente, fa riferimento a La Stampa). I termini vicini all'ordinamento del nuovo sistema universitario appartengono tutti alla metà del grafico contraddistinta dalla presenza del Sole-24Ore (*3+2*, *triennali*, *autonomia*), mentre i termini relativi ai diversi personaggi protagonisti del mondo universitario (*ricercatori*, *studenti*, *rettore*) sono sbilanciati verso il terzo-quarto quadrante.

Sovrapponendo i due grafici, possiamo notare come *Moratti* sia una parola molto correlata con *fondi* e *ricerca*.

Questa prima analisi del vocabolario di un sub-corpus ci può dare alcune indicazioni su come procedere nei passi successivi. In particolare:

- ci sono indizi a riguardo di una presunta similarità tra Corriere e Repubblica;
- il Sole-24Ore sembra un quotidiano dai contenuti nettamente differenti rispetto agli altri quotidiani considerati;
- la distinzione temporale è piuttosto netta, non solamente per gli argomenti a notiziabilità breve (personaggi o politici) ma anche per argomenti più strutturali, come la ricerca o il mondo del lavoro;

Nel grafico di figura 5.2 possiamo vedere come l'effetto principale sia quello del tempo. Il primo fattore separa infatti l'anno 2001 dal 2002, mentre il secondo fattore identifica soprattutto Repubblica e Sole-24Ore, che occupano posizioni diametralmente opposte nella scala del secondo fattore.

Gran parte della variabilità che contraddistingueva il Sole-24Ore viene ora assimilata dalla variabile 2001 (che prende il posto della testata economica nell'identificare l'argomento riforma degli albi) e il Sole-24Ore diventa il quotidiano che fa più leva sugli stages e sul mondo del lavoro.

Il Corriere invece viene a collocarsi in una posizione molto vicina all'origine degli assi: i titoli che usa non sono aggressivi nel lessico utilizzato quanto quelli delle altre testate.

Per una breve descrizione della terminologia e le procedure utilizzate nell'analisi delle corrispondenze fare riferimento all'appendice B.

5.2 Analisi fattoriale sul corpus esteso: alcune considerazioni preliminari

Abbiamo visto nel paragrafo precedente come utilizzando un sub-corpus di dimensioni ridotte quale quello formato dai titoli si possano già intuire alcune differenze tra i quotidiani analizzati. D'ora in avanti considereremo nelle analisi il corpus intero: titoli e testi veri e propri assieme. I titoli non costituiscono che una minima parte del corpus, nondimeno, aumentando le dimensioni dei dati in analisi, ci aspettiamo di individuare strutture latenti dei dati che presentino maggiori complessità.

Ma cominciamo pure con la costruzione del nostro piano fattoriale. I fattori verranno costruiti a partire dai segmenti individuati nella fase di lessicizzazione in quanto questi sono molto meno soggetti ad ambiguità rispetto alle semplici forme grafiche. Ai segmenti individuati nel corpus verranno anche affiancate alcune forme grafiche (delle quali parecchie riconosciute da Taltac come non ambigue nella fase di normalizzazione) non ambigue e che sono state ritenute essenziali per l'informazione che possono veicolare.

Le variabili che verranno proiettate sul grafico generato dai fattori sono le interazioni tra testata e anno: intuitivamente la variabile *Corriere 2001* è relativo agli articoli pubblicati dal Corriere della Sera nel 2001. Il totale delle variabili considerate ammonta perciò a 8: per ognuno dei quattro quotidiani considereremo due annate. Il calcolo degli autovalori e degli autovettori, al-

la base dell'analisi delle corrispondenze, ci consente di scomporre la tabella di contingenza di partenza in una serie di variabili che possiamo considerare esterne e che sono correlate con quelle di partenza. Ricordiamo che le dimensioni che otteniamo sono 7, in quanto uno degli autovettori ottenuti viene generalmente definito "banale".

L'inerzia² totale della tabella di contingenza ammonta a 0,900 , ed è distribuita tra le variabili come nella tabella seguente:

Corriere 2001	Corriere 2002	Sole 24Ore 2001	Sole 24Ore 2002	La Stampa 2001	La Stampa 2002	Repubblica 2001	Repubblica 2002
0,122	0,088	0,197	0,106	0,101	0,097	0,084	0,105

La variabile che cattura la maggior parte della variabilità interna è la prima annata del Sole 24 ore (22% del totale). A seguire vengono il Corriere del 2001 (14%), il Sole-24Ore del 2002, la Repubblica del 2002 e la Stampa del 2001. Il Corriere del 2002 (10% del totale) e la Repubblica del 2001 (9%) sono le variabili che catturano meno variabilità.

Questi risultati possono anche essere visti nella tabella seguente, che riporta la distanza chi-quadro delle variabili-colonna dal centroide della tabella di contingenza.

Corriere 2001	Corriere 2002	Sole 24Ore 2001	Sole 24Ore 2002	La Stampa 2001	La Stampa 2002	Repubblica 2001	Repubblica 2002
0,902	1,021	0,866	0,719	1,207	1,246	1,380	0,932

5.3 Un primo piano fattoriale

Una volta costruita la soluzione fattoriale, vediamo come il primo e il secondo fattore assieme assorbano il 50% della variabilità totale, i primi tre il 65% (vedi anche lo screeplot di figura 5.3). Inoltre, valutando il peso delle variabili sui fattori possiamo individuare alcuni elementi che consentano di classificare fra loro le testate e le annate:

- le testate nel 2002 sono molto vicine fra loro;

²Possiamo considerare l'inerzia come una misura della variabilità totale espressa dalla tabella di contingenza. Le dimensioni calcolate grazie all'analisi delle corrispondenze spiegheranno interamente questa variabilità, ma saranno soprattutto poche di esse a contribuire a spiegare un grosso ammontare di variabilità.

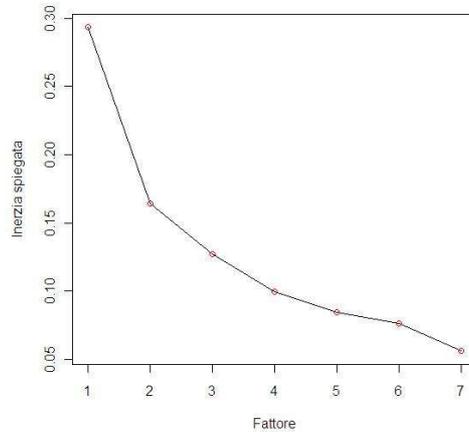


Figura 5.3: Screeplot

- il Sole-24Ore del 2001 è molto correlato con il primo fattore, ma quasi per nulla con il secondo;
- il Corriere del 2001 non ha alcun peso nel determinare il primo fattore, ma caratterizza molto il secondo;
- la Repubblica del 2002 ha un certo peso nel primo fattore (correlazione opposta al Sole-24Ore del 2001), e poco peso nel secondo.

Guardando in figura 5.4 vediamo come la disposizione dei punti-colonna possa far pensare soluzione ottimale mono-fattoriale, in quanto è ben visibile la disposizione “a mezzaluna” delle variabili. Il fattore che influenzerebbe la disposizione dei punti sarebbe quello temporale (annata 2001 vs. annata 2002). A staccarsi dall’andamento generale è solamente il Sole-24Ore del 2001 (per i motivi già evidenziati in precedenza, vedi 3.2.3).

Possiamo identificare, anche aiutandoci con la correlazione tra variabili e fattori, tre zone ben distinte tra loro. La parte destra del grafico è monopolizzata dal Sole-24Ore del 2001, la parte superiore è influenzata dall’anno 2001, mentre il terzo quadrante è definito dall’anno 2002.

Il Sole-24Ore ha un peso preponderante nel determinare la configurazione dei punti nello spazio: il linguaggio adottato nei mesi del 2001 è stato tal-

mente peculiare rispetto agli altri giornali da “schiacciare” nella sola parte sinistra del grafico tutte le altre variabili.

Un’ulteriore dato da presentare è la correlazione³ tra fattori e variabili. Dalla tabella che segue vediamo come i primi due fattori riassumano in una maniera abbastanza esauriente le variabili originali. Da notare che il Sole-24Ore del 2002, la Stampa e la Repubblica del 2001 presentano delle correlazioni molto alte rispettivamente con i fattori 3, 4 e 5.

Insomma, la soluzione fattoriale, nonostante possa essere condensata in maniera soddisfacente dai primi due fattori, presenta alcuni problemi.

	Fatt.1	Fatt.2	Fatt.3	Fatt.4	Fatt.5	Fatt.6
Corriere 2001	0	838	17	47	79	6
Corriere 2002	288	77	65	142	62	42
Sole-24Ore 2001	924	17	54	4	1	0
Sole-24Ore 2002	13	270	695	7	1	0
Stampa 2001	154	55	9	716	4	19
La Stampa 2002	236	45	136	22	145	388
Repubblica 2001	30	121	4	48	511	235
Repubblica 2001	413	26	193	7	113	119

5.4 Un secondo piano fattoriale

Abbiamo visto come la soluzione fattoriale adottata risenta molto della presenza di un “outlier” come il Sole-24Ore, che, data la sua distanza dal baricentro dei dati, influenza la disposizione sul piano fattoriale delle altre variabili. Quello che però, a giudicare dai calcoli e dalle proiezioni, potrebbe essere un buon risultato in quanto discrimina in maniera molto demarcata un quotidiano dagli altri, è probabilmente più una pecca che un successo. Infatti questo distacco del quotidiano economico dagli altri è prevedibile, ed è stato già notato e analizzato, portando alla conclusione che è la tematica originale del rinnovo degli albi a determinare l’originalità del giornale. In conclusione la presenza del Sole-24Ore non fa altro che rendere meno evidenti le differenze tra gli altri quotidiani.

Si rende pertanto necessario ricalcolare la soluzione fattoriale al netto del Sole-24Ore, in modo poter cogliere al meglio le eventuali differenze tra i quotidiani residui. Tale ricalcolo richiede però la modifica dei dati di partenza:

³Il pacchetto ca del software statistico R restituisce le correlazioni tra variabili e fattori in valore assoluto, e moltiplicata per mille, in modo da fornire non tanto un indicatore di correlazione puntuale, ma soprattutto un indicatore della forza della relazione tra variabili

tra i segmenti utilizzati molti erano delle parole o segmenti esclusivi del Sole-24Ore, oppure erano espressioni che comparivano, al netto del Sole-24Ore, con una frequenza troppo bassa. Questa configurazione delle frequenze rende impossibile il calcolo della soluzione fattoriale o comunque poco affidabile l'interpretazione delle lessie inserite nell'analisi in quanto la loro frequenza cala drasticamente. Delle 344 righe della tabella di contingenza ne utilizzeremo 180.

I dati relativi al Sole-24Ore verranno proiettati sul grafico come righe e colonne supplementari (vedi appendice B.2, ma non contribuiranno alla costruzione dei fattori. Ciò elimina di fatto la possibilità di visualizzare sul piano le originalità della testata (come proiettare dei dati appartenenti solamente al Sole-24Ore su fattori non determinati da esso?) e limita leggermente le analisi, escludendo però tutta quella parte così vituperata nel precedente paragrafo perché disturbatrice (la riforma degli ordini professionali).

L'inerzia totale in questo caso si attesta molto vicino al valore 0,75, con la distribuzione percentuale mostrata nella tabella sottostante, con una distribuzione percentuale molto omogenea e oscillante tra il 20% del totale per il Corriere 2001 e il 14% sempre del Corriere, ma nel 2002. Per quel che riguarda l'inerzia spiegata dai fattori, la situazione è meno definita rispetto alla precedente: i primi due fattori spiegano il 54,8% dell'inerzia totale, i primi tre arrivano al 73%. Questo significa che il secondo e il terzo fattore spiegano all'incirca la stessa quota di variabilità (21 e 18 percento rispettivamente). Infatti, se sul primo fattore quasi tutti i quotidiani (tranne la Repubblica 2001 e la Stampa 2001) hanno un certo peso, il secondo fattore isola principalmente la Stampa del 2001 e il Corriere 2002. Il terzo fattore, per finire, identifica soprattutto la Stampa del 2002 e la Repubblica 2002.

Nella tabella seguente sono riportate le inerzie relative alle variabili che hanno contribuito al calcolo della soluzione fattoriale:

Corriere 2001	Corriere 2002	Repubblica 2001	Repubblica 2002	La Stampa 2001	La Stampa 2002
0,152	0,108	0,111	0,110	0,136	0,132

La situazione è illustrata efficacemente anche dalla prossima tabella, relativa alle correlazioni tra le variabili-colonna e i fattori. Possiamo vedere come

nessuno dei fattori sia totalmente “trasversale”, e cioè correlato significativamente con tutte le variabili. L’unico fattore che si avvicina in modo significativo ad un fattore comune è il primo. Il secondo, il terzo e il quarto fattore sono molto influenzati rispettivamente da Stampa del 2001, la Stampa 2002 e Repubblica 2001.

	Fatt.1	Fatt.2	Fatt.3	Fatt.4	Fatt.5
Corriere 2001	847	49	12	62	29
Corriere 2002	259	268	12	121	340
Repubblica 2001	138	0	50	720	92
Repubblica 2002	408	0	346	3	243
Stampa 2001	0	918	1	42	38
Stampa 2002	241	4	677	57	21
(*)Sole-24Ore 2001	80	10	3	1	0
(*)Sole-24Ore 2002	0	3	0	0	23

5.4.1 La soluzione fattoriale: i primi e tre fattori

La figura 5.5 mostra una configurazione leggermente diversa dalla figura 5.4. Per prima cosa notiamo come lo sviluppo bidimensionale sia più accentuato, e come il Sole-24Ore diventi, improvvisamente il quotidiano che si colloca più vicino all’origine degli assi. Come a dire: una volta tolto il lessico economico, il Sole-24Ore diventa un quotidiano molto neutro.

Il primo fattore

Sul primo quadrante di figura 5.5 sono proiettati termini apparentemente molto contrastanti tra loro: *Confindustria*, *mondo del lavoro*, *mercato del lavoro*, *sbocchi professionali*, ma anche *Berlinguer*, *Forza Italia*, *campagna elettorale*. Sicuramente la componente “anno 2001” è forte nel caratterizzare il quadrante, e si nota soprattutto dai termini politici, che hanno caratterizzato la campagna elettorale del 2001 (NB: il ministro Moratti, come personaggio politico è peculiare soprattutto del 2002). Più difficile è spiegarsi la presenza di tanti termini legati al mondo del lavoro. I termini che con minor margine di errore legato all’ambiguità possiamo legare al mondo del lavoro sono *mondo del lavoro*, *mercato del lavoro*, *sbocchi professionali*, che hanno effettivamente specificità positiva per il 2001. Molto probabilmente si tratta di termini legati alle paure e alle speranze (insomma, ai dibattiti) più prossimi all’impatto della riforma, e che in seguito hanno lasciato spazio ad un atteggiamento di attendismo sull’argomento.

Il secondo quadrante di figura 5.5 (caratterizzato soprattutto dal Corriere del 2001), contiene soprattutto termini “tecnici” relativi al mondo dell’università (*Ministro dell’Università, sistema dei crediti, corsi di laurea triennali, nuovi titoli universitari, attuazione della riforma, autonomia degli atenei*) ma solamente un termine che rimandi al mondo politico, *Polo*. L’atteggiamento del Corriere è spesso quello di quotidiano informativo e imparziale, e lo si nota anche dalle proiezioni dei segmenti sul piano fattoriale. Questo risultato conferma in parte le analisi effettuate nel paragrafo 4.2, rinnovando il dato di obsolescenza precoce del lessico tecnico.

In una zona influenzata anche dal secondo fattore troviamo due termini legati alla voce degli studenti: *mila studenti*, e *contro la riforma*.

Il secondo fattore

La parte sinistra del grafico 5.5 (correlazioni negative delle variabili con il secondo fattore) è influenzata maggiormente dagli argomenti trattati nel 2002 e in particolare il terzo quadrante contiene gran parte delle variabili-testata relative a quell’anno. Il fatto che siano così concentrate fa intuire come gli argomenti trattati dai giornali in quell’anno si allineino ad una politica comune: e proprio la parola politica è il termine che contraddistingue questo quadrante. Possiamo chiamarlo quadrante dei personaggi (*Tremonti, Moratti, Buttiglione, Fassino, Silvio Berlusconi* e anche *Berlusconi* senza il nome), ma ci troveremmo anche Ciampi. In senso più lato questo è un quadrante politico anche per la presenza di un partito come i Ds, o di termini legati al finanziamento pubblico come *Finanziaria, miliardi di lire, tassa sul fumo, disegno di legge*.

Come il primo quadrante, anche il quarto è a “bassa intensità”. Infatti i punti proiettati in questo settore non risentono dell’alta concentrazione di variabili che può verificarsi nel secondo e terzo quadrante. La tematica prevalente non è facilmente individuabile, ma in ogni caso se c’è un settore in particolare ove si parli di ricerca, è questo. Termini come *formazione universitaria, ricerca scientifica, comunità scientifica, ricerca di base* e *Cnr* compaiono tutti in questo quadrante.

La presenza della Stampa del 2001 in prossimità di queste tematiche, ma soprattutto a cavallo tra 2001 e 2002 colloca il quotidiano in una posizione differente dagli altri due. La Stampa è il quotidiano meno “del 2001” e quello più “del 2002” nel senso che nella scala individuata dai nostri fattori si trova sempre più a sinistra degli altri quotidiani. L’interpretazione temporale del primo fattore è abbastanza immediata, ma il porre la Stampa come “precursore” di ideali potrebbe essere una forzatura, tanto più che nel grafico precedente (figura 5.4) nel 2002 compare nel raggruppamento comune di tutti i quotidiani.

Certamente bisogna dare atto della diversità di punto di vista di questo quotidiano rispetto al duo Corriere-Repubblica.

Il terzo fattore

Le variabili-colonna molto correlate con il terzo fattore sono la Repubblica 2002 e la Stampa 2002. Nelle figure 5.6 e 5.7 la contrapposizione tra queste due testate poste agli estremi dell’asse verticale sono evidenti. Soprattutto in figura 5.7 si nota il distacco che c’è tra la Stampa e gli altri quotidiani. Certamente il primo fattore coglie la parte maggiore della variabilità dei dati, che nel nostro caso è imputabile all’effetto-tempo, ma il secondo e il terzo fattore sono nettamente caratterizzati dalla Stampa, e dato che i primi tre fattori spiegano quote di devianza totale molto simili, possiamo affermare che la Stampa si distingue per contenuti da Corriere e Repubblica. I termini utilizzati dal quotidiano torinese sembrano essere piuttosto neutri.

5.5 Conclusioni e spunti

Come poteva essere pronosticato sin dall’inizio, anche i dati confermano che Corriere e Repubblica facciano corsa a sè. I due quotidiani di opinione per eccellenza non si distinguono ne’ per informazioni fornite, ne’ per colore politico, ne’ per attenzioni maggiori o minori a questa o quella categoria di persone: l’unica piccola differenza può essere l’utilizzo di un lessico legger-

mente più tecnico da parte del Corriere, cosa che lo accosta leggermente al Sole-24Ore.

La Stampa e il Sole-24Ore invece si differenziano sia dai due di cui sopra, sia fra di loro. Se il Sole-24Ore lo fa in maniera prevedibile utilizzando un lessico prettamente economico, ma al momento di valutare i suoi contenuti politico-didattici si riveli piuttosto inconsistente, la Stampa si differenzia nel suo proporre argomenti alternativi a tutti gli altri e soprattutto nel suo approccio didattico-scientifico all'università. Se non sembra lasciarsi trascinare dal dibattito politico, lascia qualche piccolo dubbio riguardo ad un suo possibile "ancoraggio al passato", deducibile dall'utilizzo delle vecchie *lire*, e nel non volersi troppo appropriare del "nuovo" lessico universitario, che invece è stato abbracciato molto più volentieri dal Corriere.

Ciò che però non è stato ancora reso palese l'opinione che tracciano questi quotidiani della riforma. Basta citare le rivolte studentesche e la fuga di cervelli (Corriere, Repubblica) per palesarsi contrari alla riforma? oppure, la Stampa parla della ricerca scientifica in toni allarmati oppure con una nota di fiducia nelle istituzioni? o, non citando la politica, cerca di mantenersi neutrale? sono queste le domande cui cercheremo di rispondere nel prossimo capitolo.

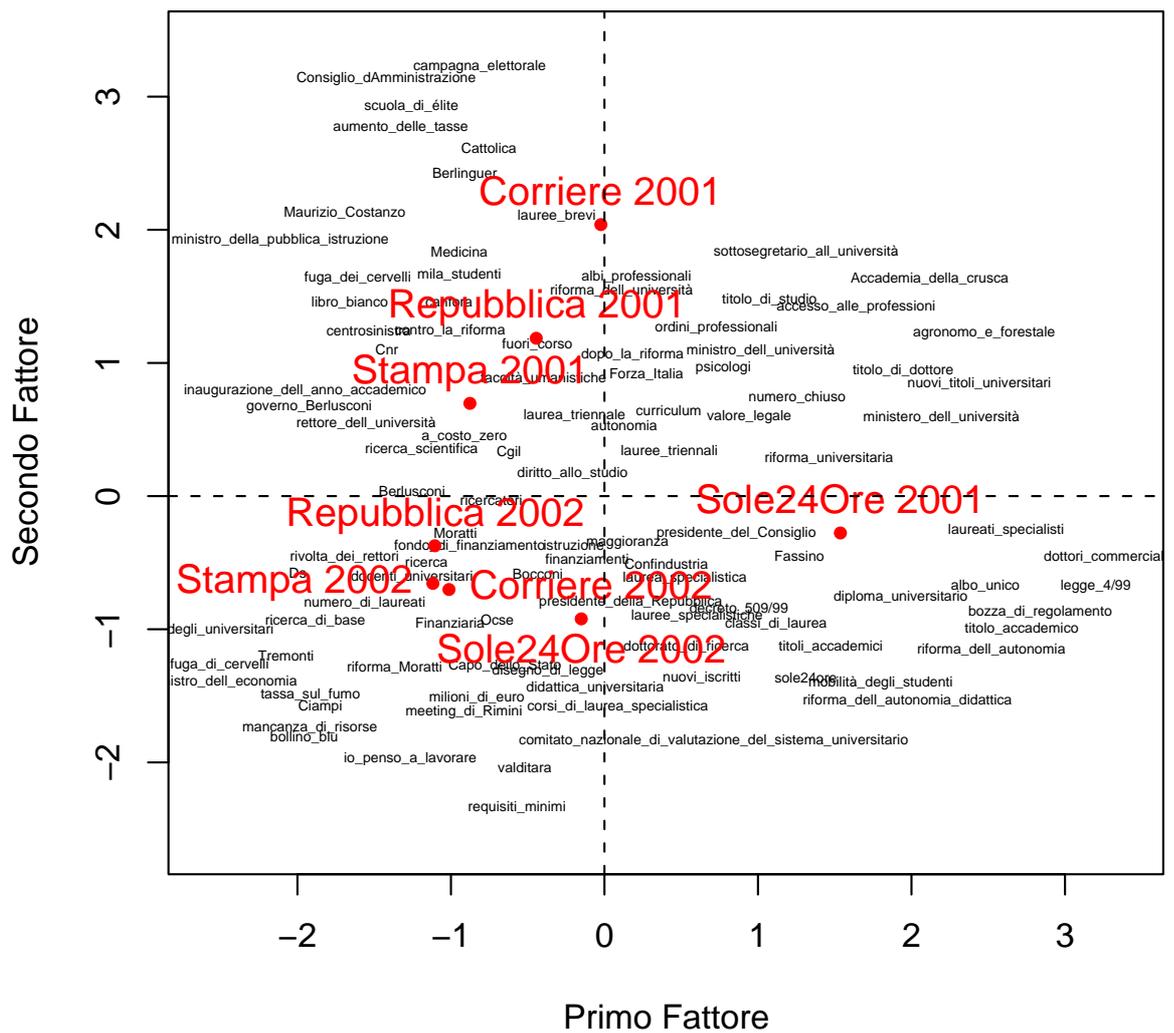


Figura 5.4: Piano fattoriale I° e II° fattore, con Sole-24Ore

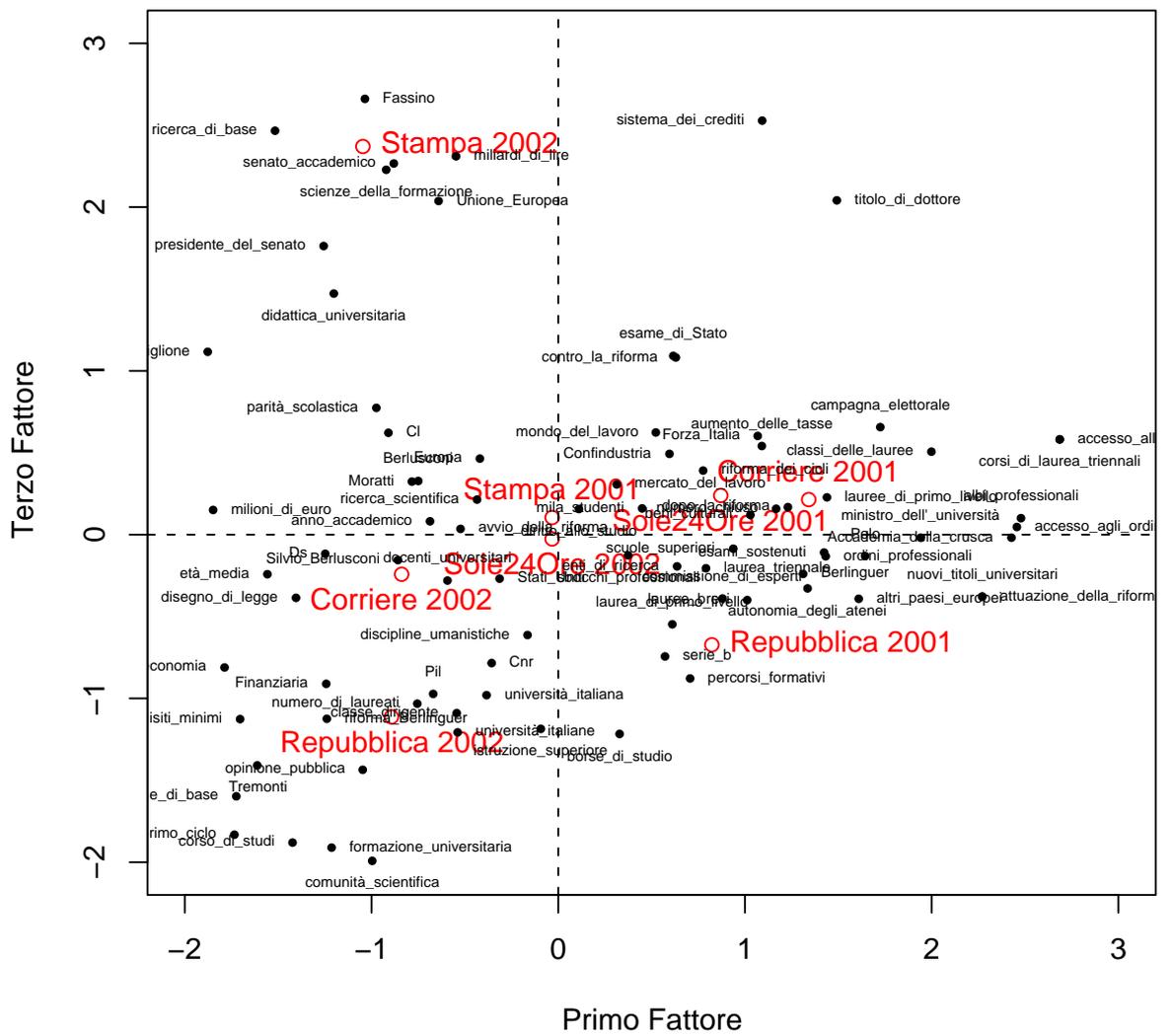


Figura 5.6: Piano fattoriale I° e III° fattore, Sole-24Ore variabile supplementare

Capitolo 6

E le opinioni?

Il corpus a nostra disposizione non si è dimostrato sinora di facile analisi. Di dimensioni elevate, non si presta ad una lettura esaustiva ed analitica dei singoli testi che possa integrare o prendere avvio dalle analisi sin qui effettuate.

Un altro elemento di difficoltà risiede nella tipologia di quotidiani utilizzati, che per loro natura sono quotidiani di opinione e non apertamente schierati. Pertanto le differenze non sono da ricercare in un contenuto palese dei testi emerso dal vocabolario utilizzato, ma anche in quella rete di relazioni che lega fra loro le parole e che non può essere ricostruita se non ad un livello molto generale tramite l'analisi delle corrispondenze.

Fino ad ora lo strumento utilizzato nelle analisi è paragonabile ad un setaccio, che per le sue finalità è più adatto a catalogare testi fra loro differenti, piuttosto che a qualificarne i contenuti. Infatti il *main goal* delle analisi finora effettuate non è stato tanto rispondere alla domanda “cosa dice questo quotidiano in questa annata?”, quanto rispondere alla domanda “questo quotidiano in questo anno dice cose differenti dagli altri?”.

L'analisi delle corrispondenze effettuata nel capitolo precedente ha risposto solo parzialmente al problema di far capire il pensiero espresso dai giornali, non andando a cogliere tutta quella parte di discorso che connota (e non solo denota) il testo. Dovremo indagare più a fondo i testi, affinando gli strumenti di ricerca, passando dall'utilizzo del setaccio separatore ad un bisturi

che consenta un intervento mirato. Il problema dell'utilizzo di un piccolo bisturi nel nostro caso è quello di dover sminuzzare una mole considerevole di dati.

La tecnica che ci consentirà un'analisi qualitativa del corpus si basa sulla selezione di una parte significativa di articoli relativa ad argomenti mirati, in modo da poter poi procedere ad un'analisi più accurata degli articoli ottenuti, senza doversi riferire all'intero corpus.

Se ipotizziamo che i quotidiani sostengano opinioni diverse sugli stessi argomenti, dobbiamo per prima cosa capire quale debba essere l'oggetto dell'analisi qualitativa: i capitoli 1 e 2 forniscono alcune tracce concrete per individuare alcuni argomenti di interesse in base ai quali cercheremo di aggredire il corpus.

Per rendere operativa questa considerazione dobbiamo pensare al corpus come una collezione di parole appartenenti a categorie semantiche differenti: il nostro obiettivo è cercare di rintracciare tutti quegli articoli che sono caratterizzati in maniera particolare dalle parole appartenenti alla categoria semantica. Al software verrà fornita una lista di forme grafiche o segmenti dei quali verrà calcolato l'indice TFIDF: in base all'indice verrà poi creato un ranking degli articoli. Il risultato è intuitivo: potremo ordinare gli articoli in nostro possesso in base all'indice TFIDF calcolato sulla lista relativa, e pertanto in base alla maggiore o minore caratterizzazione data all'articolo dalla categoria semantica oggetto di analisi.

L'indice TFIDF viene calcolato per ogni parola del vocabolario in base alla funzione:

$$TFIDF_i = f_i \times \lg \frac{N}{n} \quad (6.1)$$

dove f_i è la frequenza totale della parola i , N è il numero totale di documenti nel corpus (articoli nel nostro caso) e n è il numero di articoli che contengono la parola i . Questo indice va a pesare la frequenza con cui compare la parola nel testo in base ad un fattore tanto più grande quanto la parola è concentrata nelle partizioni del corpus. Per un approfondimento sulle tecniche utilizzate fare riferimento all'appendice A.5.

6.1 I diritti ed i doveri dell'Università

Sembra naturale cominciare le nostre analisi dall'argomento più basilare della riforma, e formulando le ipotesi più semplici. Che aspettative c'erano al momento dell'inaugurazione della riforma? I quotidiani hanno veicolato l'idea che l'Università ha di se stessa?

Alcuni i campi di intervento operativi della riforma sono riportati nel paragrafo 1.1.1 dove sono elencati i diritti ed i doveri dell'Università. Il ranking è stato effettuato in base alle parole chiave estratte dalla Dichiarazione dei diritti e doveri dell'Università; in sostanza per realizzare la lista di parole in base alla quale calcolare l'indice *TFIDF* si utilizzeranno gran parte dei sostantivi, dei verbi e degli aggettivi utilizzati nella redazione dei diritti e doveri dell'Università, parole presentate di seguito:

corporative, diritti, docenti, doveri, efficienza, formazione, formazione di base, formazione professionale, formazione superiore, formazione universitaria, imprese, innovative, innovativi, innovativo, innovazione, innovazioni, investimenti, investimento, qualità, reclutamento, ricerca, ricerca di base, ricerca scientifica, ricercatore, ricercatori, società, studenti, trasparenza, valutazione

Il calcolo restituisce, ordinati, 407 dei 431 articoli del corpus, ovvero quelli che contengono almeno una delle parole utilizzate per la creazione del ranking. Per questa prima categoria semantica da analizzare verranno scelti 14 articoli, selezionati in base alla loro posizione occupata nella distribuzione dell'indice, come da figura 6.1, dalla quale possiamo vedere, come ci sia un netto distacco dei primi 14 articoli dai seguenti.

Dei 14 articoli solamente 3 sono stati scritti nel 2001. Sul totale 5 sono di Repubblica, 4 della Stampa, 3 del Sole-24Ore e 2 del Corriere.

Nonostante nel criterio di ricerca siano presenti molti termini tecnici cari anche al Sole-24Ore (che ricordiamo schiera più del doppio degli articoli sia della Repubblica che del Corriere, e quasi il triplo della Stampa), il quotidiano più segnalato è la Repubblica. Tra i primi 50 articoli (soglia scelta anche in base all'inclinazione del grafico 6.1, che aumenta circa in corrispondenza del

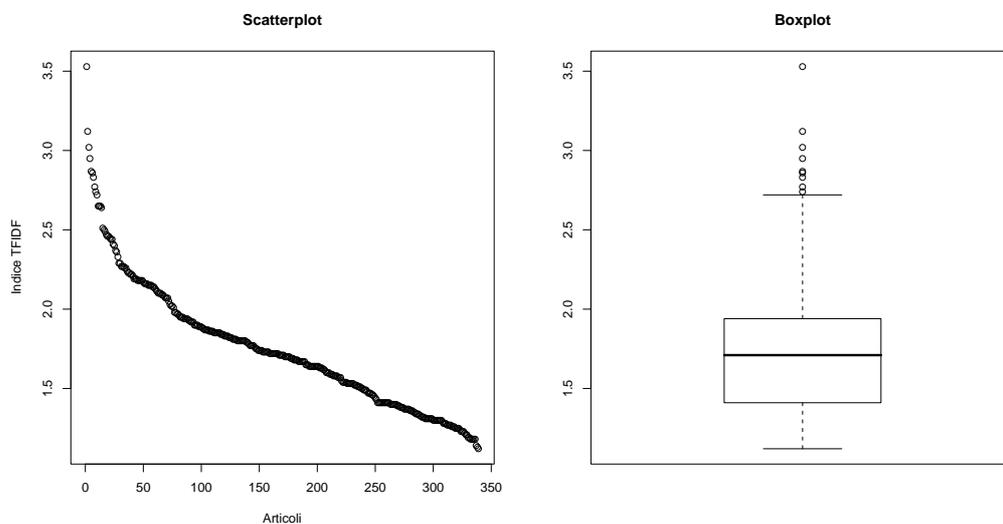


Figura 6.1: Distribuzione dell'indice TFIDF

valore dell'indice pari a 2) 25 sono del Sole-24Ore e 15 sono di Repubblica. Nonostante sul totale del corpus questi due quotidiani costituiscano il 62% dei testi, nella selezione considerata occupano l'80% degli articoli. Inoltre se il rapporto tra numero di articoli del Sole-24Ore e della Repubblica è di più del doppio, nella selezione è di 1,7 a 1. Questi due quotidiani, e soprattutto Repubblica, nella lista ordinata secondo il TFIDF sono sovrarappresentati, e pertanto sembrano riferirsi in maniera più esplicita al lessico utilizzato nella Dichiarazione.

Possiamo segnalare alcune note particolari sui 14 articoli che analizzeremo. Due dei 14 sono lettere aperte di personaggi del mondo universitario e politico ai quotidiani e altri due sono delle lunghe interviste: si tratta insomma di dichiarazioni di intento o spot pubblicitari più che articoli giornalistici. Cercheremo di capire se non altro quali quotidiani sono stati più disponibili alla pubblicizzazione della riforma o più serrati nel condurre le interviste.

Un'altra chiave di lettura per interpretare l'estrazione degli articoli dal corpus è da individuare nel fattore temporale. L'unico articolo tra i primi 14 pubblicato durante il governo di centrosinistra è quello che presenta l'indice più alto: siamo a marzo 2001.

La Repubblica

La Repubblica è il quotidiano più rappresentato nella selezione dall'indice. Per trovare un articolo pubblicato nella prima metà del 2001 dobbiamo scendere la graduatoria fino al 19° posto (luglio), poi al 42° posto c'è un articolo di giugno, al 45° posto poi troviamo un articolo sempre di marzo. Al 69° posto un altro articolo di giugno.

La grande massa dei dati selezionati dall'indice *TFIDF* riguarda il 2002 e soprattutto la sua parte finale. Facendo riferimento anche ai restanti quotidiani, 5 articoli sui primi 14 sono stati scritti tra l'ottobre ed il dicembre 2002 (più di un terzo degli articoli in un ottavo del tempo). Quei mesi hanno visto il crescere della polemica sui finanziamenti al mondo universitario, che hanno messo in pericolo il mondo della ricerca e dell'istruzione. Sono riportati qui di seguito alcuni estratti dal corpus.

- Novembre 2002, dal titolo “L'istruzione italiana è in rovina e il governo chiude la borsa” (TFIDF rank 10):

L'opinione pubblica, anche quella più avvertita, non sa o non sa fino a che punto sia grave la crisi dell'istruzione in Italia

E poi anche:

meritocrazia significa innanzitutto sostegno ai meritevoli che non hanno i mezzi e strutture di formazione di alto livello. Si dica con franchezza che la stessa corporazione è corresponsabile di questo inquietante stato di cose: ma di fronte al baratro che si è aperto dinanzi all'università italiana chiudere i lacci della borsa è una misura inquietante, è come tagliare l'ossigeno a chi sta morendo per asfissia.

- Dicembre 2002, dal titolo “È sempre emergenza-università e la battaglia dei rettori continua” (TFIDF rank 7):

(Intervista ad Adriano de Maio, NdA) I: Siete stati anche criticati per la gestione degli atenei e per la qualità della didattica.

AdM: “La percezione che ho avuto in questo periodo, anche da ambienti stranieri, è che l’università italiana e i suoi professori sono apprezzati. E che la nostra produzione scientifica è buona rispetto alle medie europee e mondiali.” Allora va tutto bene? “Dobbiamo affrontare molti problemi, ma ne abbiamo preso coscienza e stiamo cercando di risolverli, introducendo la cultura della valutazione e accettando che, sulla base di questa, ci siano riconoscimenti o penalizzazioni.”

- Dicembre 2002, dal titolo “Ma achille non prende la tartaruga” (TFIDF rank 11):

Gli studenti, che non sono solo teen-ager, ma adulti e lavoratori attratti dalle prospettive di migliorarsi attraverso le possibilità offerte da una laurea breve, sono cresciuti nelle immatricolazioni, ma sono in una situazione di disincanto. Neppure le lauree di secondo livello e i master saranno sufficienti ad accorciare del tutto le distanze tra formazione e mercato.

Mostriamo di seguito alcuni estratti dagli articoli della prima metà del 2001:

- Marzo 2001 dal titolo “Un patto con famiglie e studenti così vinceremo la sfida del futuro” (TFIDF rank 1):

I 77 atenei italiani sono decisamente sottofinanziati rispetto agli standard europei ma non hanno paura di impegnarsi in una competizione sempre più dura per avere più finanziamenti e più studenti, si muovono con decisione e con sempre maggiore successo nel mercato dei finanziamenti privati e pubblici della ricerca, sanno sperimentare nuovi percorsi formativi, nuove tecnologie didattiche, nuove linee di ricerca in collaborazione con il mondo delle imprese e del lavoro.

- Marzo 2001 dal titolo “I rettori: La riforma non si tocca” (TFIDF rank 45):

(Parla Luciano Modica, presidente della CRUI, NdA) Ma, ripeto, la riforma dell'Università non si tocca, perché bloccarla provocherebbe un disastro clamoroso. Gli Atenei si sono già attrezzati per la riforma della didattica, la laurea triennale e quella specialistica. Tutto può essere migliorato, ma l'impianto generale che uniforma la struttura italiana a quella degli altri paesi europei è un caposaldo imprescindibile per uscire dalle difficoltà endemiche dell'università.

- Giugno 2001, dal titolo “Riforma universitaria il timbro finale del Polo” (TFIDF rank 42):

Come tutte le riforme di grande complessità anche questa, soprattutto nella fase conclusiva, ha dato l'impressione a molti di ingabbiare lo spirito originario in un troppo rigido carapace burocratico.

E poi anche:

Il nuovo governo finirà per trovarsi nella fortunata condizione di portare a casa la realizzazione di un provvedimento di enorme portata, da inserire nell'elenco delle annunciate grandi modernizzazioni del paese, capitalizzando sulla grande mole di lavoro già fatto ed essendo nella posizione assai vantaggiosa di apportare quelle migliorie che si renderanno necessarie, come sempre avviene, nel periodo di rodaggio.

- Novembre 2001 dal titolo “Università, i buchi della riforma” (TFIDF rank 41):

Se gli stimoli all'innovazione e alla progettazione di lungo periodo non sono assicurati nella riforma Berlinguer Zecchino, essi tanto meno provengono dal governo Berlusconi, che ha pesantemente tagliato le già esigue risorse del bilancio statale per la ricerca e gli studi. Dopo le eclatanti promesse elettorali, regolarmente rimangiate anche in campo fiscale, riemerge l'atteggiamento di chi non crede nella scienza e nella qualificazione delle giovani generazioni.

Il messaggio che si vuole mandare nella prima parte del 2001 insomma è riassumibile nella formula “le università sono pronte al cambiamento, ma molto probabilmente incontreranno delle difficoltà evidenti e non dipendenti da loro”. In ogni caso l’opinione espressa e veicolata sulla riforma non è negativa.

Nel 2002 invece, nonostante Adriano de Maio (nel quinto estratto) cerchi di spegnere le fiamme che sembrano bruciare attorno all’Università, la stampa comincia a soffiare sul fuoco, non a caso l’intervistatore gli porge delle domande sibilline, ed alla prima risposta fiduciosa di de Maio, l’intervistatore chiede conferma con la domanda dal tono molto ironico “allora tutto bene?”.

Se gli articoli del 2001, nonostante un tono cauto, promettevano un impegno dell’Università nell’affrontare i bivi di una riforma difficile ma necessaria e le critiche più violente sollevate erano relegate ad articoli molto brevi (si trovano critiche ad esempio nel marzo 2001, articolo di 355 parole in ventinovesima pagina, o nel novembre 2001, 394 parole in ventiduesima pagina), nel 2002 viene lasciato ampio spazio a qualsivoglia opinione e contributo negativo sull’Università.

Enigmatica a questo punto risulta la lettera aperta di Alberto Asor Rosa alla Repubblica nel marzo 2001, in cui si denunciano “la violenza inusitata, con cui si sono moltiplicati nelle ultime settimane gli attacchi alle riforme del centrosinistra nel campo della scuola e dell’Università”. Attacchi che, nel sub-corpus della Repubblica a nostra disposizione, prima del marzo 2001, sono molto limitati. La Repubblica, se non altro nel 2001, si mostra essere uno strenuo difensore della riforma.

La Stampa

Analizzeremo quattro articoli de La Stampa (uno del 2001 e tre del 2002), altro quotidiano sovra-rappresentato nella selezione dell’indice TFIDF. Come per la Repubblica, le tematiche dell’innovazione, della qualità e della ricerca vengono sviluppate soprattutto nel 2002. Ad onor del vero, la Stampa pubblica nel 2001 un articolo in controtendenza con gli altri quotidiani

(ricordiamo il piano fattoriale di figura 5.5, che vedeva la Stampa del 2001 in posizione molto isolata) e che denuncia lo stato precario della ricerca in Italia.

Ecco di seguito alcuni estratti dai quattro articoli della Stampa selezionati dall'indice TFIDF:

- Ottobre 2001, dal titolo “Più fondi agli atenei” (TFIDF rank 5):

Un clima di preoccupazione ha caratterizzato ieri la riunione del Comitato di presidenza della Conferenza dei Rettori universitari, per la prima volta in trasferta a torino

Poi anche:

(parla Luciano Modica, NdA) “Ogni governo, non importa di quale colore, prima dichiara che la formazione superiore e la ricerca sono le risorse strategiche del paese. Poi, ogni volta, i fondi diminuiscono.”

- Ottobre 2002, dal titolo “Fassino: la riforma Moratti è regressiva e prevede solo tagli” (TFIDF rank 8):

Infine Fassino denuncia il tentativo di ridurre l'autonomia delle università, ricordando l'appello di 15 mila ricercatori e la minaccia di dimissioni in massa dei 77 rettori.

- Febbraio 2002, dal titolo “La Moratti all'inaugurazione della Cattolica” (TFIDF rank 13):

Poco prima il cardinale Martini aveva celebrato una messa nella basilica di Sant'Ambrogio, e durante l'omelia aveva formulato per la Cattolica una preghiera ed un augurio: “Il Signore doni a tutte le componenti dell'università quell'abbondanza di Spirito Santo che permetta di fare sintesi tra cultura e Vangelo, e approfondire con tutti i mezzi della ricerca scientifica e critica la verità sull'uomo tenendo aperti gli orizzonti della trascendenza e della sua destinazione eterna”.

Ed anche:

Il governo preparerà “un sistema di monitoraggio della qualità dei corsi”, per misurare “l’adeguatezza dei mezzi messi a disposizione dei docenti e degli studenti. Porremo gli studenti nelle condizioni di fare un confronto trasparente tra università e singole facoltà quanto a strumenti didattici, strutture funzionanti, libri utilizzabili e consultabili”. Sarà rafforzato il preorientamento universitario nelle scuole, e verrà introdotta “la certificazione della qualità dei servizi”, per favorire “il confronto competitivo tra le diverse offerte didattiche”. Criteri manageriali insomma, per l’università che verrà.

- Giugno 2002, dal titolo “Scienza in pericolo, danno per l’Europa” (TFI-DF rank 14):

Un gruppo di ricercatori e di Premi Nobel ha inviato un documento all’Unione europea affinché siano evitati, con adeguati finanziamenti e con urgenza, il declino della scienza e la fuga dei cervelli.

Da questi quattro articoli possiamo trarre delle importanti considerazioni. L’atteggiamento tenuto nei confronti della riforma non è sempre positivo, ma oscilla e dipende molto dalle dichiarazioni rilasciate dai personaggi intervistati. Gli articoli non hanno quasi mai un taglio connotabile, e riportano piuttosto fedelmente le risposte date dagli intervistati. Inoltre si nota il tentativo di separare il mondo universitario e quello della ricerca. Cnr ed Università a differenza dei quotidiani generalisti sono riconosciuti come mondi a parte, che funzionano in base a logiche diverse.

Se però consideriamo che dopo i quattro articoli rintracciati nei primi 14 posti del ranking il prossimo articolo de la Stampa si trova al 57° posto, questo può indicare una certa polarizzazione degli articoli del quotidiano: ovvero, il tentativo di dare un taglio molto ben definito agli articoli.

Il Sole-24Ore

Il primo articolo del Sole-24Ore selezionato dal ranking è dell'agosto del 2001, dal titolo "Concorsi, l'interno vince facile", che traccia un bilancio della riforma dei concorsi del 1998.

- Giugno 2002 dal titolo "La prima sfida è la qualità"

(Intervista a Silvio Fortuna, all'epoca *responsabile dell'Education per Confindustria*) La riforma dei cicli e dell'autonomia didattica e dell'università è un'opportunità storica per recuperare efficienza in un campo in cui abbiamo accumulato gravi ritardi rispetto ai partner europei. Sugerirei di non parlare più di 3+2: è una semplificazione che mortifica le novità vere della riforma. [...] La rigidità è il vero nemico della riforma. Occorre recuperare lo spirito di quella straordinaria esperienza rappresentata dai diplomi universitari professionalizzanti, che ha garantito per alcuni corsi sbocchi occupazionali nel 100% dei casi.

- Gennaio 2001, dal titolo "La sfida è guidare i mutamenti in atto" (TFIDF ranking 15)

La discussione in atto tra critici e sostenitori della riforma degli ordinamenti didattici universitari ha il difetto di mantenersi a un livello molto astratto. Si discute come se la nuova normativa fosse destinata a cambiare la natura dell'istruzione superiore, abbassandone irrimediabilmente il livello per gli uni, migliorandone il rendimento, o adeguandola all' Europa, per gli altri. Essa però interviene su una realtà che, fin dagli anni 60, si è profondamente trasformata per effetto di spinte economiche e sociali, in gran parte ignorate dalla legislazione universitaria. L'apparente immobilità del sistema è quindi più formale che reale.

Ovviamente il focus è molto centrato sui riflessi della riforma sul mondo economico. Il contributo che esula da questo campo è fornito da un articolo

di Alessandro Figà Talamanca del gennaio 2001. In ogni caso i toni utilizzati non sono ne' di critica, ne' di lode nei confronti della riforma. Insomma, anche analizzando nel dettaglio gli articoli, come confermavano i piani fattoriali del capitolo 5, il Sole-24Ore si delinea come quotidiano imparziale, oggettivo e molto analitico.

Il Corriere

Per il Corriere gli unici articoli di interesse segnalati dall'indice sono una lettera ed un'intervista dell'allora ministro Letizia Moratti (1 e 21 marzo 2002), che ovviamente presentano contenuti rassicuranti e orientati a segnalare le innovazioni portate dal governo di quel periodo. Dei due è interessante soprattutto la lettera, in quanto si tratta di una risposta ad un feroce articolo (26 febbraio 2002) scritto da Angelo Panebianco, che concludeva il suo pezzo così:

Forse è il caso che il presidente del Consiglio e i ministri dell'Economia e dell'Università trovino il tempo per fare, con i loro collaboratori, un seminario a porte chiuse. Per dirsi a quattr'occhi: "Fin qui abbiamo proprio sbagliato. Dotiamoci subito di ciò che non abbiamo: una "vera" politica dell'Università (diversa da quella del centrosinistra) e una "vera" politica della ricerca scientifica (diversa da quella di tutti i governi precedenti)". [...] Ma in politica, secondo un noto adagio, ci sono cose peggiori dei crimini: gli errori. Il governo ne sta giusto commettendo uno.

Mentre Letizia Moratti conclude in questa maniera:

Il nostro programma si chiama sviluppo. I conti si faranno (li faranno i cittadini) alla fine della legislatura.

Il materiale estratto non è sufficiente a delle valutazioni. Le motivazioni possono essere diverse, ma legate soprattutto al lessico utilizzato. I piani fattoriali di figura 5.4, 5.5 e 5.6 correlano il Corriere con un lessico molto tecnico rispetto agli altri quotidiani generalisti. Pertanto possiamo aspettarci che le informazioni che veicola il Corriere siano rintracciabili costruendo il ranking in base ad altre liste di parole.

6.2 Il vil denaro

A quanto pare uno dei nodi da sciogliere per spiegare le estrazioni di dati appena effettuate sta nel denaro: la maggior parte dei problemi riguardanti l'Università hanno avuto luogo a fine 2002, il periodo di varo della Finanziaria. Prendiamo in considerazione i primi 11 articoli indicati dall'indice (vedi figura 6.2), selezionati in base alle seguenti parole:

aumento delle tasse, battaglia, economia, economica, economico, euro, finanziamenti, finanziamento, finanziamento ordinario, finanziare, Finanziaria, finanziaria, finanziario, fondi, fondo di finanziamento, investimenti, investimento, legge finanziaria, lire, miliardi, miliardi di lire, milione, milioni, milioni di euro, ministro dell' economia, risorsa, risorse, risorse finanziarie, soldi, stanziamenti, stanziamento, tagli, tagliare, taglio, tassa, tassa sul fumo, tasse

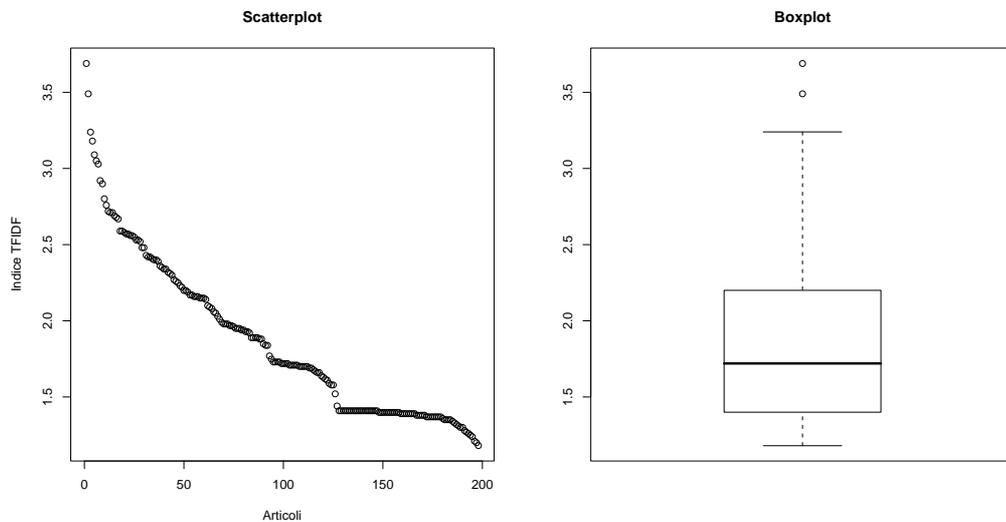


Figura 6.2: Distribuzione dell'indice TFIDF

Il Sole-24Ore

Ecco una vera notizia: il Sole-24Ore non parla di soldi. Ebbene si, non nomina la Finanziaria, non i milioni di euro, non i fondi, non le risorse, non

lo stanziamento ordinario o straordinario. Rispetto alla classificazione per questa lista di parole, il primo articolo del Sole-24Ore si trova in quindicesima posizione. Certamente, approfondendo leggermente la questione si può scoprire che l'atteggiamento è totalmente coerente con la mission del quotidiano, soprattutto per quel che riguarda la già annunciata tendenza a non sbilanciarsi su dati non ancora acquisiti e soprattutto il non entrare nel merito delle critiche o sostegni dati a decisione che in larga parte hanno a che fare con la politica. Tutto ciò porta il Sole-24Ore a non commentare alcuna scelta finanziaria in ambito universitario. Che sia un'ulteriore conferma sulla politicizzazione dell'argomento?

Il Corriere

Il Corriere schiera cinque articoli tra i primi posti: molti più rispetto all'estrazione precedente. "Marzano e Moratti, l'ultima battaglia contro i tagli" (settembre 2002), "Tremonti insensibile sulla scuola" (novembre 2002), "Ricerca, un impegno del governo" (marzo 2002), "Il governo corre ai ripari: 350 milioni all'università" (ottobre 2002), "Diciotto mesi di tagli, poi lo sfogo: questa manovra è contro di noi" (novembre 2002) sono titoli molto esplicativi dell'ottica battagliera che viene assunta.

La Repubblica e la Stampa

Scorriamo i titoli proposti dai due quotidiani: "Tremonti sordo su scuola e ricerca" (Repubblica, novembre 2002), "E Tremonti apre la borsa" (Repubblica, dicembre 2002), "L'università perde 15 milioni di euro" (Stampa, gennaio 2002), "Ora l'università ha paura del crac" (Repubblica, dicembre 2002), "Più fondi agli atenei" (Stampa, ottobre 2001), "Il mondo dell'istruzione tra trasformazioni e critiche" (Stampa, gennaio 2002).

Come possiamo vedere anche in questo caso il fronte è compatto: i tagli dei fondi sollevano un vero vespaio.

Possiamo comunque chiederci come ci sia finito un articolo del 2001 in un'estrazione che, per costruzione, intendeva individuare esclusivamente articoli di fine 2002. "Più fondi agli atenei" (potremmo menzionare anche

l'articolo di gennaio 2002 “L'università perde 15 milioni di euro”) è un titolo che non ci saremmo aspettati in quel periodo dell'anno. Il sommario dell'articolo è il seguente:

Il direttivo nazionale preoccupato per la manovra finanziaria
“L'Università penalizzata nel delicato momento della riforma”
Ricerca, chiesti maggiori investimenti per evitare fughe all'estero

Ma scopriamo immediatamente il motivo dell'allarme procurato: l'articolo è il resoconto di una riunione della CRUI tenutasi proprio a Torino.

L'articolo del gennaio 2002, anch'esso pubblicato con una tempistica differente rispetto agli altri quotidiani, dato l'argomento trattato, infatti parla dei problemi di finanziamento dell'Università di Torino. L'articolo del 2001 non era altro che un allarme locale, al quale si è cercato di dare eco nazionale.

6.3 La ricerca e la cultura

Il ranking effettuato sui diritti e doveri dell'Università e sui fondi ha lasciato alcune domande senza risposta. Soprattutto vorremmo approfondire le opinioni dei quotidiani nel 2001, ovvero il periodo più a ridosso dell'avvio della riforma. La nostra ricerca continua, esplorando in questo paragrafo otto articoli (vedi figura 6.3) selezionati in base alla loro affinità con un lessico correlato con la ricerca. La lista di parole utilizzate per l'estrazione è la seguente:

comunità scientifica, coscienza, cultura, culturale, culturali, dottorato, dottorato di ricerca, enti di ricerca, formazione, formazione di base, formazione superiore, formazione universitaria, fuga dei cervelli, intellettuale, intellettuali, preparazione, ricerca, ricerca di base, ricerca scientifica, ricercatore, ricercatori, ricerche, sapere, scientifica, scientifiche, scientifici, scientifico, scienza, scienziati

Degli otto articoli che verranno analizzati, tre sono della Stampa, due della Repubblica e uno sia per Corriere e per Sole-24Ore. Titoli come “Cnr, i

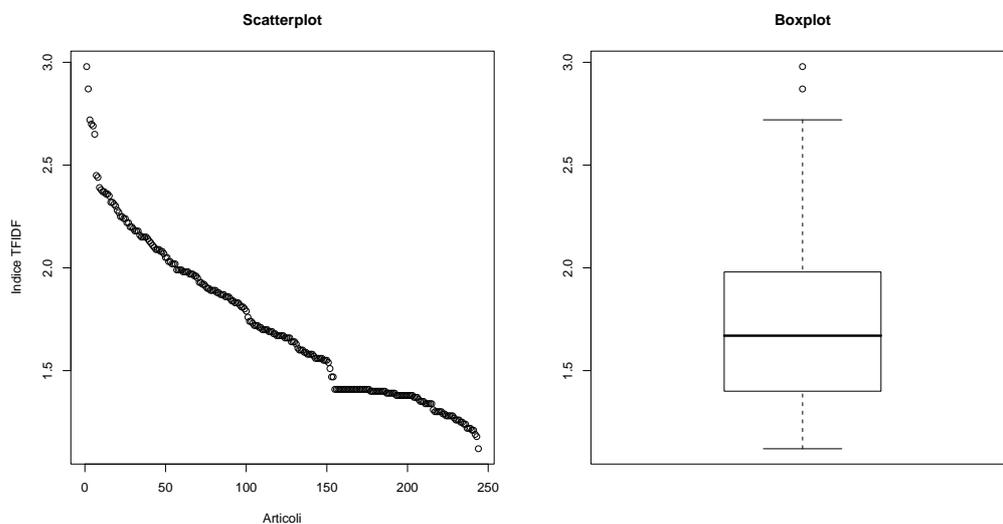


Figura 6.3: Distribuzione dell'indice TFIDF

guai di una riforma realizzata a metà”, “Ma Achille non prende la tartaruga”, “Ricerca indietro tutta”, “Scienza in pericolo, danno per l’Europa”, “Il tempo degli oracoli”, “Il ritardo da colmare” e “Ricercatori, una specie a rischio” non lasciano dubbi sull’orientamento mantenuto dai quotidiani.

Avevamo già intuito la particolare sensibilità della Stampa nei confronti della ricerca, e questa estrazione conferma le osservazioni effettuate in precedenza. È spiccata soprattutto la sensibilità alle problematiche dei ricercatori e del Cnr.

Purtroppo, come per nel paragrafo precedente, l’unico problema è che di otto articoli, due sono del 2001 e ben sei del 2002.

6.4 E il 2001?

Ricercando delle critiche puntuali, o meglio ricercando dei riferimenti alla riforma universitaria utilizzando un lessico non direttamente legato al mondo universitario, non abbiamo ottenuto se non dei risultati relativi. In più questi risultati si sono rivelati temporalmente limitati al 2002. In pratica, la

ricerca di giudizi “a posteriori” sulla riforma, ovvero il cercare di considerare i possibili effetti della riforma non ha fornito dei risultati apprezzabili.

Il 2001 insomma potrebbe venire a delinarsi come anno di transizione, non solo nella realtà, ma anche sulla carta stampata. Per supportare questa ipotesi basiamoci su una lista (più estesa di quelle utilizzate sinora) che raggruppa uno spaccato di terminologia specificamente universitaria di natura tecnico-didattico-burocratica:

applicazione della riforma, attività formative, attuazione della riforma, autonomia, autonomia degli atenei, autonomia didattica, autonomia universitaria, avvio della riforma, biennio specialistico, classi di laurea, concorrenza, contro la riforma, corsi di laurea, corsi di laurea specialistica, corsi di laurea triennali, corsi triennali, corso di laurea, crediti, crediti formativi, didattica, didattica universitaria, didattiche, didattici, didattico, dopo la riforma, formativi, formazione, formazione di base, formazione superiore, formazione universitaria, laurea breve, laurea di primo livello, laurea specialistica, laurea triennale, laurearsi, laureati, laureati specialisti, laureati triennali, lauree specialistiche, lauree triennali, numero di laureati, nuovi corsi di laurea, nuovo ordinamento, offerta didattica, offerta formativa, ordinamenti, ordinamenti didattici, ordinamenti professionali, percorsi formativi, percorso formativo, piani di studio, post-laurea, regolamenti didattici, riforma Berlinguer, riforma degli ordinamenti didattici, riforma dei cicli, riforma dei cicli scolastici, riforma del 3+2, riforma dell'autonomia, didattica, riforma dell'università, riforma Moratti, riforma universitaria, scienze della formazione, sistema dei crediti, specialistica, specialistiche, specialistici, triennale, triennali, triennio, vecchia laurea, vecchio ordinamento

Dei primi 30 articoli “tecnici” 20 sono pubblicati dal Sole-24Ore: ovviamente la precisione del lessico di questo quotidiano è sicuramente superiore a quella degli altri giornali, ma anche il suo livello di attenzione verso le realtà locali, le nuove normative e il supporto a nuove facoltà o università è molto alto (insomma, è un quotidiano-sponsor per le realtà locali, nonché tra i più recettivi rispetto agli uffici stampa universitari).

Finalmente, degli altri 10 articoli che considereremo (vedi figura 6.4) 8 sono del 2001 e 2 del 2002. Ma non solo: 8 sono del Corriere, uno per la Stampa e uno per la Repubblica. Per questo motivo, ci spingeremo nelle analisi anche oltre la decima posizione del ranking ottenuto, in modo da coinvolgere nelle analisi un numero sufficiente di testi per tutti e tre i quotidiani.

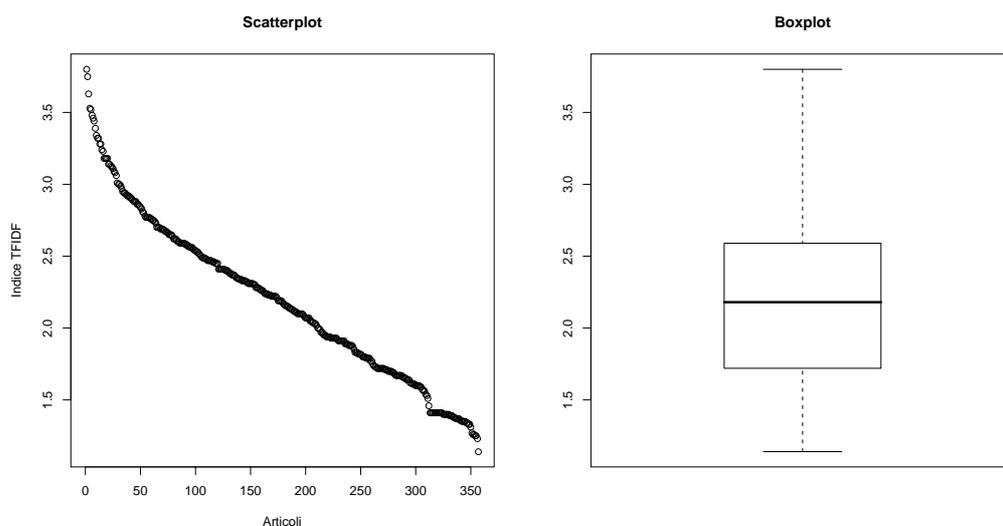


Figura 6.4: Distribuzione dell'indice TFIDF

Il Corriere

Si propongono di seguito alcuni estratti dagli articoli del Corriere:

- Maggio 2001, dal titolo “Università, il Polo bloccherà la nuova laurea” (TFIDF rank 3):

Il “blocco istantaneo” della laurea triennale e della laurea specialistica biennale figura tra i primi impegni del programma di governo, ma in campagna elettorale i leader del Polo non hanno affrontato quasi mai l'argomento. [...] Ora, probabilmente sull'ondata di fax e raccolte di firme, qualcosa sta cambiando. “In uno dei primi consigli dei ministri” spiega infatti il responsabile università di Forza Italia “approveremo un provvedimento che ci consentirà

di prendere tempo per rivedere molte cose”. Per Luciano Guerzoni, sottosegretario del Murst, però, sospendere l’applicazione della riforma significa gettare le università nel caos. [...] “Siamo alla pura restaurazione”, commenta l’ex ministro della Pubblica istruzione, Luigi Berlinguer.

E poi anche:

Nel mondo universitario [...] sta circolando un nuovo appello contro la riforma. Le firme sono circa 40, compresa quella dello storico Franco Cardini, dell’ateneo di Firenze, presidente Rai nel primo governo Berlusconi. Per i firmatari “la riforma è nata in modo arbitrario, al di fuori di una verifica seria in ambiente accademico e addirittura di un controllo da parte del Parlamento”

- Agosto 2002, dal titolo “Università, a rischio la riforma delle lauree” (TFIDF rank 9):

Il 3+2 [...] rischia di trasformarsi in un quinquennio secco. Esattamente il contrario di ciò che si era prefissa la riforma dell’Università: offrire agli studenti un percorso più breve e flessibile.

- Dicembre 2001, dal titolo “Laurea possibile anche fuori tempo massimo” (TFIDF rank 13):

L’autonomia non si discute. Sull’onda dell’entusiasmo autonomista, molti atenei si adeguano e aprono ai fuori corso. “Cerchiamo di venire incontro agli studenti, solo il 30% degli studenti si laurea”, spiegano. Ma i maligni ribattono: “C’è una rincorsa agli iscritti e si tenta di sfornare più laureati possibili per ottenere gli incentivi previsti, in questo caso, dal ministero”.

- Giugno 2001, dal titolo “Bloccare la riforma? Meglio correggerla” (TFIDF rank 24):

(Scriva Angelo Panebianco, Nda) C'è un ministero di cui si parla poco mentre si attende l'annuncio della formazione del governo. È quello della Pubblica Istruzione [...]. Eppure, è una poltrona che rischia di essere caldissima. Scuola e Università sono, in questo momento, "cantieri aperti" a causa delle radicali riforme innescate dal centrosinistra. A seconda dell'atteggiamento che il centrodestra assumerà rispetto a queste riforme cambierà in un senso o nell'altro il destino delle nostre istituzioni educative. [...] essa (la nuova classe di governo) dovrà mettere particolare impegno per dimostrare di essere capace di fare una seria ed efficace politica nel settore educativo.

E poi anche:

flessibilità, libertà di sperimentazione e gradualità sono state spazzate via man mano che l'iter decisionale si avviava alla conclusione. Per responsabilità dei politici, certo, ma anche di altri soggetti, come il Consiglio Universitario Nazionale e, forse, anche settori della burocrazia ministeriale. Compito del nuovo ministro dovrebbe essere quello di ridisegnare queste istituzioni. Così come sono, e come prova proprio la vicenda della riforma, non aiutano le università, ne ledono gravemente l'autonomia, sono produttrici di vincoli assurdi e di forme di burocratizzazione esasperata. Proprio ciò di cui l'università non ha bisogno.

È il secondo intervento di Angelo Panebianco sul Corriere. Ricordiamo come l'estrazione automatica di informazione sia stata ottenuta proponendo un lessico "tecnico", e per nulla connotato.

La Repubblica

- Giugno 2002, dal titolo "La Facoltà di Lettere non è morta" (TFIDF rank 26):

Al contrario oggi si registra un boom d'iscrizioni, e sono sicuro che moltissimi vorranno proseguire anche nei due anni successivi.

Ma non più come fuori corso. La verità è che la vecchia laurea non esisteva più da tempo. Non esisteva per il 70 per cento degli immatricolati perché essi non finivano gli studi e non conseguivano nessuna laurea. Erano universitariamente dei falliti, per colpa d' un sistema che non si curava mai di loro.

- Giugno 2001, dal titolo “Riforma universitaria il timbro finale del Polo” (TFIDF rank 41):

In questa situazione di passaggio, un piccolo gruppo di docenti, si è mobilitato invece con l'obiettivo di bloccare del tutto la riforma, senza una proposta alternativa, ma obbedendo a quel riflesso condizionato di rifiuto dei cambiamenti che Pareto definiva misonismo.

- Marzo 2001, dal titolo “Dove nasce la malattia degli Atenei dequalificati” (TFIDF rank 59):

Tre anni sembrano pochi ai nostri critici, che tuttavia non se la sentono di scegliere un secco ritorno al passato (meno male)

E poi anche:

L'essenza del meccanismo dei crediti contro cui i nostri autori non sembrano aver nulla da far valere consiste proprio nell'individuazione di una misura oggettiva del tempo di lavoro dei discenti, importante ai fini di una razionale programmazione didattica. Ciò porta inevitabilmente a una quantificazione, che non può essere calcolata se non attraverso un criterio generale e convenzionale. Altrimenti, salta tutto.

Inoltre, il ranking ripropone l'articolo del marzo 2001 dal titolo “Un patto con famiglie e studenti così vinceremo la sfida del futuro” (vedi paragrafo 6.1).

Come possiamo vedere la Repubblica si pone come netta sostenitrice della riforma, se non altro nel 2001, dato che abbiamo già chiarito come a partire da fine 2001, dopo il cambiamento di governo, ed allo stringersi del cordone della borsa statale, il quotidiano cambi la sua opinione.

La Stampa

La Stampa pubblica il seguente articolo:

- Gennaio 2001, dal titolo “Riforma universitaria, un modello europeo” (TFIDF rank 11):

L’antico sistema presenta difetti assai gravi, da tempo sotto gli occhi di tutti: basso numero di laureati, alto numero di abbandoni, altissimo numero di studenti che non riescono a concludere il percorso nei tempi previsti dai piani di studio, entrata ritardata nel mercato del lavoro, con insopportabili costi aggiuntivi individuali e collettivi. Questi mali sono conseguenza di molteplici fattori. Certamente negativa è stata la indiscriminata libertà di accesso a ogni facoltà con il possesso di un qualunque titolo medio superiore e senza alcuna verifica di adeguata preparazione.

Dobbiamo fare notare come gli articoli della Stampa che vengono selezionati dall’indice TFIDF siano tutti di portata locale, e rispecchino la visione torinese della riforma. Il taglio degli articoli è molto pratico, e troviamo titoli del tipo:

- L’Università cerca i figli di Quark
- Lauree brevi aumentano gli iscritti all’Università
- Il deserto delle lauree specialistiche
- Ingegneri bonsai con laurea breve: l’ordine degli ingegneri è in subbuglio.

Tutti riferiti ai riflessi della riforma su Torino.

6.5 Alcune considerazioni

Avevamo cominciato il percorso che ci ha guidato attraverso questo capitolo spinti dalla necessità di fare luce su alcuni punti che erano rimasti bui e nascosti tra le pieghe di una struttura che si era delineata solamente nei suoi tratti essenziali. Avevamo cominciato con una metafora, quella del bisturi come strumento in grado di sezionare il nostro corpus in modo da renderlo concretamente analizzabile in alcuni aspetti che concernessero anche le connotazioni del testo.

Un risultato, confrontando le diverse estrazioni effettuate grazie all'indice *TFIDF* è stato senz'altro raggiunto. Possiamo tirare le fila del discorso risalendo all'indietro quest'ultimo capitolo:

1. Diritti e doveri dell'Università:

- la Repubblica è il quotisiano che fa più leva su questo lessico positivo, molto legato ad una visione strategica della riforma ma allo stesso tempo di natura generica. Inoltre nel passaggio tra 2001 e 2002 sembra cominciare a connotare meno positivamente i propri articoli;
- anche la Stampa fa largo uso del lessico relativo a quest'area semantica, ma le interpreta in chiave abbastanza personale. Viene dato spazio al dibattito, e, come abbiamo visto nella fase di analisi delle corrispondenze, c'è una forte tendenza all'appellarsi all'autorità;
- il Sole-24Ore si mantiene neutro, non si sbilancia dare valutazioni strategiche sulla riforma;
- il Corriere non pesca proprio da questo bacino semantico: questo lessico (abbastanza caro ai politici) viene rappresentato soprattutto dalla lettera di Letizia moratti in risposta all'articolo di Panebianco. Ma tale lettera non è accompagnata da nessuna forma di commento successivo;

2. Fondi e denaro:

- abbiamo visto come il Sole-24Ore non parli di soldi, almeno non nei termini utilizzati dai quotidiani di opinione;
- la Stampa è l'unica a mettere la testa fuori dal gruppo. Denuncia in modo vigoroso i tagli dei fondi alla ricerca, ma ha una visione territorialmente limitata, e le sue critiche sono mirate soprattutto alla salvaguardia della realtà universitaria torinese;

3. Ricerca e cultura:

- questo argomento viene toccato da tutte le testate in analisi, e la tendenza è uniformemente orientata all'allarmismo per una ricerca che è vista come vicina al collasso;

4. Il lessico tecnico:

- è il Corriere che sembra assorbire i concetti della riforma, è lui che abbraccia il nuovo lessico, e che nel 2001 dà ampio spazio al dibattito sulla riforma;
- la Repubblica evidenzia un altro cambiamento di opinione
- la Stampa sempre molto locale
- Sole-24Ore da rivedere!!!!!!

Conclusioni

Si è cercato di dare a questo lavoro un taglio che portasse a dei risultati graduali, andando a strutturare le informazioni per raffinamenti successivi.

La mole di dati analizzata di volta in volta nel corso dei capitoli è andata costantemente diminuendo, partendo da una considerazione del vocabolario nella sua totalità, per terminare con scelte mirate e determinate grazie alle analisi precedenti. L'approccio *bottom-up* ha permesso di pescare dal corpus quelle parti di testo che maggiormente caratterizzano i diversi argomenti trattati sui giornali, sfruttando nella selezione tutta una serie di vincoli che sono stati ricavati mano a mano nel corso delle pagine.

Nel capitolo 3 avevamo considerato il corpus come un insieme omogeneo di parole, con poche o nessuna connessione tra loro se non l'appartenenza o meno a uno specifico quotidiano o a una annata. Pur facendo a meno di considerare tutta quella forza di attrazione tra le parole che è la struttura della nostra lingua, abbiamo potuto osservare come ci fossero forti differenze soprattutto tra le due annate prese in considerazione, e come il Sole-24Ore utilizzasse un lessico nettamente differente dagli altri quotidiani. Anche in rapporto all'annata le differenze si sono notate palpabili, nel proporre nel 2001 una gamma più ampia di termini esclusivi e specifici.

Una volta individuato nel tempo una delle chiavi di lettura per estrarre informazioni sul corpus abbiamo cercato di “mettere in fila” gli articoli in analisi, cercando di capire quali parole subissero un effettivo decadimento nel loro utilizzo e quali fossero sopra o sotto utilizzate grazie a picchi di frequenza. Già l'imposizione di un vincolo temporale alla massa dei dati ci ha consentito di individuare un preciso trend per alcune categorie semantiche specifiche: quelle relative al lessico del mondo universitario sembrano

rispondere a logiche abbastanza dipendenti da trend, e non si rivelano stabili nell'utilizzo se non come "livello globale". Il lessico peculiare della riforma cala, quello generico sopperisce al calo.

Un metodo per considerare in maniera globale i legami tra le parole del corpus consiste nell'analizzare la tabella di contingenza vocabolario×subtesti con la tecnica dell'analisi delle corrispondenze. La forza di associazione tra i diversi quotidiani è emersa così in modo più evidente, portando alla luce in modo lampante la diversità del Sole-24Ore, ma anche la sua quasi banalità una volta eliminato il suo lessico tecnico. La diversità ad emergere per seconda dal gruppo è quella della Stampa, lasciando così soli, appaiati Repubblica e Corriere.

Per giungere ad una distinzione che focalizzasse in maniera più puntuale le differenze evidenziate dall'analisi delle corrispondenze, si è effettuato un *ranking* semantico utilizzando l'indice *TFIDF*.

Questa analisi ci ha fornito alcune conferme decisive sui fatti emersi nei capitoli precedenti. Il Corriere si è rivelato essere il quotidiano di opinione che ha abbracciato la riforma non solamente nei suoi aspetti più di immagine, ma anche secondo gli aspetti *core* della riforma.

Quel che più preoccupa è quel segnale di "dimenticanza" della riforma. La riforma viene trattata alla stregua di una notizia qualunque, per quanto lunga e di impatto, e non sembra che il rinnovamento strutturale e strategico attechisca sulla stampa. Certamente le università nel tempo si sono attrezzate, si sono dotate di uffici stampa per comunicare al meglio, anche la concorrenza e la flessibilità, nuove frontiere da esplorare, hanno facilitato questo processo di pubblicizzazione; ma passando oltre quella che è stata un'evoluzione graduale e che andava a modificare soprattutto gli organi più periferici degli atenei, l'immagine dell'Università come casa della cultura, come portatrice di nuova conoscenza non è cambiata.

L'Università non è riuscita a ritagliarsi uno spazio per dare risalto a quello che è stato un cambio di direzione di grande portata, non ha trasmesso alla stampa, se non in modo breve e temporaneo, i cambiamenti didattici e di obiettivi che l'hanno investita. Una laurea rimane sempre una laurea, anche se passa dall'essere triennale all'essere semplicemente breve; la ricerca

rimane sempre tale, e come tale sottofinanziata, nonostante i tentativi di cambiamento strutturali che hanno coinvolto CNR ed il mondo accademico.

Della riforma sembra rimanere solamente un'eco, che si spegne lentamente affossata da altre problematiche di natura finanziaria che comunque non tirano in ballo i cambiamenti avvenuti pochi mesi prima. Non si tenta di dare una spiegazione alla carenza dei fondi, non si cerca di imputare alla riforma alcun tipo di influenza di lungo periodo sul sistema universitario, non viene invocato un rinnovamento e non viene nemmeno vituperato. Insomma la riforma universitaria scompare assieme alla laurea triennale dal lessico utilizzato, e non mette radici nel cuore dei lettori e dei giornalisti.

Cuore che ha sempre posto per mancanza di fondi, occupazioni e per ricercatori in fuga.

Appendice A

Analisi del contenuto: cenni di base

La trattazione estesa dell'argomento, dalla quale è tratta anche quest'appendice, è contenuta in Tuzzi, 2003.

A.1 Corpus, vocabolario e unità del discorso

Il corpus è in generale l'insieme dei testi che vengono presi in esame. Ogni testo è formato da una serie di unità di base separate da spazi che non sono ulteriormente scomponibili. Queste unità possono prendere vari nomi: *parole, forme grafiche, word token*.

Queste unità di base possono essere accorpate in modo da andare a formare il corpus: sono raggruppate in *testi*, che a loro volta possono contenere più *subtesti*. Nel nostro caso un testo assume la forma di un articolo, mentre un subtesto può diventare uno fra occhiello, titolo e sottotitolo.

Il corpus viene misurato nella sua lunghezza da N , che rappresenta il numero di forme grafiche che lo costituiscono (*word token*). Il vocabolario, indicato come V , si ottiene a partire dal corpus selezionando solamente (una sola volta) tutte le forme grafiche distinte. Le forme grafiche distinte prendono nome di *word type*.

Ogni *word type* è legato al corpus dal numero di volte che è ripetuto in

esso: il vocabolario perciò può essere descritto come la lista dei *word type* accompagnanti dalla relativa frequenza (indicata con f_i) con cui compaiono nel corpus.

Ordinando il vocabolario secondo la frequenza delle *word type* possiamo ottenere una serie di informazioni sul corpus. Il software Taltac raggruppa le *word type* a seconda del loro *rango*: una classe di rango è composta da *word type* che condividono la frequenza. Il rango perciò parte dal valore 1 (per la *word type* a frequenza massima) ed arriva a catalogare le *word type* a frequenza 1 (*hapax*) alle quali viene assegnato il valore relativo al totale di frequenze distinte del vocabolario.

In base alla frequenza (e perciò anche in base al rango) vengono costruite le *fasce di frequenza*: la fascia *alta* raduna le *word type* dalla frequenza massima a quella frequenza che si ripete per due (o più) *word type*. La fascia *media* raggruppa le *word type* dalla fascia alta fino a quel punto in cui le frequenze si saturano, ovvero, partendo dal basso, dagli *hapax* a salire ogni possibile classe di frequenza conta almeno un *word type*.

A.2 Variabili e specificità

Il calcolo delle specificità è affidato ad un modello probabilistico. Se chiamiamo X la variabile casuale che rappresenta la frequenza della parola (o del segmento) w_p nel testo g_j , allora tale variabile aleatoria segue una distribuzione ipergeometrica di parametri N, n_i, n_j .

La tabella seguente permette di esemplificare la forma tipo di un vocabolario organizzato in una matrice `formeXtesti`.

	g_1	g_1	...	g_j	...	g_p	Totale
w_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$
w_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
w_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
$w_{V(N)}$	$n_{V(N)1}$	$n_{V(N)2}$...	$n_{V(N)j}$...	$n_{V(N)p}$	$n_{V(N).}$
Totale	$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.p}$	N

$$X \sim Hyp(N, n_{i.}, n_{.j})$$

$$\Pr\{X = x | N, n_{i.}, n_{.j}\} = \frac{\binom{n_{i.}}{x} \binom{N-n_{i.}}{n_{.j}-x}}{\binom{N}{n_{.j}}}$$

Assumere tale distribuzione ci consente di calcolare la probabilità che la variabile X si discosti dalla sua media. Le forme grafiche la cui frequenza si accosta alla media della distribuzione vengono definite *banali*. Misurando invece sulle code della distribuzione la probabilità che X sia minore o uguale ad una soglia di frequenza k definita a priori in base ad una probabilità prefissata, consente di determinare se la forma grafica sia sopra o sotto-utilizzata per quel particolare testo.

A.3 L'indice IS

Acronimo di *Indice di Significatività*, l'indice IS rappresenta il grado di assorbimento del segmento rispetto alle parole che lo costituiscono¹. Come già visto tale indice viene calcolato utilizzando la formula $IS = \left[\sum_{i=1}^L \frac{f_{sgm}}{f_{fg_i}} \right] \times P$ dove L è il numero di forme grafiche del segmento, f_{sgm} è la frequenza del segmento del quale calcolare l'indice, f_{fg_i} è la frequenza della forma grafica i che fa parte del segmento, e P è il numero di parole piene che costituiscono il segmento. Più alto è il valore dell'indice, tanto più le forme grafiche che

¹Guida TaLTac²

fanno parte del segmento compaiono nel corpus inserite nel segmento. Quando l'indice IS vale L^2 , che è il suo valore teorico massimo, significa che tutte le parole che formano il segmento vengono utilizzate solamente all'interno del segmento stesso. Questa costruzione dell'indice IS mette in evidenza i segmenti più lunghi. Per evidenziare invece i segmenti più corti (Giuliano, 2008), possiamo utilizzare l'*indice IS relativo*:

$$IS_{rel} = \left[\sum_{i=1}^L \frac{f_{segm}}{f_{fg_i}} \right] \times \frac{P}{L^2}$$

che varia tra 0 e 1, ed è ottenuto semplicemente dividendo l'indice IS per il suo valore teorico massimo.

I primi segmenti ordinati secondo il relativo indice *IS* teoricamente compaiono nel testo sempre come unità di parole coese: cioè le parole che formano il segmento non compaiono mai da sole. Nella tabella di seguito sono presentati alcuni dei segmenti che compaiono nel testo almeno 10 volte ordinati secondo il relativo indice *IS*. Non tutti questi segmenti individuati nel corpus sono poi stati lessicalizzati: soprattutto i nomi propri sono stati mantenuti separati al fine di non separare tutte i diversi acronimi utilizzati per riferirsi ad essi. Basti pensare ai segmenti *Letizia Moratti*, *ministro Moratti*, *ministro Letizia Moratti*.

Segmento	Occorrenze totali	Numero di forme grafiche	Indice IS	Indice IS relativo
Rinaldo Bertolino	15	2	2,94	0,73
Ortensio Zecchino	17	2	2,83	0,71
Letizia Moratti	129	2	2,70	0,68
dottori commercialisti	70	2	2,65	0,66
Giampaolo Rossi	20	2	2,48	0,62
Angelo panebianco	11	2	2,45	0,61
assistenti sociali	14	2	2,38	0,59
offerta formativa	49	2	2,29	0,57
Piero tosi	16	2	2,25	0,56
valore legale	34	2	2,20	0,55
periodo transitorio	13	2	2,18	0,55
requisiti minimi	28	2	2,14	0,54

A.4 Il ciclo di vita delle parole

La divisione in semestri e non in unità di tempo più piccole è stata necessaria al fine di ottenere dei subcorpora sufficientemente ampi da poter essere analizzati con tecniche statistiche. In alcuni mesi dell'anno si registrano infatti dei “vuoti” di informazione. Come noto [Tuzzi, 2003] infatti il numero di occorrenze di una parola di un corpus non cresce in maniera direttamente proporzionale all'ampiezza del subcorpus. Se per l'analisi delle corrispondenze la soglia per ritenere una parola inseribile nelle analisi può essere piuttosto bassa, nella stima di un modello lineare si sono scelte parole con frequenza superiore o uguale a 40.

Le frequenze vengono normalizzate rispetto al semestre, in modo da eliminare eventuali differenze tra i subcorpora, sia in media che in varianza, dove quest'ultima riveste una particolare importanza, in quanto, come già affermato, all'aumento delle dimensioni di un testo non corrisponde un aumento proporzionale delle occorrenze delle singole parole.

Il rapporto tra indice IT e IT_0

Come illustrato nel capitolo 4.1 l'indice IT viene calcolato a partire dal fattore IT_0 , come da equazione 4.1. È importante notare come l'indice IT_0 contempli nel suo campo di esistenza unicamente due valori: il -1 e lo 0, in quanto generato dal rapporto degli stessi elementi, posti senza vincoli al numeratore e al valore assoluto al denominatore. L'unico scopo dell'indice IT_0 è di identificare quali parole siano in trend crescente/decescente, e quali abbiano un trend "misto". Tale scopo viene raggiunto contando il numero di volte che le occorrenze normalizzate oltrepassano la media relativa alle occorrenze della parola in questione.

Pertanto l'intensità, o meglio il valore dell'indice IT preso nei razionali tra -1 e 1, è dato unicamente dal termine $\frac{(Occ_1 - Occ_n)}{(Occ_1 + Occ_n)}$: è per tale motivo che l'indice IT non si presta ad individuare situazioni di irregolarità nella serie storica delle occorrenze, e rappresenta solamente una buona statistica descrittiva nell'individuare trend monotoni. Dipende infatti solamente dal valore iniziale e quello finale.

Il modello lineare stimato

Le stime ottenute con il metodo dei minimi quadrati non sono da ritenersi altro che statistiche riassuntive di un possibile trend seguito dalla serie storica delle occorrenze normalizzate. La stima del modello lineare viene utilizzata principalmente per colmare eventuali "punti bui" non esplorabili tramite l'utilizzo dell'indice IT .

Molte parole che vengono classificate come rapidamente obsolescenti dall'indice IT presentano un trend lineare negativo e parametro quadratico positivo: ovviamente l'adattamento migliore per questa combinazione è far adattare localmente una parabola alle osservazioni.

A.5 L'indice TFIDF

La sigla TFIDF è il risultato dell'unione di due diversi acronimi:

- TF: che sta per *term frequency* e rappresenta la frequenza della generica forma grafica k_j all'interno del documento d_i ;
- IDF: che sta per *inverse document frequency* ed è dato dall'inverso della frazione di documenti che contengono k_j rispetto al totale di documenti che compongono il corpus.

La formula di calcolo dell'indice rispetto alla forma grafica (o segmento) i perciò diventa: $TFIDF_i = f_i \times \lg \frac{N}{n}$ dove f_i è la frequenza totale della parola i , N è il numero totale di documenti nel corpus (articoli nel nostro caso) e n è il numero di articoli che contengono la parola i .

La parte $\lg \frac{N}{n}$ vale 0 quando la parola per la quale si sta calcolando l'indice è utilizzata in tutti gli articoli del corpus, mentre ha valore massimo ($\lg N$, che nel nostro corpus equivarrebbe a circa 6) se la parola è utilizzata solamente in un documento. Il concetto di dispersione alla base dell'indice $TFIDF$ ha interpretazione opposta a quello comunemente utilizzato nella fase di analisi delle specificità. Nel calcolo dell'indice $TFIDF$, tanto più la parola è concentrata in un articolo, tanto più questo articolo tratta dell'argomento di cui fa parte la parola. Insomma, se per l'analisi delle specificità la concentrazione di una parola in un solo articolo è indice di pessima caratterizzazione del subtesto da parte di quella parola, nel caso dell'indice $TFIDF$ la concentrazione della forma grafica aiuta nell'estrazione di informazione.

L'indice così calcolato può essere utilizzato sia per pesare le frequenze delle parole in base alla loro dispersione nel testo, sia per calcolare per ogni sub-testo un indice che consenta di ordinare i sub-testi in base all'utilizzo di determinate unità di discorso. La possibilità di calcolare l'indice $TFIDF$ su parole appartenenti ad aree semanticamente delimitate consente poi di disporre di (nel nostro caso) articoli all'interno dei quali tutte le parole (o quasi tutte) appartenenti all'area semantica scelta saranno indicizzate grazie al $TFIDF$. Sommando gli indici ottenuti si otterrà il valore relativo all'articolo. TalTac, per ovviare problemi di stima dovuti all'eccessiva lunghezza di alcuni sub-testi, normalizza l'indice-somma così ottenuto in base alla lunghezza dell'articolo.

Diritti e Doveri	Fondi	Ricerca	Tecnico
corporative	approfondimento	comunità scientifica	applicazione della riforma
diritti	aumento delle tasse	coscienza	attività formative
docenti	battaglia	cultura	attuazione della riforma
efficienza	economia	culturale	autonomia
formazione	economica	culturali	autonomia degli atenei
formazione di base	economico	dottorato	autonomia didattica
formazione professionale	euro	dottorato di ricerca	autonomia universitaria
formazione superiore	finanziamenti	enti di ricerca	avvio della riforma
formazione universitaria	finanziamento	formazione	biennio specialistico
imprese	finanziamento ordinario	formazione di base	classi di laurea
innovative	finanziare	formazione superiore	concorrenza
innovativi	Finanziaria	formazione universitaria	contro la riforma
innovativo	finanziaria	fuga dei cervelli	corsi di laurea
innovazione	finanziario	intellettuale	corsi di laurea specialistica
innovazioni	fondi	intelletuali	corsi di laurea triennali
investimenti	fondo di finanziamento	preparazione	corsi triennali
investimento	investimenti	ricerca	corso di laurea
qualità	investimento	ricerca di base	crediti
reclutamento	legge finanziaria	ricerca scientifica	crediti formativi
ricerca	lire	ricercatore	didattica
ricerca di base	miliardi	ricercatori	didattica universitaria
ricerca scientifica	miliardi di lire	ricerche	didattiche
ricercatore	milione	sapere	didattici
ricercatori	milioni	scientifica	didattico
società	milioni di euro	scientifiche	dopo la riforma
studenti	ministro dell'economia	scientifici	formativi
trasparenza	risorsa	scientifico	formazione
valutazione	risorse	scienza	formazione di base
	risorse finanziarie	scienziati	formazione superiore
	soldi		formazione universitaria
	stanziamenti		laurea breve
	stanziamento		laurea di primo livello
	tagli		laurea specialistica
	tagliare		laurea triennale
	taglio		laurearsi
	tassa		laureati
	tassa sul fumo		laureati specialisti
	tasse		laureati triennali
	tasso		lauree specialistiche
			lauree triennali
			numero di laureati
			nuovi corsi di laurea
			nuovo ordinamento
			offerta didattica
			offerta formativa
			ordinamenti
			ordinamenti didattici
			ordinamenti professionali
			percorsi formativi
			percorso formativo
			piani di studio
			post-laurea
			regolamenti didattici
			riforma Berlinguer
			riforma degli ordinamenti didattici
			riforma dei cicli
			riforma dei cicli scolastici
			riforma del 3+2
			riforma dell'autonomia
			riforma dell'autonomia didattica
			riforma dell'università
			riforma della scuola
			riforma Moratti
			riforma universitaria
			scienze della formazione
			sistema dei crediti
			specialistica
			specialistiche
			specialistici
			triennale
			triennali
			triennio
			vecchia laurea
			vecchio ordinamento

Tabella A.1: Liste di parole utilizzate per il calcolo dell'indice TFIDF

Appendice B

Analisi delle corrispondenze, cenni di base

B.1 Le tabelle di contingenza

L'inerzia

La qualità delle soluzioni fattoriali¹ ottenute viene valutata tramite l'*inerzia* spiegata dai fattori utilizzati. L'inerzia dipende direttamente dalla statistica χ^2 , che, in termini piuttosto semplici, viene calcolata a partire da una tabella di contingenza come segue:

$$\chi^2 = \sum \frac{(\text{valori_osservati} - \text{valori_attesi})^2}{\text{valori_attesi}}$$

La quantità $\frac{\chi^2}{n}$, dove n è la somma totale delle frequenze della tabella di partenza prende il nome di *inerzia*.

L'interpretazione geometrica dell'inerzia considera questa quantità come la media pesata delle distanze χ^2 tra i profili di riga (perciò i profili per ogni elemento del vocabolario) e il loro profilo medio. Perciò, tanto più “dispersi” sono i profili di riga, tanto maggiore sarà l'inerzia totale della tabella.

L'inerzia viene utilizzata anche per valutare la qualità della soluzione

¹Per i dettagli consultare M.Greenacre, *Correspondence analysis in practice*, volume da quale è principalmente tratto questo paragrafo

fattoriale prescelta: operativamente ogni fattore riassume in se' una parte dell'inerzia, e pertanto l'inerzia colta dai fattori può essere interpretata esattamente come la percentuale di varianza spiegata dei modelli di regressione.

B.2 Il calcolo della soluzione fattoriale

Variabili attive e passive

L'analisi delle corrispondenze ci consente di proiettare le variabili in nostro possesso su di uno spazio ottenuto tramite la generazione dei fattori. Tale proiezione può allora riguardare non solo le variabili utilizzate per generare i fattori, ma anche delle variabili dette *supplementari*, definite sempre sullo spazio delle righe della tabella.

L'interpretazione dei fattori

Ricordiamo brevemente come la rappresentazione sul piano fattoriale sia frutto della visualizzazione congiunta di punti che giacciono su spazi differenti: pertanto la vicinanza dei punti può essere interpretata in termini di correlazione solamente se si paragonano i profili di riga fra di loro, oppure i profili colonna fra di loro, mentre l'interpretazione congiunta delle posizioni dei punti-riga e dei punti-colonna può spesso portare a forzature.

Bibliografia

- [1] Conferenza dei rettori delle università italiane, *Sullo stato delle università italiane*, 2001
- [2] Abbott A.& Schiermeier Q., *Italy's bright spots*, Nature 438, 1046-1047, Dicembre 2005
- [3] Abbott A., *Special Report Saving: Italian science*, Nature 440, 264-265, Marzo 2006
- [4] Baldrini, Morcellini (a cura di), *Un'idea di università. Comunicazione universitaria e logica dei media*, 2005
- [5] Harald Baayen R., *Word frequency distributions*, Kluwer Academic Publishers, 2001
- [6] Bernardi L., *Prospettive per l'alta formazione*, in Marini D. (a cura di), *Nord Est 2002. Rapporto sulla società e l'economia*, Fondazione Nord Est, 2002
- [7] Bernardi L., *Percorsi di ricerca sociale*, Carocci, 2005
- [8] Bolasco S., *Analisi multidimensionale dei dati*, Carocci Editore, 1999
- [9] Bolasco S., Bisceglia B., Baiocchi F., *Estrazione automatica d'informazione dai testi*, in Mondo digitale n°1, marzo 2004
- [10] Bolasco S., *Statistica testuale e text mining: alcuni paradigmi applicativi*, da Quaderni di statistica Vol. 7, 2005

- [11] Bolasco S., Canzonetti A., *Sguardi sull'evoluzione dell'italiano standard degli anni novanta, grazie al Text Mining e alla categorizzazione automatica del lessico del quotidiano "La Repubblica"*, 2003
- [12] Cammelli M., Merloni F., *Università e sistema della ricerca Proposte per cambiare*, Il mulino, 2006
- [13] Capecchi V., Livolsi M., *La stampa quotidiana in Italia*, 1971, Milano, Bompiani
- [14] Castronovo V., Tranfaglia N., (a cura di), *La stampa italiana nell'era della TV*, Bari, Laterza, 1994
- [15] Consorzio interuniversitario Almalaurea, *L'università in transizione: laureati vecchi e nuovi alla luce della riforma*, Il mulino, 2006
- [16] Dardano M., *Il linguaggio del quotidiano*, in D'Amico N., Della Seta L. (a cura di), *Il quotidiano in classe*, Bologna, Zanichelli 1980
- [17] Derrida J. & Rovatti P.A., *L'università senza condizione*, Raffaello Cortina Editore, Milano 2002
- [18] Fabbris L., *Statistica multivariata - Analisi esplorativa dei dati*, McGraw-Hill, 1997
- [19] Ferraris M., *Una ikea di università*, Milano, Cortina, pp. 117, 2001
- [20] Froio F., *Le mani sull'università*, Editori Riuniti, Roma 1996
- [21] Findalh O., Hoijer B., *Media content and human comprehension*, in Rosengren K. (ed.) *Advances in content analysis*, Beverly Hills, Sage, 1981
- [22] Giannessi F., *La Riforma Universitaria: evoluzioni e effetti*, in Centro Studi CRUI (a cura di), *Approfondimenti*, novembre 2006
- [23] Capano G., *L'università in Italia*, Il Mulino 2000

- [24] Greenacre M., *Correspondence analysis in practice*, Chapman&Hall/CRC, 1993
- [25] Greenacre M., Blasius J., *Correspondence analysis in the social sciences: recent developments and applications*, Academic press, 1994
- [26] Livolsi, M., (a cura di), *La fabbrica delle notizie*, Milano, Angeli, 1984
- [27] Losito G., *L'analisi del contenuto nella ricerca sociale*, Franco Angeli, Milano, 1993
- [28] Marvulli R., *Analisi statistica di contenuto sul messaggio elettorale*, G.Giappichelli Editore, Torino, 1989
- [29] Marvulli R., *Distribuzione degli spazi sui quotidiani e tipologia del pubblico*, Giappichelli Editore, Torino, 1977
- [30] Murialdi, P., *Breve storia del "Corriere della Sera" e de "La Repubblica"*, in Livolsi, M., (a cura di)
- [31] CNVSU, *Settimo Rapporto sullo Stato del Sistema Universitario*, Ministero dell'Università e Ricerca, 2006
- [32] Nascimbeni G., *Il linguaggio del giornalismo*, Mursia, 1992
- [33] Chiaberge R., *Cervelli d'Italia*, Baldini e Castoldi, Milano 1996
- [34] Roidi V., *La fabbrica delle notizie*, 2001
- [35] Semeraro R., *Valutazione e qualità della didattica universitaria*, Franco Angeli, 2006
- [36] Sorrentino C., *I percorsi della notizia. La stampa quotidiana italiana tra politica e mercato*, Baskerville, 1995
- [37] Tuzzi A.(a cura di), *Dall'intervista alla notizia*, Sapere, Padova, 2001
- [38] Tuzzi A., *L'analisi del contenuto*, Carocci editore, Roma, 2003

- [39] Tuzzi A., Pauli F., *Identifying specific textual units of documents taken from large corpora. Comparing methods*, JADT 2006
- [40] Tuzzi A., Cristoferi V., Cemin M., Castagna M., *Molto sono emozionato. Il linguaggio autistico nei testi prodotti con la Comunicazione Facilitata*, in *Applicazioni di analisi statistica dei dati testuali*, Aureli Cutillo E., Bolasco S. (a cura di), Casa editrice Università La Sapienza, 2004
- [41] Verbeke G., Molenberghs G., *Linear mixed models for longitudinal data*, Springer Series in Statistics. New-York: Springer-Verlag, 2000
- [42] Wolf M., *Gli effetti sociali dei media*, Bompiani, 1992
- [43] Wolf M., *Teorie delle comunicazioni di massa*, 1985
- [44] www.adsnotizie.it per i dati sulla tiratura dei quotidiani
- [45] <http://www.ilsole24ore.com>