



UNIVERSITÀ DI PISA

Facoltà di Lettere e Filosofia
Facoltà di Scienze, Matematiche, Fisiche e Naturali

Corso di Laurea in Informatica Umanistica

RELAZIONE

Analisi linguistica di blog giornalistici Confronto fra post di blog e articoli

Candidato:

Marzia Giardiello

Relatore:

Prof. Mirko Tavoanis

Correlatore:

prof. Franco Turini

Anno accademico 2011-2012

*A Costantino, Milena,
Ilaria e Serena*

INDICE

Introduzione	4
1.Capitolo I: Il blog.....	6
1.1 Struttura.....	6
1.2 Tipologia.....	7
1.2.3 Blog diario	8
1.2.2 Blog tematico.....	9
1.2.3 Blog letterario.....	10
1.3 Piattaforme blog	10
2. Capitolo II: I giornalisti e i rispettivi blog	11
2.1 I profili.....	12
2.2 Dati relativi ai blog.....	15
2.2.1 Frequenza di aggiornamento	15
2.2.2 Numero di commenti dei lettori	22
3. Capitolo III: Analisi blog vs articoli	28
3.1 I corpora: criteri di scelta.....	28
3.2 Analisi linguistica.....	29
3.2.1 Lunghezza dei post e lunghezza degli articoli.....	29
3.2.2 Numero di proposizioni complesse	32
3.2.3 Lunghezza delle proposizioni complesse	33
3.2.4 Analisi delle proposizioni subordinate	35
4. Capitolo IV: Conclusioni	41
5. Capitolo V: Bibliografia e webgrafia	45

Introduzione

Il presente lavoro è dedicato all'analisi linguistica del blog giornalistico tramite il confronto di articoli di blog e articoli di giornale scritti dallo stesso autore.

In particolare la tesi verterà sull'analisi di cinque blog selezionati in base a criteri personali: particolare interesse per gli argomenti trattati e preferenza di stile.

Dal punto di vista linguistico, in generale, i blog sono erroneamente considerati raccolte di testi poco sorvegliati (una ricerca del 2007 di Stubbe, Ringlstetter e Schulz fa ritenere i blog, insieme ai forum, il genere testuale con più alto numero di errori – 6,5 errori circa ogni mille parole – analisi effettivamente non realistica se si considera la scelta degli autori di classificare errore tutto ciò che non compare nei grandi dizionari). Altre indagini invece permettono di giudicare diversamente questo genere testuale: Tavosanis (2006) conduce uno studio che si basa sull'analisi numerica di “errori più frequenti” dell'italiano scritto ritrovati nel web, nei blog e nei quotidiani online. La lista di parole scelte è il risultato della scrematura eseguita sulla lista di 103 parole comprese nell'elenco ritrovato sotto la voce *errore* del vocabolario Zingarelli (2005). Dopo aver scartato errori di pronuncia, e quindi non utilizzabili per l'analisi che riguarda il linguaggio scritto, parole esposte a problemi di codifica e omografi, il numero di errori finale si è ridotto a 19 parole. Queste parole sono state ricercate attraverso Google nel web in generale, su un gruppo di siti che ospitavano blog (blog.excite.it, clarence.it, splinder.it, splinder.com) e su un gruppo di quotidiani online (corriere.it, repubblica.it, ilmattino.caltanet.it, unita.it) e per ciascuno dei tre gruppi è stato calcolato il rapporto fra la forma errata e la corrispondente forma corretta.

Il risultato ha messo in luce come la percentuale totale di errori risulti nei blog leggermente superiore a quella dei quotidiani online, ma fortemente inferiore rispetto a quella del web nel suo complesso; questo risultato, dunque suggerisce l'idea che il linguaggio blog non è da considerarsi così poco revisionato e, facendo riferimento alle

quattro le categorie convenzionali del parametro tempo di scrittura, descritto da Tivosanis (2011)¹

- scrittura rapida e non revisionata: testi scritti di getto senza nessun tipo di pianificazione o revisione;
- scrittura rapida e revisionata: testi scritti rapidamente, senza progettazione preliminare ma con un qualche grado di revisione;
- scrittura convenzionale: testi scritti all'interno di un processo pianificato con revisione accurata;
- scrittura progettata per altri tipi di pubblicazione: testi scritti per altri mezzi (tipicamente, per la diffusione a stampa) e copiati o pubblicati meccanicamente sul web,

le più adatte a descrivere la maggior parte dei blog sono la seconda e la terza categoria.

I blog giornalistici in particolare sono caratterizzati da un linguaggio abbastanza simile a quello dei giornali. La sintassi è articolata, l'ortografia e la sintassi rispettate e in generale il lessico si avvicina a un "uso medio" ; ovviamente vanno considerate le dovute eccezioni che sono poi alla base di questo studio basato sul confronto linguistico fra i due generi testuali (post e articolo).

L'analisi si baserà essenzialmente sui seguenti confronti:

- La lunghezza media degli articoli con la lunghezza media dei post di blog.
- La quantità di frasi complesse degli articoli con la quantità di frasi complesse dei blog
- La lunghezza medie delle frasi complesse
- L'analisi delle frasi subordinate: quantità e tipo.

¹ Tivosanis 2010 *L'italiano del web* cap. 2.1 "Parametro I: tempo di scrittura" (pag. 46-47)

Capitolo I: Il Blog

Prima di procedere all'analisi linguistica è necessaria una panoramica su cos'è un blog ed in particolare un blog giornalistico o d'attualità.

Blog, parola attestata nell'OED (Oxford English Dictionary) nel 1999, e per l'italiano nel 2000, è l'abbreviazione di *weblog* (= diario in rete); il termine oggi è usato per indicare una pagina web ordinata cronologicamente e aggiornata di frequente. Si tratta di un sito web, o uno spazio all'interno di un sito, autogestito, nel quale vengono pubblicate in tempo reale notizie, informazioni, opinioni o storie.

Gli utenti che leggono i blog hanno di regola la possibilità di commentare i post inseriti dall'autore generando spesso delle discussioni con altri lettori sul tema trattato.

Il fenomeno dei blog nasce nel 1997 in America (il primo blog vero e proprio nasce il 23 dicembre grazie a Jorn Barger, un appassionato di caccia deciso ad aprire uno spazio personale su cui condividere riflessioni riguardo al suo sport) e si afferma in Italia nel 2001.

Creare un blog è un'operazione abbastanza semplice. La procedura più utilizzata consiste nella registrazione dell'utente-autore su siti specifici che automaticamente creano lo spazio virtuale usando impostazioni predefinite e pochi parametri di configurazione senza che sia necessario conoscere il linguaggio HTML. Questo comunque non è l'unico modo per creare un blog: l'utente può anche scrivere direttamente su pagine Html statiche oppure personalizzare un CMS creato per altri scopi.

1.1 Struttura

I tratti strutturali comuni ai blog sono i seguenti: i testi sono forniti di data e sono presenti sulla pagina web in ordine anticronologico (prima i messaggi più recenti) e la maggior parte delle volte sono introdotti da un titolo. Generalmente i blog hanno anche

un archivio (ordinato per mesi o per anni) dei messaggi già pubblicati cosa che rende semplice la ricerca di post in base alla data in cui sono stati pubblicati. Per agevolare la ricerca di post spesso è anche presente una sezione “categorie” dove è possibile rintracciare gli argomenti chiave dei testi.

Altra caratteristica è che spesso i sistemi di pubblicazione consentono di inserire all'interno dei testi materiali di vario genere: dai file PDF alle immagini, dai video ai file audio.

Molti blog offrono la possibilità di commentare i post e in alcuni casi di blog molto popolari i lettori lasciano centinaia o migliaia di commenti caratterizzati (tranne i casi in cui la qualità del commento supera quella del post) da un linguaggio informale più accomunabile ad altri generi testuali come il forum che al blog.(Tavosanis 2011)².

1.2 Tipologie

I blog non sono testi unici ma raccolte di post; risulta quindi difficile etichettare un blog in quanto si pone spesso il dubbio se assegnare un'etichetta di genere al blog in generale o ai singoli post. Nonostante ciò in genere la maggior parte dei blogger risulta coerente nella stesura dei post, cioè se si tratta di un blog tematico difficilmente vi si troveranno post in cui si raccontano esperienze personali e viceversa.

Sul web infatti sono presenti molti elenchi di categorie (in alcuni casi addirittura una tipologia specifica di blog viene etichettata con due nomi diversi). Si elencano su Wikipedia molte categorie: social blog, blog collettivo, blog aziendale, blog tematico, photoblog, blog vetrina, blog politico, m-blog, video blog, blog letterario, blog didattico, audioblog, blog novel, splog (finti blog riempiti di testi riprodotti solo per dare la sensazione che un determinato argomento sia particolarmente rilevante).

² *L'italiano del web*. Paragrafo 7.2 “Caratteristiche attuali” (pag. 144 – 146)

Una suddivisione più semplificata e basata sull'analisi linguistica è fornita da Tavosanis (2011) e si basa su tre categorie :

- blog tematico (inclusi I blog giornalistici) 59 %
- blog diario 31%
- blog letterario 10%³

1.2.1 Blog diario

Nel blog personale (diario) l'autore scrive le proprie esperienze giornaliere, racconti, desideri. I commenti da parte dei lettori sono spesso molto apprezzati e danno spesso vita a scambi d'opinioni e addirittura a *flaming* (dall'inglese flame per "fiamma": è l'atto di inviare messaggi deliberatamente ostili e provocatori da parte di un utente che commenta i post del blog); spesso nascono delle vere e proprie *flame war* ("guerre" di messaggi) tra utenti diversi. Questo tipo di blog è usato spesso da studenti di scuola superiore o universitari, con un gran numero di collegamenti incrociati tra un blog e l'altro ma c'è da dire che questa tipologia oggi è abbastanza in declino.

Dal punto di vista linguistico i blog diario risultano quelli in cui si concentra maggiormente l'espressività grafica: lo anticipa Canobbio (2005) nel suo articolo tratto da *Lid'O Lingua Italiana d'oggi* (2005) in cui analizza un corpus di 50 post (pubblicati da novembre 2002 ad aprile 2005) presi dal sottogenere "diario personale".

Con espressività grafica si intende quell'insieme di deviazioni dall'ortografia italiana standard che è riconducibile principalmente a tre meccanismi di alterazione:

- lettura endofasica, in cui un simbolo viene inserito in modo che il lettore lo interpreti in base al suo nome (es. '6' per *sei*, seconda persona dell'indicativo presente del verbo essere; oppure 'x' per *per*);

³ I dati percentuali sono risultati dall'analisi fatta su un campione di 100 blog presi in esame nell'aprile del 2007 e ricercati casualmente su *Google Blog Search* (Tavosanis, 2007)

- abbreviazioni non utilizzate abitualmente nella lingua parlata (es. ‘nn’ per *non*, ‘qlk’ per *qualche*);
- variazioni grafiche vere e proprie (es. ‘k’ per indicare /k/ quando l’ortografia italiana richiede ‘ch’).⁴

La motivazione principale è il risparmio di tempo nella scrittura. Altri tipi di deviazioni non riguardano la grafia ma piuttosto l’espressività come l’uso di *emoticons* (o faccine), l’uso di maiuscole e minuscole in modi diversi dallo standard e i giochi di punteggiatura.

1.2.2 Blog tematico

I blog tematici (o giornalistici) sono i blog dedicati ad un argomento specifico e per questo sono accostabili agli articoli di giornale. Non a caso molti blog tematici sono creati da giornalisti che utilizzano questo genere testuale per sintetizzare, riprodurre o ampliare i contenuti dei loro articoli.

Dal punto di vista linguistico si può notare come un giornalista che utilizzi il blog come canale di comunicazione (differente dal canonico articolo di giornale) a prima vista non si discosti più di tanto dalla scrittura per la pubblicazione su carta (o per giornale online), e l’analisi portata avanti da questa tesi approfondirà l’argomento nei capitoli successivi.

Non mancano le eccezioni⁵ ed è chiaro che l’ambiente del blog permette di “lasciarsi andare”: molti blog fanno uso di espressioni confidenziali e linguaggio colloquiale. La tendenza generale (si vedrà nei confronti su base linguistica nel cap. 3) tuttavia è quella di conservare, nei post tematici un linguaggio che fa riferimento all’italiano giornalistico.

⁴ Definizioni fornite da *L’italiano del web* di M. Tavosanis, cap. 4 “Espressività grafica” (pag. 71-87)

⁵ [*Neppure la notizia del (previstissimo e inevitabile) pateracchione sfornato dagli azzecagarburgli al servizio del Pdl, ma a spese nostre, con l’ennesima leggina porcellina, è riuscita a rovinarmi la giornata di ieri, illuminata dall’incontro con la nuova donna della mia vita, la bellissima Francesca in tutti i suoi tre chili e seicento grammi di gloria.*] estratto di un post del blog “Tempo Reale” di Vittorio Zucconi.

1.2.3 Blog letterario

Per quanto riguarda questa categoria, spesso è difficile interpretare se si è in presenza di un blog che tratta di letteratura e/o di poesia perché è facilmente confondibile con il blog diario (in cui l'autore può decidere di inserire post composti solo da poesie o testi letterari).

Genericamente i post di tali blog sono caratterizzati dalla presenza, al loro interno, del lessico letterario che si manifesta in quantità maggiori rispetto ai quotidiani. Da una ricerca di Tavosanis (2006) basata sull'analisi di 72 verbi identificati come "letterari" oppure "obsoleti" nel vocabolario De Mauro (2000), infatti, è risultato che nei blog la presenza di tali verbi è di circa quattro volte superiore rispetto al sito de "la Repubblica".⁶(Non tutte le 72 forme all'infinito dei verbi sono state ritrovate, ma solo 25)

I blog letterari più famosi in realtà sono blog tematici dedicati alla letteratura: *Carmilla on line*, *Nazione Indiana*, *Lipperatura*; tra i più noti, non esistono allo stato attuale blog costituiti esclusivamente da testi.

In questa categoria possono rientrare anche le "fan fiction" ossia blog dedicati ai romanzi, ai cartoni animati, ai telefilm e ai film (Lipperini, 2008).

1.3 Piattaforme blog

Le piattaforme blog si dividono in due categorie: i servizi web pronti all'uso e i software da installare su un server. Nel primo caso, semplice, basta registrarsi per creare il blog; al contrario, scegliendo una piattaforma da installare su un server o in hosting (noleggiare uno spazio con un abbonamento), il lavoro è più lungo ma concede la possibilità di personalizzare molto il proprio blog. (D'Ottavi e Sorchiotti, 2008).

Dal 2001 ad oggi sono nati molti servizi in italiano che permettono di gestire un blog gratuitamente. Tra i più utilizzati citiamo: *Blogger*, *Wordpress.com*, *Blogsome*, *Il*

⁶ Il totale di occorrenze nei blog di verbi "letterari" è di 230, mentre il totale sul sito de "La Repubblica" è di 65

Cannocchiale, Io Bloggo, Blogdrops, NoiBlogger, Libero, LiveJournal, Windows Live Spaces, MySpace, Skyrock.

Blogger.com è una delle più semplici piattaforme di blog online. Si stima che ospiti tra i 15 e i 20 milioni di blog in tutto il mondo (su blogger si trova anche il blog ufficiale di Google); Ilcannocchiale è una piattaforma italiana ed ospita più di 74000 blog.

Wordpress è una piattaforma realizzata in PHP, che utilizza MySQL come database. Lanciata nel 2003 è distribuita con licenza GNU (*General Public License*) che permette di utilizzare liberamente il software. Caratteristiche principali sono velocità, semplicità d'utilizzo e un'ottima interfaccia utente.

Italiana e “defunta” è la piattaforma Splinder.com creata nel 2002 da Marco Paolombi e ufficialmente chiusa il 31 gennaio 2012; Splinder fu la prima piattaforma ad offrire in lingua italiana un servizio blog in hosting.⁷

Capitolo II: I giornalisti e i rispettivi blog

I giornalisti selezionati per questo lavoro di confronto fra gli articoli di giornale e i post di blog sono, in ordine alfabetico: Ernesto Assante, Alessandro Gilioli, Paolo Ottolina, Federico Rampini, Vittorio Zucconi.

Il criterio con cui sono stati scelti i giornalisti è prettamente personale (Assante, Rampini e Zucconi sono ad esempio giornalisti che scrivono su Repubblica, il quotidiano che leggo maggiormente); in particolar modo il tentativo è stato quello di scegliere dal punto di vista tematico una certa varietà di materie e dal punto di vista linguistico chiarezza e semplicità nello stile.

⁷ La panoramica sulle piattaforme è fornita dal testo “Come si fa un blog 2.0” di D’Ottavi e Sorchiotti, cap.3)

2.1 Profili

Nome	Ernesto ASSANTE
Occupazione	Giornalista e scrittore per vari editori, critico e storico musicale, collabora come critico musicale per <i>La Repubblica</i> dal 1978.
Nascita	Napoli, 1958
Formazione	Università degli Studi di Roma “La Sapienza”
Blog	Media Trek http://assante.blogautore.repubblica.it/ Temi trattati: musica ed informatica. archivio blog dal gennaio 2005
Temi	Si occupa principalmente di musica. Il suo interesse, però va anche aldilà del campo musicale, estendendosi ai <i>new media</i> e alle <i>new technology</i> , con particolare riferimento alla loro relazione con il giornalismo, al loro impatto sulle nuove generazioni e ai meccanismi psicosociali che si vanno instaurando nelle nuove frontiere del web fino ai <i>videogames</i> .

Nome	Alessandro GILIOLI
Occupazione	Giornalista professionista dal 1977 e autore di diversi libri; dal 2002 vive e lavora a Roma come caporedattore del <i>L'Espresso</i>
Nascita	Milano, 28 febbraio 1962
Formazione	- Università degli Studi di Milano, laurea in Filosofia - Istituto per la Formazione al Giornalismo, Milano - Columbia School of Journalism, NYC
Blog	Piovono rane http://gilioli.blogautore.espresso.repubblica.it/ archivio blog dal dicembre 2005 Temi trattati: politica (soprattutto estera), blog dal tono provocatorio e brillante.
Temi	Si occupa di comunicazione, rete, politica e attualità italiana, scrivendo principalmente di temi legati alla politica contemporanea e alla libera informazione (libertà di stampa, diritto al web...). Come egli stesso ha dichiarato in un'intervista a <i>Next Open Innovation</i> di <i>Telecom Italia</i> “Della tecnologia e della Rete mi piace occuparmi per le conseguenze che hanno sul piano culturale, politico, economico e sociale”. Nel 2009 è stato fra i promotori della <i>Carta dei cento per il libero Wi-Fi</i> , per la liberazione del <i>Wi-Fi</i> italiano, soffocato dal decreto Pisanu, ed ha parlato sul palco del <i>No-Berlusconi-Day</i> .

Nome	Paolo OTTOLINA
Occupazione	Giornalista del Corriere dal 2001, vice caposervizio del sito web.
Nascita	Asti, 1974
Formazione	- Università degli studi di Torino, laurea in Scienze delle Comunicazioni - Istituto per la Formazione al Giornalismo "Carlo De Martino"
Blog	Mal di tech http://malditech.corriere.it/ archivio blog dal novembre 2009 Temi: tecnologia e scienza, in particolare informatica e telefonia. Molto ricco di video e immagini, è caratterizzato da post molto corposi, di lunghezza atipica per essere un diario in rete.
Temi	Si occupa principalmente di tecnologia, dichiara che <i>il suo primo computer è stato un Vic 20, che bazzica la rete dal 1996 e che la sua casa è una sorta di magazzino delle cose perdute in cui vanno e vengono gadget hi-tech. Soprattutto smartphone, tablet, fotocamere e pc⁸</i> , recensisce per il Corriere della Sera e per il sito.

Nome	Federico RAMPINI
Occupazione	Giornalista e scrittore italiano, è stato inviato del quotidiano <i>La Repubblica</i> a Parigi, Bruxelles, San Francisco e Pechino, dove ha modo di entrare in contatto e approfondire la conoscenza della cultura e del modello economico asiatico. Attualmente vive a New York dove è editorialista e corrispondente dagli Stati Uniti per <i>La Repubblica</i> .
Nascita	Genova, 25 marzo 1956
Formazione	- Università Commerciale Luigi Bocconi, Milano - École des hautes études en sciences sociales, Parigi
Blog	Estremo Occidente http://rampini.blogautore.repubblica.it/ Temi: politica estera e attualità
Temi	Si occupa principalmente di temi di Economia e attualità, con particolare riferimento ai modelli macroeconomici degli Stati Uniti e dell'Estremo Oriente.

⁸ Dal suo blog <http://malditech.corriere.it/>

Nome	Vittorio ZUCCONI
Occupazione	Giornalista e scrittore italiano, è attualmente direttore e corrispondente estero dagli Stati Uniti del quotidiano <i>La Repubblica</i> .
Nascita	Bastiglia (Modena), 16 agosto 1944
Formazione	Università degli Studi di Milano
Blog	Tempo reale http://zucconi.blogautore.repubblica.it/ Tratta argomenti molto vari coprendo la maggior parte degli eventi di attualità di Italia e non.
Temi	Si occupa principalmente di attualità e di politica

attraverso colori della stessa gamma, ma di intensità superiore in base al numero, sono stati segnati quanti interventi sono stati fatti e in che giorno del mese.

Dall'analisi fatta risulta che il giornalista che aggiorna più frequentemente il suo blog è Alessandro Gilioli (il suo blog *Piovono Rane* risulta al 19 gennaio 2010 al secondo posto dei 10 blog italiani più visitati dopo quello di Beppe Grillo, secondo una classifica stilata dal sito Blogbabel⁹) che pubblica uno o più post tutti i giorni con poche eccezioni.

A seguire troviamo Federico Rampini e Vittorio Zucconi che nell'arco dei tre mesi hanno pubblicato complessivamente tra gli 80 e i 90 messaggi. Rampini è più costante nell'aggiornare il suo spazio, Zucconi invece, molto più discontinuo, raggiunge le cifre del collega pubblicando, poche volte al mese, anche fino a 7 post.

Ernesto Assante tra tutti è quello più metodico: come si può vedere dal grafico, ha un atteggiamento in un certo modo costante (nei confronti del blog): massimo un post una decina di volte, in media, al mese.

Paolo Ottolina è un esempio di giornalista invece che aggiorna molto poco il suo blog: nei tre mesi analizzati in totale ha pubblicato solo 13 post e nel mese di giugno, in particolare, solo 3.

Per un'analisi più approfondita le prossime tabelle riassumono la frequenza con cui i giornalisti hanno aggiornato i loro blog in un lasso di tempo più lungo (un anno).

⁹ **Blogbabel:** <http://it.blogbabel.com/classifica-blog/>

assante												
frequenza post												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1	0	0	1	1	0	1	0	0	1	0	1	1
2	0	1	1	0	0	1	1	0	25	1	0	1
3	1	0	1	1	0	0	1	0	6	1	0	0
4	0	0	0	0	0	1	0	1	0	0	0	0
5	0	0	0	0	1	1	1	0	0	1	1	0
6	1	0	1	1	2	0	0	1	1	0	0	0
7	0	0	0	0	0	0	1	0	1	1	1	1
8	1	0	1	1	0	1	0	1	0	0	1	2
9	0	0	1	0	1	1	1	0	1	1	1	1
10	1	0	0	0	0	1	0	0	0	0	0	0
11	0	1	0	1	0	0	0	1	0	0	0	1
12	0	1	0	1	1	0	1	0	0	0	0	1
13	1	1	1	0	0	0	0	0	0	1	0	0
14	0	0	1	1	1	0	0	1	1	0	1	1
15	0	0	0	0	1	1	1	1	1	0	0	1
16	0	1	0	0	0	1	1	0	0	2	1	1
17	1	0	1	0	0	1	0	1	1	1	0	1
18	0	0	1	0	1	1	0	1	1	0	1	2
19	0	0	0	1	0	0	0	1	0	0	0	0
20	0	1	0	1	1	1	0	0	0	1	1	0
21	1	0	1	1	0	1	0	1	1	0	1	1
22	0	0	1	0	2	0	1	0	1	0	1	1
23	0	1	0	0	0	1	1	1	1	1	1	1
24	1	0	1	1	1	0	0	0	0	0	0	0
25	0	1	0	0	0	1	1	0	0	1	0	1
26	0	1	0	1	1	0	0	0	0	0	0	1
27	0	0	0	0	0	0	1	1	1	1	1	0
28	1	0	0	0	0	0	0	1	1	0	0	0
29	0	0	1	1	0	1		1	0	0	1	1
30	0	0	0	1	1	0		0	0	1	1	1
31	0	1		0		1		0	0			0
	0,3	0,3	0,5	0,5	0,5	0,5	0,4	0,5	1,4	0,5	0,5	0,7
al giorno												0,5

Legenda nr. post						
0	1	2	3	4	5	>5

Tabella 2: Numero di post pubblicati al giorno da Assante in un anno (Ago. 2010 Set. 2009)

(Da notare il caso riguardante le date 2-3 dicembre in cui è andata in onda la finale della trasmissione X Factor su Rai2 e in cui Assante ha pubblicato un numero di post decisamente superiore alla media.)

gilioli

frequenza post												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1	0	1	0	4	2	2	2	2	2	3	4	4
2	1	3	1	0	1	5	7	1	2	2	2	5
3	3	0	0	6	4	2	5	0	2	5	2	4
4	1	1	2	2	1	5	5	1	2	4	3	4
5	2	2	1	4	1	4	4	3	1	3	6	1
6	2	1	1	3	1	1	2	4	4	3	4	2
7	0	1	5	3	1	0	4	2	1	4	9	3
8	1	2	3	2	1	2	1	2	5	2	8	5
9	0	1	1	1	2	5	3	3	2	2	5	1
10	1	1	4	4	1	5	3	1	4	2	0	3
11	2	0	1	4	1	3	6	6	2	5	2	5
12	0	4	1	5	1	5	3	4	0	2	4	1
13	0	1	1	3	3	2	1	1	3	2	3	2
14	0	2	1	3	1	4	1	2	6	2	3	4
15	0	1	1	1	3	8	4	3	4	1	2	5
16	0	1	5	2	4	2	2	1	2	5	7	3
17	1	1	1	5	1	2	6	1	3	3	2	0
18	1	2	3	3	3	4	4	3	1	3	1	2
19	0	2	0	3	6	3	2	3	2	5	3	5
20	0	1	2	1	4	4	2	7	1	1	3	3
21	1	4	2	1	2	1	2	5	3	2	3	4
22	1	2	4	1	6	3	4	3	2	2	2	6
23	0	1	2	1	2	3	1	2	1	2	3	6
24	1	0	3	3	1	2	4	1	2	2	3	4
25	0	0	2	3	1	1	4	1	0	4	2	8
26	0	2	0	3	3	5	5	4	1	2	5	1
27	2	1	0	4	3	1	2	1	0	1	3	3
28	0	1	3	2	2	1	3	3	2	2	3	6
29	1	2	2	1	1	4	0	5	4	6	3	6
30	2	2	2	0	6	2	0	3	2	2	3	6
31	0	0		0	0	2	0	4	2		3	3
	0,7	1,4	1,8	2,5	2,2	3,0	3,0	2,6	2,2	2,8	3,4	3,7
algiorno												2,5

Legenda nr. post						
0	1	2	3	4	5	>5

Tabella 3: Numero di post pubblicati al giorno da Gilioli in un anno (Ago. 2010 Set. 2009)

ottolina

frequenza post												ott	sett
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov			
1	0	0	0	0	0	0	2	0	0	0	dati non disponibili	dati non disponibili	
2	0	0	0	0	0	1	0	0	0	1			
3	0	0	1	0	0	0	0	0	1	0			
4	0	0	0	1	0	2	0	0	0	0			
5	0	1	0	1	1	0	1	0	0	0			
6	1	0	0	1	1	0	0	0	0	0			
7	0	0	1	0	1	0	0	0	0	0			
8	0	0	0	0	1	0	0	0	0	0			
9	0	0	0	1	0	0	0	0	1	0			
10	0	0	0	0	0	0	1	0	0	0			
11	0	0	0	1	0	0	1	0	0	0			
12	0	0	0	0	1	0	2	1	0	0			
13	1	0	0	1	1	0	1	0	0	1			
14	0	1	0	0	1	0	0	0	0	0			
15	0	0	0	1	0	0	2	0	0	0			
16	1	1	0	1	1	0	0	0	0	0			
17	0	0	0	0	0	0	1	0	0	0			
18	0	0	0	0	0	1	0	0	0	0			
19	0	0	0	0	0	2	3	1	0	0			
20	1	0	0	0	0	0	1	0	1	0			
21	0	0	0	0	0	0	0	0	0	0			
22	0	0	0	0	0	0	2	0	0	0			
23	0	0	0	0	0	0	1	0	0	0			
24	1	0	0	0	0	1	0	0	0	0			
25	1	0	0	1	0	0	0	1	0	0			
26	0	0	0	0	0	0	1	1	0	0			
27	0	0	0	1	1	0	0	2	0	0			
28	1	0	1	0	1	1	0	1	0	0			
29	0	0	0	1	1	0		2	0	0			
30	0	0	0	1	0	0		0	0	0			
31	0	0		0		1		0	0				
	0,2	0,1	0,1	0,4	0,4	0,3	0,7	0,3	0,1	0,1	-	-	
al giorno												0,2	

Legenda nr. post

0	1	2	3	4	5	>5
---	---	---	---	---	---	----

Tabella 4: Numero di post pubblicati al giorno da Ottolina in un anno (Ago. 2010 Sett. 2009)

rampini

	frequenza post											
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1	0	0	4	0	1	3	0	0	1	0	1	0
2	0	1	2	1	1	2	0	0	2	1	2	1
3	0	0	1	1	0	2	1	0	1	1	1	0
4	0	0	0	2	0	2	1	0	1	1	0	1
5	0	1	0	1	0	2	0	0	0	0	1	0
6	0	2	0	2	0	2	0	0	0	0	2	0
7	0	0	0	1	4	0	0	0	0	0	1	0
8	0	0	2	1	3	2	0	1	0	0	1	1
9	0	0	2	0	2	2	4	0	3	2	1	3
10	5	1	3	2	0	1	3	0	1	0	0	1
1	1	2	0	1	1	3	1	1	0	0	0	2
12	1	0	0	2	5	3	3	1	1	1	0	1
13	2	2	0	1	2	1	0	1	2	2	0	0
14	0	1	1	0	3	0	0	1	2	3	0	1
15	0	2	0	1	2	2	1	0	3	1	0	0
16	0	0	0	0	5	2	1	1	3	1	0	1
17	3	0	1	0	2	0	2	1	3	1	0	0
18	1	1	2	5	2	4	0	2	2	1	0	0
19	3	2	1	4	4	3	1	2	0	2	0	0
20	2	0	0	0	2	1	0	4	0	0	3	0
21	0	0	0	2	4	0	0	2	0	0	3	1
22	3	0	5	1	1	2	2	1	0	0	3	2
23	4	0	2	2	0	3	3	0	0	2	1	2
24	4	0	0	2	3	0	3	1	0	0	1	1
25	5	1	0	0	3	0	1	1	0	0	1	2
26	1	0	2	2	2	0	1	0	0	0	0	1
27	3	0	1	1	0	0	1	1	0	0	2	0
28	1	1	1	1	1	0	1	5	0	0	1	0
29	0	0	2	0	2	3		2	0	0	1	0
30	1	0	1	0	1	3		1	0	1	1	1
31	1	0		1		3		0	0		1	0
	1,3	0,5	1,1	1,2	1,9	1,6	1,1	0,9	0,8	0,7	0,9	0,7
al giorno	1,1											

Legenda nr. post

0	1	2	3	4	5	>5
---	---	---	---	---	---	----

Tabella 5: Numero di post pubblicati al giorno da Rampini in un anno (Ago. 2010 Set. 2009)

zucconi

frequenza post												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1	1	0	3	1	1	3	1	1	2	0	2	3
2	0	0	2	0	0	4	2	2	3	1	0	3
3	0	0	3	1	0	6	3	0	1	2	2	3
4	0	0	1	4	0	2	1	1	1	1	1	2
5	0	0	0	1	0	2	2	2	0	1	0	2
6	0	0	0	3	2	2	0	2	1	1	2	0
7	0	0	0	2	1	0	2	2	3	1	2	2
8	0	0	0	0	1	2	1	2	0	0	3	1
9	0	0	0	1	2	2	1	4	1	1	0	2
10	0	0	0	1	1	4	1	1	1	3	1	2
11	0	0	0	3	1	2	1	0	1	2	0	2
12	0	0	0	4	1	2	0	3	0	3	1	0
13	0	0	0	2	2	2	2	2	1	0	1	0
14	0	5	0	4	2	1	1	3	1	2	1	1
15	0	3	0	0	4	5	0	4	1	1	3	2
16	0	5	0	2	1	2	5	1	2	2	0	3
17	1	1	0	4	2	1	1	0	1	2	0	1
18	1	0	0	1	2	3	1	3	0	0	0	2
19	2	4	0	2	1	2	1	1	1	0	2	1
20	1	3	0	3	0	1	0	1	1	2	2	0
21	1	2	0	1	1	2	1	2	0	2	1	0
22	1	0	0	2	1	2	2	4	0	0	1	0
23	1	0	0	0	0	0	3	1	0	0	0	1
24	4	0	0	3	1	1	3	2	1	0	4	1
25	1	0	0	4	0	1	2	1	0	1	0	0
26	2	5	0	1	2	3	3	1	1	0	2	1
27	2	2	0	1	1	1	2	5	1	2	1	0
28	1	6	0	3	0	2	2	2	0	0	2	0
29	1	8	0	1	1	3		2	1	0	1	0
30	0	4	1	1	1	3		3	0	2	0	1
31	2	2		2		1		1	0		1	0
	0,7	1,6	0,3	1,9	1,1	2,2	1,6	1,9	0,8	1,1	1,2	1,2
al giorno												1,3

Legenda nr. post

0	1	2	3	4	5	>5
---	---	---	---	---	---	----

Tabella 6: Numero di post pubblicati al giorno da Zucconi in un anno (Ago. 2010 Set. 2009)

2.2.2 Commenti degli utenti

Questa analisi riguarda il numero dei commenti che gli utenti postano sui blog dei giornalisti, in particolare il numero di commenti fatti per ogni post pubblicato nel periodo agosto 2010 – settembre 2009. La prima casella in alto a sinistra rappresenta la data 01/08/2010; l'ultima casella in basso a destra rappresenta la data 31/09/2009.

assante

frequenza commenti												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1		190	343	334		43			105		227	95
2		102	201			44	89		24	169		545
3	113		153	44			68		49	105		
4						32		89				
5					23	57	43			130	138	
6	16		251	58	61			57	29			
7							117		166	56	63	85
8	88		63	38		20		59			64	85
9			61		122	16	119		151	228	695	206
10	94					142						
11		256		70								113
12		297		25	141		70					
13	128	442	71							102		
14			63	76	56			55	59		14	80
15					60	57	88	39	127			24
16		218				50	284			34	70	91
17	123		138			48		56	37	114		19
18			121		42	108		80	122		235	156
19				42				37				
20		67		88	78	20				98	107	
21	44		30	35		63		115	45		116	30
22			40		41		86		36		51	147
23		153				151	45	157	26	115	107	196
24	203		67	89	10							
25		14				49	57			112		147
26		26		26	38							81
27							31	41	8	102	83	
28	163							44	118			
29			362	70		36		96			59	154
30				66	30					53	83	109
31		799				56						
	108	233	140	76	59	58	91	71	73	109	141	131
per post												107,6

Legenda nr. commenti							
0	50	100	150	200	250	300	>300

Tabella 7: Numero di commenti per post in un anno nel blog di Assante (Ago. 2010 Sett. 2009)

gilioli

frequenza commenti												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1		31		27	42	22	20	67	47	34	64	14
2	25	33	63		30	64	30	119	62	24	31	15
3	13			13	36	20	30		38	27	7	27
4	20	14	25	34	13	32	30	53	126	71	37	28
5	59	56	23	16	30	56	12	31	97	64	27	56
6	38	35	29	32	26	134	28	44	45	34	29	19
7		63	17	21	30		27	47	109	24	33	14
8	61	25	34	35	81	45	33	102	30	52	40	25
9		13	38	22	34	27	27	43	19	38	22	20
10	66	56	46	17	26	36	23	68	50	64		18
11	48		54	40	40	15	21	32	123	31	46	20
12		23	35	23	18	17	21	39		48	35	40
13	64	39	22	31	30	13	18	36	130	17	33	38
14		48	16	18	40	12	26	67	60	43	32	13
15		42	37	21	73	16	23	56	86	58	29	91
16	27	37	14	21	41	24	22	53	44	39	29	33
17	32	42	13	18	30	28	26	52	71	30	21	18
18	62	21	35	18	43	17	21	28	187	28	78	28
19		21		62	27	26	59	18	45	23	30	14
20		65	10	35	29	23	28	18	79	25	34	24
21	11	29	36	22	15	108	28	20	57	15	21	34
22	22	40	58	13	18	30	15	25	84	26	51	21
23		62	15	20	31	28	21	31	60	59	55	20
24	61		13	49	35	35	21	23	22	84	42	27
25			21	18	17	84	17	63		22	23	21
26		33		35	14	16	21	35	68	61	30	19
27	20	28		33	25	43	26	20		75	37	26
28		32	43	86	54	20	25	42	62	15	73	13
29	32	11	52	32	15	39		75	67	16	22	13
30	40	47	40		23	100		18	45	42	28	21
31						36		29	38		26	37
	39	36	31	29	32	39	25	45	70	40	35	26
per post												37,2

Legenda nr. commenti

0	50	100	150	200	250	300	>300
---	----	-----	-----	-----	-----	-----	------

Tabella8: Numero di commenti per post in un anno nel blog di Gilioli (Ago. 2010 Sett. 2009)

ottolina

frequenza commenti												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1							8				dati non disponibili	dati non disponibili
2						16				0		
3			12						0			
4						2						
5		11		3	43		0					
6	9			0	7							
7			1	1	0							
8					4							
9									0			
10				20			10					
11							9					
12				8	1		13	2				
13	28				3		1			0		
14		3		19	38							
15							1					
16	3	2		9	2							
17				13			0					
18			8			21						
19						5	0	0				
20	2		6				8		0			
21												
22							3					
23							11					
24	3					0						
25	12,5							0				
26				9			0	2				
27					10			20				
28	0			0	5	2		0				
29					29			7				
30				3								
31				34		12						
	8	5	7	10	13	8	5	4	0	0	-	-
per post												5

legenda nr. commenti							
0	50	100	150	200	250	300	>300

Tabella9: Numero di commenti per post in un anno nel blog di Ottolina (Ago. 2010 Sett. 2009)

rampini

frequenza commenti												
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1			8		0	1			4		1	
2		7	6	5	1	3			2	4	5	3
3			9	8		6	6		14	1	1	
4				3		4	6		7	8		10
5		4		3		7					31	
6		2		3		8					4	
7				5	15						1	
8			1	54	1	0		14			4	
9			1		1	1	2		7	3	7	0
10	0	1	3	5		1	0		52			3
11	7	9		1	1	0	1	2				1
12	2	2		14	3	1	2	3	9	3		10
13	7	4		9	7	2		1	3	2		
14		0	5		1			15	5	2		0
15		7		10	1	1	0		3	32		
16					2	1	0	1	1	3		1
17	2		3		0		11	27	3	4		
18	5	3	2	8	1	1		4	10	8		
19	6	5	8	1	1	3	22	1		18		
20	2				2	2		2			9	
21				5	4			4			0	2
22	1		7	0	0	5	2	6			1	4
23	2		12	1		2	0			2	1	7
24	1			6	5		1	5			0	4
25	13	2			1		1	7			2	8
26	4		7	11	2		3					3
27	7		8	1			0				2	
28	0	0	1	13	3		1	2			0	
29			2		6	0		2			0	
30	7		12		6	2		19			3	4
31	8			2		1					1	
	4,3	3,5	5,5	7,5	2,7	2,3	3,4	6,7	9,1	6,8	3,8	3,8
per post												5,0

Legenda nr. commenti

0	50	100	150	200	250	300	>300
---	----	-----	-----	-----	-----	-----	------

Tabella10: Numero di commenti per post in un anno nel blog di Rampini (Ago. 2010 Sett. 2009)

zucconi

	frequenza commenti											
	ago	lug	giu	mag	apr	mar	feb	gen	dic	nov	ott	set
1			53	264	80	31	232	128	137		202	107
2			110			95	41	139	104	143		147
3			48	124		52	96		193	144	150	93
4			87	108		190	223	170	269	170	497	272
5				43		106	139	109		164		243
6				125	50	345		145	380	174	177	
7				196	149		224	105	107	279	426	144
8					40	131	118	133			311	91
9				44	80	111	62	73	312	191		176
10				197	35	72	239	187	239	118	374	158
11				69	240	148	262		414	128		304
12				95	80	165		66		197	520	
13				98	37	74	173	80	245		122	
14		13		64	127	143	183	66	273	97	208	294
15		15			84	50		80	191	112	256	128
16		29		55	60	101	66	295	234	57		238
17	58	84		69	144	56	81		357	102		305
18	25			109	77	59	204	148				174
19	241	33		82	125	158	530	55	219		250	976
20	146	46		55		150		97	575	139	78	
21	77	159		114	35	106	86	97		85	169	
22	204			119	261	164	106	64			329	
23	56						69	160				561
24	37			63	52	241	73	151	163		215	444
25	29			48		181	90	91		124		
26	80	63		186	44	111	73	100	183		177	136
27	127	36		146	32	194	78	68	131	77	388	
28	171	31		109		97	179	65			187	
29	150	30		124	110	97		105	231		289	
30		44	25	99	260	150		72		167		222
31	91	33		113		352		100			211	
	107	47	64	108	100	135	151	112	248	140	263	263
per post												145

Legenda nr. commenti

0	50	100	150	200	250	300	>300
---	----	-----	-----	-----	-----	-----	------

Tabella 11: Numero di commenti per post in un anno nel blog di Zucconi (Ago. 2010 Sett. 2009)

Capitolo III: Analisi blog vs articoli

3.1 Corpora: criteri di scelta dei testi

Per rendere possibile questa analisi comparata fra i due tipi di testo, sono stati creati due corpus testuali distinti in formato xml di 50.000 parole ognuno: il primo composto da post di blog, il secondo da articoli di giornale, raccolti in ordine cronologico a partire dalla fine del mese di maggio 2010. I testi raccolti nei corpora sono stati scelti seguendo un criterio meccanico: sia per gli articoli di blog, sia per quelli giornalistici è stato sempre scelto (e poi inserito nel corpus) il primo di ogni pagina web risultata dalla ricerca su quel determinato autore. Per gli articoli in particolare la scelta degli articoli è avvenuta secondo questa modalità: deciso il giornalista sono stati cercati, negli archivi delle testate web, tutti gli articoli di cui questi è autore, e per ogni pagina di risultato è stato selezionato il primo.

Questo tipo di ricerca ha determinato la scelta di articoli scritti in un lasso di tempo di 2-7 anni, il primo articolo preso in esame di Assante è datato maggio 2003, l'ultimo maggio 2010; il primo articolo di Gilioli è di maggio 2006 l'ultimo è di maggio 2010; i termini per Ottolina sono giugno 2005, maggio 2010; per Rampini agosto 2008, maggio 2010 e infine per Zucconi luglio 2007 maggio 2010.

L'immagine seguente mostra il risultato della ricerca: di articoli di Assante sul sito di Repubblica, l'articolo scelto è il primo della seconda pagina di risultati, nello specifico, l'articolo numero 21 in ordine cronologico: *“Tutto pronto per il tablet pc l'ultima mania di casa Apple”*.



Figura 5: Articoli selezionati: il primo di ogni pagina di risultato. In questo caso la seconda.

3.2 Analisi linguistica

3.2.1 Lunghezza degli articoli e lunghezza dei post

Questo primo confronto fra gli articoli di giornale e i post di blog giornalistici mette in luce in maniera piuttosto netta la differenza tra i due tipi di testo per quanto riguarda, appunto, la loro lunghezza.

Prendendo in considerazione i corpora in cui sono stati raccolti tutti i testi utili allo studio, sono stati selezionati i primi 10 articoli e i primi 10 post di blog di ogni autore,

sono state conteggiate le parole di ognuno, è stata calcolata la lunghezza media sia degli articoli sia dei post (attraverso prima una funzione di conteggio parole e poi calcolando la media) e infine sono stati confrontati i risultati.

Lunghezza (in numero di parole) dei post blog dei 5 giornalisti

Assante	Gilioli	Ottolina	Rampini	Zucconi
55	472	260	729	152
420	217	965	340	182
806	16	272	198	206
87	54	187	92	52
123	73	733	52	444
1271	622	33	35	255
161	35	1085	200	124
314	224	625	38	266
219	31	74	35	149
110	18	431	465	312
Totale=3566	Totale=1762	Totale=4665	Totale=2184	Totale=2142
Media=355	Media=175	Media=465	Media=220	Media=215

Lunghezza (in numero di parole) degli articoli dei 5 giornalisti

Assante	Gilioli	Ottolina	Rampini	Zucconi
1194	742	557	767	1558
645	616	380	614	1014
391	728	1019	641	783
219	1563	265	64	1002
352	1171	350	632	996
460	1027	755	602	721
251	1133	306	870	1075
248	5582	679	204	991
1040	957	879	904	620
651	201	304	144	997
Totale=5481	Totale=13720	Totale=5494	Totale=5442	Totale=9757
Media=550	Media=1370	Media=550	Media=545	Media=975

Lunghezza media blog (n. parole)	Lunghezza media articolo (n. parole)
Media = 1430/5 = 286 ~ 285	Media = 3990/5 = 798 ~ 800

Mettendo a confronto ognuno dei risultati per ogni giornalista già è chiaro che, nonostante nei blog la lunghezza dei post sia piuttosto altalenante (si alternano post molto brevi, alcuni anche costituiti da meno di 20 parole, a post lunghi), la lunghezza media degli articoli è circa tre volte maggiore della media dei post di blog.

Anche facendo la somma delle medie per ogni autore il risultato propende verso la maggiore lunghezza degli articoli giornalistici. (Nel caso di Gilioli il risultato è eclatante:

175 la lunghezza media dei post contro 1370 che è la lunghezza media degli articoli; quasi dieci volte più lunghi).¹⁰

3.2.2 Numero di frasi complesse

Il secondo confronto fra articoli e post di blog riguarda la sintassi; in particolare si evidenzierà il numero di proposizioni complesse che compongono i testi.

La frase, o proposizione è l'unità minima di comunicazione dotata di senso compiuto. Tradizionalmente si considera dotata di “senso compiuto” una frase in cui si trovi almeno un predicato nella forma di un verbo di modo finito, che può essere accompagnato da un soggetto.¹¹

Un periodo, o frase complessa è sempre strutturato da una proposizione principale o reggente, cui sono collegate altre proposizioni. Ogni periodo grammaticale termina con un segno di punteggiatura forte: punto fermo(.), punto esclamativo (!) punto interrogativo (?).¹²

Innanzitutto sono state individuate le frasi complesse nei due corpora e marcate attraverso un elemento xml (<s> per inizio frase e </s> per fine frase); infine sono state contate automaticamente grazie ad una semplice funzione di ricerca.

Il risultato di questa analisi è visibile nella seguente tabella.

Numero frasi complesse corpus “blog”		Numero frasi complesse corpus “articoli”	
Totale	2113	Totale	1663

I post di blog risultano essere composti da un numero maggiore di frasi complesse (1994 nel primo caso, 1677 nel secondo).

¹⁰ Le grandezze sono state arrotondate alla metà più vicina

¹¹ **Frase** definizione da *Grammatica Italiana. Italiano comune e lingua letteraria*. L. Serianni. Par.2.9 “Frase, periodo, sintagma” (pag. 85)

¹² **Periodo** definizione da *Grammatica Italiana. Italiano comune e lingua letteraria*. L. Serianni. Par.2.9 “Frase, periodo, sintagma” (pag. 86)

3.2.3 Lunghezza delle proposizioni complesse

Per procedere alla prossima analisi è stato valutato un campione rappresentativo per le frasi: per ogni corpus sono state analizzate le prime 20 proposizioni dei testi scritti da ognuno dei cinque giornalisti, per un totale quindi di 100 proposizioni a corpus. Di ogni frase è stata conteggiata la lunghezza ed infine, dopo aver calcolato quella media, sono stati messi a confronto il numero di parole componenti una frase di articolo e quello di parole componenti una frase di post. I risultati nella seguente tabella.

Lunghezza media frasi del corpus "blog"						Lunghezza media frasi del corpus "articoli"					
assante	7	5	37	6	41	assante	8	37	31	10	5
	7	15	18	8	16		4	4	45	39	25
	22	23	21	4	8		23	30	43	32	32
	42	16	6	30	31		41	23	24	17	20
valore medio					18.5	valore medio					24.6
gilioli	65	29	23	32	27	gilioli	9	14	15	12	15
	58	37	27	58	37		7	20	11	3	14
	17	23	28	11	38		55	17	17	32	36
	23	3	16	17	15		25	37	54	48	11
valore medio					29.2	valore medio					22.6
ottolina	9	10	21	13	19	ottolina	36	24	30	34	58
	18	24	25	36	38		18	27	28	6	19
	26	24	15	5	9		45	41	41	19	17
	12	30	44	31	24		24	17	25	23	27
valore medio					21.6	valore medio					27.9
rampini	16	48	21	48	14	rampini	21	15	7	15	27
	18	13	29	35	21		17	20	40	27	28
	12	13	13	8	16		14	53	34	22	29
	36	48	28	61	23		26	16	14	20	17
valore medio					26	valore medio					23.1
zucconi	73	23	24	31	1	zucconi	15	60	14	37	78
	69	26	53	34	92		6	48	49	68	24
	45	11	59	39	14		60	38	41	49	9
	32	20	3	28	35		3	37	56	9	3
valore medio					35.6	valore medio					35.2
valore medio "blog"					26	valore medio "articoli"					27

Dall'analisi svolta sulle lunghezze delle frasi, la lunghezza media delle proposizioni complesse degli articoli e quella dei post di blog risulta essere, con piccolo scarto, simile in entrambe le tipologie testuali. Nei giornalisti Ernesto Assante, Alessandro Gilioli e Paolo Ottolina: tra la lunghezza media delle frasi dei loro articoli e quella dei loro post di blog c'è una differenza di 6-8 parole, mentre in Rampini il numero di parole delle proposizioni degli articoli è minore in media di 3 parole rispetto alle proposizioni dei post; In Zucconi invece la lunghezza media delle frasi è uguale.

Il valore totale medio nei blog, per quanto riguarda questa analisi, è molto vicino a quello degli articoli; si tratta di un dato molto interessante perché mette in luce la sostanziale somiglianza fra le due tipologie testuali. (Anche quanto la differenza è di 6-8 parole è chiaro che le frasi complesse dei blog e degli articoli hanno una lunghezza simile).

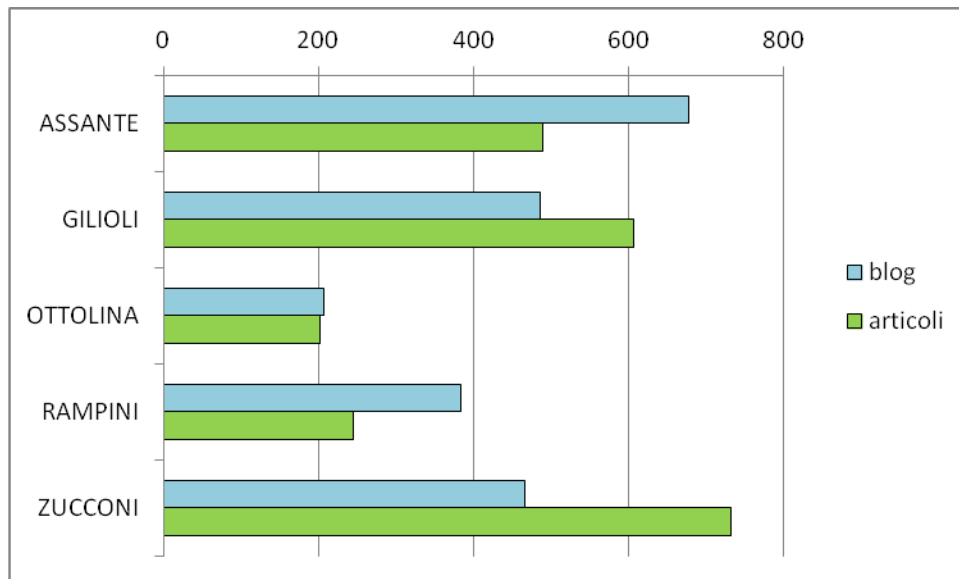
3.2.4 Analisi delle proposizioni subordinate

Quest'analisi riguarda il numero e il tipo di proposizioni subordinate presenti nei testi.¹³

Una proposizione subordinata si caratterizza per la mancanza di autonomia, ossia per la necessità di dipendere da una proposizione reggente che può essere una frase semplice o, a sua volta, rinviare ad un'altra proposizione.¹⁴

Dai corpora presi in esame sono risultati, dopo un'analisi preliminare i seguenti dati , ma si tratta di dati in fase di revisione.

Autore	Subordinate nei post	Subordinate negli articoli
ASSANTE	677	489
GILIOLI	485	606
OTTOLINA	206	201
RAMPINI	383	245
ZUCCONI	465	731



¹³ Non è stata condotta un'analisi sulle proposizioni principali e sulle proposizioni coordinate

¹⁴ **Proposizione subordinata** da *Grammatica Italiana*. Italiano comune e lingua letteraria. L. Serianni Par.15.30 "Subordinazione" (pag. 545)

Nel caso di Assante vi è una maggiore presenza di proposizioni subordinate nei post del suo blog piuttosto che negli articoli, così come Rampini. Gilioli, al contrario utilizza meno subordinate nel suo blog e in maggior numero nei suoi articoli; allo stesso modo procede Zucconi.

Il giornalista Ottolina invece resta costante nell'uso delle subordinate conservando la stessa percentuale in entrambi i generi testuali.

Inoltre, sono state analizzate le singole subordinate, valutandone il tipo.

Sono stati individuati 17 tipi di subordinate: **avversative, causali, comparative, concessive, consecutive, dichiarativa, finali, interrogative dirette e indirette, ipotetiche, limitative, modali, oggettive, relative, soggettive, strumentali e temporali.**

Di seguito vi sono le comparazioni (in percentuale) tra il numero di subordinate presenti nel blog e negli articoli esaminati per ogni giornalista

Tipo di subordinate	Assante		Gilioli		Ottolina		Rampini		Zucconi	
	Blog	Articoli	Blog	Articoli	Blog	Articoli	Blog	Articoli	Blog	Articoli
AVVERSATIVA	4,9	3,9	0,6	4,3	3,9	5,9	1,0	0,8	3,2	2,7
CAUSALE	6,5	3,9	9,7	3,8	1,9	6,4	4,2	2,4	6,2	4,4
COMPARATIVA	1,9	1,4	1,6	1,2	5,3	2,0	0,8	1,2	2,4	1,1
CONCESSIVA	0,4	0,6	0,4	0,5	0,0	0,5	0,3	0,4	0,9	0,0
CONSECUTIVA	1,6	0,6	0,4	0,2	0,0	0,0	0,8	0,4	0,4	0,1
DICHIARATIVA	2,2	6,1	4,7	3,3	1,0	5,4	4,2	2,8	3,2	2,3
FINALE	11,5	16,1	15,3	12,4	8,7	9,9	15,4	17,8	11,6	11,5
INTERR. DIR.	0,0	0,2	0,6	0,8	0,0	1,5	0,5	0,8	1,1	0,3
INTERR. INDIR.	0,1	0,8	0,2	0,7	0,5	0,5	0,3	0,8	0,4	0,0
IPOTETICA	3,5	2,0	11,8	7,1	7,8	4,5	6,0	3,6	6,5	4,2
LIMITATIVA	1,6	1,0	0,6	2,6	1,9	1,0	0,5	1,6	0,6	0,5
MODALE	4,1	2,2	2,3	1,2	1,5	0,0	3,4	0,4	3,4	3,0
OGGETTIVA	16,1	12,8	11,8	13,0	3,9	6,9	8,4	13,0	10,1	8,8
RELATIVA	32,2	38,1	31,8	30,7	49,5	45,5	41,0	41,3	40,6	47,5
SOGGETTIVA	4,6	2,6	2,5	2,8	3,4	1,5	3,1	1,2	2,4	2,1
STRUMENTALE	4,3	3,3	2,3	6,9	4,4	5,0	4,4	4,0	0,6	5,2
TEMPORALE	4,3	4,3	3,5	8,6	6,3	3,5	5,7	7,3	6,2	6,2

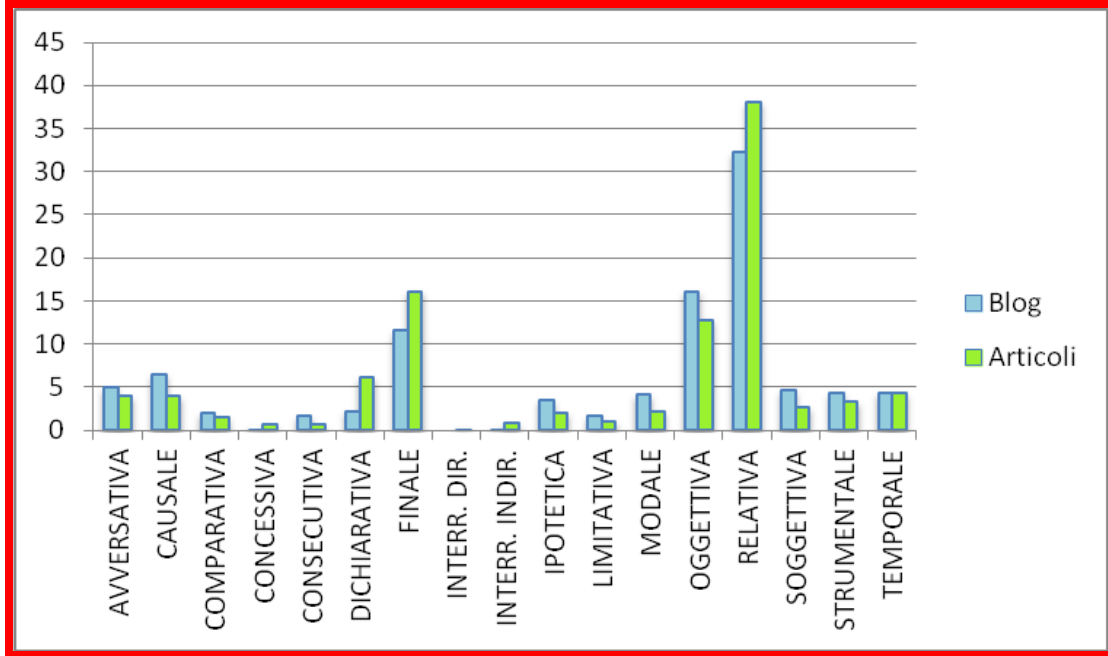


Grafico 1. Confronto post/articoli sui tipi di subordinata in Assante

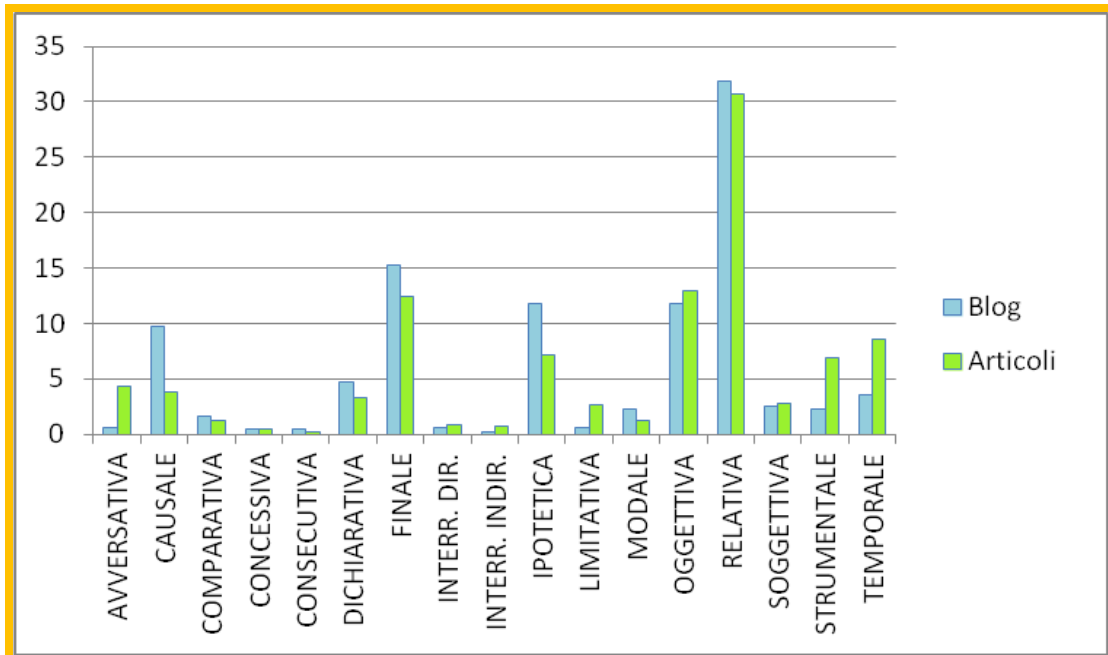


Grafico 2. Confronto post/articoli sui tipi di subordinata in Gilioli

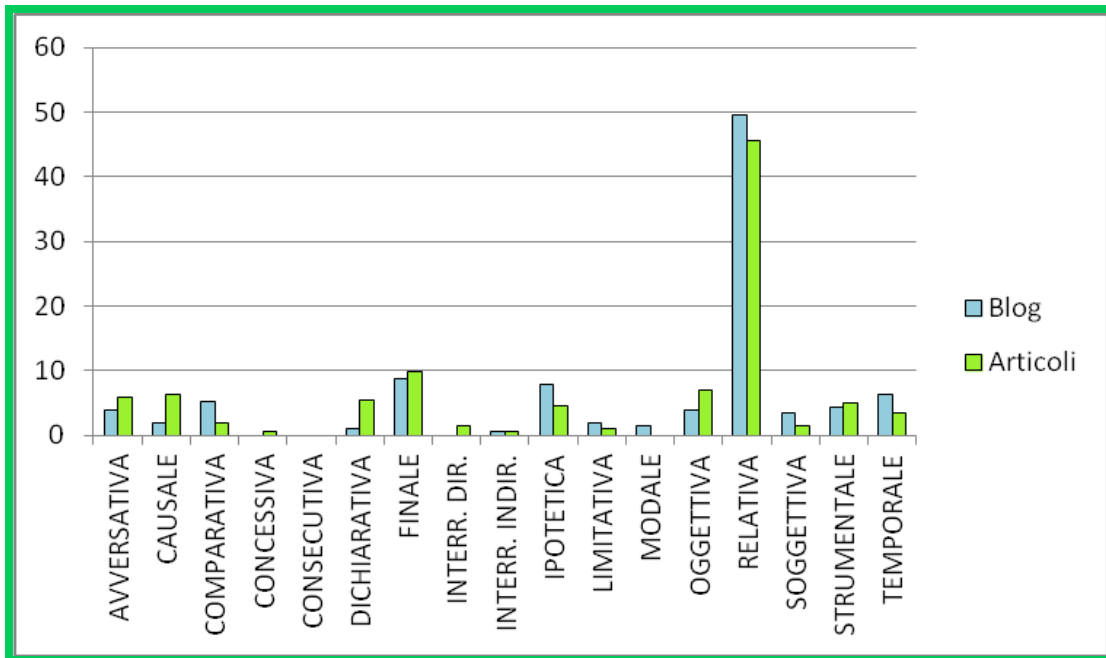


Grafico 3. Confronto post/articoli sui tipi di subordinata in Ottolina

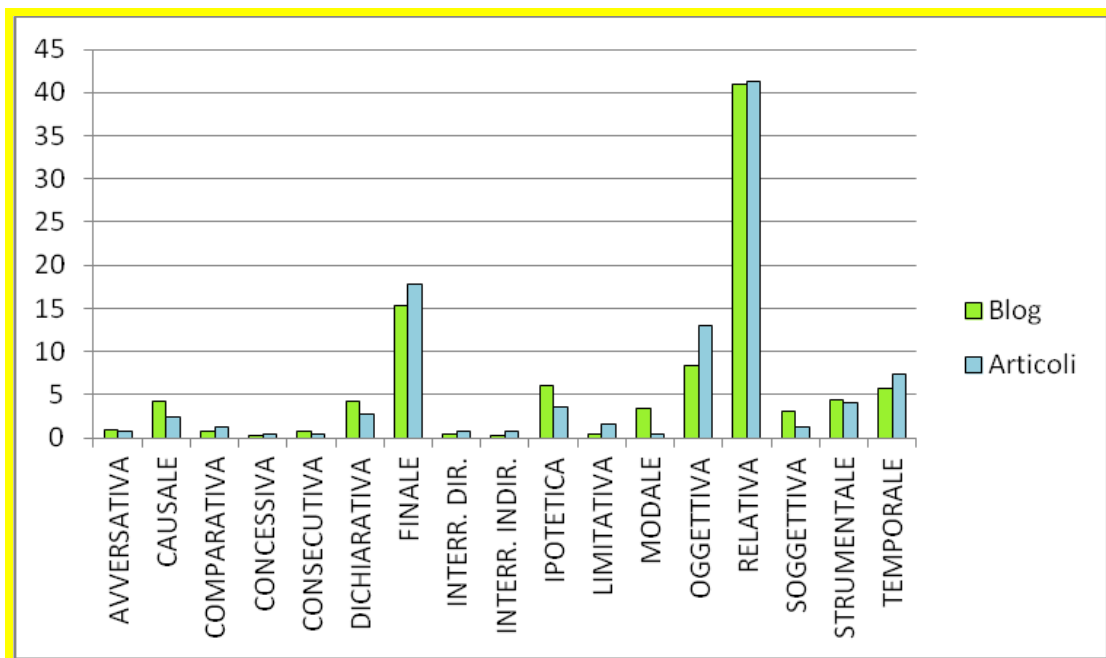


Grafico 4. Confronto post/articoli sui tipi di subordinata in Rampini

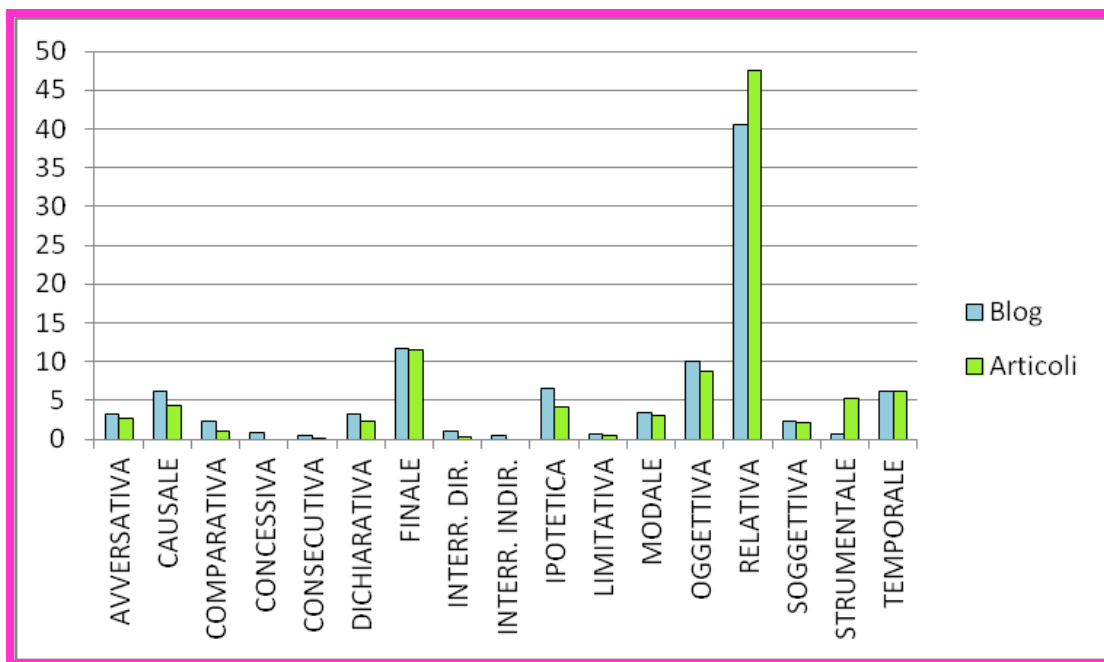


Grafico 5. Confronto post/articoli sui tipi di subordinata in Zucconi

In Assante si può rilevare un alto numero di proposizioni relative (tra il 30 e il 40% del totale), finali e oggettive, sia nei post che negli articoli. Le dichiarative negli articoli sono il triplo rispetto a quelle nei post. Al contrario, le ipotetiche sono in maggior numero nei post (rapporto 1 a 1,5); tranne nel caso delle dichiarative, finali e relative tutte le altre subordinate si trovano nei post in numero superiore rispetto agli articoli; la percentuale delle temporali è uguale in entrambi i generi testuali.

In Gilioli si nota la presenza di molte proposizioni relative, finali, ipotetiche e oggettive. Le avversative sono in maggior numero negli articoli (più di quattro volte maggiore), così come le strumentali e le temporali. Le causali del blog invece sono quasi il doppio di quelle degli articoli.

In Ottolina vi sono molte proposizioni relative nel blog esaminato; anche le avversative risultano parecchio presenti. Il dato interessante, nel caso di questo giornalista, è che il numero delle subordinate nel blog è sostanzialmente uguale a quello degli articoli.

In Rampini e Zucconi si trovano molte relative (nel caso di Zucconi quasi il 50% del totale, per quanto riguarda gli articoli). In Rampini c'è una leggera maggioranza dell'uso delle subordinate nel blog, mentre in Zucconi c'è una leggera predominanza delle suddette negli articoli, ma sostanzialmente l'andamento è pressoché costante con una percentuale simile di subordinate in entrambi i generi testuali.

Capitolo IV : Conclusioni

L'analisi condotta sui due corpus, quello composto da post di blog e quello composto di articoli di giornali, redatti da cinque giornalisti, ha permesso di delineare determinate conclusioni riguardo questi due generi testuali.

Nell'introduzione era stato anticipato come, erroneamente, i blog vengano considerati raccolte di testi poco sorvegliati; dalle analisi, basate sui confronti fra i post e gli articoli è emerso quanto determinate caratteristiche testuali e il linguaggio sia compatibile tra l'uno e l'altro genere di testo.

Analizzando uno per volta i confronti svolti durante questo lavoro si evidenzia però, una differenza sostanziale che riguarda la lunghezza del singolo testo: quella degli articoli è nettamente superiore a quella del post di blog.

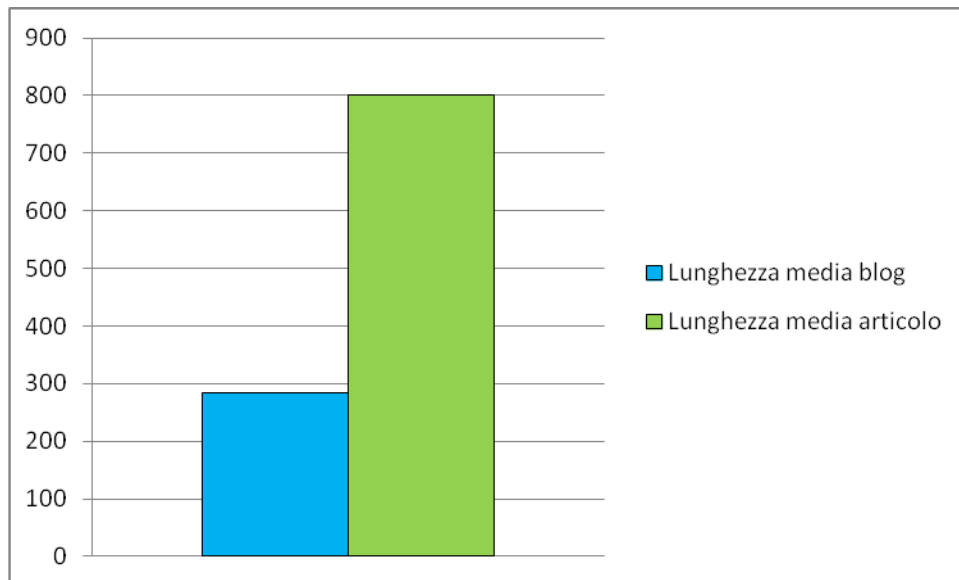


Grafico 6. Confronto fra la lunghezza media di un post e la lunghezza media di un articolo

Si tratta dell'unica differenza evidente nel confronto blog/articolo.

Proseguendo nella visualizzazione dei risultati il secondo confronto riguarda il numero di frasi complesse. Il risultato è visibile nel seguente grafico:

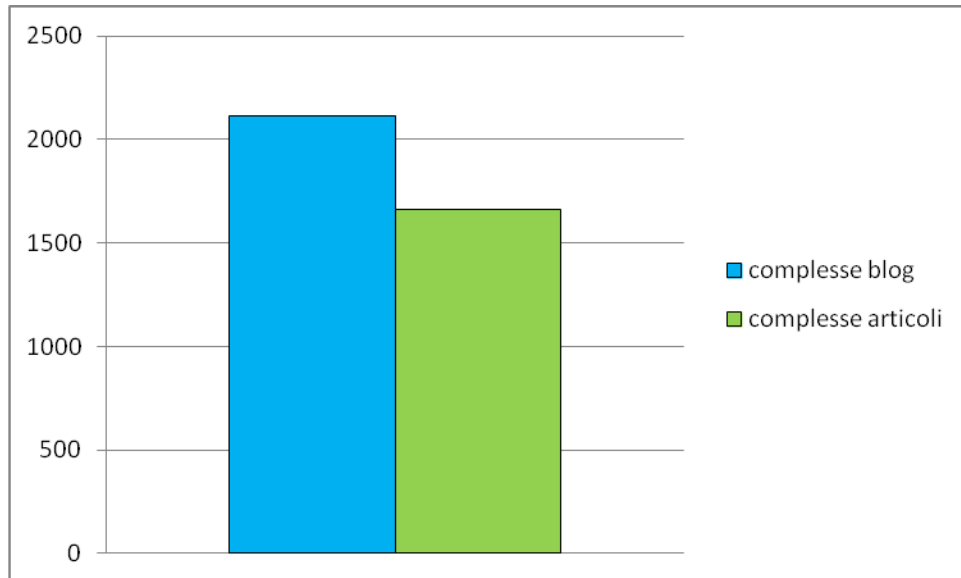


Grafico 7. Confronto fra numero di frasi complesse di un post e numero di frasi complesse di un articolo

La differenza non è molto alta (le frasi complesse negli articoli sono $\frac{3}{4}$ rispetto a quelle dei blog) ma comunque basta a sottolineare la caratteristica dei blog di essere composti da un numero di frasi che terminano col punto (.) o col punto esclamativo (!) o col punto interrogativo (?) piuttosto alto.

Il terzo risultato è quello più interessante e riguarda la lunghezza della frase complessa nei due generi testuali. L'analisi ha portato alla luce un risultato uguale (seppure con un piccolissimo scarto) fra il numero di parole che compongono la complessa dei post e il numero di parole che compongono la complessa dei blog.

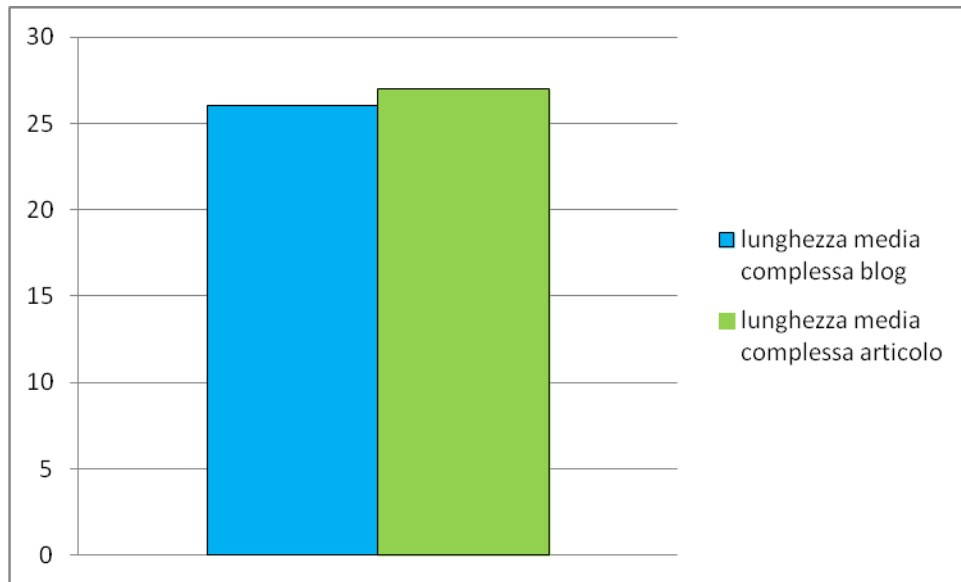


Grafico 8. Confronto fra lunghezza media delle complesse di un post e lunghezza media delle complesse di un articolo

Ed infine l'ultimo confronto riguarda l'analisi delle subordinate (per il dettaglio si rimanda al paragrafo 3.2.4). Il risultato generale è che il numero delle subordinate totali nel corpus del blog è molto vicino a quelle delle subordinate totali nel corpus degli articoli con uno scarto di circa 2,5%.

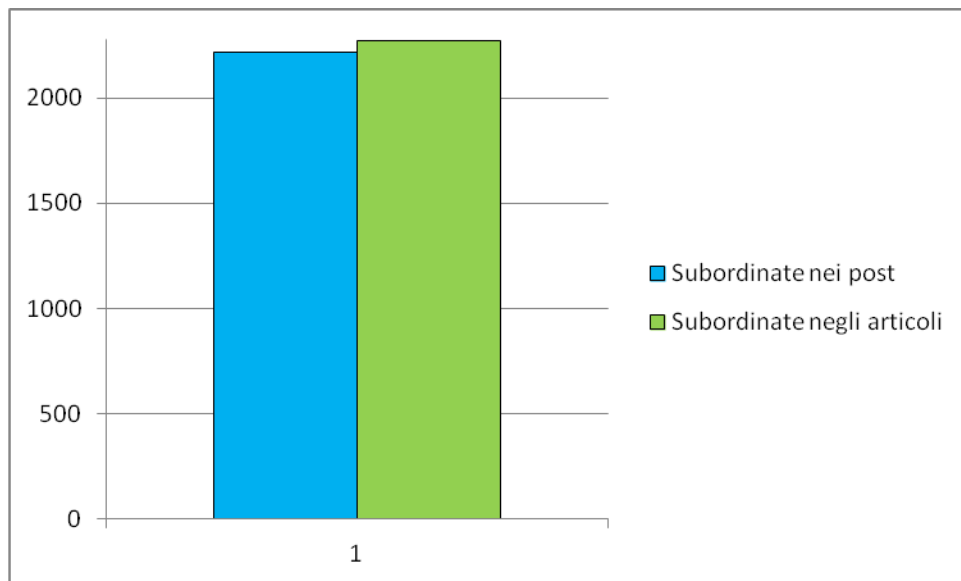


Grafico 9. Confronto sul numero di subordinate fra i post e gli articoli

La conclusione di questa tesi, dunque, mette in luce come i blog tematici, quelli presi in considerazione in questo lavoro, sintatticamente sono simili agli articoli di giornale, e tranne per la lunghezza del pezzo (il blog è molto più corto) le caratteristiche permettono di considerare il blog un genere testuale programmato, revisionato e adatto alla pubblicazione (sul web) proprio come un articolo di giornale.

Capitolo V: Bibliografia

Canobbio A.T. 2005. *Blog: la lingua che uccide*, in “LID’O – Lingua italiana d’oggi”

D’Ottavi A., Sorchiotti T. 2008. *Come si fa un blog 2.0*. Tecniche nuove.

Lipperini L. 2004. *La notte dei blogger. La prima antologia dei nuovi narratori della rete*. Einaudi.

Rossetti S. 2010. *La frase complessa nel Corriere della Sera. Codifica sintattica e analisi diacronica (1946-2006)*, tesi di laurea quadriennale, Università di Pisa.

Serianni L. 1988. *Grammatica Italiana: italiano comune e lingua letteraria*. UTET.

Stubbe A., Ringlsetter C, Schulz K.U. 2007. *Genre as Noise-Noise in Genre*, in *Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-2007) – Workshop on Analytic for Noisy Unstructured Text Data*, IAPR-IBM Research, Hyderabad.

Tavosanis M. 2006a. *Are blogs edited? A linguistic survey of Italian blogs using search Engines* in *Proceedings of the Computational Approaches to Analyzing Weblogs Conference*, AAAI.

Tavosanis M. 2006b. *Linguistic features of Italian blogs: literary language in Proceedings of the workshop on NEW TEXT: Wikis and blogs and other dynamic text sources*. EACL.

Tavosanis M. 2007. *Juvenile Netspeak and Subgenre Classification Issues in Italian Blogs*, in *Proceedings of the International Workshop towards Genre-Enable Search Engines: The Impact of NLP*, workshop held in conjunction with RANLP-2007, Borovets. RANPL.

Tavosanis M. 2011. *L’italiano del web*. Carocci Editore.

Webgrafia per il profilo degli autori

E. Assante:

- LinkedIn *Ernesto Assante*

<http://it.linkedin.com/pub/ernesto-assante/3/87b/7a9>

- Armando Adolghiso

http://www.adolghiso.it/enterprise/ernesto_assante.asp

- Wikipedia, voce *Gino Castaldo*

- http://it.wikipedia.org/wiki/Gino_Castaldo

A. Gilioli

- Piovono rane
<http://gilioli.blogautore.espresso.repubblica.it/chi-sono/>
- Wikipedia, voce *Alessandro Gilioli*
http://it.wikipedia.org/wiki/Alessandro_Gilioli
- Nextopeninnovation
<http://www.nextopeninnovation.it/site/events/innovationcafe/intervista27.html>
- Digicult
<http://www.digicult.it/digimag/article.asp?id=1542>

P. Ottolina

- LinkedIn *Paolo Ottolina*
<http://www.linkedin.com/pub/paolo-ottolina/36/469/702>
- Malditech
<http://malditech.corriere.it/>

F. Rampini

- Wikipedia, voce *Federico Rampini*
http://it.wikipedia.org/wiki/Federico_Rampini
- Il sussidiario
<http://www.ilsussidiario.net/Intervistati/R/434/Luca-Ricolfi/R>
- Ustation
<http://www.ustation.it/contenuti/audio/107-scientiati-d-orient>
- Quadrifoglio
<http://www.quadrifoglio.org/oggetti/37142.pdf>

V. Zucconi

- Wikipedia, voce *Vittorio Zucconi*
http://it.wikipedia.org/wiki/Vittorio_Zucconi