

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN  
Economia e Statistica Agroalimentare

Ciclo XXVII

**Settore Concorsuale di afferenza:** 13/D2 Statistica Economica

**Settore Scientifico disciplinare:** SECS – S/03 Statistica Economica

MERCATO DELL'INFORMAZIONE, POSIZIONAMENTO  
DEI MEDIA E DISTORSIONI: UN'ANALISI ATTRAVERSO  
L'ALGORITMO WORDFISH

**Candidato:**  
**Dott. Matteo Masotti**

**Relatore**  
**Prof. Mario Mazzocchi**

**Coordinatore del Corso di Dottorato**  
**Prof.ssa Alessandra Luati**

**Esame finale anno 2015**

# Indice

---

Introduzione .....	3
Parte 1 - Aspetti teorici del mercato dell'informazione e dell'analisi automatizzata dei testi	
1. Cattive notizie e negativity bias: come l'avversione al rischio influenza la percezione degli eventi negativi .....	8
2. Un modello di mercato dell'informazione e il ruolo della domanda.....	19
2.1 Il mercato delle notizie e il ruolo della concorrenza .....	19
2.2 Il modello di Mullainathan e Shleifer.....	19
2.3 Impostazione del modello .....	20
2.3.1 Primo caso: lettori razionali avversi allo slanting .....	24
2.3.2 Secondo caso: lettori non perfettamente razionali.....	25
3. Elementi di analisi automatica dei testi .....	33
3.1 L'analisi automatica dei testi: le metodologie principali .....	33
3.2 Un approccio scalare alla categorizzazione dei testi: l'algoritmo Wordfish .....	36
Parte 2 - La stampa italiana e il mercato dell'informazione: due casi recenti	
4. Il modello Wordfish in pratica .....	42
5. C'è sempre polarizzazione? Due esempi dalla stampa italiana.....	48
5.1 I flussi migratori provenienti dalla sponda sud del Mediterraneo e le reazioni di due importanti quotidiani italiani .....	48
5.1.1 Il posizionamento dei quotidiani .....	50
5.1.2 Il lessico utilizzato.....	54
5.2 Un esempio di informazione omogenea: l'epidemia di influenza suina raccontata da tre quotidiani italiani.....	56
5.2.1 Il posizionamento dei tre quotidiani.....	58
5.2.2 Il lessico utilizzato.....	60
Conclusioni .....	63
Riferimenti Bibliografici .....	65

*“Per godere degli inestimabili benefici che la libertà della stampa assicura,  
è necessario sottomettere gli inevitabili mali che provoca.”*

*A. de Toqueville, La democrazia in America, 1835*

## Introduzione

---

Una stampa libera e plurale è un elemento fondante di ogni sistema democratico ed è fondamentale per la creazione di un'opinione pubblica informata e in grado di esercitare controllo e pressione sulle classi dirigenti, siano esse appartenenti al contesto politico-sociale o a quello economico.

La nascita di pubblicazioni simili alla stampa contemporanea risale al periodo successivo all'invenzione della stampa a caratteri mobili da parte del tedesco Johann Gutenberg, avvenuta nel 1450. La diffusione di questa tecnica, che permette la produzione in serie di stampe uguali tra loro, ha dato il via alla diffusione di fogli stampati, all'inizio in modo discontinuo, poi pubblicati a intervalli sempre più regolari, contenenti informazioni di cronaca, economia, politica, e altri eventi locali.

Per giungere alla creazione dei primi giornali pubblicati con regolarità è necessario attendere il XVIII secolo quando a Strasburgo venne fondato il periodico *Relation aller Fürnemmen und gedenckwürdigen Historien* ("Resoconto di tutte le notizie importanti e memorabili") che veniva stampato a cadenza settimanale. Verso la metà del 1600 iniziano a comparire i primi giornali stampati a cadenza quotidiana: il primo quotidiano europeo, sempre stampato in lingua tedesca, è la *Einkommende Zeitungen* fondata a Lipsia nel 1650, mentre per assistere alla nascita del primo quotidiano italiano bisognerà attendere fino alla fine del 1800 e più precisamente il 1866, quando *La gazzetta di Mantova*, fondata nel 1664 e pubblicata mensilmente, diventa un quotidiano a tutti gli effetti.

Dal momento della loro creazione i quotidiani si sono imposti come un'importantissima fonte di informazione per l'opinione pubblica e si sono dimostrati un importantissimo strumento di diffusione delle idee politiche, economiche e sociali più diverse. Per questo motivo, nel corso del XX secolo, con la nascita della società di massa causata dall'avvento della rivoluzione industriale, i periodici e i quotidiani conobbero un'estensione su scala globale e contribuirono al diffondersi presso l'opinione pubblica delle ideologie che hanno caratterizzato il secolo scorso.

La seconda metà del Novecento ha conosciuto innovazioni tecnologiche che hanno portato grandi cambiamenti nel ruolo della carta stampata come veicolo di trasmissione di informazioni e di opinioni. La prima innovazione tecnologica che ha riguardato il mondo

dell'informazione è stata la diffusione della televisione, che ha contribuito a rendere più frequente e capillare la diffusione delle notizie e delle opinioni, lasciando ai giornali, più "lenti" rispetto alla TV, un ruolo più incentrato sull'approfondimento e sull'elaborazione di opinioni, piuttosto che sulla semplice cronaca.

La velocità di trasmissione delle informazioni è poi aumentata esponenzialmente con l'avvento e la diffusione di Internet, che ha, da un lato ridotto il costo in termini monetari di acquisizione delle notizie, portandolo praticamente a zero, e dall'altro ha dato la possibilità a chiunque disponesse di un dispositivo collegato alla rete di produrre testi da condividere con gli altri utenti sparsi per il mondo. In questo contesto i giornali, così come tutti i mezzi di informazione "tradizionali" si sono dovuti evolvere, per far fronte alle nuove forme che il mercato dell'informazione ha assunto negli ultimi anni. Sono nate così le versioni on-line dei giornali, che mettono a disposizione gratuitamente sia i loro archivi sia notizie e opinioni aggiornate con grande frequenza.

La rivoluzione tecnologica degli ultimi venti anni, pertanto, ha reso disponibile al pubblico e ai ricercatori una quantità enorme di dati che, se analizzati correttamente, possono fornire grandi quantità di informazioni relative alle idee e ai contenuti proposti dai diversi autori nel corso del tempo.

Produrre notizie, tuttavia, ha un costo. Anche se grazie a internet e alla rivoluzione digitale i costi materiali dei periodici (stampa, distribuzione, resi eccetera) si sono notevolmente ridotti, gli editori devono comunque far fronte a spese di tipo diverso, che vanno dalla gestione dei siti web al pagamento dei giornalisti che producono materialmente i testi da pubblicare. Per questi motivi, pertanto, la produzione di informazioni si inserisce in un contesto di mercato, sottoposto alle logiche della domanda e dell'offerta.

In questo lavoro verrà analizzato in particolar modo il ruolo della domanda nel mercato delle notizie. La base di partenza dell'analisi è costituita dal lavoro presentato nell'articolo del 2005 di Mullainathan e Schleifer dal titolo *The market for news*. In questo paper gli autori mettono in discussione la teoria classica del mercato dell'informazione, che descrive il consumatore di notizie come perfettamente razionale e interessato solamente ai fatti: ciò farebbe sì che i produttori di notizie siano spinti sempre più a riportare gli eventi senza distorsioni e omissioni. Mullainathan e Schleifer si concentrano invece sul ruolo della domanda e sulla non perfetta razionalità dei lettori, che sono per loro natura inclini a

acquisire notizie in linea con le proprie opinioni. Secondo questo modello, la differenza di opinioni dei consumatori di notizie spinge gli editori a regolare la propria linea editoriale, per venire incontro alla domanda di mercato: un'audience omogenea spingerà i giornali presenti sul mercato a rendere simili le loro linee editoriali (fenomeno definito spin) e a competere solo sul lato del prezzo. Dall'altro lato, un pubblico con opinioni diversificate porterà le testate a diversificare i contenuti proposti, spingendosi verso posizioni estreme: si verificherà quindi una polarizzazione dell'offerta editoriale, che potrà rivelarsi ancora più radicale rispetto alle differenze presenti tra le opinioni dei lettori.

Lo scopo di questo lavoro è di capire se questo modello di mercato dell'informazione è applicabile al contesto italiano. A tal proposito si è analizzato il comportamento di alcuni importanti quotidiani nazionali, che dichiarano essere collocati su linee editoriali molto distanti tra loro, in occasione di due eventi che hanno profondamente interessato l'opinione pubblica italiana. Il primo caso di studio riguarda il fenomeno dei flussi migratori provenienti dalla sponda sud del Mediterraneo nel mese di ottobre 2013, mese in cui un grave naufragio al largo delle coste siciliane è costato la vita a centinaia di migranti. Questa tragedia ha acceso un forte dibattito politico sul tema della gestione dei flussi migratori e sulle normative allora vigenti che lo regolavano. Il secondo evento analizzato è l'epidemia di influenza H1N1 (definita anche come "influenza suina"), che nel 2009 si è diffusa su scala globale, dando vita anche a vasti fenomeni di panico.

L'analisi delle differenze e delle similitudini presenti tra i quotidiani considerati avverrà utilizzando l'algoritmo Wordfish (Slapkin e Proksch, 2008). Questo algoritmo, elaborato per l'analisi di testi non strutturati, produce una classificazione dei testi basandosi sulla frequenza con cui le parole vi compaiono. L'algoritmo, tuttavia, non dà indicazioni sulle variabili latenti alla base della classificazione che produce: queste variabili devono essere individuate dal ricercatore, in base alle parole che maggiormente caratterizzano i testi analizzati.

Questo lavoro si articola in quattro capitoli. Il primo contiene un'analisi della letteratura riguardante il negativity bias, una distorsione cognitiva molto frequente e che condiziona fortemente la percezione della realtà e, di conseguenza la domanda di notizie. A causa di questo bias cognitivo viene posta più attenzione verso i fatti percepiti come negativi e pericolosi (non serve che il pericolo sia reale o immaginario). Per questo motivo, le notizie

riguardanti questo genere di avvenimenti hanno un forte impatto sull'opinione pubblica, che tende a considerarli in modo non razionale.

Nel secondo capitolo verrà presentato il modello di mercato dell'informazione proposto da Mullainathan e Schleifer (2005), che propone una panoramica dell'impatto delle diverse tipologie di domanda di notizie sul mercato delle news.

Nel terzo capitolo, dopo un breve excursus dei metodi di analisi automatizzata dei testi maggiormente presenti in letteratura, verrà presentato più in dettaglio l'algoritmo Wordfish, utilizzato per analizzare il posizionamento dei quotidiani italiani considerati nel descrivere le reazioni del pubblico e della politica all'emergenza legata ai flussi migratori provenienti dal Nordafrica e l'epidemia di influenza H1N1 nel corso del 2009.

Il quarto capitolo contiene un'analisi più approfondita del significato e del ruolo dei parametri stimati dall'algoritmo Wordfish e un'applicazione pratica dell'algoritmo stesso, partendo dai dati messi a disposizione da Slapin e Proksch.

Infine, nel quinto capitolo verranno presentati i due casi studio sopra citati. Per ciò che riguarda l'analisi dei fenomeni migratori, verranno considerati gli articoli sul tema pubblicati nel mese di ottobre 2013 da *La Repubblica* e *Liberò* e scaricati dagli archivi on-line delle testate, mentre, per ciò che riguarda la copertura mediatica data all'epidemia di influenza suina, verranno analizzati gli articoli sul tema pubblicati da aprile a dicembre 2009 da *La Repubblica*, *Corriere della Sera* e *Liberò*, e, come nel caso precedente, presenti negli archivi on-line dei tre quotidiani.

# Parte 1 - Aspetti teorici del mercato dell'informazione e dell'analisi automatizzata dei testi

---



## 1. Cattive notizie e negativity bias: come l'avversione al rischio influenza la percezione degli eventi negativi

---

*Una goccia di catrame rovina una botte di miele, ma un goccio di miele non ha nessun effetto in una botte di catrame – proverbio russo*

Con il termine “negativity bias” si indica la propensione degli individui a dare maggior peso agli eventi negativi, anche in presenza di eventi positivi di pari entità ed intensità. Rozin e Royzman definiscono il negativity bias come “il principio [...] per cui] nella maggior parte delle situazioni gli eventi negativi sono più salienti, potenti, dominanti in caso di combinazioni di eventi e generalmente più efficaci rispetto agli eventi positivi” (Rozin e Royzman, 2001).

I concetti di *bene*, *male* e *forza* sono basilari nella teoria psicologica e sono universali, in quanto presenti con lo stesso significato in tutte le culture (Osgood e Tzeng, 1990). Con *bene* si definisce la presenza di esperienze, sensazioni, situazioni desiderabili e piacevoli; il concetto di *male* è invece l'opposto, cioè la presenza di situazioni spiacevoli e indesiderabili. La *forza*, infine, si riferisce all'impatto degli effetti di determinate esperienze: più un evento è forte, più sono consistenti e durevoli i suoi effetti (Baumeister et al. 2001)

Gli autori individuano quattro caratteristiche principali del negativity bias: la maggior forza degli eventi negativi (*negative potency*); la maggiore pendenza della funzione di utilità in presenza di eventi negativi (*greater steepness of negative gradients*); la predominanza degli eventi negativi (*negative dominance*); la maggiore differenziazione degli eventi negativi (*greater negative differentiation*).

Il principio della *negative potency* afferma che, in presenza di eventi positivi e negativi di uguale intensità, gli eventi negativi sono, soggettivamente, più influenti, più duraturi e più potenti di quelli positivi. Questo concetto è stato ripreso ed applicato al contesto del comportamento del consumatore da Daniel Kahneman e Amod Tversky nella *prospect theory* da loro elaborata (Kahneman, Tversky, 1979) attraverso il concetto di avversione alle perdite, o *loss aversion*.

La presenza di questo principio, inoltre, è molto forte in ambito politico, dove popolarità e intenzioni di voto sono molto influenzati dagli eventi negativi. Ne sono degli esempi lo scarso effetto dei buoni risultati economici sulle intenzioni di voto, come analizzato da

Bloom e Price nel caso delle elezioni americane (Bloom e Price, 1975) e le reazioni ad eventi catastrofici. La gestione dell'emergenza seguita all'uragano Sandy, che ha colpito New York alla fine del 2012, ha portato ad un aumento dei consensi per Barack Obama, che ha ricevuto da quegli eventi una spinta importante per la sua rielezione<sup>1</sup>, mentre le elezioni di medio termine del 2014, seppur avvenute in un periodo di ripresa economica, hanno punito il governo in carica, dimostrando così l'incapacità degli aspetti positivi di cancellare gli effetti delle situazioni negative presenti (nel caso degli Stati Uniti, la situazione internazionale e le potenziali minacce alla sicurezza nazionale). Allo stesso modo, la reazione del governo francese agli attentati terroristici del gennaio 2015 ha fatto sì che il consenso nei confronti del Presidente della Repubblica Hollande e del Primo Ministro Valls risalissero consistentemente, dopo un lungo periodo di declino<sup>2</sup>.

Rientra in questo aspetto del negativity bias anche la teoria del livello di adattamento, proposta negli anni '60 del '900 dallo psicologo americano Herry Helson (Helson, 1964). Secondo Helson, gli impatti che i cambiamenti hanno sulla percezione degli individui sono temporanei, anche se hanno più effetto delle situazioni stabili. Il perdurare nel tempo degli effetti dei cambiamenti, tuttavia, è decrescente e dipende dal tipo di conseguenze che questi cambiamenti hanno prodotto, tanto che sul lungo periodo gli effetti sperimentati vengono percepiti come parte integrante del nuovo contesto stabile che si è creato<sup>3</sup>. Nel tentativo di dimostrare questo meccanismo, che Brickman e Campbell (Brickman e Campbell, 1971) definiscono *hedonic treadmill*, Brickman et al. (Brickman et al., 1978) hanno dimostrato come gli eventi negativi vengano "assimilati" più lentamente ed abbiano quindi effetti più duraturi nel tempo. Gli autori sono giunti a queste conclusioni intervistando tre gruppi di individui: un gruppo di vincitori di un premio alla lotteria, un gruppo di persone che hanno subito un incidente invalidante e un terzo gruppo, di controllo, i cui appartenenti non hanno ancora avuto esperienze fuori dal comune.

---

1 Come riportato, ad esempio, in Pew Research Center, 2012

2 Si veda, ad esempio, Stefano Montefiori, *Dopo Charile Hebdo, Hollande riacquista punti – 21 per la precisione – perché ha dato l'impressione di non volere sfruttare la tragedia. Il presidente più impopolare della V Repubblica francese straccia ogni record, e batte persino Mitterrand, che con la Guerra del Golfo ne riprese solo 19*, Corriere della Sera 20 gennaio 2015

3 Questo meccanismo sta, tra le altre cose, alla base del *paradosso della felicità* o *paradosso di Easterlin* (Easterlin, 1974), secondo il quale il benessere degli individui non aumenta proporzionalmente al variare della ricchezza o dei beni posseduti, ma è fortemente influenzata, in senso negativo, dalle aspettative crescenti indotte dal maggior livello di benessere

Come ci si aspettava, i vincitori della lotteria intervistati non hanno dichiarato livelli di felicità molto diversi da quelli degli appartenenti agli altri due gruppi: nello specifico, solo il 23% degli intervistati ha dichiarato di aver cambiato radicalmente, e in meglio, il proprio stile di vita dopo la vincita e gli effetti negativi di quest'ultima, se presenti, sono sempre stati affiancati da effetti positivi. Quando intervistati a proposito del loro livello di felicità passato, presente e atteso per il futuro, hanno dato risposte simili a quelle degli appartenenti al gruppo di controllo.

La situazione è invece diversa per gli intervistati che hanno subito un incidente invalidante. Questi individui, quando sono stati interpellati a proposito del loro tasso di felicità passata e presente, hanno dichiarato livelli di felicità passata superiori a quelli del gruppo di controllo e tassi di felicità presente inferiori agli individui del gruppo di controllo; i livelli di felicità attesi per il futuro erano in linea con quelli indicati dalle altre persone intervistate. Gli autori definiscono questa situazione come effetto nostalgia, dove gli stati emotivi precedenti all'incidente vengono ricordati come più positivi di quanto lo fossero realmente stati. In definitiva, pertanto, si può affermare che gli eventi negativi hanno effetti più intensi, duraturi nel tempo e articolati rispetto agli avvenimenti positivi.

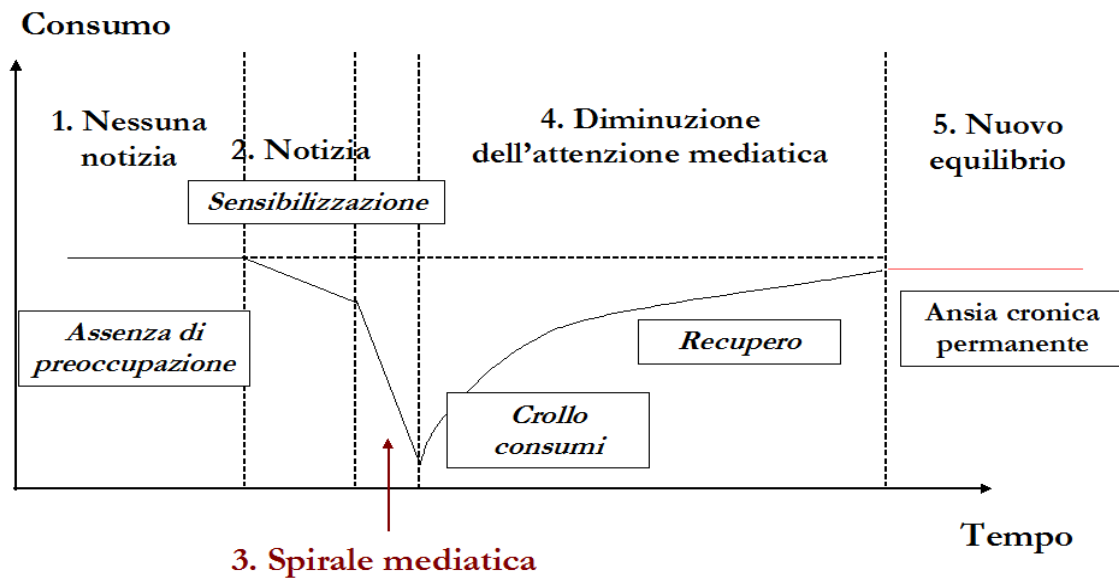
L'asimmetria dell'adattamento agli eventi negativi è riscontrabile anche nell'ambito delle crisi alimentari. Il perdurare di un livello di preoccupazione diffuso al seguito di una crisi alimentare, vera o mediatica che sia, è rappresentato dal modello della “spirale mediatica” elaborato da Beardsworth e Keil (Beardsworth e Keil, 1997). Questo modello descrive le cinque fasi in cui si articola nel tempo il comportamento dei consumatori in presenza di una crisi a cui viene data una consistente copertura mediatica, rappresentate dalla figura seguente, tratta da Mazzocchi, 2005:

Le cinque fasi del modello sono le seguenti:

1. equilibrio iniziale, dove l'opinione pubblica non è a conoscenza o non si interessa di possibili crisi;
2. iniziano a diffondersi le prime notizie sulla crisi e un pubblico più vasto comincia ad interessarsi al problema;
3. la preoccupazione dell'opinione pubblica aumenta, così come il numero delle notizie riguardanti la crisi presenti nei mezzi di comunicazione di massa ed all'interno del dibattito pubblico;

4. inizia la risposta del pubblico alla crisi ed i consumi diminuiscono, spesso in maniera sproporzionata rispetto al rischio oggettivo;
5. la preoccupazione dell'opinione pubblica rientra gradualmente e cala l'attenzione nei confronti della crisi, facendo sì che si instauri un nuovo equilibrio; persiste tuttavia un livello di ansia latente, che può riemergere in presenza di eventi simili a quelli appena trascorsi.

Figura 1.1: il modello di Beardsworth e Keil



Fonte: Mazzocchi, 2005.

Il secondo aspetto del negativity bias, ovvero la maggiore pendenza della funzione di utilità in presenza di eventi negativi, si riferisce invece al fatto che la percezione degli eventi negativi, dal momento in cui vengono alla luce, venga amplificata più velocemente rispetto a quando si verificano eventi positivi. Questo effetto è collegato al principio della negative potency, ma ne è distinto dal fatto che, per ciò che riguarda stimoli di bassa entità, la forza di quelli negativi non sembra essere più grande di quella di eventi positivi di uguale entità. Anche questa caratteristica del negativity bias è presente nella *Prospect theory* di Tversky e Kahneman (Tversky e Kahneman, 1979), che dimostra come l'avversione agli eventi negativi cresca in modo più rapido rispetto alla soddisfazione data da eventi positivi di pari valore.

Il terzo principio alla base del negativity bias è quello della predominanza degli eventi negativi, o *negative dominance*. In base a questo concetto, la percezione complessiva di un insieme di elementi positivi e negativi risulta essere più negativa rispetto alla semplice somma algebrica dei valori soggettivi di questi elementi. Questo concetto, pertanto si riferisce alla *valutazione* degli elementi positivi e negativi considerati e non alle loro conseguenze reali. Un esempio di predominanza degli eventi negativi è dato dalla situazione in cui il valore soggettivo in termini di negatività della perdita di 100€ equivale, in “valore assoluto” alla utilità derivante dalla vincita di 150€ e, nel complesso, l'esperienza viene vissuta come negativa.

Un caso di studio riguardante questo aspetto del negativity bias è quello proposto da Moosmayer (Moosmayer, 2012), che riguarda la disponibilità a pagare (*willingness to pay*) dei consumatori in presenza di problematiche di tipo etico legate ai prodotti acquistati. I risultati di questo studio dimostrano come, mentre la disponibilità a pagare aumenta in presenza di beni che rispettano caratteristiche di tipo etico, la valutazione complessiva dei prodotti è soggetta a negativity bias, dato che le informazioni negative riguardanti gli aspetti etici dei prodotti sono giudicate con molta più severità da parte degli acquirenti rispetto alle valutazioni positive: ciò viene confermato dal fatto che, in presenza di stimoli negativi, la disponibilità a pagare ha un impatto diretto e negativo sulle decisioni di acquisto rispetto al prezzo atteso, mentre in presenza di stimoli positivi avviene l'esatto contrario: il consumatore è disposto a spendere di più se i prodotti rispettano le sue aspettative riguardo gli aspetti etici dei prodotti considerati.

Gli effetti della predominanza degli eventi negativi, inoltre, sono indipendenti da quelli della maggior forza degli eventi negativi, i quali, rimanendo all'esempio precedente, si verificano quando all'utilità di una vincita di 100€ corrisponde una disutilità maggiore in presenza di una perdita della stessa somma di denaro. Per essere individuati, inoltre, questi effetti non richiedono necessariamente la presenza di stimoli misurati in modo oggettivo, in quanto si basano, per l'appunto, sulle percezioni degli individui. Questo aspetto del *negativity bias* è, secondo gli autori, quello più robusto e quello che viene identificato più facilmente e nella maggior parte delle casistiche analizzate.

Un altro aspetto fondante della predominanza degli eventi negativi, individuato da Kanouse e Hanson (Kanouse e Hanson, 1987), è costituito dal fatto che la presenza di eventi negativi

all'interno di esperienze composite, che comprendono quindi eventi sia positivi che negativi, interferisce con la capacità dell'individuo di godere appieno dell'utilità data dagli aspetti positivi dell'esperienza che sta provando. Questa sovrastima degli eventi negativi, tuttavia, è presente solo quando gli individui si trovano di fronte ad un'esperienza composta, che contiene al suo interno sia aspetti positivi che negativi, che non possono essere separati: in questi casi, pertanto, gli aspetti che possono essere valutati negativamente vengono percepiti in modo ancora peggiore (per fare un esempio, il concetto di “medico incapace” tende ad essere considerato peggiore della semplice idea di “incapace”).

Questo processo cognitivo è evidente anche quando ci si trova in presenza di notizie relative ad eventi negativi che colpiscono i soggetti percepiti come più deboli, come ad esempio bambini, anziani e animali domestici. Le storie riguardanti eventi negativi che colpiscono queste tipologie di soggetti ricevono una copertura mediatica molto più estesa delle notizie che riguardano altre categorie, come dimostra l'attenzione che i mezzi di informazione, in particolare quelli ascrivibili alla categoria dell'*infotainment*<sup>4</sup> prestano a eventi quali il recente “caso Yara Gambirasio”<sup>5</sup> o il “delitto di Cogne”<sup>6</sup>

Infine, la *negativity dominance* può manifestarsi sia in modo sincronico sia in modo diacronico. Si parla di sincronia quando le parti positive e negative che compongono l'esperienza sono percepite dall'individuo in modo simultaneo: in questo caso, le componenti negative saranno molto più influenti di quelle positive nella valutazione complessiva dell'esperienza. Se la percezione di caratteristiche positive e negative di un'esperienza avviene in modo diacronico, invece, si verifica una cancellazione delle percezioni positive da parte di quelle negative e vice-versa. Il “costo” richiesto per la cancellazione di queste ultime è tuttavia molto più elevato rispetto alla situazione opposta e non sempre può essere quantificabile. Complessivamente pertanto, per rendere più piacevole un'esperienza composta, è più efficace un de-potenziamento degli aspetti negativi piuttosto che un potenziamento di quelli negativi.

---

4 L'Enciclopedia Treccani definisce l'*infotainment* come: “termine (derivato dalla fusione di *information* ed *entertainment*) che indica l'ibridazione tra informazione e intrattenimento. [...] L'*infotainment*. è ormai una modalità di tutto il giornalismo televisivo, anche se con toni e stili diversi: dai talk show ai programmi contenitore mattutini e pomeridiani fino a certi telegiornali. “

5 Per una ricostruzione giornalistica dell'omicidio della giovane Yara Gambirasio: Emanuele Menietti, *Il caso Yara Gambirasio, dall'inizio*, Il Post 17 giugno 2014

6 Una cronistoria del delitto è presente in, [http://www.criminologia.it/delitto\\_di\\_cogne/crime/cronologia.htm](http://www.criminologia.it/delitto_di_cogne/crime/cronologia.htm)

L'ultimo aspetto che caratterizza il negativity bias è costituito dalla maggiore differenziazione degli eventi negativi, o *greater negative differentiation*. Questo concetto indica come la percezione degli stimoli negativi sia più complessa ed elaborata rispetto a quella degli stimoli positivi. Anche a livello linguistico, i termini e i concetti usati per descrivere sensazioni spiacevoli tendono ad essere più numerosi di quelli riferiti a esperienze positive, come evidenziato da Peeters (Peeters, 1971). Una situazione simile è quella relativa alle emozioni: quelle positive sono meno numerose di quelle negative, che vengono percepite e descritte in modi più precisi rispetto alle prime (Baumeister et al. 2001) .

La ragione alla base dell'esistenza di questi bias che, portando a sovrastimare la portata di alcuni eventi a discapito di altri inficia la razionalità dell'individuo, ha, secondo Baumeister (Baumeister et al.2001) e Royzman (Royzman et al. 2001), anche se quest'ultimo ha qualche cautela in più, una ragione evolutiva. Durante l'evoluzione, gli individui che meglio si sono adattati a gestire gli eventi negativi sono quelli che hanno avuto la maggiore probabilità di sopravvivere e, di conseguenza, di diffondere il proprio patrimonio genetico: se non riuscire ad avere un'esperienza positiva può precludere sensazioni piacevoli, l'ignorare i pericoli può avere conseguenze anche letali. Questo istinto si è quindi radicato nella parte più primitiva del cervello umano, che è quella che governa gli istinti primari, passando dall'ambito della sopravvivenza pura e semplice a quello della maggiore sensibilità per gli eventi negativi.

Il principio del negativity bias, elaborato in principio in psicologia, è stato esteso a numerosi ambiti del comportamento umano, tra cui l'economia, l'analisi linguaggio, lo studio dei meccanismi attraverso i quali si formano e consolidano le opinioni ed altre situazioni.

Rozin e Royzman e Baumeister e i suoi collaboratori hanno svolto, in modo indipendente, una classificazione degli ambiti in cui è stato riscontrato il negativity bias. Essendo questi lavori molto citati in letteratura, è opportuno riportare alcuni elementi delle tassonomie che gli autori presentano, che possono essere utili per l'analisi del mercato delle informazioni e per avere un quadro più chiaro degli effetti del negativity bias in questo ambito.

*Formazione delle impressioni:* il dominio della formazione delle impressioni è quello in cui il negativity bias è stato studiato maggiormente e in modo più approfondito dagli psicologi.

Il lavoro fondamentale e che ha posto le basi di questa branca dello studio del comportamento umano è quello di Asch (Asch, 1946), dal titolo *Forming impressions of personality*. In questo studio ai partecipanti è stato richiesto di fornire delle valutazioni riguardanti altri individui sulla base di elenchi di caratteristiche sia positive che negative; questi elenchi venivano poi modificati leggermente, togliendo o modificando una voce per volta, per valutare il cambiamento nelle percezioni degli intervistati. Sebbene questo lavoro non fosse direttamente finalizzato allo studio del negativity bias, con esso si è dimostrato come, in presenza di stimoli complessi, le valutazioni date dagli individui non corrispondano alla semplice somma algebrica di queste caratteristiche.

Il lavoro di Asch ha dato il via ad una estesa letteratura riguardante lo studio della formazione delle impressioni e del ruolo che ricopre il negativity bias. Come esempio, Rozin e Royzman prendono i lavori di Kanouse e Hanson (1987) e di Skowronski e Carlston (1989), che analizzano il ruolo della frequenza e della capacità di definire una categoria di individui o di oggetti (capacità diagnostica) degli eventi negativi. Gli eventi negativi sono, secondo Kanouse e Hanson, più rari e, di conseguenza, tendono a collocarsi in posizioni estreme rispetto alla condizione emotiva iniziale dell'individuo, risultando così più influenti nella formazione delle opinioni. Per ciò che riguarda la capacità diagnostica, Skowronski e Carlston sostengono che gli aspetti negativi siano più efficaci nel definire una categoria rispetto a quelli positivi e che, per questa ragione, diano origine a distorsioni nei giudizi (il principio della capacità diagnostica può tuttavia verificarsi anche per gli eventi positivi, nel caso questi siano percepiti come dominanti e sufficientemente rari).

Un altro aspetto importante del negativity bias riferito alla formazione delle opinioni è la maggior forza dei tratti più estremi all'interno di descrizioni complesse, come sostenuto negli studi di Wojciszke (Wojciszke et al. 1993) e Fiske (Fiske, 1980). Questo aspetto conferma la presenza di una maggior pendenza della funzione di utilità degli eventi negativi, che vengono quindi percepiti come più incisivi rispetto a quelli positivi.

Infine, Lupfer et al. (Lupfer et al. 2000) hanno verificato che il negativity bias si verifica solo se, al momento dell'osservazione, i soggetti si trovano di fronte a soggetti o oggetti che presentano sia caratteristiche positive che caratteristiche negative, mentre le distorsioni nel giudizio sono assenti se ci si confronta con descrizioni totalmente positive o totalmente negative. Questa caratteristica fa sì che, nell'ambito della formazione delle impressioni, il



negativity bias appartenga al dominio della predominanza degli eventi negativi (*negative dominance*) e non a quello della maggior forza di questi ultimi (*negative potency*).

*Linguaggio ed emozioni:* questi due domini sono molto legati tra loro (il linguaggio è fondamentale per interpretare, comunicare e interpretare le emozioni complesse) e sono fortemente influenzati dalle distorsioni causate dagli eventi negativi.

Prima di analizzare distintamente il ruolo del negative bias nei due domini, è opportuno considerare che, come evidenziato tra gli altri da Bohner (Bohner et al., 1988) e da Peeters e Czapinski (Peeters e Czapinski, 1990), agli eventi negativi sono attribuite un maggior numero di cause rispetto a quelli positivi e i primi vengono percepiti come più complessi rispetto a questi ultimi. Inoltre, come mostrato in Morewedge 2009, le negatività tendono ad essere attribuite ad agenti esterni, mentre gli eventi positivi sono maggiormente percepiti come frutto delle proprie azioni.

Kahneman e Miller (Kahneman e Miller, 1986) hanno poi dimostrato come gli individui tendano a ragionare in termini controfattuali in presenza di situazioni sfavorevoli, al contrario di ciò che avviene in presenza di eventi positivi.

Passando al ruolo del negativity bias nella formazione del linguaggio, Rozin e Royzman hanno constatato come il vocabolario utilizzato per descrivere gli eventi negativi sia molto più ricco, preciso e articolato di quello utilizzato in caso di aspetti positivi: ciò vale sia per l'inglese (Rozin e Royzman, 2001), che per altre 20 lingue, tra cui le neolatine francese e spagnolo, analizzate dagli autori in un lavoro successivo (Rozin et al., 2010). Osgood e Hoosain (Osgood e Hoosain, 1983), inoltre, hanno osservato come i tempi di reazione necessari ad assegnare la valenza negativa ad un aggettivo siano più alti rispetto all'interpretazione degli aggettivi positivi, ciò a riprova della maggior complessità delle esperienze negative.

Per ciò che riguarda il ruolo del negativity bias nell'elaborazione delle emozioni, sempre Rozin e Royzman affermano che le cause che originano i sentimenti negativi sono classificabili in un numero maggiore di modi rispetto a quelle che scatenano emozioni positive; di conseguenza, anche l'elaborazione delle emozioni negative e le reazioni a queste ultime sono più elaborate e complesse di quelle derivanti da emozioni positive.

*Elaborazione delle informazioni:* le risorse cognitive degli individui sono limitate e devono essere focalizzate sull'elaborazione degli stimoli e delle informazioni che sono percepite come più importanti. La presenza di un bias verso gli eventi negativi, pertanto, fa sì che questi ultimi vengano elaborati dagli individui in modo più approfondito ed accurato, utilizzando più risorse cognitive, come avviene nel caso della maggiore persistenza dei ricordi collegati ad essi.

In Baumeister et al. (2001) vengono presentati alcuni esempi di esperimenti che supportano questa tesi, da cui risulta evidente la maggiore attenzione posta verso gli eventi negativi. Quattro studi in particolare sono utili per la comprensione del ruolo del negativity bias nel mercato delle notizie: Fiske (Fiske, 1980) ha dimostrato come, in presenza di fotografie rappresentanti eventi positivi e negativi, gli intervistati a cui erano sottoposte le immagini passassero più tempo ad analizzare quelle appartenenti alla seconda categoria.

La maggior attenzione nei confronti delle esperienze negative, inoltre, fa sì che gli individui tendano ad individuarne le cause ed il senso con maggiore determinazione rispetto a ciò che accade in presenza di eventi piacevoli o positivi, come dimostra l'analisi condotta da Weiner (Weiner, 1985) su diciassette studi riguardanti l'attribuzione spontanea di causalità<sup>7</sup>.

Sempre rimanendo in questo ambito, Pratto e John (Pratto e John, 1991) hanno dimostrato come gli eventi negativi vengano ricordati più a lungo ed attraggano maggiormente l'attenzione, in modo automatico e indipendente dalla volontà dell'individuo, rispetto agli eventi positivi.

In Marshall e Kidd, 1981, infine, viene verificata la maggiore attenzione posta nei confronti degli eventi negativi chiedendo ai partecipanti a diversi studi se preferissero ricevere prima una notizia negativa poi una positiva o vice-versa. In percentuali di casi che variavano dal 77% all'88% gli intervistati hanno scelto di ascoltare prima la notizia negativa, manifestando quindi una maggiore attenzione nei confronti degli eventi negativi.

Per concludere, quindi, questi studi confermano la presenza del negativity bias nei processi di elaborazione delle informazioni, in quanto quelle negative sono processate in modo più approfondito e preciso, attivano un maggior numero di processi cognitivi e catturano più facilmente l'attenzione rispetto a quelle positive.

---

<sup>7</sup> Weiner (1985) definisce la attività spontanea di attribuzione come lo sforzo degli individui per comprendere gli eventi di cui hanno esperienza e le cause che li hanno generati

Alla luce di quanto emerso finora, si può affermare che la presenza del negativity bias, porti gli individui ad essere maggiormente attratti dagli eventi negativi. Questa distorsione, perciò, aumenta la propensione dei consumatori di notizie ad acquisire news dal contenuto negativo, sia che riguardino eventi realmente accaduti, come catastrofi naturali, conflitti, malattie e altri eventi percepiti come molto gravi, sia che facciano riferimento a previsioni negative, siano esse in campo economico, sociale, sanitario o altri ambiti che possono avere ripercussioni negative sia sui singoli che sui gruppi di individui.

Proprio a questo aspetto del mercato delle informazioni è dedicato il lavoro di Trussler e Soroka (2013), che analizza il mercato dell'informazione politica e i meccanismi che lo compongono. In questo articolo gli autori dimostrano come il ruolo della domanda sia più importante di quello dell'offerta nello spiegare la maggior presenza di notizie negative nel campo dell'informazione politica. Gli individui maggiormente interessati alle notizie negative, inoltre, sono proprio quelli più interessati alle notizie riguardanti la politica.

Nel prossimo capitolo verrà proposto un modello di mercato dell'informazione, presentato da Mullainathan e Schleifer nel 2005 e incentrato sul ruolo della domanda nella caratterizzazione del mercato delle informazioni.

## 2. Un modello di mercato dell'informazione e il ruolo della domanda

### 2.1 Il mercato delle notizie e il ruolo della concorrenza

All'interno del dibattito sul ruolo della concorrenza nel mercato dell'informazioni uno dei problemi principali è stabilire se e come la competizione tra media, siano essi giornali, canali televisivi, siti internet o altri attori, porti a una maggiore obiettività delle testate e una maggiore attinenza delle notizie riportate ai fatti realmente accaduti.

Il modello elaborato da Mullainathan e Shleifer nel loro articolo del 2005 intitolato *The market for news* rappresenta una dimostrazione teorica di come la domanda influenzi fortemente il comportamento dei media nel riportare le notizie. La concorrenza tra testate, inoltre, non porta ad una maggiore obiettività dei singoli media, ma al contrario, spinge verso una radicalizzazione dei contenuti, portando ad un mercato sempre più polarizzato. Per avvicinarsi il più possibile alla verità, il consumatore di news dovrà quindi informarsi dal maggior numero possibile di fonti.

### 2.2 Il modello di Mullainathan e Shleifer

Nell'articolo del 2005 intitolato *The market for news*, Mullainathan e Shleifer propongono un'analisi teorica del mercato dell'informazione, partendo dal lato della domanda, che secondo i due autori è profondamente influenzata sia dalle opinioni dei consumatori di notizie che dalla natura stessa dei fatti riportati.

Il punto di partenza dello studio sta nel contestare la concezione del mercato dell'informazione data dagli economisti classici, secondo la quale i lettori/ascoltatori ricercano l'accuratezza delle informazioni e che, pertanto, i media siano spinti a fornire notizie il più possibile veritiere: le forze di mercato spingerebbero quindi i produttori di notizie a eliminare il più possibile i bias, o distorsioni.

Questa visione del mercato delle news si differenzia radicalmente dai risultati degli studi non di tipo economico condotti nel campo dell'informazione giornalistica. Secondo questi studiosi, i produttori di news hanno come scopo principale quello di vendere copie, programmi televisivi e spazi pubblicitari e per fare questo mettono sul mercato sia grandi quantità di prodotti di intrattenimento che notizie commentate e interpretate, così come chiedono i consumatori. Per questo motivo si parla di "imperativo narrativo" (come

analizzato, tra gli altri, da Lee ,2005; Hamilton 2003; Graber, 1984): i media non propongono notizie pure, ma le interpretano e ne omettono parti per venire incontro alla domanda dei consumatori, che sono spesso affetti da *confirmation bias*<sup>8</sup> (Kim, 2015; McFadden e Lusk, 2015; Nickerson, 1998; Severin e Tankard 1992). Questa operazione di interpretazione della realtà viene attuata attraverso un procedimento definito *slanting*. Con questo termine, proposto da Hayakawa (Hayakawa, 1940) indica il “processo di selezione dei dettagli che possono mettere in buona o cattiva luce il soggetto descritto” dalla notizia. Questa operazione ha un costo per il produttore di news, che varia in base alle preferenze degli utenti (maggiore radicalità comporta costi di approvvigionamento di notizie più alti).

Il modello presentato da Mullainathan e Shleifer propone una teoria del mercato delle informazioni che tenga conto di questi due aspetti: la preferenza per news coerenti con le proprie opinioni e la propensione dei produttori di notizie a fornire news reinterpretate in modo da venire incontro alle opinioni dei lettori/ascoltatori.

Per fare ciò, gli autori individuano due tipologie di mercato dell’informazione e due tipologie di lettori: mercati monopolistici e mercati duopolistici; lettore medio, che attinge le informazioni da una sola fonte, e lettore definito *coscienzioso*, che si informa utilizzando più fonti. La molteplicità delle opinioni dei lettori/ascoltatori, come si dimostrerà in seguito, è la spinta principale verso il raggiungimento di un sistema dell’informazione che, se preso nel suo complesso, riproduce il più fedelmente possibile la realtà, anche se le singole testate tendono a spingersi sempre più verso posizioni polarizzate, mentre la semplice competizione tra produttori di notizie non è sufficiente ad eliminare i bias dell’informazione, ma solamente a ridurre i prezzi delle testate (un buon esempio può essere quello delle riviste “di gossip”, che hanno un’audience con aspettative omogenee e che, pertanto, competono tra di loro solamente abbassando i prezzi e/o allegando gadget).

### 2.3 Impostazione del modello

Si suppone che i lettori siano interessati a conoscere l’andamento di una variabile  $t$ , che può essere lo stato dell’economia, l’occupazione, lo stato dei mercati internazionali ecc.

---

<sup>8</sup> Questo bias fa sì che l’utilità di notizie che confermano le opinioni preesistenti sia maggiore di quella data da notizie che contrastano con le idee di chi legge e/o ascolta.

La variabile  $t$ , inoltre, ha una distribuzione normale, con media zero e varianza pari a  $\vartheta_t$ .

Formalizzando, la distribuzione di  $t$  è definita come:

$$t \sim N(0, \vartheta_t)$$

Definendo  $p = (1/\vartheta_t)$  come precisione.

I lettori/ascoltatori hanno delle opinioni riguardo a  $t$  che possono essere affette da bias. Questi bias si distribuiscono nella stessa modalità di  $t$ , ma traslati di un valore  $b$  e con la stessa varianza:

$$b \sim N(b, \vartheta_t)$$

Le testate giornalistiche, quindi, ricevono dati  $d$  riferiti all'evento  $t$  e li riportano sotto forma di notizie. Le notizie che ricevono sono viziate da un errore  $\varepsilon$ , che si distribuisce come:

$$\varepsilon \sim N(b, \vartheta_\varepsilon)$$

Nel riportare la notizia, i media applicano un valore di slant  $s$ : si ottiene quindi:  $n=d+s$ , da cui  $n=t+\varepsilon+s$ .

Per ciò che riguarda l'utilità del fruitore delle notizie, il modello descrive il comportamento di due tipologie di lettori/ascoltatori: quelli razionali e imparziali e quelli affetti da bias.

Nel primo caso i consumatori di news sono alla ricerca di informazioni pure e sono avversi allo slant, che risulta essere un costo, sia in termini di tempo impiegato per estrarre i fatti veri dalle notizie parziali, sia in termini di denaro speso per ottenere queste news. L'utilità di un lettore completamente razionale sarà inversamente proporzionale al valore dello *slant* e corrisponderà a:

$$U_r = \bar{u} - \chi s^2 - P$$

Dove  $\bar{u}$  rappresenta l'utilità media del lettore,  $P$  il prezzo di acquisto della notizia e  $\chi$  il peso dello slant.

I lettori/ascoltatori non razionali e affetti da bias, invece, vedono diminuire la propria utilità ricevendo notizie che non sono consistenti con le loro opinioni. Si definisce consistenza la distanza tra la notizia  $n$  e le opinioni  $b$  dei lettori:  $(n-b)^2$ . Il rapporto diretto tra bias e maggiore utilità percepita non è però infinito, in quanto anche il lettore/ascoltatore più schierato non è

attratto dagli slant estremi, almeno sul lungo periodo: mantenendo costante la consistenza con le proprie opinioni, questo tipo di consumatore di news preferisce le notizie con slant minore.

La funzione di utilità di questa tipologia di consumatore, nel caso in cui decida di acquistare l'informazione è:

$$U_r = \bar{u} - \chi s^2 - \varphi(n - b)^2 - P$$

dove  $\varphi > 0$  è il parametro che regola la preferenza nel ricevere notizie coerenti con le proprie opinioni.

Il terzo aspetto del modello consiste nelle strategie adottate dai produttori di notizie. Questi ultimi, prima di ricevere i dati  $d$  dichiarano sia la propria strategia di slanting  $s(d)$  che il prezzo  $P$  della notizia. Il potenziale acquirente, pertanto, acquista la notizia solo se il prezzo è inferiore all'utilità attesa dalla fruizione della news, definita  $E_d[U(s(d))]$ . Questa utilità attesa si basa sulle aspettative del lettore/ascoltatore sul dato  $d$  e non è affetta da bias:

$$d \sim N(t, \vartheta_d).$$

Infine, una volta che il consumatore decide di acquistare la notizia, la testata osserva i dati  $d$  e riporta la news come  $n=d+s(d)$ .

Questo approccio, quindi, analizza il mercato delle news sul lungo periodo, che può essere scomposto in cinque fasi:

- I media annunciano la loro strategia di slant  $s(d)$ : questa può essere, ad esempio, il posizionamento politico di un giornale, dichiarato al momento della sua nascita e ribadito attraverso le scelte editoriali;
- Viene annunciato il prezzo di vendita delle notizie. Nel caso dei media on line questo prezzo è più basso rispetto a quello dei media cartacei ed è dato dal costo della connessione dati, dal tempo impiegato per cercare le notizie e dall'eventuale presenza di annunci pubblicitari all'interno delle pagine web e dei video;
- Il consumatore decide se acquistare o meno la notizia (acquistando il giornale o navigando nel sito web del quotidiano), in base alle aspettative che si è formato a riguardo dell'evento  $t$ , al prezzo e alla strategia di slant della testata;

- I media riportano la notizia con il relativo slant:  $n=d+s(d)$  se ci si trova in una situazione di monopolio;  $n=d+s_j(d)$  con  $j=1,2$  in caso di duopolio
- Se il consumatore ha deciso di acquistare la notizia, la riceve e ne ricava un'utilità.

Il modello di mercato dell'informazione proposto prende in considerazione, come già accennato in precedenza, due tipologie di domanda e due tipologie di offerta: l'audience di riferimento può essere omogenea o eterogenea e l'offerta di notizie può esistere in un contesto di monopolio o di duopolio.

In caso di pubblico omogeneo, tutti i lettori/ascoltatori condividono le stesse opinioni  $b$  riguardo all'evento  $t$ , che conoscono con un livello di precisione  $p$  (ad esempio, tutti consumatori di un determinato alimento possono condividere lo stesso timore in caso di problematiche ad esso collegate). In questo caso, la funzione di utilità dei consumatori di notizie è definita come  $u_i(d)$

Un pubblico eterogeneo, invece, vede la presenza di differenti opinioni (come spesso accade per i fatti di natura politica), distribuite tra due estremi  $b_1$  e  $b_2$ , con  $b_1 < b_2$  e  $b_2 > 0$  e  $\bar{b}$  come media tra le opinioni estreme. La funzione di utilità in caso di domanda eterogenea è definita come  $u_{b_i}(d)$ .

Per ciò che riguarda il lato dell'offerta, vengono prese in considerazione due tipologie di mercato: monopolio e duopolio. Nel primo caso, l'unico media presente attuerà politiche di slant e di prezzo ottimali e che dipendono dalla platea di utenti a cui offre le notizie:  $s_{hom}^*$  e  $P_{hom}^*$  in caso di lettori con opinioni omogenee e  $s_{het}^*$  e  $P_{het}^*$  in caso di opinioni diverse.

In situazioni di duopolio, invece, i media  $j=1,2$  sviluppano le proprie strategie di slanting e di prezzo, sempre dipendenti dalle tipologie di opinioni diffuse tra i consumatori di news. In caso di pubblico omogeneo, le strategie saranno definite come  $s_{j,hom}^*$  e  $P_{j,hom}^*$ , mentre con pubblico dalle opinioni differenziate saranno  $s_{j,het}^*$  e  $P_{j,het}^*$ . La struttura del mercato delle informazioni proposta dai due autori riprende quella proposta da Hotelling (Hotelling (1929) e d'Aspremont et al. (1979)) per descrivere un contesto in cui sono presenti due industrie che vendono beni simili, ma in cui i consumatori hanno diversi costi di trasporto, che, nel caso delle notizie, corrispondono alle diverse opinioni e, di conseguenza, agli sforzi che i lettori/ascoltatori compiono per evitare gli slant. Dall'altro lato, le strategie di slanting corrispondono, nel modello di Hotelling, alla scelta della locazione delle imprese.



L'ultimo concetto da definire prima di poter analizzare l'andamento del mercato delle notizie è quello di *bias del mercato*, calcolato come il bias medio delle fonti di informazione presenti sul mercato stesso.

In una situazione di audience omogenea, il bias medio dei lettori (*average reader bias*) è definito come:

$$ARB_{hom} = E_d[(n - d)^2]$$

Con  $n$  che corrisponde alla notizia recepita dal lettore/ascoltatore. In caso di audience omogenea, dunque, il bias totale è definito come il valore medio della distanza tra dato e notizia nella percezione dei lettori.

In caso di lettori dalle opinioni eterogenee, definita  $n_i$  la notizia ricevuta dal soggetto  $i$  (con  $i \in [1,2]$ ), il bias è definito come:

$$ARB_{het} = \int_i E_d[(n_i - d)^2]$$

Corrisponde quindi alla differenza media tra notizia e dato reale che tutti i lettori/ascoltatori percepiscono.

Avendo definito gli elementi che compongono il modello, è possibile passare all'analisi del mercato dell'informazione, a seconda della tipologia di pubblico e di mercato.

### 2.3.1 Primo caso: lettori razionali avversi allo slanting

Questa tipologia di consumatori di notizie rappresenta quella su cui si basa il modello classico di mercato dell'informazione. I lettori, indipendentemente dall'eterogeneità delle opinioni sono interessati solamente ai fatti "reali" e, pertanto, sono avversi allo slanting. I produttori di notizie non applicano slant e propongono lo stesso prezzo. In caso di monopolio, strategia di slanting e prezzi saranno rispettivamente:

$$S_{hom}^* = S_{het}^* = 0$$

$$P_{hom}^* = P_{het}^* = \bar{u}$$

E in caso di duopolio

$$s_{j,hom}^* = s_{j,het}^* = 0$$

$$P_{j,hom}^* = P_{j,het}^* = \bar{u}$$

L'unico risultato della competizione è quello dell'abbassamento dei prezzi di vendita delle notizie.

### 2.3.2 Secondo caso: lettori non perfettamente razionali

L'analisi diventa più complessa in presenza di lettori affetti da bias. In caso di monopolio, le strategie di slanting e prezzo adottate dal monopolista sono:

$$s_{hom}^* = \frac{\phi}{\chi + \phi} (b - d)$$

$$P_{hom}^* = \bar{u} - \frac{\chi\phi}{\chi + \phi} [b^2 + v_d]$$

se  $\bar{u} > \chi\phi/(\chi + \phi)[b^2 + v_d]$  allora una strategia di slanting è possibile; in caso contrario non esiste un slant che convinca il consumatore ad acquistare la notizia, in quanto il prezzo ottimale sarebbe negativo. Essendo il monopolista l'unico a vendere informazioni, tutto il suo guadagno deriva dall'aumento del prezzo, che deve corrispondere ad un aumento della utilità attesa  $\bar{u}$ : la notizia  $n$  riportata sarà quindi:

$$n = \frac{\phi}{\chi + \phi} b + \frac{\chi}{\chi + \phi} d$$

La notizia riportata, pertanto, è una combinazione di bias e dati, con i relativi pesi costituiti da parametri di utilità.

In questo caso si verifica quindi uno slanting lineare del media verso l'opinione  $b$  dell'audience, che, come ricordato in precedenza, è omogenea e propensa a ricevere notizie che concordano con l'opinione stessa. La strategia lineare di slanting  $s_B(d)$  è ricorrente nel mercato delle news e può essere formalizzata come:

$$s_B(d) \equiv \frac{\phi}{\chi + \phi} (B - d)$$

Questa notazione corrisponde a quella della strategia di slanting ottimale adottata dal monopolista, che attua una strategia lineare per massimizzare l'utilità del consumatore per ogni dato  $d$  di cui dispone<sup>9</sup>: si tratta quindi di uno slant verso  $b$ .

La notizia riportata sarà quindi una combinazione di bias e dati reali, che avranno come peso il valore dei parametri dell'utilità ( $\phi/(\chi + \phi)$  e  $\chi/(\chi + \phi)$ )

L'intensità dello slanting, inoltre, dipende sia dalla preferenza dei lettori per notizie coerenti con le proprie idee che dal costo dello slanting stesso.

Si ha quindi che:

$$\frac{\partial |s_{hom}^*(d)|}{\partial \phi} > 0$$

$$\frac{\partial |s_{hom}^*(d)|}{\partial \chi} < 0$$

Questa impostazione suggerisce la presenza di uno *spin* nel mercato dell'informazione. Se un'autorità esterna, come può essere un'autorità di controllo sulla sicurezza alimentare, possiede un nuovo dato  $d$  ha un forte vantaggio iniziale, in quanto può decidere se e come comunicarlo ai media. A questo punto i media slanteranno verso le opinioni dell'audience ma, come descritto in precedenza, il dato  $d$  ha un impatto significativo su ciò che i media riportano. I lettori/ascoltatori, pertanto, domanderanno notizie con uno slant ancora più accentuato verso il loro valore di riferimento  $b$  e i media offriranno notizie con uno slant maggiore, innestando così una spirale che coinvolgerà anche i media che all'inizio sono neutrali. Un esempio di questo meccanismo verrà presentato nei paragrafi successivi, dove verranno analizzate le strategie di tre influenti quotidiani italiani in occasione dell'epidemia di influenza H1N1, o febbre suina, verificatasi nel corso del 2009.

Infine, la condizione  $u^- > \chi\phi/(\chi + \phi)[b^2 + v_d]$  garantisce che l'utilità di riserva<sup>10</sup> dei consumatori sia sufficientemente alta per far sì che le notizie con un livello di bias ottimale

---

<sup>9</sup> Ciò è possibile in quanto le funzioni di utilità attesa sono separabili per ogni valore di  $d$ , contengono cioè sottoinsiemi numerabili e densi per ogni valore di  $d$

<sup>10</sup> Livello di utilità minimo per il quale un contratto è ritenuto accettabile dall'agente.

siano preferibili all'assenza di notizie; questo concetto rimane costante in tutte le tipologie di mercato dell'informazione.

Introducendo la competizione tra testate in questo tipo di mercato (con un pubblico che ha opinioni omogenee) si ottiene un equilibrio per cui:

$$s_{j,hom}^*(d) = \frac{\phi}{\chi + \phi} (b - d)$$

$$P_{j,hom}^* = 0$$

Si può vedere come, in caso di pubblico con idee omogenee, la competizione tra testate si sposti solo sul lato del prezzo, in quanto ognuno dei duopolisti applica una strategia di slanting uguale a quella del monopolista considerato in precedenza.

La quantità media di bias prodotta nei due contesti è quindi la stessa:

$$ARB_{mon}(v_d) = ARB_{duo}(v_d)$$

La competizione, pertanto, non elimina i bias, ma li asseconda e porta solo ad una riduzione di prezzo e *non* ad un aumento dell'accuratezza delle notizie prodotte.

Rimane ora da analizzare la forma di mercato dell'informazione più complessa, quella con lettori dalle opinioni eterogenee. In questo contesto, i media devono scegliere la parte di pubblico a cui fare riferimento nel trasmettere le notizie.

In caso di monopolio, si suppone che l'unica testata presente si trovi di fronte ad un pubblico con un bias medio  $\bar{b}=0$  ed abbia lo scopo di massimizzare i profitti. Esisterà quindi un valore  $C_m$ , dipendente dai parametri del modello ( $\chi$  e  $\Phi$ ) che determina il comportamento del monopolista: se  $b_2 - b_1 < C_m$  allora il monopolista sceglierà la strategia che massimizza i suoi profitti<sup>11</sup>:

$$s_{het}^* = s_{\bar{b}}(d) = \frac{\phi}{\chi + \phi} (\bar{b} - d) = -\frac{\phi}{\chi + \phi} d$$

$$P_{het}^* = \bar{u} - \frac{\chi\phi}{\chi + \phi} v_d - \phi^2 b_2^2$$

---

<sup>11</sup> la differenza tra le opinioni deve essere inferiore al valore di  $C_m$  considerato

Se invece  $b_2 - b_1 > C_m$  il monopolista sceglie di non vendere in quel mercato e nessun lettore/ascoltatore ottiene le notizie. In caso di monopolio, quindi, i lettori riceveranno notizie riguardanti un certo dato  $d$  solo se la dispersione delle opinioni non è troppo grande (i due estremi dell'asse ideologico delle opinioni non devono essere troppo distanti).

La risposta alla dispersività delle opinioni cambia radicalmente in caso di duopolio. Supponendo che i duopolisti adottino una strategia di slant lineare e che il bias medio dell'audience sia nullo, come affermato in precedenza, esisterà un valore costante  $C_d$  pari a:

$$C_d = \sqrt{\frac{4}{33} \left[ \frac{\phi + \chi}{\phi^2} \bar{u} - \frac{\chi}{\phi} v_d \right]}$$

tale per cui se  $b_2 < C_d$  i duopolisti adotteranno le strategie:

$$s_{1,het}^*(d) = \frac{\phi}{\phi + \chi} \left( \frac{3}{2} b_1 - d_1 \right)$$

$$s_{2,het}^*(d) = \frac{\phi}{\phi + \chi} \left( \frac{3}{2} b_2 - d_2 \right)$$

$$P_{j,het}^* = \frac{6\phi^2}{\phi + \chi} b_j^2$$

Si può quindi affermare, generalizzando, che il media 1 slanta verso l'estremo  $b_1$  e il media 2 si sposta verso l'estremo  $b_2$ : tutti i lettori, pertanto, ottengono le notizie relative al dato  $d$ . Ognuno dei due media si posiziona quindi il più lontano possibile sull'asse ideologico rispetto al concorrente: la notizia riportata, dunque corrisponderà a:

$$n_j = s_{j,het}^* = \frac{\phi}{\phi + \chi} \frac{3}{2} b_j + \frac{\chi}{\chi + \phi} d_j$$

La notizia riportata sarà quindi una media pesata del dato  $d$  e di  $3/2b_j$ , con  $b_j$  che rappresenta il punto più estremo della distribuzione del bias dei lettori. I duopolisti, pertanto, calibrano le loro strategie di slanting verso il valore  $3/2b_j$ , mirando quindi ad una posizione *più estrema* di quella dei loro lettori più estremi. In questo modo il mercato delle news si polarizza sempre di più, in quanto sia i produttori che i consumatori di notizie tendono a localizzarsi in posizioni sempre più estreme. Questo comportamento dei media è assimilabile a quello delle imprese basato sul modello di Hotelling in presenza di distribuzione uniforme della

domanda e di costi di trasferimento quadratici: in caso di monopolio l'intero mercato è coperto, a patto che le opinioni più estreme non siano troppo distanti tra loro, mentre in presenza di duopolio, a differenza della situazione descritta da Hotelling, i competitor possono posizionarsi anche al di fuori della distribuzione dei bias dei lettori e, in situazione di equilibrio, si collocano su posizioni estreme.

Questa strategia di slanting viene utilizzata perché l'effetto prezzo domina il mercato fino a che i competitor non sono abbastanza distanziati tra loro: per massimizzare i profitti, pertanto, le testate giornalistiche devono estremizzare i contenuti proposti, per spostare la competizione sui contenuti proposti e non sulla semplice riduzione del prezzo. Un esempio di queste strategie di posizionamento verrà presentato nel capitolo 5.1, dove verranno analizzate le strategie editoriali di due quotidiani italiani a proposito di un tema molto divisivo quale quello della gestione dei flussi migratori.

L'intensità dello slant del duopolista, inoltre, è maggiore di quella del monopolista e ciò porta, come detto in precedenza, a una polarizzazione dell'offerta e di conseguenza ad un aumento del bias medio dei lettori:

$$E[|(s_{j,het}^*(d))|] \geq E[|(s_{het}^*(d))|]$$

Quindi:

$$ARB_{mon,het}(v_d) < ARB_{duo,het}(v_d)$$

L'impatto della competizione sull'accuratezza delle notizie, pertanto, deve essere valutato attentamente, in quanto l'ingresso di nuove testate su un mercato già inficiato da bias potrebbe favorire ulteriormente la polarizzazione delle opinioni, anziché aumentare l'obiettività delle notizie trasmesse.

Per analizzare le conseguenze del duopolio in un contesto di pubblico eterogeneo, è necessario studiare il comportamento di un ipotetico lettore/ascoltatore "coscientoso", che acquisisce notizie da entrambe le testate presenti sul mercato. Questa strategia dovrebbe permettere al lettore di evitare lo slanting, in quanto entrambi i media trattano lo stesso dato  $d$ .

Per prima cosa è necessario formalizzare la definizione di slant data da Hayakawa e citata in precedenza. Si suppone che il media riceva, anziché un unico dato  $d=t+\epsilon$ , una sequenza di

“particelle” di notizie riferite a  $d$ , che possono avere connotazioni positive o negative (nel caso di un evento legato alla salute pubblica, queste particelle possono essere ad esempio, i dati sull'estensione del fenomeno, le opinioni di diversi esperti, le notizie su possibili rimedi eccetera). Queste parti di  $d$  sono rappresentati come una stringa  $f$  di lunghezza  $L$  che contiene parti positive (+1), negative (-1) o non inerenti a  $d$  ( $\emptyset$ ). La probabilità che la parte di notizia in posizione  $i$  ( $f(i)$ ) sia positiva, negativa o non inerente a  $d$  è data dalla funzione di probabilità:

$$\Pr(f_i) = \begin{cases} +1 = qg(d) \\ -1 = q(1 - g(d)) \\ \emptyset = (1 - q) \end{cases}$$

Con  $g(\cdot)$  funzione continua e crescente compresa tra 0 in senso stretto e 1 in senso lato ( $0 < q \leq 1$ ). Con probabilità  $(1-q)$  la posizione  $i$  non conterrà una notizia inerente a  $d$ ; se invece la particella contiene una notizia utile, questa sarà positiva con probabilità  $g(d)$  o negativa con probabilità complementare[controlla].

A questo punto, se il media decide di non applicare una strategia di slanting, riporta l'intera stringa  $f$ , senza alterazioni e il lettore potrà informarsi utilizzando il numero di particelle di  $L$  positive (definito  $N_+(f)$ ) e negative (definite  $N_-(f)$ ). Per la legge dei grandi numeri:

$$\frac{N_+(f)}{N_-(f) + N_+(f)} = g(d) + \eta \rightarrow g(d)$$

Con  $\eta$  che definisce un rumore che tende a zero quando la lunghezza  $L$  della stringa  $f$  tende all'infinito. Di conseguenza, se il lettore acquisisce un maggior numero di informazioni attingendo da diversi media, aumenterà il valore di  $L$  e di conseguenza il rumore diminuirà, facendo sì che ci si avvicini al dato reale  $d$ .

I media applicano lo slant omettendo selettivamente di riportare alcune parti della stringa  $f$ : se si intende enfatizzare i lati positivi si ometteranno le particelle di notizie negative e viceversa. Formalizzando:

$$g(d)_s = \frac{N_+(f')}{N_-(f') + N_+(f')} \approx d + s$$

Infine, più aumenta la lunghezza  $L$  della stringa, migliore diventa la strategia di slanting adottata dai media, in quanto hanno più possibilità di rimuovere selettivamente parti di notizia contrarie alla loro linea editoriale.

Una volta formalizzate le forme dello slanting, si passa ad analizzare le modalità con cui un lettore coscienzioso può, attingendo notizie da tutte le fonti disponibili, ridurre l'impatto di queste strategie.

Supponendo che ogni testata riceva la stringa  $f$  relativa ai dati  $d=t+\varepsilon$ , essa verrà rielaborata dai media per essere messa sul mercato: ogni giornale  $j$  presente sul mercato proporrà la propria stringa  $f_j$ , con il relativo slant. A questo punto si presentano due opzioni: se i due media applicano lo slant nella stessa direzione ma con intensità diversa ( $s_1>s_2>0$  oppure  $0>s_2>s_1$ ) allora il lettore coscienzioso otterrà il massimo possibile delle informazioni utilizzando una sola fonte, quella, cioè, che omette meno particelle di notizia, applicando uno slant più debole.

L'altra opzione si verifica quando i due media si posizionano sui lati opposti dell'asse ideologico, applicando quindi uno slant di direzione opposta alla notizia che riportano, ovvero  $s_1>0>s_2$ . Avverrà quindi che uno dei due media ometterà particelle di notizie di segno positivo, mentre l'altro non riporterà le parti della stringa  $f$  dal tono negativo. Per arrivare a conoscere il dato  $d$  reale alla base delle notizie  $n_1$  e  $n_2$  riportate dai media, il lettore coscienzioso attingerà da entrambe i media compiendo una verifica incrociata (*cross checking*) dei dati. Questo comportamento viene definito dalla funzione:

$$xc(s_1, s_2) \begin{cases} \min\{s_1, s_2\} & \text{se } s_1 > 0, s_2 > 0 \\ \max\{s_1, s_2\} & \text{se } s_1 < 0, s_2 < 0 \\ 0 & \text{negli altri casi} \end{cases}$$

Il soggetto che ottiene le notizie da entrambe i media presenti sul mercato riceverà la news  $n_c$  definita come:

$$n_c = \begin{cases} n & \text{con una fonte} \\ d + xc(s_1, s_2) & \text{con due fonti} \end{cases}$$

Definendo il bias del lettore coscienzioso come:

$$CRB = E_d[(n_c - d)]^2$$



si può notare come questo valore non sia direttamente collegato alla diversità delle opinioni dell'audience. Il valore di CRB dipende invece dalla notizia che i giornali riportano in condizione di equilibrio ( $n_c$ ), la quale, a sua volta, è influenzata dalle opinioni del pubblico a cui i media si riferiscono. Da questa considerazione deriva il punto di arrivo del modello, cioè il ruolo fondamentale dell'*eterogeneità* delle opinioni, unito alla *concorrenza* dal lato dell'offerta, nel garantire la maggiore accuratezza possibile delle notizie presenti sul mercato.

Formalizzando, si può affermare che in presenza di opinioni variegata, il bias dei lettori coscienti è inferiore in situazioni di duopolio rispetto alla presenza di un unico offerente:

$$CRB_{het,duo} < CRB_{het,mon}$$

Così come, in situazione di duopolio, il bias subito dal pubblico in presenza di opinioni omogenee è inferiore al bias che si riscontra in presenza di eterogeneità delle opinioni. Formalizzando:

$$CRB_{het,duo} < CRB_{hom,duo}$$

Queste due disuguaglianze rappresentano le conseguenze più importanti delle caratteristiche della domanda nell'ambito del mercato dell'informazione: la competizione di per se non porta alla presenza sul mercato di notizie più vicine alla verità, ma conduce solo ad un abbassamento dei prezzi delle testate. In presenza di lettori dalle opinioni differenti, invece, si assiste ad una polarizzazione del mercato, con i media che tendono a proporre notizie sempre meno neutrali. Questa segmentazione del mercato, tuttavia, favorisce i lettori che acquisiscono le informazioni da più fonti differenti e che riescono, in questo modo, a ottenere più dettagli riguardo ai fatti su cui si informano, avendo così un quadro della realtà più vicino alla verità dei fatti. L'eterogeneità delle opinioni, pertanto, è il mezzo principale per ottenere un mercato dell'informazione in cui le notizie siano caratterizzate da un alto livello di accuratezza. Argomenti che non generano opinioni differenziate, come possono essere quelle relative ad una crisi sanitaria o alimentare, ottengono una copertura meno accurata da parte dei singoli mezzi di informazione, che presenta storture più marcate rispetto ad altre tipologie di notizie.

## 3. Elementi di analisi automatica dei testi

---

### 3.1 L'analisi automatica dei testi: le metodologie principali

Lo studio delle idee e delle opinioni è un elemento fondamentale in molteplici ambiti delle scienze sociali ed economiche, sia per chi deve prendere decisioni che per chi voglia capire l'impatto di queste decisioni sui diretti interessati, siano essi consumatori o elettori.

La continua evoluzione degli strumenti di calcolo e immagazzinamento di dati, oltre alla diffusione sempre più capillare di internet e dei social media ha fatto sì che la quantità di documenti disponibili per gli studiosi sia costantemente aumentato. In tempi più recenti, questa maggiore disponibilità di dati e di strumenti di calcolo ha portato all'evoluzione di molteplici metodi di analisi automatizzata dei testi, che possono essere utilizzati in base alle esigenze dei ricercatori e alla tipologia dei documenti analizzati.

L'analisi automatizzata dei testi è quindi finalizzata ad individuare la presenza di pattern ricorrenti e di variabili latenti all'interno di un insieme di testi, detto *corpus*. Questi elementi latenti possono essere di vario tipo e, nel caso specifico di questo lavoro, riguardano il posizionamento dei testi analizzati lungo un asse ideologico, che verrà individuato ex-post, per verificare la validità per la stampa italiana del modello proposto da Mullainathan e Shleifer ed analizzato nel capitolo precedente.

Per individuare lo strumento di analisi automatizzata dei testi più adatto allo scopo di questo lavoro, è possibile partire da una breve classificazione dei metodi di analisi presenti in letteratura. A tal proposito, Grimmer e Stewart (2013) propongono un'analisi dei principali metodi di analisi automatizzata dei testi, ripresa anche in Ceron et al. (2013).

Prima di passare in rassegna i vari metodi di analisi del testo, è però necessario elencare quattro principi di base dell'analisi testuale presentati dagli autori:

- Tutti i metodi quantitativi sono sbagliati, ma qualcuno di essi può essere utile: i processi mentali che portano alla formazione di un discorso o di un testo sono estremamente complessi e non possono mai essere completamente codificati. Ogni lingua, inoltre, presenta sfumature e ambiguità che fanno sì che frasi apparentemente molto simili abbiano in realtà significati diversi o addirittura opposti (per l'italiano, si pensi al classico esempio della parola "pesca", che sta ad indicare sia il frutto che

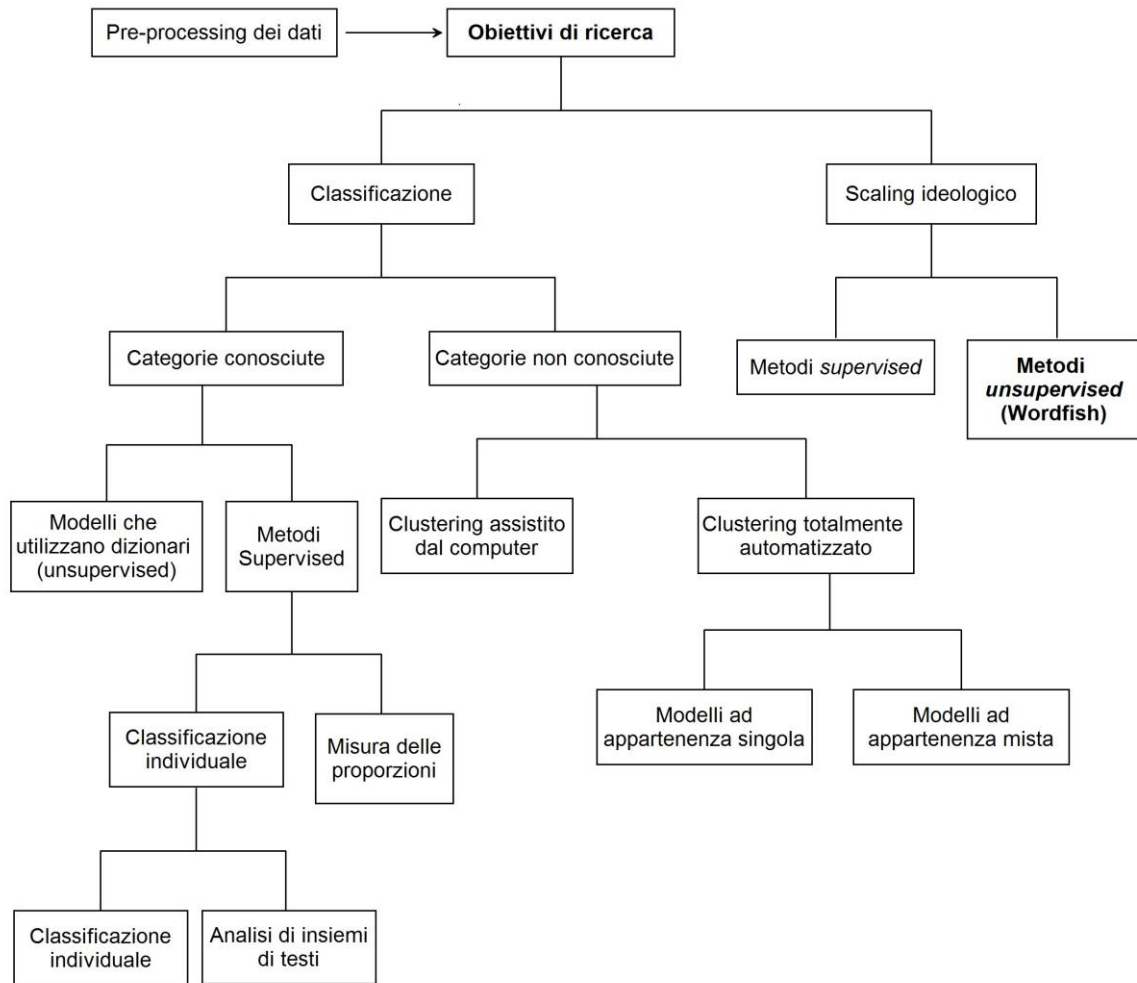
la terza persona del participio presente del verbo pescare; in inglese, la parola “fly” può indicare l’insetto mosca o il verbo “volare”). Questa estrema complessità fa sì che qualsiasi metodo o algoritmo che tenti di codificare il processo generativo di un testo scritto, o della trascrizione di un documento orale, sarà quindi sempre inaccurato. Per ciò che riguarda l’analisi del testo, inoltre, aggiungere più variabili al modello utilizzato non produrrà necessariamente strumenti di inferenza più accurati, così come la riduzione del numero degli elementi all’interno del modello non porterà sempre a modelli più efficienti.

I modelli di analisi automatizzata del testo, quindi, devono essere valutati in base alla loro capacità di fornire risultati utili a seconda del tipo di ricerca che si sta compiendo.

- I metodi quantitativi di analisi testuale aumentano gli strumenti disponibili per i ricercatori, non li sostituiscono: anche se i metodi di analisi del testo si sono dimostrati utili in molteplici situazioni, non possono sostituire il lavoro del ricercatore, sia esso precedente (la scelta del modello da utilizzare, la scelta e la categorizzazione dei testi con cui istruire gli algoritmi utilizzati) o successivo all’analisi (interpretazione dei risultati, validazione del modello)
- Non esiste un metodo di analisi automatizzata dei testi migliore in assoluto: diversi dataset e diversi interessi di ricerca richiedono differenti modelli di analisi testuale. Nel caso trattato in questo lavoro, è necessario collocare i testi lungo una scala di valori, e per questo motivo è stato scelto l’algoritmo Wordfish, elaborato dagli autori proprio per questo scopo. Ogni ipotesi di ricerca, quindi, richiede l’utilizzo di modelli (o di famiglie di modelli) adeguati, così come diversi metodi di validazione dei risultati.
- È fondamentale validare il modello utilizzato, in quanto, a seconda dei dati analizzati e dello scopo della ricerca, è possibile che modelli, seppur formalmente corretti, portino a risultati fuorvianti o errati. I metodi di validazione sono differenti a seconda dei metodi utilizzati e dei risultati che si ricercano.

Tenendo presente questi quattro principi, Grimmer e Stewart presentano uno schema dei diversi modelli di analisi automatizzata dei testi, mostrando le differenze in termini di tipologie di documenti e di tipo di dati disponibili.

Figura 3.1: le principali categorie di metodi automatici di analisi del testo



Fonte: tradotto da Grimmer e Stewart (2013)

Il passaggio precedente alla scelta del modello da utilizzare consiste nel *pre-processing* dei testi: queste operazioni preliminari possono essere utili per rendere più veloci ed efficienti le elaborazioni. L'operazione più importante è quella dello "stemming", che consiste nel ridurre le parole del testo alle loro radici semantiche, al fine di semplificare le analisi (si ha un numero minore di parole da inserire nei modelli) ed evitare ripetizioni che potrebbero falsare i risultati o quantomeno rendere più complessa la stima dei parametri da individuare.

Una prima distinzione è tra metodi che classificano testi in categorie e metodi che mirano a collocare i testi lungo una scala di valori. Questi modelli comprendono metodi *supervised*, che richiedono l'attribuzione di un contenuto semantico definito in precedenza ai testi di un corpus (Ceron et al. 2013, pag. 34) e metodi *unsupervised*, che non richiedono la definizione

ex-ante dei contenuti semantici dei testi del corpus. L'algoritmo Wordfish ricade in quest'ultima categoria.

I modelli che hanno come scopo la classificazione dei testi, inoltre, si dividono in modelli utilizzabili quando le categorie sono conosciute e modelli per l'individuazione di categorie sconosciute. La trattazione approfondita di queste famiglie di modelli esula dallo scopo di questo lavoro; è necessario solamente constatare che anche questi modelli possono essere suddivisi in *supervised* e *unsupervised*.

### 3.2 Un approccio scalare alla categorizzazione dei testi: l'algoritmo Wordfish

Questo algoritmo, finalizzato all'analisi di testi non strutturati<sup>12</sup>, è stato presentato da Slapin e Proksch nel 2008 nel lavoro dal titolo *A scaling model for estimating time-series party positions from texts* e nasce per fornire uno strumento che permetta di stabilire in modo efficace la posizione politica di uno o più partiti, considerando anche le incertezze collegate a questa classificazione; il posizionamento viene calcolato analizzando la frequenza delle parole contenute nei loro manifesti elettorali.

L'unico requisito per poter utilizzare Wordfish è che il lessico dei documenti analizzati rimanga sufficientemente stabile durante l'arco temporale considerato e che i termini utilizzati, quindi, mantengano lo stesso significato durante tutto il periodo considerato. Ciò rende possibile l'utilizzo di questo algoritmo, anche per la classificazione di articoli di quotidiani. L'algoritmo Wordfish, quindi, si basa su un approccio scalare all'analisi dei testi, che ha l'obiettivo di classificare i documenti lungo un asse ideologico che il ricercatore dovrà identificare a posteriori.

Gli autori identificano tre punti forti di questo modello:

- Produce stime partendo da serie storiche, laddove altri algoritmi come Wordscores hanno più difficoltà a produrre risultati utilizzabili;

---

<sup>12</sup> Testi che non presentano strutture predefinite che consentono di ritrovare un certo tipo di dato sempre nella stessa posizione del documento. I testi come gli articoli di giornale, i libri e i post sui social network rientrano in questa tipologia di documenti.

- Non richiede l'utilizzo di testi di riferimento da cui iniziare le analisi, in quanto ricava il posizionamento dell'articolo dalla distribuzione statistica delle parole che lo compongono;
- Utilizza tutte le parole dei documenti e calcola l'importanza di ognuna di essa: ciò rende potenzialmente superflue diverse operazioni di pre-trattamento del testo quali l'eliminazione dei nomi propri, dei numeri ecc.

Il concetto fondamentale proposto dagli autori è quello che, nonostante il processo di creazione dei testi sia estremamente complesso, sia possibile svolgere analisi empiriche su questi ultimi utilizzando un approccio bayesiano naif, che presuppone un modello di probabilità basato sull'indipendenza della posizione delle singole parole da quella delle altre che compongono il testo. Questa teoria, anche se non sempre applicabile alla realtà, si è dimostrata utile ed efficace per la classificazione automatica dei testi (McCallum e Nigam, 1998).

I testi vengono quindi rappresentati come vettori di valori che stanno ad indicare il numero delle volte che una parola vi compare, indipendentemente dalla sua posizione nel testo: si tratta perciò di una *bag of words*.

A questo punto la frequenza delle parole all'interno dei testi è analizzata assumendo che queste ultime si distribuiscano all'interno della *bag of words* con la forma di una distribuzione di Poisson. Ciò significa che il numero delle volte in cui il media  $i$  cita la parola  $j$  all'interno dei suoi articoli è derivato da una distribuzione di questo tipo. La distribuzione di Poisson offre il vantaggio di avere un solo parametro  $\lambda$  che rappresenta sia la media che la varianza ed è quindi possibile stimarla in modo più efficiente rispetto ad altre distribuzioni statistiche.

La funzione del modello è quindi la seguente:

$$y_{ijt} \sim \text{Poisson}(\lambda_{ijt})$$

$$\lambda_{ijt} = \exp(\alpha_{it} + \psi_j + \beta_j * \omega_{it})$$

Dove  $y_{ijt}$  rappresenta il numero di volte in cui la parola  $j$  compare nel documento  $i$  al tempo  $t$ ,  $\alpha_{it}$  indica gli effetti fissi relativi alla lunghezza del documento  $i$  scritto al tempo  $t$ ;  $\psi_j$  rappresenta il valore dell'effetto fisso della parola  $j$  ed è legato alla frequenza con la quale

quest'ultima compare nei testi del corpus;  $\beta_j$  indica il peso specifico della parola  $j$  e la sua importanza per la classificazione dei testi e  $\omega_{it}$  è la stima del posizionamento del soggetto  $i$  nel periodo  $t$ . I testi analizzati, pertanto, verranno classificati stimando i valori di  $\beta_j$  e  $\omega_{it}$ .

Wordfish considera ogni documento come espressione di una singola posizione lungo l'asse delle opinioni e stima tutte le posizioni presenti simultaneamente: ciò significa che le posizioni espresse dai singoli testi pubblicati al tempo  $t$  non dipendono dalle posizioni politiche espresse al tempo  $t-1$ . Per questo motivo, un partito o un giornale che mantiene costanti le proprie posizioni tenderà ad usare le parole con frequenze relative simili, così come in caso di avvicinamento delle posizioni, le frequenze relative delle parole utilizzate dai due schieramenti tenderanno ad assumere valori sempre più vicini tra loro.

La stima del modello, che riguarda tutta la parte destra dell'equazione, viene calcolata utilizzando un algoritmo di Expectation-Maximization, o EM, che richiede cinque step sequenziali. Questo algoritmo è composto da due parti: la parte E (expectation) calcola il valore delle variabili latenti che sottostanno ai testi analizzati, mentre la parte M (maximization) massimizza la log-verosimiglianza condizionata alle aspettative.

I passaggi necessari per la stima del modello sono i seguenti:

1. Calcolo dei valori iniziali: il valore iniziale degli effetti fissi delle parole ( $\psi_j$ ) è ottenuto calcolando la media logaritmica<sup>13</sup> del numero delle volte in cui ogni parola compare nei documenti. Il valore dell'effetto fisso legato alla lunghezza dei documenti ( $\alpha_{it}$ ) è calcolato come il rapporto logaritmico tra il numero medio di parole di ogni documento e il numero di parole contenuto nel documento pubblicato dal primo giornale nel primo periodo analizzato ( $\alpha_{11}$ ): in questo modo si ottiene un valore iniziale di  $\alpha_{it}$  pari a zero. Infine, per ottenere i valori iniziali dei pesi delle parole ( $\beta_j$ ) e della posizione del partito/giornale utilizzando la frequenza relativa delle parole ( $\omega_{it}$ ) si sottrae il valore iniziale di alfa e psi dal logaritmo delle frequenze delle parole contenute nei documenti, quindi si vettori singolari destri e vettori singolari sinistri derivanti dalla scomposizione ai valori singolari della matrice del conteggio delle parole come valori iniziali di  $\omega_{it}$  e  $\beta_j$ .

---

<sup>13</sup>La formula della media logaritmica è  $\bar{x}_{log} = e^{\frac{1}{n} \sum_{i=1}^n \ln(x_i)}$

2. Stima dei parametri relativi al posizionamento dei partiti ( $\omega_{it}$  e  $\alpha_{it}$ ):

$\omega_{it}$  e  $\alpha_{it}$  sono stimati condizionatamente ai valori attesi dei parametri riferiti al ruolo delle parole contenute nei documenti analizzati che, nella prima iterazione, hanno un valore pari al loro valore iniziale. Bisogna pertanto massimizzare la log-verosimiglianza che segue, relativa a ogni documento  $i$  per ogni periodo  $t$  considerato:

$$\sum_{j=1}^m (-\lambda_{ijt} + \ln(\lambda_{ijt}) * y_{ijt})$$

Con:

$$\lambda_{ijt} = \exp(\alpha_{it} + \psi_j^{iniz} \beta_j^{iniz} \omega_{it})$$

Dove i valori iniziali di  $\omega_{it}$  e  $\alpha_{it}$  sono i valori di partenza dello step di massimizzazione. Per stimare il modello, inoltre, si pone  $\alpha_1$  uguale a zero, la media di tutte le posizioni degli attori pari a zero e la deviazione standard pari a 1: questa soluzione permette di considerare i cambiamenti delle posizioni politiche nel tempo, considerando la varianza delle posizioni nel tempo fissa.

3. Stima dei parametri relativi alle parole ( $\beta_j$  e  $\psi_j$ ):

i parametri legati al peso delle parole sono stimati condizionatamente ai valori attesi di quelli relativi al posizionamento dei documenti ( $\alpha_{it}$  e  $\omega_{it}$ ) ottenuti nel passaggio precedente. Per ogni parola  $j$  viene massimizzata la log-verosimiglianza seguente:

$$\sum_{it=1}^n (-\lambda_{ijt} + \ln(\lambda_{ijt}) * y_{ijt})$$

Con:

$$\lambda_{ijt} = \exp(\alpha_{it}^{step2} + \psi_j + \beta_j * \omega_{it}^{step2})$$

4. Calcolare la log-verosimiglianza complessiva:

la log-verosimiglianza complessiva del modello è costituita dalla somma delle log-verosimiglianze delle singole parole, che a loro volta dipendono dai parametri legati alle caratteristiche del partito/giornale. Si ha quindi:



$$\sum_{j=1}^m \sum_{it=1}^n (-\lambda_{ijt} + \ln(\lambda_{ijt}) * y_{ijt})$$

5. Ripetere i passaggi 2-4 fino alla convergenza:

utilizzando i nuovi valori si ricalcolano i valori attesi dei parametri  $\alpha_{it}$  e  $\omega_{it}$  poi, successivamente, i parametri relativi alle parole,  $\beta_j$  e  $\psi_j$ ; questo passaggio viene ripetuto fino ad ottenere un livello accettabile di convergenza, misurato attraverso la differenza tra i valori della log-verosimiglianza ottenuti nel quarto passaggio tra l'iterazione corrente e quella precedente.

Si passa quindi al calcolo degli intervalli di confidenza delle stime dei parametri, che viene effettuato attraverso un bootstrap parametrico<sup>14</sup>. Una volta applicato l'algoritmo EM a tutti gli elementi del dataset per ricavare il valore di  $\lambda_{ijt}$ , vengono generati 500 nuovi dataset, costituiti da valori casuali estrapolati da una distribuzione di Poisson con parametro  $\lambda_{ijt}$ . Utilizzando quindi le stime date dal modello EM come valori di partenza, si stimano 500 nuovi parametri relativi al posizionamento dei documenti sulla scala delle opinioni. Vengono utilizzati i risultati compresi tra i quantili 0,025 e 0,975 per ottenere un intervallo di confidenza del 95%.

---

<sup>14</sup> Procedura di ricampionamento con ripetizione che consente di ricavare dalla distribuzione empirica del campione un certo numero di campioni da cui ricavare una stima del parametro da analizzare, dalle quali ottenere misure di variabilità quali errore standard, distorsione e intervalli di confidenza.

## Parte 2 - La stampa italiana e il mercato dell'informazione: due casi recenti

---

## 4. Il modello Wordfish in pratica

---

Come descritto nel capitolo precedente, il modello Wordfish, utilizzato per l'analisi di testi non strutturati, si basa su due assunti principali: la posizione delle parole all'interno di un testo non influenza la posizione delle altre (ogni testo può quindi essere trattato come un insieme di parole indipendenti tra loro, definito come *bag of word*) e il numero delle volte in cui le parole si presentano all'interno di un testo può essere descritta con una distribuzione di Poisson. Questa particolare distribuzione permette, attraverso la stima di un singolo parametro  $\lambda$ , di analizzare allo stesso tempo la media e la varianza della distribuzione dei dati da analizzare.

Ogni testo, pertanto, viene descritto come un vettore che contiene il numero delle volte in cui le singole parole vi compaiono. La matrice composta da questi vettori, che rappresentano ciascuno un documento del corpus di testi considerati, costituisce l'oggetto dell'analisi, definito Word Frequency Matrix<sup>15</sup>, che viene poi processato in R, utilizzando il pacchetto *austin* (Lowe, 2011).

Tabella 4.1: esempio di Word Frequency Matrix

WFM	influenza	suina	...
Testo 1	24	4	...
Testo 2	15	0	...
...	...	...	...

Lo scopo del modello è quello di ordinare i testi analizzati collocandoli su una dimensione ideologica ipotizzata, ex-ante o ex-post, dal ricercatore (ad esempio la classica distinzione destra-sinistra nel caso di documenti prodotti da soggetti politici, oppure pro-contro una determinata idea). L'individuazione della posizione dei documenti analizzati si basa sulla frequenza relativa con cui le singole parole compaiono all'interno dei singoli testi ed è tanto più accurata quanto più è grande il numero di parole contenuto nel corpus che si vuole analizzare.

Il modello Wordfish, inoltre, presenta diversi vantaggi rispetto ad altri metodi di analisi automatica dei testi: può produrre serie storiche di stime, analizzando la variazione nel tempo del posizionamento dei documenti, non richiede l'utilizzo di testi campione, in quanto si

---

<sup>15</sup> La word frequency matrix è stata estratta utilizzando l'applicazione JFreq di Will Lowe (Lowe 2012)

basa sull'assunzione che la frequenza con cui le parole compaiono nei documenti del corpus analizzato abbia una distribuzione statistica ben definita (in particolare, una distribuzione di Poisson) e utilizza tutte le parole presenti nel corpus per l'analisi del posizionamento dei documenti, stimando l'importanza di ognuna di esse.

Per poter utilizzare questo algoritmo, tuttavia, è necessario fissare alcune ipotesi a priori, che il ricercatore deve tenere in considerazione al momento di selezionare i testi da analizzare, proprio a causa del fatto che Wordfish non viene istruito attraverso documenti di riferimento. I testi utilizzati devono riguardare esplicitamente l'argomento trattato (nel caso dei quotidiani gli articoli selezionati devono trattare esplicitamente l'argomento considerato, come ad esempio la reazione all'arresto di Berlusconi per il caso Ruby) e il significato delle parole utilizzate deve rimanere lo stesso in tutti i periodi temporali considerati.

La distribuzione delle parole all'interno dei testi è stimata attraverso la massimizzazione della funzione di log-verosimiglianza:

$$y_{ijt} \sim \text{Poisson}(\lambda_{ijt})$$
$$\lambda_{ijt} = \exp(\alpha_{it} + \psi_j + \omega_{it} * \beta_j)$$

dove  $y_{ijt}$  corrisponde al numero delle volte in cui la  $j$ -esima parola compare nell' $i$ -esimo documento scritto al tempo  $t$ .

Come si può notare dalla formula, la distribuzione alla base di questo modello si basa sulla stima di quattro parametri, due relativi alle singole parole (definiti  $\beta_j$  e  $\psi_j$ ), che rimangono fissi nel tempo<sup>16</sup>, e due relativi ai singoli documenti presenti nel corpus analizzato ( $\alpha_{it}$  e  $\omega_{it}$ ). I parametri utilizzati per collocare i documenti nelle diverse posizioni sono due,  $\omega$  e  $\beta$ .

Entrando più nel dettaglio:

- $\alpha_{it}$ : è il parametro che considera l'effetto fisso legato alla lunghezza dei documenti prodotti nei diversi istanti di tempo. Viene calcolato come il logaritmo del rapporto tra il numero medio del conteggio delle parole di ogni documento  $i$  al periodo  $t$  e lo stesso valore relativo al primo documento del corpus, in modo da avere il primo valore di  $\alpha_{it}$  pari a zero. Questo parametro controlla il fatto che alcuni documenti presenti nel corpus possano essere considerevolmente più lunghi di altri.

---

<sup>16</sup> Ciò è dovuto al fatto che, come da ipotesi del modello, il significato delle parole rimane lo stesso durante tutto il periodo di tempo considerato.

- $\psi_j$ : questo parametro considera l'effetto fisso legato alla frequenza con cui la  $j$ -esima parola viene utilizzata nei diversi documenti e se alcune parole vengono utilizzate più frequentemente di altre, indipendentemente dal documento. Il suo valore iniziale viene calcolato come la media logaritmica del numero di volte in cui ogni parola  $j$  compare in tutti i documenti presenti nel corpus.
- $\beta_j$ : rappresenta il peso specifico di ogni singola parola  $j$  e, con il suo segno, ne descrive l'importanza relativa nel distinguere il posizionamento dei diversi documenti. Per far sì che il valore di  $\beta$  sia utilizzabile, viene posto un vincolo tale per cui  $\beta$  è distribuito come una normale con media 0 e deviazione standard  $\sigma$ . In questo modo si evita che il peso delle parole assuma valori infiniti e riduce il peso dato alle parole usate meno di frequente in tutti i documenti o degli eventuali refusi o errori di battitura.
- $\omega_{it}$ : è il parametro di maggiore interesse, in quanto rappresenta il posizionamento dei documenti presenti nel corpus e dipende dalla frequenza con cui le diverse parole sono presenti all'interno di ogni documento analizzato.

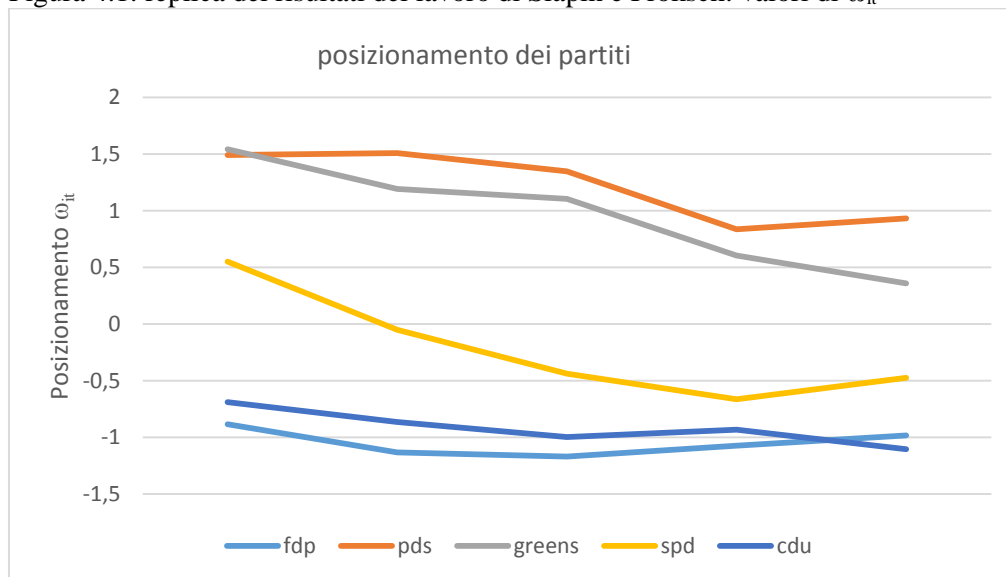
Come si è visto nel capitolo precedente, per ottenere la stima dei parametri si utilizza un algoritmo Expectation-Maximization, o E-M, che, attraverso un procedimento iterativo calcola la massima verosimiglianza delle variabili latenti del modello. Questo algoritmo è diviso in due parti: lo step E, expectation, calcola i valori attesi dei parametri da stimare, mentre lo step M, maximization, massimizza la funzione di log-verosimiglianza, i cui parametri sono condizionati ai valori calcolati nello step E.

Le stime dei valori iniziali di  $\omega_{it}$  e  $\beta_j$  sono calcolate in due step, partendo dai valori iniziali di  $\alpha_{it}$  e  $\psi_j$ . Per prima cosa si sottrae il valore di partenza del parametro  $\psi_j$ , relativo alla frequenza media con cui le parole vengono usate nei documenti del corpus, e del parametro  $\omega_{it}$ , relativo alla lunghezza dei testi, dal logaritmo della frequenza con cui le parole vengono utilizzate nei documenti. A questo punto si esegue una decomposizione ai valori singolari della matrice risultante e, come valori iniziali di  $\omega_{it}$  e  $\beta_j$  si utilizzano i vettori singolare sinistro e singolare destro derivati dalla decomposizione.

Il grafico seguente rappresenta la replica dei risultati del lavoro di Slapin e Proksch, con i dati messi a disposizione dagli autori, per ciò che riguarda il posizionamento dei documenti analizzati, che consistono nei programmi elettorali dei maggiori partiti tedeschi. Gli autori

usano come testi iniziali, per calcolare i valori di partenza del modello, i documenti prodotti dal partito FDP e dal partito dei Verdi del 1994. Per la precisione, il valore di  $\omega_{it}$  dei manifesti elettorali dell’FDP viene posto forzatamente maggiore del valore di  $\omega_{it}$  dei manifesti dei Verdi e, successivamente, vengono stimati i valori dei quattro parametri del modello

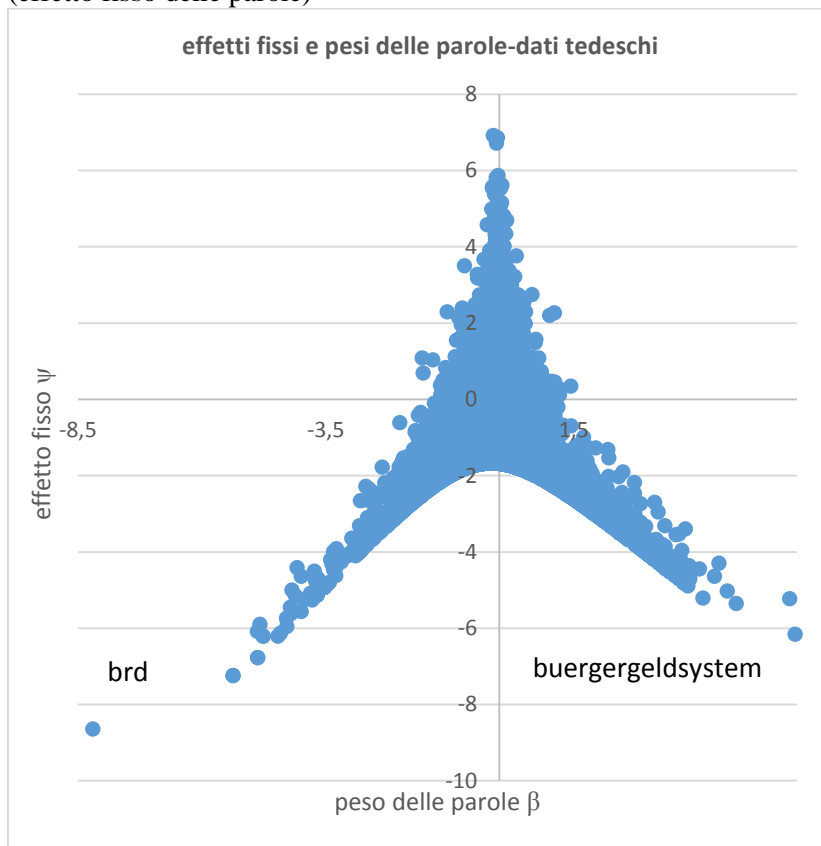
Figura 4.1: replica dei risultati del lavoro di Slapin e Proksch: valori di  $\omega_{it}$



Fonte: elaborazione personale su dati forniti dagli autori

Una volta ottenute le stime di questi quattro parametri è possibile individuare sia il posizionamento sull’asse ideologico individuato dei soggetti coinvolti (nel nostro caso i giornali) analizzando i valori assunti da  $\omega$ , sia quali sono le parole che maggiormente contraddistinguono le due parti in causa. Quest’ultima analisi viene svolta attraverso il grafico che mette in relazione i valori assunti dai parametri relativi alle parole, ossia  $\beta_j$  e  $\psi_j$ . Mettendo in relazione il peso di ogni singola parola con il valore della media logaritmica del conteggio delle volte in cui questa compare all’interno di tutti i documenti (rappresentato da  $\psi_j$ ) si possono individuare i termini che più caratterizzano gli schieramenti, che sono quelle situate alla base del grafico (III e IV quadrante), il quale risulta avere una forma simile a quella della Tour Eiffel.

Figura 4.2: replica dei risultati del lavoro di Slapin e Proksch: valori di  $\beta_j$  (peso delle parole) e  $\psi_j$  (effetto fisso delle parole)



Fonte: elaborazione personale su dati forniti dagli autori

Questo grafico evidenzia due aspetti della distribuzione delle parole all'interno del corpus. Il primo aspetto è dato dal fatto che le parole più frequenti, come le congiunzioni e gli articoli, non contribuiscono a definire il posizionamento degli articoli, non avendo un significato politico preciso: esse sono posizionate nella parte centrale e superiore del grafico, in quanto hanno un valore di  $\psi_j$  (effetto fisso) elevato associato ad un peso  $\beta_j$  prossimo allo zero. Il secondo aspetto è costituito dal fatto che le parole usate meno di frequente, e che pertanto sono più indicative del posizionamento del documento in cui sono presenti e hanno con più probabilità un qualche significato politico, sono collocate alla base del grafico, più precisamente nei quadranti III e IV, dato che presentano un basso valore di  $\psi_j$  associato a valori positivi o negativi di  $\beta_j$ , che definiscono il ruolo della parola  $j$  nel collocare l'articolo lungo lo spettro ideologico che sta alla base della distinzione tra i documenti del corpus analizzato.

Dopo aver analizzato il funzionamento dell'algoritmo Wordfish applicato a documenti reali quali i programmi elettorali di alcuni partiti tedeschi, si utilizzerà questo strumento per

descrivere il comportamento di alcuni influenti quotidiani italiani. Wordfish verrà utilizzato per capire come i quotidiani italiani presi in considerazione hanno descritto due eventi di vasta portata che hanno interessato il nostro paese (e non solo): l'epidemia di influenza H1N1 del 2009, che ha scatenato reazioni di panico molto spesso ingiustificate, e la gestione dei flussi migratori provenienti dalla sponda sud del Mediterraneo. Quest'ultimo evento sarà analizzato per ciò che riguarda il mese di ottobre del 2013, quando, a seguito del naufragio di un barcone carico di migranti, morirono centinaia di persone al largo delle coste dell'isola di Lampedusa; questo fatto aprì un dibattito nel paese e portò una parte dell'opinione pubblica e dei soggetti politici a chiedere una profonda revisione delle leggi che regolavano l'accoglienza dei migranti, con particolare riferimento all'abolizione della legge 30 luglio 2002, n. 189, conosciuta come "Bossi-Fini", dal nome dei due promotori.



## 5. C'è sempre polarizzazione? Due esempi dalla stampa italiana

---

### 5.1 I flussi migratori provenienti dalla sponda sud del Mediterraneo e le reazioni di due importanti quotidiani italiani

Il tema dell'immigrazione di massa è un argomento presente nel dibattito pubblico italiano già da diversi anni, in particolare da quando, a metà anni '90, imponenti flussi di migranti provenienti dai Balcani, ed in particolar modo dall'Albania, cominciarono ad arrivare sulle coste adriatiche. Negli anni il fenomeno migratorio è profondamente mutato, a causa delle diverse condizioni internazionali: negli ultimi anni si è passati da flussi di migranti che arrivavano sulle coste italiane partendo dall'Europa dell'Est a flussi migratori che partono dalle coste del nord-Africa, e sono costituiti da migranti provenienti da Medio Oriente, Africa settentrionale e Africa Sub-Sahariana.

La gestione dei flussi migratori e la gestione delle emergenze ad essi legate si presenta ciclicamente sulla stampa italiana, sia in caso di sbarchi che non causano incidenti e vittime, sia in concomitanza con eventi drammatici, quali i naufragi di barconi al largo delle coste italiane. Essendo il tema migrazioni molto sentito da parte dei cittadini e fortemente legato a istanze politiche ed economiche, è legittimo aspettarsi una polarizzazione nella copertura data dai mass media agli eventi ad esso legati, delle interpretazioni date al fenomeno e delle soluzioni proposte dai giornalisti, dai politici e dagli addetti ai lavori.

Per verificare l'effettiva presenza di una polarizzazione tra l'offerta di notizie da parte dei quotidiani italiani, verranno analizzati gli articoli a tema immigrazione pubblicati da due quotidiani di area politica opposta, *La Repubblica* e *Liberò* nel mese di ottobre del 2009.

Questo periodo è stato scelto in quanto, il 3 ottobre 2009, un'imbarcazione con a bordo oltre 500 migranti, un carico nettamente superiore alla capacità del mezzo, naufraga al largo delle coste dell'Isola dei Conigli, situata in prossimità dell'isola di Lampedusa, in acque territoriali italiane, causando un numero di vittime altissimo: 366 morti accertate e 20 dispersi.

Questo naufragio ha dato origine ad un ampio dibattito politico, riguardo all'opportunità di modificare la normativa sull'immigrazione, la cosiddetta legge Bossi-Fini (legge 30 luglio

2002, n. 189), ritenuta dai suoi detrattori troppo restrittiva, ed ha dato origine ai negoziati per l'istituzione dell'operazione *Mare Nostrum*, missione di salvataggio in mare dei migranti in transito nel Canale di Sicilia, provenienti dalle coste libiche diretti verso i territori italiano e maltese, attuata dalle forze della Marina Militare dell'Aeronautica Militare italiane. Proprio il dibattito sull'opportunità o meno di mantenere in vigore la legge Bossi-Fini ha generato forti contrapposizioni tra le forze politiche e di conseguenza, tra i mezzi di comunicazione di massa, che hanno a loro volta preso posizione pro o contro l'abolizione della normativa.

Nei paragrafi successivi si passerà all'analisi della copertura mediatica data da *La Repubblica* e *Libero* al problema immigrazione nel periodo sopra citato.

Analizzando il numero di parole dedicato dai due quotidiani al tema immigrazione è possibile farsi una prima idea dell'intensità con cui il tema è stato trattato dalle due testate concorrenti, per capire fin da subito se possono esserci differenze marcate tra loro.

I dati raccolti dimostrano come, con l'eccezione del 3 ottobre, il quotidiano che ha riservato il maggior numero di parole al tema immigrazione sia *La Repubblica*, mentre *Libero* si è mantenuto su numeri di parole spese quotidianamente sui temi dei migranti sempre inferiori a quelli del quotidiano concorrente. Questa differenza è particolarmente evidente nei giorni che vanno dal 7 al 9 ottobre, periodo immediatamente successivo al naufragio che ha causato centinaia di morti e ha dato il via a un dibattito pubblico sulla regolamentazione nazionale dei fenomeni migratori, con le due testate che si caratterizzano per posizioni molto distanti: *La Repubblica* mette in discussione la normativa vigente e promuove una raccolta firme per chiedere l'abolizione della legge "Bossi-Fini", mentre *Libero* si schiera a difesa di questa norma.

Figura 5.1: numero di parole utilizzate dai due quotidiani in articoli sull'immigrazione - ottobre 2009.

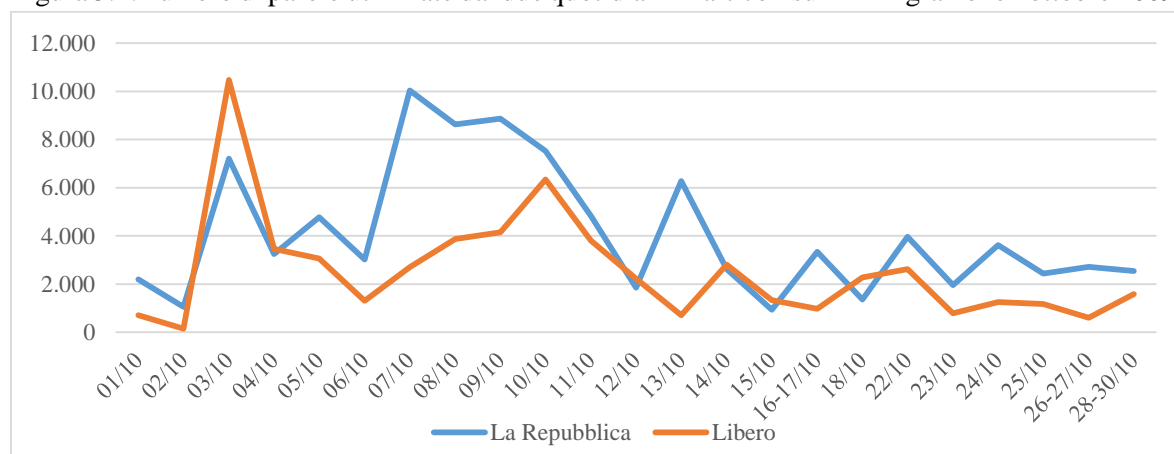


Tabella 5.1: numero di parole dedicate dai due quotidiani al tema immigrazione – Ottobre 2009

<i>Data</i>	<i>La Repubblica</i>	<i>Libero</i>
1 ottobre	2.201	706
2 ottobre	1.055	154
3 ottobre	7.205	10.472
4 ottobre	3.239	3.461
5 ottobre	4.779	3.055
6 ottobre	3.015	1.308
7 ottobre	10.038	2.703
8 ottobre	8.622	3.864
9 ottobre	8.861	4.151
10 ottobre	7.525	6.342
11 ottobre	4.832	3.800
12 ottobre	1.851	2.223
13 ottobre	6.281	708
14 ottobre	2.617	2.807
15 ottobre	929	1.328
16-17 ottobre	3.332	975
18 ottobre	1.360	2.281
22 ottobre	3.965	2.615
23 ottobre	1.960	788
24 ottobre	3.613	1.252
25 ottobre	2.433	1.167
26-27 ottobre	2.710	600
28-30 ottobre	2.536	1.580

### 5.1.1 Il posizionamento dei quotidiani

Dopo aver analizzato l'intensità della copertura mediatica data dai due quotidiani alla problematica dell'immigrazione nel mese di ottobre del 2009, si passa a ricercare le differenze tra le due fonti nell'analisi del fenomeno migratorio, processando attraverso l'algoritmo Wordfish gli articoli che *La Repubblica* e *Libero* hanno dedicato esplicitamente a questa tematica. Come si vedrà in seguito, l'analisi di questa tipologia di testi presenta alcuni elementi di ambiguità, dovuti alla tipologia di articoli che vengono prodotti dai giornali, che possono essere sia di normale cronaca, i quali non danno origine a meccanismi di differenziazione e di concorrenza tra testate, che di opinione, portatori, pertanto, di differenziazione e polarizzazione. Queste ambiguità verranno superate adottando un'analisi dei testi che prevede alcuni passaggi intermedi, allo scopo di individuare i momenti in cui i quotidiani si sono collocati su posizioni differenti.

Gli articoli sono stati estratti dagli archivi on-line delle due testate in base a due criteri: che presentassero al loro interno il termine *immigrazione* e che, ad una successiva analisi,

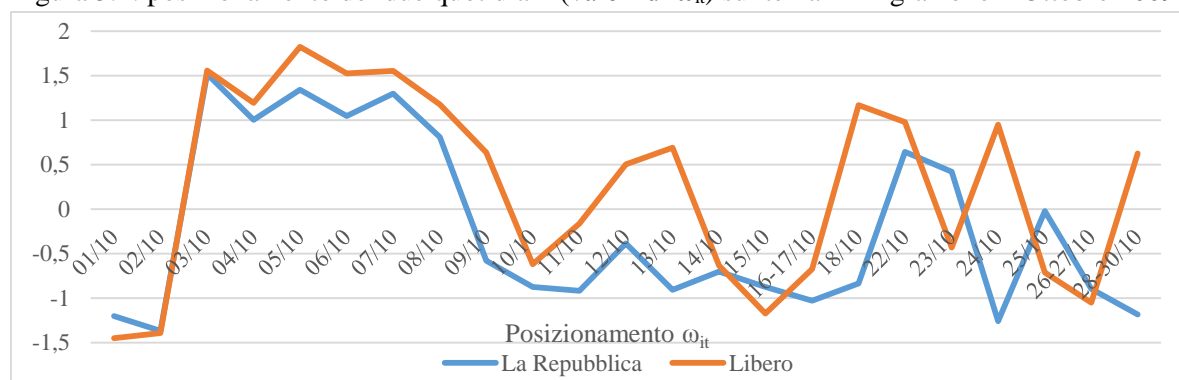
riguardassero esplicitamente il tema. Una volta individuati i testi da utilizzare, essi sono stati raccolti in corpus che comprendono tutti gli articoli pubblicati ogni giorno, tranne nei casi del 16 e 17 ottobre e dei giorni che vanno dal 28 al 30 ottobre, dove gli articoli sono stati accorpati per armonizzare i corpora comprendenti gli articoli dei due quotidiani e per avere testi sufficientemente ampi.

Una volta ottenuti i testi da analizzare si è proceduto all'estrapolazione del parametro  $\omega_{it}$  attraverso l'algoritmo Wordfish. La prima analisi è stata effettuata su tutti e 24 i documenti ed ha prodotto i risultati riportati nella tabella 5.1 e nella figura 5.1.

Tabella 5.2: posizionamento dei due quotidiani sul tema immigrazione – Ottobre 2009

Data	$\omega_{it}$ La Repubblica	$\omega_{it}$ Libero
1 ottobre	-1,201	-1,450
2 ottobre	-1,369	-1,394
3 ottobre	1,517	1,559
4 ottobre	1,005	1,193
5 ottobre	1,342	1,823
6 ottobre	1,048	1,524
7 ottobre	1,299	1,554
8 ottobre	0,810	1,178
9 ottobre	-0,580	0,636
10 ottobre	-0,874	-0,618
11 ottobre	-0,917	-0,163
12 ottobre	-0,385	0,502
13 ottobre	-0,909	0,691
14 ottobre	-0,701	-0,634
15 ottobre	-0,874	-1,173
16-17 ottobre	-1,030	-0,673
18 ottobre	-0,836	1,168
22 ottobre	0,645	0,979
23 ottobre	0,420	-0,434
24 ottobre	-1,259	0,950
25 ottobre	-0,021	-0,715
26-27 ottobre	-0,895	-1,052
28-30 ottobre	-1,183	0,624

Figura 5.2: posizionamento dei due quotidiani (valori di  $\omega_{it}$ ) sul tema immigrazione – Ottobre 2009



Il grafico riportato nella figura 5.2 evidenzia come il posizionamento dei due quotidiani non segua un andamento costante, ma veda l'alternarsi di fasi in cui i testi presentano un valore di  $\omega_{it}$  simile a fasi di profonda differenziazione.

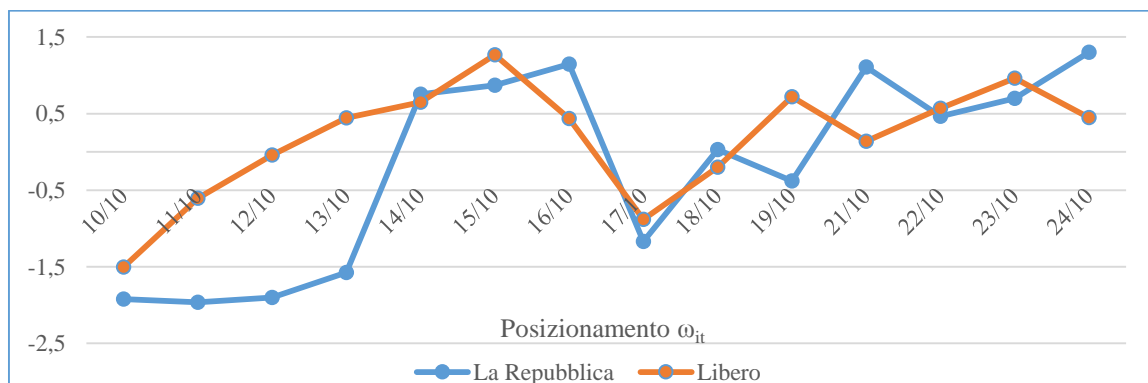
Ciò è dovuto al fatto che, come già affermato in precedenza, gli articoli riguardanti il tema immigrazione sono sia di semplice cronaca sia di commento e di opinione. Come secondo passaggio, pertanto, si è ridotto il numero dei giorni considerati, per individuare meglio i periodi in cui le due testate hanno proposto temi differenti. Per fare ciò sono stati esclusi dall'analisi i giorni che vanno dal primo ottobre al 9 ottobre inclusi, dove le due testate hanno riportato solamente i fatti di cronaca relativi all'immigrazione, compresi quelli sul naufragio del 3 ottobre, e commenti a riguardo che i vari esponenti della politica e delle associazioni hanno rilasciato.

Elaborando i documenti restanti, che vanno pertanto dal 10 al 30 ottobre, si ottengono i valori di  $\omega_{it}$  riportati nella tabella 5.2 e nel grafico riportato nella figura 5.2.

Tabella 5.3: posizionamento dei due quotidiani sul tema immigrazione (valori di  $\omega_{it}$ ) – Ottobre 2009

Data	$\omega_{it}$ La Repubblica	$\omega_{it}$ Libero
10 ottobre	-1,921	-1,506
11 ottobre	-1,965	-0,605
12 ottobre	-1,902	-0,040
13 ottobre	-1,577	0,444
14 ottobre	0,754	0,647
15 ottobre	0,869	1,267
16-17 ottobre	1,145	0,434
18 ottobre	-1,171	-0,880
22 ottobre	0,028	-0,202
23 ottobre	-0,380	0,720
24 ottobre	1,110	0,138
25 ottobre	0,463	0,568
26-27 ottobre	0,700	0,961
28-30 ottobre	1,300	0,446

Figura 5.3: posizionamento dei due quotidiani sul tema immigrazione (valori di  $\omega_{it}$ ) – 10-30 Ottobre 2009



Come si può intuire chiaramente dal grafico riportato nella figura 5.2, gli articoli di *Libero* e *La Repubblica* presentano una marcata differenziazione nel lasso di tempo che va dall'11 al 13 ottobre, mentre dopo quel periodo i due trend tornano ad essere più simili, probabilmente a causa del ripresentarsi di altri eventi che hanno riportato le linee editoriali dei due quotidiani sui binari della cronaca.

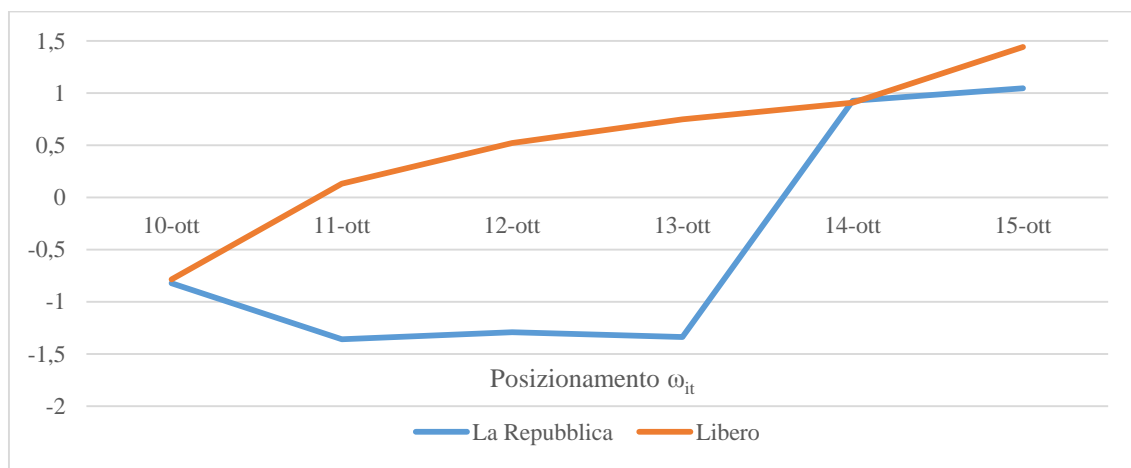
Come ultimo passaggio, pertanto, si sono analizzati i documenti che riuniscono gli articoli dei due quotidiani nei giorni che vanno dal 10 al 15 ottobre.

I valori di  $\omega_{it}$  ottenuti sono riportati nella tabella 5.4 e nel relativo grafico:

Tabella 5.4: posizionamento dei due quotidiani sul tema immigrazione (valori di  $\omega_{it}$ ) – 10-15 Ottobre 2009

Data	$\omega_{it}$ La Repubblica	$\omega_{it}$ Libero
10 ottobre	0,782	0,767
11 ottobre	1,302	-0,152
12 ottobre	1,218	-0,551
13 ottobre	1,543	-0,758
14 ottobre	-0,918	-0,896
15 ottobre	-1,028	-1,421

Figura 5.4: posizionamento dei due quotidiani sul tema immigrazione – (valori di  $\omega_{it}$ ) 10-15 Ottobre 2009



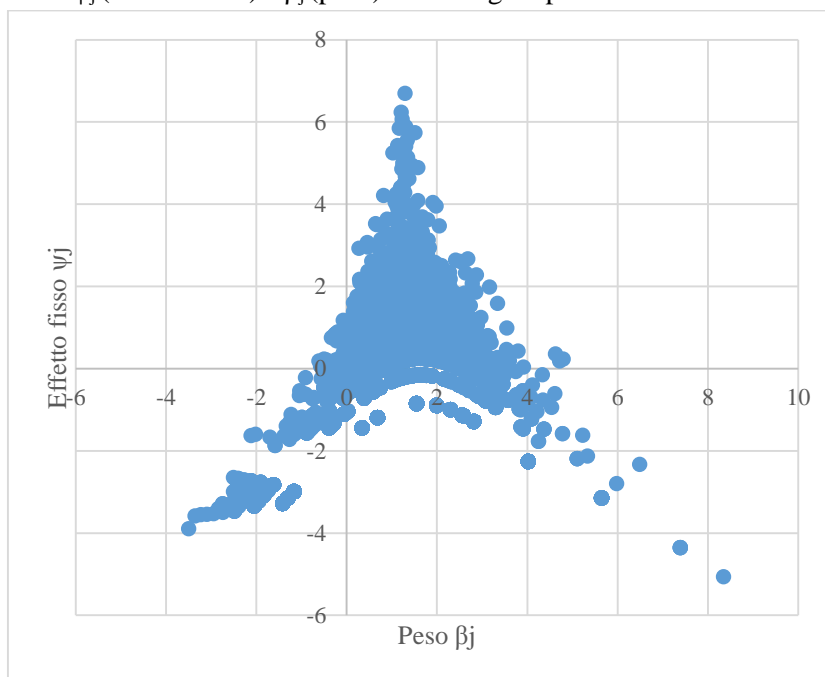
Il grafico in figura 5.4 mostra l'evidente distanza che gli articoli proposti dai due quotidiani nei giorni che vanno dall'11 al 13 ottobre; distanza che, dal 14 ottobre 2009 torna a ridursi, portando il valore di  $\omega_{it}$  di entrambi le testate su valori simili.

Una volta appurata la presenza di una differenza nelle posizioni dei due quotidiani, l'analisi del lessico utilizzato dalle due testate permette di capire quali sono gli elementi che danno origine a tale differenziazione.

## 5.1.2 Il lessico utilizzato

In questa sezione verrà analizzato il lessico utilizzato dai due quotidiani per descrivere i fatti relativi ai fenomeni migratori. Il grafico 5.5 rappresenta i valori di peso  $\beta_j$  e effetto fisso  $\psi_j$  calcolati dal modello Wordfish per ogni parola contenuta nel corpus analizzato.

Figura 5.5 valori di  $\psi_j$  (effetto fisso) e  $\beta_j$  (peso) delle singole parole di tutti i documenti del corpus



Come illustrato nel capitolo 4, le parole che hanno un valore di effetto fisso basso sono quelle che contribuiscono maggiormente a collocare il testo in cui sono presenti su un determinato punto della scala ideologica lungo la quale si dipanano i vari documenti presenti all'interno del corpus analizzato.

Tabella 5.5 - I 15 termini più significativi presenti nei quadranti III e IV

Parole quadrante III	$\beta_j$	$\psi_j$	Parole quadrante IV	$\beta_j$	$\psi_j$
carceraria	-3,497	-3,894	lecce	8,350	-5,065
crescita	-3,358	-3,584	siena	7,395	-4,354
iacchetti	-3,227	-3,552	piccola	7,395	-4,354
giovani	-3,095	-3,541	colonnello	7,395	-4,354
perfino	-2,942	-3,530	leuca	7,395	-4,354
urgente	-2,841	-3,404	otranto	6,489	-2,331
enzo	-2,744	-3,283	ricerche	5,982	-2,799
naufrago	-2,744	-3,283	rimini	5,651	-3,146
urla	-2,736	-3,496	evitata	5,651	-3,146
hollande	-2,649	-3,388	dda	5,651	-3,146
inedito	-2,629	-3,363	roan	5,651	-3,146
venditti	-2,629	-3,363	tenente	5,651	-3,146
vegli	-2,507	-2,651	carburante	5,651	-3,146
continente	-2,506	-2,988	motore	5,651	-3,146

Il valore del peso  $\beta_j$  determinerà poi il ruolo effettivo delle singole parole; a valori negativi di  $\beta_j$  corrisponderanno valori bassi di  $\omega_{it}$  e, viceversa, valori positivi del peso  $\beta_j$  saranno collegati a valori più alti del parametro  $\omega_{it}$ , che determina il posizionamento dei singoli testi.

I primi termini da prendere in considerazione sono quelli che presentano valori di peso  $\beta_j$  ed effetto fisso  $\psi_j$  tale per cui sono collocati nel quadrante III. Queste parole, che più delle altre portano i testi che le contengono ad ottenere valori di  $\omega_{it}$  bassi, sono presenti con una frequenza relativa più alta all'interno degli articoli pubblicati da *La Repubblica* e ne caratterizzano il posizionamento nel periodo di tempo considerato. Queste parole fanno riferimento all'aspetto più "umanitario" del problema immigrazione, con riferimenti anche alla abrogazione della legge "Bossi-Fini". Parole come *urgente*, *naufrago* e *urla*, che registrano valori di  $\beta_j$  tra i 15 più bassi dell'intero corpus, sono riferiti alle condizioni dei migranti giunti sulle coste italiane e alla necessità di intervenire in fretta per evitare che si verificano altre stragi nel Canale di Sicilia. I termini *Iachetti*, *Venditti* e *Hollande* fanno invece riferimento al dibattito sulla legislazione vigente in termini di immigrazione: i primi due sono i cognomi di due artisti italiani che hanno sottoscritto un appello a per la cancellazione della legge "Bossi-Fini", mentre il terzo è il cognome del presidente della Repubblica francese in carica all'epoca, nel cui paese è in vigore una legislazione sull'immigrazione ritenuta molto restrittiva e usata dalle due "fazioni" come termine di paragone con le leggi vigenti in Italia.

Per quanto riguarda le parole collocate nel IV quadrante, che pertanto presentano valori di  $\beta_j$  che spingono verso l'alto il valore del parametro  $\omega_{it}$  dei testi che le contengono, si può notare come siano legati agli aspetti più prettamente di cronaca del fenomeno immigrazione e sono presenti con una frequenza relativa maggiore all'interno degli articoli pubblicati da *Liberò*. Termini come *Lecce*, *Siena*, *Rimini*, *ricerche* e *motore* sono riferiti al racconto dei fatti, spesso sotto forma di rilancio di agenzie di stampa, piuttosto che ad un'analisi critica del fenomeno delle migrazioni e a una presa di posizione verso la normativa vigente (la riproposizione di fatti di cronaca, al contrario, può essere intesa come un invito a mantenere, o a inasprire, le leggi in vigore).

Per concludere, l'analisi della copertura giornalistica data ai temi dell'immigrazione, in un momento storico caratterizzato da un forte dibattito pubblico riguardo a questi temi, dimostra



come i due quotidiani analizzati abbiano seguito linee editoriali differenti e concorrenziali, dando origine ad una forte differenziazione dei contenuti proposti, per andare incontro alle differenti tipologie di domanda dei loro lettori. In particolare, *La Repubblica* si concentra più sugli aspetti socio-politici legati al tema immigrazione, mentre *Liberò* propone articoli più incentrati sui fatti di cronaca riferiti al tema, legati alla difesa della legge Bossi-Fini.

## 5.2 Un esempio di informazione omogenea: l'epidemia di influenza suina raccontata da tre quotidiani italiani

Dopo aver analizzato il comportamento di due influenti testate giornalistiche italiane, che hanno una linea politica dichiaratamente opposta, nell'analizzare una situazione fortemente divisiva come quella legata all'immigrazione si è verificato come i due giornali abbiano assunto, in determinati periodi di tempo, posizioni decisamente diverse tra loro.

Questa polarizzazione non si è verificata nel caso della copertura giornalistica data alla pandemia di influenza A, definita anche "influenza suina" o "febbre suina" che, a partire dalla fine di aprile del 2009, si è diffusa dal Messico a tutto il mondo. L'influenza A, causata da una mutazione del virus influenzale H1N1 e trasmessa all'uomo da alcuni suini allevati nel paese centroamericano, ha dato origine ad una psicosi a livello mondiale, rivelatasi poi priva di basi razionali, in quanto il tasso di mortalità di questa nuova forma influenzale è stato di gran lunga inferiore rispetto, ad esempio, a quello della normale influenza invernale. I dati relativi all'Italia parlano di un tasso di mortalità, definito come il numero di individui uccisi dal virus in relazione all'intera popolazione, pari allo 0,05 per mille<sup>17</sup>, contro il 2 per mille dell'influenza stagionale. Nonostante questi dati, l'attenzione del pubblico verso questa nuova ed "esotica" forma influenzale è stata altissima e ha raggiunto livelli quasi da psicosi, con risultati anche bizzarri, come ad esempio il boom di vendite dei disinfettanti in gel per le mani che, pur non essendo particolarmente efficaci per il contrasto del virus, hanno registrato picchi di vendite durante il periodo di tempo considerato.

Come già anticipato, l'influenza suina ha assunto il carattere di pandemia a partire dall'aprile del 2009 e la copertura mediatica data a questo evento ha registrato quattro picchi del numero

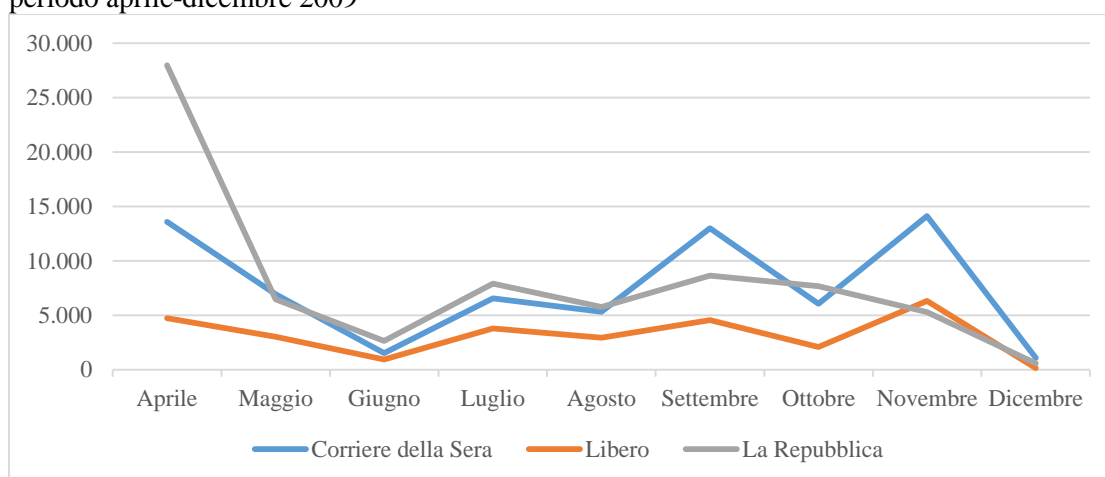
---

<sup>17</sup> Al 4 febbraio 2010 il Ministero della Salute Italiano, con il comunicato n° 27/2010 stima 229 casi di morte dovuti al virus H1N1, a fronte dei 4.391.000 casi di contagio

di notizie ad esso dedicate: questo fenomeno è verificabile osservando il numero di parole che i giornali hanno utilizzato nella stesura di articoli riferiti esplicitamente a tematiche relative all'influenza A.

I tre quotidiani analizzati, *La Repubblica*, *Il Corriere della Sera* e *Libero* si sono occupati, al pari delle altre testate, in modo approfondito della febbre suina, dedicando al tema maggiore attenzione soprattutto in tre periodi: in aprile 2009, quando la febbre suina esce dai confini del Messico e viene riconosciuta come malattia pandemica, a luglio 2009, quando cominciano ad arrivare notizie di italiani che hanno contratto la febbre suina, a settembre 2009, quando i primi casi di morte legati al virus A/H1N1 cominciano a verificarsi anche in Italia e infine a novembre 2009, quando si registrano picchi di contagi da virus A, che si affiancano ai contagi dovuti al virus della normale influenza invernale. L'interesse delle testate nei confronti dell'influenza suina cessa quasi di colpo a dicembre 2009, quando un deciso calo nel numero dei contagi rilevati fa sì che l'interesse dei media nei confronti della (presunta) emergenza si affievolisca decisamente. Nei mesi successivi al dicembre 2009 la presenza sui giornali di articoli relativi al virus A è sporadica (a conferma della teoria della *news spiral*, che prevede un persistente stato di attenzione al termine del picco delle notizie relative ad un'emergenza<sup>18</sup>) e, pertanto, l'analisi del posizionamento dei quotidiani farà riferimento al periodo aprile-dicembre 2009.

Figura 5.6: numero di parole utilizzate dai tre quotidiani in articoli riguardanti la febbre suina nel periodo aprile-dicembre 2009



<sup>18</sup> Come descritto dal modello di Beardsworth e Keil citato nel capitolo 1

Tabella 5.6: numero di parole utilizzate dai tre quotidiani in articoli riguardanti la febbre suina nel periodo aprile-dicembre 2009

Numero di parole	Aprile	Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre
Corriere della Sera	13.588	6.941	1.527	6.562	5.316	13.000	6.058	14.117	1.066
Libero	4.720	3.029	918	3.782	2.936	4.551	2.078	6.317	133
La Repubblica	27.991	6.458	2.653	7.920	5.770	8.643	7.682	5.299	577

I dati sopra riportati dimostrano che il quotidiano che ha dedicato più parole all'argomento "Influenza A" è *La Repubblica*, forte anche del fatto di avere un'edizione cartacea costituita da un maggior numero di pagine rispetto a quello delle altre due testate. *Libero* e *Il Corriere della Sera* dedicano un numero di parole decisamente inferiore all'argomento, anche se il trend del numero di parole spese da queste due testate per le notizie riguardanti la febbre suina è simile a quello di Repubblica, come si può vedere dalla figura 5.6.

Una volta analizzata l'entità e la durata della copertura mediatica che i tre quotidiani hanno riservato al tema dell'influenza suina, si può passare a studiarne il posizionamento ed il lessico utilizzato nel corso del tempo.

### 5.2.1 Il posizionamento dei tre quotidiani

Come nel caso precedente, il primo e più importante elemento da considerare è quello del posizionamento che i giornali hanno assunto nel trattare il problema dell'influenza suina nel corso del periodo di tempo considerato, per capire se e come questo tema sia stato descritto in modo diverso dalle singole testate.

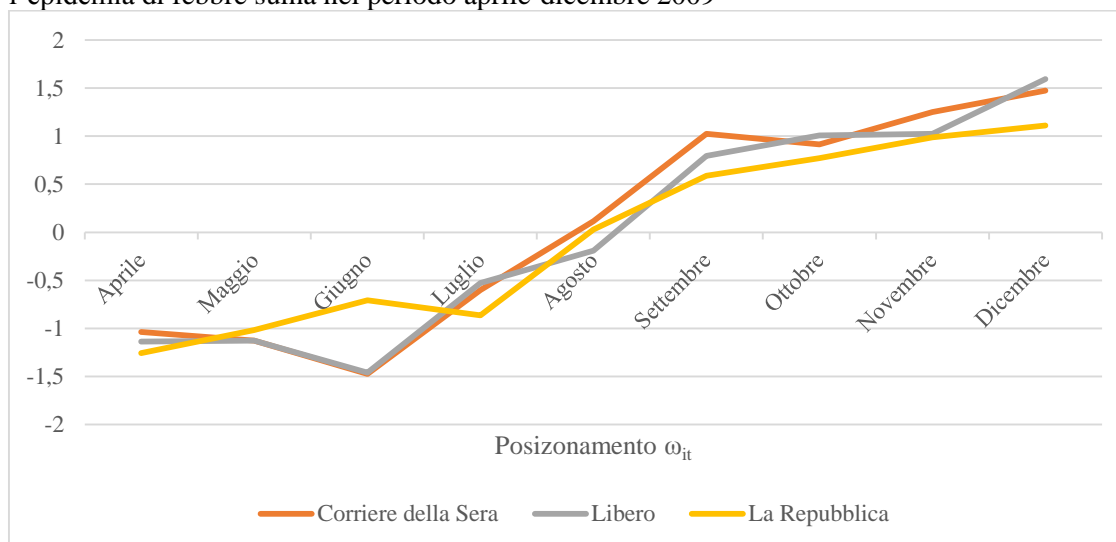
Questa dimensione è, come per il caso degli articoli sull'immigrazione analizzati in precedenza, descritta dal parametro  $\omega_{it}$  derivante dall'analisi degli articoli selezionati effettuata utilizzando l'algoritmo Wordfish; i singoli testi sono stati raccolti in documenti che contengono gli articoli dedicati ogni mese al tema dell'influenza A.

L'andamento del valore di  $\omega_{it}$  sarà, prevedibilmente, diverso rispetto a quello registrato per gli articoli che trattavano il tema immigrazione, in quanto l'epidemia di influenza suina ha dato origine ad una fobia collettiva che ha, tra le altre cose, modificato la domanda di informazioni da parte dei lettori.

Tabella 5.7: valore di  $\omega_{it}$  degli articoli di Repubblica, Libero e Corriere della Sera riguardanti la febbre suina nel periodo aprile-dicembre 2009

Posizionamento $\omega$	Aprile	Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre
Corriere della Sera	-1,036	-1,128	-1,473	-0,602	0,116	1,025	0,915	1,249	1,475
Libero	-1,136	-1,128	-1,460	-0,529	-0,193	0,794	1,008	1,024	1,594
La Repubblica	-1,258	-1,019	-0,709	-0,865	0,028	0,589	0,770	0,986	1,112

Figura 5.7: valore di  $\omega_{it}$  degli articoli di Repubblica, Libero e Corriere della Sera riguardanti l'epidemia di febbre suina nel periodo aprile-dicembre 2009



La figura 5.7 rappresenta la variazione nel tempo dei valori di  $\omega_{it}$  degli articoli analizzati e dimostra chiaramente come tutte e tre le testate considerate abbiano mantenuto una linea editoriale molto simile per trattare il tema dell'influenza suina. Come si può notare, infatti, con l'eccezione del mese di giugno, dove i valori di  $\omega_{it}$  degli articoli proposti da *La Repubblica* si discostano in modo abbastanza marcato da quelli di *Libero* e *Il Corriere della Sera*, tutti e tre i quotidiani hanno trattato l'epidemia di influenza H1N1 utilizzando, sostanzialmente gli stessi argomenti e gli stessi termini. L'analisi del fenomeno influenza suina da parte di queste tre testate è, pertanto, uniforme, nonostante i tre giornali abbiano linee editoriali molto diverse riguardo a numerosi altri temi.

Dopo aver individuato i valori e il trend del posizionamento  $\omega_{it}$  degli articoli è possibile analizzare il lessico utilizzato dalle due testate, per capire in che direzione si è evoluta la descrizione della epidemia di influenza suina nel corso del tempo.

Come per il caso illustrato nel capitolo precedente, l'analisi del lessico utilizzato verrà effettuata tramite il confronto dei valori relativi all'effetto fisso ( $\psi_j$ ) e al peso delle parole ( $\beta_j$ ), concentrando l'attenzione sulle parole collocate nei quadranti III e IV del piano

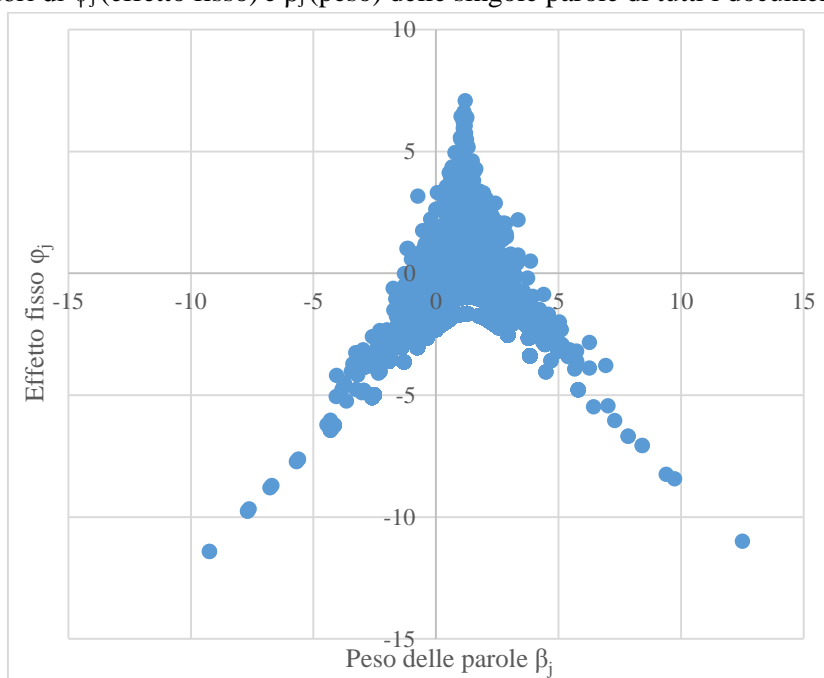
cartesiano, che presentano, quindi, un basso effetto fisso. Saranno i valori di  $\beta_j$  a stabilire il ruolo delle singole parole nel determinare il posizionamento (definito dal parametro  $\omega_{it}$ ) dei testi.

Nel paragrafo successivo verranno individuati i termini più significativi per l'individuazione dei due "estremi" dell'asse ideologico sui quali si collocano gli articoli.

## 5.2.2 Il lessico utilizzato

La figura seguente rappresenta i valori di peso  $\beta_j$  e effetto fisso  $\psi_j$  calcolati dal modello Wordfish per ogni parola contenuta nel corpus di testi. I termini alla base della "torre Eiffel" (che risiedono quindi nei quadranti III e IV) sono quelli che permettono di definire il lessico utilizzato nei diversi testi e, di conseguenza, gli argomenti trattati al loro interno.

Figura 5.8 valori di  $\psi_j$  (effetto fisso) e  $\beta_j$  (peso) delle singole parole di tutti i documenti del corpus



Le parole più significative presenti nel quadrante III del grafico, che sono quindi relativi a valori di  $\omega_{it}$  più bassi, fanno riferimento a fatti accaduti all'estero. Tra le prime posizioni, infatti, si trovano i termini *ocean*, *dream*, (in riferimento alla nave da crociera Ocean Dream, a bordo della quale si verificò una epidemia di influenza H1N1 mentre navigava nelle acque

sudamericane), *Grenada*, *Barbados*, *venezuelana* ed altri termini riferiti a eventi che si sono verificati al di fuori dei confini nazionali.

Ciò significa che, nei primi periodi dell'epidemia, i quotidiani si sono concentrati sul racconto dei fatti riguardanti l'epidemia di influenza suina provenienti dall'estero, facendo pochi, o nulli, riferimenti a possibili ripercussioni sull'Italia (nei primi mesi dell'epidemia non si sono verificati contagi in Italia e in Europa).

Tuttavia, man mano che il virus H1N1 si è diffuso nel resto del mondo, i giornali hanno spostato la loro attenzione sui fatti legati all'epidemia provenienti prima dall'Europa poi dall'Italia, come dimostrano le parole presenti nel quadrante IV del grafico, ovvero quelle che sono collegate a valori di  $\omega_{it}$  maggiori.

Tabella 5.8 i 15 termini più significativi presenti nei quadranti III e IV

Parole quadrante III	$\beta_j$	$\psi_j$	Parole quadrante IV	$\beta_j$	$\psi_j$
l'oms	-9,233	-11,410	lecce	12,506	-10,985
dell'equipaggio	-9,233	-11,410	fazzi	9,736	-8,428
ocean	-7,687	-9,750	manduria	9,404	-8,247
dream	-7,687	-9,750	murphy	8,418	-7,060
grenada	-7,620	-9,664	brittany	8,418	-7,060
margarita	-6,764	-8,798	pancrazio	7,849	-6,676
barbados	-6,692	-8,707	salentino	7,849	-6,676
l'uomo	-5,679	-7,724	attrice	7,296	-6,036
dell'arrivo	-5,679	-7,724	appreso	7,025	-5,437
l'organizzazione	-5,602	-7,628	pregresse	6,936	-3,777
venezuelana	-5,602	-7,628	taranto	6,442	-5,470
moderata	-4,445	-6,203	autopsia	6,272	-2,838
firma	-4,302	-6,028	aglio	6,267	-3,889
dell'oms	-4,300	-6,432	licastro	5,813	-4,772

Le parole più a destra del quadrante IV, a cui corrispondono i valori di  $\beta_j$  più alti sono tutte riferite a notizie sull'influenza suina provenienti dall'Italia: *Lecce*, *Fazzi*, *Manduria*, sono tutti luoghi in cui si sono verificati casi di influenza H1N1, mentre *attrice*, *Brittany* e *Murphy* sono termini riferiti all'attrice italo-americana Brittany Murphy, che si sospettava essere morta proprio a causa della nuova influenza (fatto rivelatosi poi non vero). Dall'analisi dei termini presenti in questo quadrante, inoltre, si può intuire che gli articoli con valori di  $\omega_{it}$  elevati siano caratterizzati anche portatori di allarmismo: il riferimento all'*aglio* è dovuto al fatto che esso sia un rimedio della medicina tradizionale per combattere l'influenza, mentre i termini *autopsia* e *pregresse* indicano un tentativo di minimizzare la pericolosità dell'influenza H1N1, essendo riferiti a decessi attribuiti in un primo momento alla febbre

suina e rivelatisi poi dovuti ad altre cause; lo stesso si può dire della morte dell'attrice citata in precedenza.

L'analisi degli articoli relativi all'influenza suina, pertanto, ha evidenziato come i tre quotidiani presi in considerazione, che hanno linee editoriali solitamente diverse tra loro, abbiano tenuto un comportamento molto simile nel raccontare fatti e problematiche legate all'epidemia di influenza H1N1, inseguendo la domanda dei lettori, sempre più interessati a notizie riguardanti lo sviluppo della malattia in Italia e meno propensi ad acquisire notizie su come la malattia si sia evoluta al di fuori dei confini nazionali.

Le strategie coincidenti dei tre quotidiani, pertanto, dimostra come, in caso di eventi che non portano i lettori ad avere opinioni differenti, i quotidiani tendano a non differenziare la loro offerta di contenuti, adeguandosi alla domanda e dando una copertura limitata del fenomeno considerato (nel caso dell'influenza suina, viene gradualmente abbandonato il racconto dell'evoluzione dell'epidemia all'estero).

## Conclusioni

---

Il mondo dell'informazione è fortemente dipendente dai meccanismi di mercato, dalla domanda dei lettori e dall'offerta di notizie proposta dagli editori per venire incontro alle richieste di chi acquista i giornali, guarda i programmi di informazione e naviga sul web alla ricerca di notizie e opinioni. Se questa affermazione era vera al tempo in cui i giornali esistevano solo come pubblicazioni stampate, tanto più lo è oggi. La nascita e la diffusione di internet hanno profondamente modificato il mercato dell'informazione, amplificando a dismisura l'offerta di articoli a disposizione dei lettori e costringendo gli editori ad aggiornare le proprie strategie per far fronte ad un pubblico molto esigente in termini di quantità e velocità di aggiornamento delle informazioni.

Proprio la domanda di notizie ha un ruolo fondamentale nel definire e plasmare l'offerta di informazioni e, di conseguenza, il mercato delle news. Come si è visto nelle pagine precedenti, è proprio la diversità delle tipologie di domanda di informazioni a spingere i giornali a proporre contenuti diversi a seconda del tipo di eventi trattati.

Nel caso dei quotidiani italiani analizzati si è verificato come testate giornalistiche che dichiarano di collocarsi su posizioni molto distanti abbiano trattato due temi che hanno avuto un impatto molto forte sull'opinione pubblica in modo molto diverso.

Nel primo caso si dimostra un importante assunto del modello di Mullaihatan e Shleifer: un'audience composta da lettori non razionali e con opinioni diversificate porterà le testate a differenziare fortemente la loro offerta di notizie e a collocarsi su posizioni assai differenti tra loro. Questo è quanto si è verificato nell'ottobre del 2013 a proposito del tema della gestione dei flussi migratori che ha molto colpito la politica e l'opinione pubblica del nostro paese. In questo caso le due testate analizzate, *La Repubblica*, che si posiziona nell'area progressista, e *Libero*, che si colloca nettamente nell'area più conservatrice del panorama politico italiano, hanno confermato la loro distanza, assumendo posizioni molto diverse all'interno del dibattito sul tema della gestione dei flussi migratori, proponendo punti di vista diversi e spesso contrastanti. Questa polarizzazione dei contenuti proposti ha fatto sì che i lettori disposti ad informarsi attingendo notizie da più fonti potessero avere un quadro più completo della situazione e delle proposte avanzate dai diversi attori per affrontare il problema dei flussi migratori provenienti dal Nord Africa.



Nel secondo caso presentato, relativo all'epidemia di influenza suina, i tre quotidiani considerati, seppur collocati su posizioni politiche distanti tra loro (oltre a *La Repubblica* e *Liberò* sono stati considerati gli articoli proposti dal *Corriere della Sera*, che ha posizioni moderate e centriste) hanno trattato il tema dell'epidemia del virus H1N1 sostanzialmente nello stesso modo. Le notizie riportate da tutti e tre i giornali nel corso del 2009 a proposito di questo tema, infatti, si sono rivelate molto simili tra loro: all'inizio hanno riguardato gli eventi internazionali relativi all'epidemia, per poi spostarsi sempre più verso fatti relativi all'epidemia di influenza suina avvenuti sul territorio nazionale.

Questo comportamento dei giornali conferma un aspetto fondamentale del modello di mercato delle informazioni presentato nel capitolo 2: la presenza di lettori non razionali e con idee omogenee, a causa, in questo caso, di un bias cognitivo che li ha portati a sovrastimare la pericolosità della malattia, spinge le testate a rendere omogenea la propria offerta, dando origine ad un fenomeno di *spin*, che porta all'annullamento delle differenze presenti tra le notizie proposte e, di conseguenza, a un minore approfondimento complessivo dei temi trattati.

Per concludere, si può affermare che i due casi analizzati confermano l'applicabilità del modello presentato in *The Market for News* (Mullainhatal e Shleifer, 2005) al mercato dell'informazione italiano e, di conseguenza, la centralità del ruolo della domanda nella definizione e diversificazione dell'offerta di notizie. Un pubblico con idee omogenee, dovute in questo caso alla presenza di un forte bias cognitivo (il *negativity bias*) che causa distorsioni nei giudizi riguardanti situazioni potenzialmente rischiose, farà sì che le testate giornalistiche non differenzino la loro offerta, ma che la standardizzino per venire incontro alla domanda dei clienti.

Un pubblico con idee eterogenee, invece, farà sì che l'offerta di notizie sia più ampia e differenziata, rendendo così possibile una migliore descrizione degli eventi e permettendo ai lettori più interessati e disposti a informarsi utilizzando più fonti di conoscere gli eventi in tutti i loro aspetti.

## Riferimenti Bibliografici

---

- Asch, Solomon E. *Forming impressions of personality*. The Journal of Abnormal and Social Psychology 41.3 (1946): 258.
- Beardsworth, Alan, and Teresa Keil. *Sociology on the Menu*. Taylor & Francis, 1997.
- Benoit, Kenneth. *Supervised Methods for Classifying and Scaling Texts*. 2013.
- Baumeister, Roy F., et al. *Bad is stronger than good*. Review of general psychology 5.4 (2001): 323.
- Bloom, Howard S., and H. Douglas Price. *Voter response to short-run economic conditions: The asymmetric effect of prosperity and recession*. American Political Science Review 69.04 (1975): 1240-1254.
- Bohner, Gerd, et al. *What triggers causal attributions? The impact of valence and subjective probability*. European Journal of Social Psychology 18.4 (1988): 335-345.
- Brickman, Philip, and Donald T. Campbell. *Hedonic relativism and planning the good society*. Adaptation-level theory (1971): 287-305.
- Brickman, Philip, Dan Coates, and Ronnie Janoff-Bulman. *Lottery winners and accident victims: Is happiness relative?* Journal of personality and social psychology 36.8 (1978): 917.
- Ceron, Andrea, Stefano M. Iacus, and Luigi Curini. *Social media e sentiment analysis*. Springer, 2014.
- Collette, Benoît, and François Pétry. *Comparing the position of Canadian political parties using French and English manifestos as textual data*. From Text to Political Positions: Text analysis across disciplines 55 (2014): 27.
- d'Aspremont, Claude, J. Jaskold Gabszewicz, and J-F. Thisse. *On Hotelling's "Stability in competition"*. Econometrica: Journal of the Econometric Society (1979): 1145-1150.
- Easterlin, Richard A. *Does economic growth improve the human lot? Some empirical evidence*. Nations and households in economic growth 89 (1974): 89-125.

- Fiske, Susan T. *Attention and weight in person perception: The impact of negative and extreme behavior*. *Journal of Personality and Social Psychology* 38.6 (1980): 889.
- Graber, Doris Appel. *Processing the news: How people tame the information tide*. University Press of America, 1988.
- Grimmer, Justin, and Brandon M. Stewart. *Text as data: The promise and pitfalls of automatic content analysis methods for political texts*. *Political Analysis* (2013): mps028.
- Hamilton, James. *All the news that's fit to sell: How the market transforms information into news*. Princeton University Press, 2004.
- Hotelling, Harold. *Stability in competition*. Springer New York, 1990.
- Kahneman, Daniel, and Amos Tversky. *Prospect theory: An analysis of decision under risk*. *Econometrica: Journal of the Econometric Society* (1979): 263-291.
- Kahneman, Daniel, and Dale T. Miller. *Norm theory: Comparing reality to its alternatives*. *Psychological review* 93.2 (1986): 136.
- Kanouse, David E.; Hanson Jr., L. Reid *Negativity in evaluations*. capitolo di Jones, Edward E. (Ed); Kanouse, David E. (Ed); Kelley, Harold H. (Ed); Nisbett, Richard E. (Ed); Valins, Stuart (Ed); Weiner, Bernard (Ed), *Attribution: Perceiving the causes of behavior*. , (pp. 47-62). Hillsdale, NJ, England: Lawrence Erlbaum Associates, (1987).
- Kim, Mihee. *Partisans and Controversial News Online: Comparing Perceptions of Bias and Credibility in News Content From Blogs and Mainstream Media*. *Mass Communication and Society* 18.1 (2015): 17-36.
- Lee, Julian CH. *The narrative imperative*. *Anthropological Forum*. Vol. 15. No. 2. Routledge, 2005.
- Lo, James, Sven-Oliver Proksch, and Jonathan B. Slapin. *Ideological Clarity in Multiparty Competition: A New Measure and Test Using Election Manifestos*. *British Journal of Political Science* (2014): 1-20.
- Lupfer, Michael B., Matthew Weeks, and Susan Dupuis. *How pervasive is the negativity bias in judgments based on character appraisal?*. *Personality and Social Psychology Bulletin* 26.11 (2000): 1353-1366.

- Marshall, Linda L., and Robert F. Kidd. *Good news or bad news first?* *Social Behavior and Personality: an international journal* 9.2 (1981): 223-226.
- Mazzocchi, M. *La reazione del consumatore a notizie ricorrenti sui rischi legati alle carni.* *Economia e Diritto Agroalimentare* 10.3 (2005): 117-128.
- McFadden, Brandon R., and Jayson L. Lusk. *Cognitive biases in the assimilation of scientific information on global warming and genetically modified food.* *Food Policy* 54 (2015): 35-43.
- McCallum, Andrew, and Kamal Nigam. *A comparison of event models for naive bayes text classification.* AAAI-98 workshop on learning for text categorization. Vol. 752. 1998.
- Moosmayer, Dirk C. *Negativity bias in consumer price response to ethical information.* *Business Ethics: a European Review* 21.2 (2012): 198-208.
- Morewedge, Carey K. *Negativity bias in attribution of external agency.* *Journal of Experimental Psychology: General* 138.4 (2009): 535.
- Mullainathan, Sendhil, and Andrei Shleifer. *The market for news.* *American Economic Review* (2005): 1031-1053.
- Nickerson, Raymond S. *Confirmation bias: A ubiquitous phenomenon in many guises.* *Review of general psychology* 2.2 (1998): 175.
- Osgood, Charles E., and Oliver Tzeng. *Language, meaning, and culture: The selected papers of CE Osgood.* Praeger Publishers, 1990.
- Osgood, Charles E., and Rumjahn Hoosain. *Pollyanna II: Two types of negativity.* *The Journal of Psychology* 113.2 (1983): 151-160.
- Peeters, Guido, and Janusz Czapinski. *Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects.* *European review of social psychology* 1.1 (1990): 33-60.
- Pew Research Centre *Obama Gains Edge in Campaign's Final Days*, Washington D.C., 4 novembre 2012

- Pratto, Felicia; John, Oliver P. *Automatic vigilance: The attention-grabbing power of negative social information*. *Journal of Personality and Social Psychology*, Vol 61(3), Sep 1991, 380-391.
- Royzman, Edward B. *Are Experiments Possible? The Limitations of a Posteriori Control in Experimental Behavior Analysis: The Case of Clinical Process Research*. *Theory & Psychology* 10.2 (2000): 171-196.
- Rozin, Paul, and Edward B. Royzman. *Negativity bias, negativity dominance, and contagion*. *Personality and social psychology review* 5.4 (2001): 296-320.
- Rozin, Paul, Loren Berman, and Edward Royzman. *Biases in use of positive and negative words across twenty natural languages*. *Cognition and Emotion* 24.3 (2010): 536-548.
- Skowronski, John J., and Donal E. Carlston. *Negativity and extremity biases in impression formation: A review of explanations*. *Psychological bulletin* 105.1 (1989): 131.
- Slapin, Jonathan B., and Sven-Oliver Proksch. *A scaling model for estimating time-series party positions from texts*. *American Journal of Political Science* 52.3 (2008): 705-722.
- Taylor, Shelley E. *Adjustment to threatening events: A theory of cognitive adaptation*. *American psychologist* 38.11 (1983): 1161.
- Trussler, Marc, and Stuart Soroka. *Consumer Demand for Cynical and Negative News Frames*. *The International Journal of Press/Politics* 19.3 (2014): 360-379.
- Weiner, Bernard. *"Spontaneous" causal thinking*. *Psychological bulletin* 97.1 (1985): 74.
- Wojciszke, Bogdan, Hanna Brycz, and Peter Borkenau. *Effects of information content and evaluative extremity on positivity and negativity biases*. *Journal of Personality and Social Psychology* 64.3 (1993): 327.