



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

TESI DI LAUREA

**Un'analisi linguistica del microblog e social
network Twitter**

Candidato: *Lorenzo Mori*

Relatore: *Mirko Tavosanis*

Correlatore: *Franco Turini*

Anno Accademico 2010-2011

A chi c'è stato.

Indice generale

1. Introduzione	1
1.1 L'era dei microblog	2
1.2 L'obiettivo del nostro lavoro	5
2. Twitter	6
2.1 Uno sguardo a Twitter	6
2.2 Storia ed evoluzione	8
2.3 Funzioni Principali	10
2.3.1 Tweet	10
2.3.2 Retweet	13
2.3.3 Hashtag	15
2.3.4 Liste	15
3. Un'analisi linguistica di Twitter	17
3.1 Criteri per la selezione degli utenti	17
3.2 Criteri per la raccolta dei tweet	21
3.3 L'analisi del corpus	22
3.3.1 Uno sguardo approfondito ai tweet	22
3.3.2 L'espressività grafica del corpus	25
3.3.3 Le emoticon	25
3.3.4 La punteggiatura	34
3.3.5 Il lessico	35
3.3.6 Le abbreviazioni e le variazioni	35
3.3.7 Gli errori	42

4. L' analisi dei dati con Excel	47
4.1 Criteri di classificazione	47
4.2 Le correlazioni tra i diversi fenomeni	50
Conclusioni	54
Bibliografia	56

1. Introduzione

Questo lavoro di tesi è dedicato a un'analisi linguistica dell'italiano scritto utilizzato sul microblog e social network Twitter¹.

Lo scopo di questo lavoro è quello di descrivere le caratteristiche dell'italiano utilizzato all'interno di Twitter, verificando se questo appartiene all'italiano standard oppure se abbia subito l'influsso dell'evoluzione tecnologica della comunicazione, adattandosi quindi a quello che è il linguaggio della Rete, fatto di una comunicazione informale e costituita da abbreviazioni, emoticon, sostituzioni di caratteri, simboli da leggere in modo endofasico (come 6 per "sei") e così via.

I quattro capitoli di cui si compone sono così organizzati:

Nel primo sarà inizialmente presentato al lettore un quadro generale dei social media² e della loro diffusione nell'epoca di oggi, quella del Web 2.0³.

Nel secondo ci concentreremo direttamente su Twitter. Ricostruiremo brevemente la sua storia, dalla sua nascita alla sua affermazione, e vedremo le sue principali caratteristiche e funzioni, per renderci conto di che cosa offra oggi questo servizio ai suoi utenti.

Nel terzo capitolo ci concentreremo su quello che è il vero progetto di questa tesi, ovvero un'analisi linguistica di alcuni messaggi di testo raccolti all'interno di Twitter.

¹ <http://www.twitter.com>

² «Social media è un termine generico che indica tecnologie e pratiche online che le persone adottano per condividere contenuti testuali, immagini, video e audio. I professori Andreas Kaplan e Michael Haenlein definiscono social media come un gruppo di applicazioni Internet basate sui presupposti ideologici e tecnologici del Web 2.0 che consentono la creazione e lo scambio di contenuti generati dagli utenti. I social media rappresentano fundamentalmente un cambiamento nel modo in cui la gente apprende, legge e condivide informazioni e contenuti. In essi si verifica una fusione tra sociologia e tecnologia che trasforma il monologo (da uno a molti) in dialogo (da molti a molti) e ha luogo una democratizzazione dell'informazione che trasforma le persone da fruitori di contenuti ad editori. I social media sono diventati molto popolari perché permettono alle persone di utilizzare il web per stabilire relazioni di tipo personale o lavorativo». (Wikipedia, voce *Social Media* http://it.wikipedia.org/wiki/Social_media (visitato il 20 luglio 2011).

³ Con Web 2.0 si indica uno stato di evoluzione di Internet, e in particolare del World Wide Web, rispetto al precedente Web 1.0, diffuso fino agli anni novanta e composto prevalentemente da siti web statici, caratterizzati da una minima interazione con l'utente. Con Web 2.0 si vuole identificare l'insieme di tutte quelle applicazioni online che permettono un notevole livello di interazione sito-utente, come i blog, i microblog, i forum, le chat, i social network ecc.

Infine, nel quarto capitolo, verrà spiegata l'analisi dei dati svolta con Microsoft® Excel e saranno illustrate le eventuali correlazioni trovate all'interno del corpus analizzato.

Nello specifico si analizzeranno 2.895 messaggi, prelevati da 160 utenti suddivisi nelle categorie di *utenti professionisti* e *utenti non professionisti*, cercando di rispondere ad alcune domande:

esiste un linguaggio specifico di Twitter? Come ci si comporta, per quanto riguarda la scrittura, di fronte al limite dei 140 caratteri imposti dal servizio? Questo spazio influenza la modalità di scrivere e di comunicare o si tende sempre a utilizzare un italiano standard? E ci sono delle differenze ortografiche e grammaticali tra i diversi tipi di utenti? Nello specifico, ci sono differenze tra gli utenti che utilizzano il servizio come uno strumento professionale e coloro che lo utilizzano nel tempo libero? E se sì quali sono e perché possiamo trovarle?

Cercheremo di rispondere a tutte queste domande attraverso queste pagine e attraverso i numerosi esempi contenuti all'interno di esse.

1.1 L'era dei microblog

In brevissimo tempo Internet è entrato a far parte della nostra vita diventando uno strumento d'uso quotidiano che ha cambiato notevolmente le nostre abitudini e il nostro modo di comunicare. Si pensi a quante persone utilizzano giornalmente la posta elettronica, effettuano chiamate tramite la tecnologia VoIP⁴, consultano il giornale online o controllano i loro risparmi attraverso l'home banking⁵. Milioni, se non miliardi. Quello di oggi è senz'altro un mondo digitale dove la popolazione ha un rapporto così stretto con Internet che è difficile immaginare una vita senza di esso.

Secondo i dati Audiweb⁶ di Aprile 2011 sono stati 26,6 milioni gli utenti attivi italiani che si sono collegati almeno una volta a Internet, per un'audience

⁴ Voice over IP (Voce tramite protocollo Internet). È una tecnologia che permette di telefonare attraverso una connessione Internet.

⁵ Con home banking si definiscono quei servizi bancari che consentono al cliente di effettuare operazioni bancarie da casa mediante una connessione Internet.

⁶ Dati riferiti al mese di aprile 2011 e prelevati dal sito web <http://www.audiweb.it>

online giornaliero medio di 13 milioni di utenti, che hanno speso 1 ora e 18 minuti collegati alla rete consultando 158 pagine per persona.

Nel giorno medio gli uomini italiani online sono 7,4 milioni; nel dettaglio il 46,1% degli utenti attivi è tra i 35 e i 54 anni, il 18,4% è tra i 55 e i 74 anni, il 17,44% tra i 25 e i 34 anni e il 10% tra i 18 e i 24 anni.

Le donne online invece sono 5,6 milioni; nel dettaglio il 45,9% delle donne online è tra i 35 e i 54 anni, il 22,4% è tra i 25 e i 34 anni, il 20,5% è tra i 55 e i 74 anni e il 12,2% è tra i 18 e i 24 anni.

FIGURA 1.1 Riepilogo dell'Internet audience relativo al mese di aprile 2011. Dati forniti da Audiweb – Nielsen.

Internet audience, genere / Età – variazione 2011 / 2010						
	Attivi nel giorno medio, UOMINI aprile 2011 (.000)	Attivi nel giorno medio, UOMINI aprile 2010 (.000)	Var % Apr 2011 / Apr 2010	Attivi nel giorno medio, DONNE aprile 2011 (.000)	Attivi nel giorno medio, DONNE aprile 2010 (.000)	Var % Apr 2011 / Apr2010
Tutte le fasce	7.418	6.594	12,5%	5618	5.050	11,2%
2-10 anni	114	106	7,5%	106	81	30,9%
11-17 anni	405	333	21,6%	381	360	5,8%
18-24 anni	742	687	8,0%	653	570	14,6%
25-34 anni	1.294	1.261	2,6%	1370	1.213	12,9%
35-54 anni	3.426	3.006	14,0%	2.580	2.372	8,8%
55-74 anni	1.366	1.154	18,4%	517	447	15,7%
Oltre 74 anni	71	47	51,5%	12	7	71,4%

Se passiamo ad analizzare cosa fanno le persone su Internet ci accorgiamo che un ruolo fondamentale è svolto dai *social network* e dai *microblog*, poiché «(...) nel 2003 il tempo medio passato su social network e blog era di poco superiore al quarto d'ora, per superare l'ora nel 2005, arrivare a due nel 2007, toccare le tre e mezzo nel 2008⁷».

I social network sono stati definiti come «servizi *web-based* che permettono agli individui di costruire un profilo pubblico o semi pubblico all'interno di un sistema delimitato, articolare una lista di altri utenti con cui condividere una connessione e vedere e attraversare le loro liste di connessione e quelle fatte da altri all'interno del sistema». Quello che in sostanza li caratterizza, più che la capacità di promuovere ulteriori relazioni, è che «rendono capaci gli utenti di articolare e rendere visibili le loro reti sociali⁸».

I microblog invece sono dei servizi sempre basati sul web - *web-based* – dov'è possibile pubblicare «piccoli contenuti sotto forma di messaggi di testo (normalmente fino a 140 caratteri), immagini, video, audio MP3 ma anche segnalibri, citazioni, appunti. Questi contenuti vengono pubblicati in un servizio di social network, visibili a tutti o soltanto alle persone della propria community⁹».

Social network e microblog sono diventati oggi un servizio molto utilizzato dagli utenti di Internet nella comunicazione digitale, soprattutto grazie alla presenza dei nuovi media, che permettono di accedere a questi servizi anche dalla televisione o dal telefono cellulare.

Scambiarsi informazioni, notizie recenti, condividere le proprie esperienze personali, o semplicemente comunicare con gli amici, è diventato quindi molto più facile nell'era del Web 2.0, proprio grazie alla possibilità di scrivere contenuti da qualsiasi dispositivo connesso ad Internet.

Gli utenti si trovano così a raccontare e a condividere con tutto il mondo della rete numerose informazioni personali circa la loro vita, le loro opinioni sui prodotti, sulla politica, sulle istituzioni, sullo spettacolo; il tutto porta quindi alla creazione di

⁷ Giuseppe Riva, *I Social network*, pp. 7 e 8.

⁸ Journal of Computer-Mediated Communication <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html> (visitato il 20 luglio 2011).

⁹ Wikipedia, voce *Microblogging* <http://it.wikipedia.org/wiki/Microblogging> (visitato il 20 luglio 2011).

un grandissimo numero di messaggi contenenti informazioni e preferenze personali. Con questo risulta evidente che i servizi di social networking e microblogging sono una fonte appetibile per chi opera nel campo delle indagini di mercato e della sociologia; tramite questi contenuti, distribuiti liberamente sul web, è infatti possibile analizzare e trarre informazioni sui singoli utenti e sui loro comportamenti.

1.2 L'obiettivo del nostro lavoro

Come detto in precedenza l'obiettivo di questo lavoro è quello di analizzare un corpus di messaggi di testo, inseriti da due gruppi di utenti - quello degli *utenti professionisti* e quello degli *utenti non professionisti* - all'interno del microblog e social network Twitter.

Per *utenti professionisti* intendiamo coloro che utilizzano il proprio profilo Twitter per divulgare informazioni sulla propria attività e che sono conosciuti a livello nazionale o internazionale. Tutti quelli che non rientrano in questa categoria saranno invece considerati come *utenti non professionisti*.

Una volta raccolto il corpus di messaggi lo scopo è quello di procedere ad un'analisi linguistica degli stessi, al fine di delineare il linguaggio utilizzato dagli utenti all'interno di questo microblog, cercando di capire se questi facciano uso o meno del linguaggio giovanile, della scrittura endofasica - o sostituzione grafica - delle abbreviazioni, delle emoticon e di altri fenomeni di questo tipo.

Allo stesso tempo cercheremo anche gli errori ortografici, focalizzandoci su quelli dovuti al livello d'istruzione dell'utente, e quindi a dei suoi limiti, e quelli involontari, dovuti invece a errori di battitura o a un inadeguato tipo di supporto utilizzato per la scrittura.

2. Twitter

2.1 Uno sguardo a Twitter

Nato nel 2006 da un'idea di Jack Dorsey, Twitter può essere descritto come un sistema di comunicazione digitale, più propriamente chiamato microblog e social network, attraverso il quale è possibile comunicare con milioni di persone nel mondo¹⁰ inviando e ricevendo notizie sotto forma di messaggi. Il nome *Twitter* deriva dal verbo inglese *to tweet* che significa *cinguettare*.

Twitter fornisce ai propri iscritti una pagina personale aggiornabile attraverso dei messaggi di testo detti *tweet*, mostrati istantaneamente nella pagina del profilo dell'utente e non aventi lunghezza superiore a 140 caratteri.

Il limite dei 140 caratteri fa sì che Twitter venga considerato un sistema multiplatforma, poiché permette l'aggiornamento, oltre che dal computer, anche dal cellulare e dai servizi di messaggistica istantanea.

FIGURA 2.1 Alcuni tweet pubblicati dall'utente e disponibili nella sua cronologia.



¹⁰ Nel marzo del 2011 il numero delle persone presenti su Twitter era stimato intorno ai 200 milioni di utenti. (BBC <http://www.bbc.co.uk/news/business-12889048C>)

I messaggi di Twitter possono essere gestiti, ovvero è possibile scegliere le persone da cui ricevere gli aggiornamenti e a cui mostrarli, creando così una rete sociale di contatti organizzata in base alle proprie preferenze.

L'atto di seguire qualcuno è chiamato *following*, mentre le persone che hanno deciso di visualizzare i nostri aggiornamenti sono detti *follower*.

Una delle caratteristiche che differenziano Twitter dagli altri social network, spiega Federica Dardi¹¹, è che «il legame tra le persone su Twitter può non essere reciproco», ovvero è possibile seguire un utente senza che lui ci segua, così come è possibile essere seguiti da qualcuno che decidiamo di non seguire. «Non si tratta quindi né di amici, né di fan, ma di soggetti che decidono di seguire qualcuno, follower, o che vengono seguiti per la qualità dei contenuti pubblicati, following (...)»¹²

Ogni volta che qualcuno inserisce un messaggio questo compare nella Cronologia o Timeline contenuta nella pagina Home, che racchiude l'elenco di tutti i tweet pubblicati dai nostri following. Nella pagina Home è possibile visualizzare anche altre informazioni circa le menzioni, i retweet e le liste, di cui parleremo più avanti (v. 2.3.1, 2.3.2, 2.3.3, 2.3.4).

FIGURA 2.2 Alcuni tweet pubblicati dagli utenti seguiti dall'autore e disponibili nella cronologia.



¹¹ Federica Dardi, *Twitter. Guida per il professionista delle chiacchiere digitali*, p. 4

¹² Luca Conti, *Comunicare con Twitter*, p. 12

Tra le differenze con gli altri social network una che sicuramente ci capiterà di notare è l'impossibilità di scrivere sulla pagina di un'altra persona, come per esempio avviene su Facebook; questo perché su Twitter ognuno gestisce la propria pagina personale e se si vuole contattare un utente è possibile farlo o tramite la menzione, inserendo nel corpo del messaggio @nomeutente, o tramite la risposta, cliccando sul tasto *Rispondi* posto all'interno della finestra del messaggio, o tramite un messaggio diretto.

Un messaggio diretto è un tweet privato, che rispetta sempre il limite massimo dei 140 caratteri, inviato ad un determinato utente e visibile soltanto da quest'ultimo. A differenza della menzione e della risposta, che possono essere inviati sia ai nostri following che ai nostri follower, i messaggi diretti possono essere inviati soltanto ai propri follower, cioè a tutti quegli utenti che ci seguono. Nel caso volessimo comunicare con un utente che non ci segue è possibile usare, come detto in precedenza, la @menzione, che però rimane pubblica e visibile a tutti gli altri utenti.

2.2 Storia ed evoluzione

Twitter nasce nel 2006 a San Francisco, in una società chiamata Odeo e gestita da Jack Dorsey, Biz Stone ed Evan Williams.

Durante un brainstorming¹³ Jack Dorsey introdusse l'idea di creare un servizio di comunicazione via SMS per collegare tra loro gruppi di persone e consentire così la condivisione di informazioni. (Per permettere di anteporre al testo il nome dell'autore e rientrare nei 160 caratteri usati dagli SMS fu scelto di usare un limite massimo di 140 caratteri).

Questo progetto fu chiamato inizialmente Stat.us; successivamente fu rinominato in Twitter, ma dato che il team voleva adattare il servizio all'uso sul

¹³ Il brainstorming, letteralmente tempesta cerebrale, è una tecnica di creatività di gruppo utilizzata per trovare una soluzione a un determinato tipo problema. Lo scopo di questa tecnica è quello di far proporre ad ogni persona una possibile soluzione senza che questa sia criticata a priori, per poi scegliere quella maggiormente convincente.

cellulare fu rinominato in Twtr, perché, secondo Dorsey questo nome, «è facile da digitare con i tasti del cellulare 8-9-8-8-7 corrispondenti alle lettere t-w-t-t-r»¹⁴.

La prima versione, interamente basata sul web, fu inaugurata il 21 marzo 2006 quando Dorsey pubblicò il primo tweet della storia, «*just setting up my twtr*»¹⁵.

Il primo anno Twtr non riscuote molto successo perché la gente si chiede come possa essere utilizzato e quale sia la sua utilità. Nel frattempo i creatori decidono di riportare il nome del servizio a Twitter. Durante il festival musicale e cinematografico South by Southwest del 2007, tenutosi negli Stati Uniti, per Twitter arriva la svolta. Il servizio infatti conosce un fortissimo aumento di popolarità quando due grandi schermi vengono utilizzati esclusivamente per far scorrere lo stream dei tweet pubblici. I partecipanti iniziano a conoscere Twitter e a scambiarsi messaggi di informazioni e opinioni sul festival tramite i loro cellulari, facendo così una grandissima pubblicità a Twitter, che diviene in breve tempo un servizio molto popolare tra gli utenti della rete.

Il successo, e il notevole incremento di utenti, porta i creatori di Twitter ad apportare delle modifiche al loro microblog. Vengono introdotti i feed RSS¹⁶, gli hashtag¹⁷, la possibilità di accorciare i link per facilitarne l'inserimento nei 140 caratteri disponibili, la possibilità di condividere, oltre al testo, anche foto, video e musica, ma soprattutto vengono rilasciate le API¹⁸ pubbliche, per permettere lo sviluppo di numerose applicazioni open source da integrare con Twitter.

Il valore di Twitter cresce ulteriormente in seguito a recenti fatti di cronaca, come i disordini scoppiati in Iran dopo le elezioni presidenziali del 2009, in cui i manifestanti riuscirono grazie al servizio a superare il regime di censura dei media

¹⁴ «*we wanted to focus on mobile, so we dropped the vowels to get the SMS shortcode 89887 (TWTR)*» Jack Dorsey.

¹⁵ «sto configurando il mio twtr» Jack Dorsey, 21 marzo 2006.

¹⁶ Un feed RSS è uno strumento che consente di ricevere automaticamente informazioni abbonandosi ad una determinata fonte

¹⁷ Sono delle parole chiave inserite all'interno di un tweet e precedute dal simbolo #. Vengono utilizzati per facilitare la ricerca di contenuti aventi come argomento quella parola chiave.

¹⁸ Le API, acronimo di Application Programming Interface (Interfaccia di programmazione di un'applicazione) permettono di espandere le funzionalità di un programma. Rendendole pubbliche si permette quindi ad altri utenti di interagire con il software e di estenderne le sue funzioni e le sue caratteristiche.

tradizionali e a fornire testimonianze e fotografie dell'eventi; come il terremoto in Abruzzo del 2009, dove l'avvertimento della prima scossa fu comunicato prima su Twitter che dalle agenzie di stampa; come l'attuale rivolta in Libia, dove Twitter è al momento usato dai libici per fornire notizie in tempo reale sugli ultimi avvenimenti, o come la curiosa notizia della cattura di Bin Laden, dove un informatico pakistano con un tweet sul proprio profilo¹⁹ si meravigliava della presenza straordinaria di un elicottero sopra la sua abitazione, senza rendersi conto che in realtà stava descrivendo il blitz della cattura del noto terrorista.

Tutto questo mostra come Twitter, nato inizialmente per comunicare aggiornamenti di stato tra un ristretto gruppo di persone, sia diventato in breve tempo uno strumento di informazione globale e di giornalismo, dove riversare e da cui attingere informazioni pubbliche di qualsiasi genere. Inoltre non va dimenticata la sua facilità d'uso e la sua grande possibilità di accesso da qualsiasi dispositivo connesso a una rete, che ha portato la crescita degli utenti dai circa 20.000 iniziali del 2007 ai circa 200.000.000 del 2011.

2.3 Funzioni Principali

2.3.1 Tweet

Ogni messaggio che viene pubblicato su Twitter si chiama *Tweet*, che tradotto dall'inglese significa *cinguettio*. Ogni volta che un tweet viene pubblicato appare immediatamente sulla bacheca dell'autore e nella sezione cronologia²⁰ di tutti gli utenti che lo stanno seguendo.

La lunghezza massima di ogni tweet, come abbiamo visto in precedenza, è di 140 caratteri inclusi gli spazi; il tweet può essere inviato da qualsiasi dispositivo connesso alla rete, sia esso un computer, un tablet o un cellulare. Il limite di 140 caratteri non può essere superato: se si supera tale limite il contatore dei caratteri diventa rosso e di numero negativo, e non è possibile inviare il messaggio fin tanto che non si rientra nel numero di caratteri consentito. Proprio per massimizzare lo

¹⁹ <https://twitter.com/#!/ReallyVirtual/>

²⁰ In cui vengono visualizzati i tweet di tutte le persone che seguiamo (i nostri following).

spazio a disposizione sono nati numerosi servizi di riduzione di indirizzi web²¹, chiamati anche *shorten link* o accorciatori di url²², che permettono di risparmiare il numero di caratteri dovuti ai classici indirizzi URL trasformando indirizzi lunghi in indirizzi brevi.

Originariamente Twitter chiedeva di inserire il proprio messaggio nella casella di testo rispondendo alla domanda *What are you doing?*²³.

FIGURA 2.3 La vecchia interfaccia di Twitter



Inizialmente quindi le risposte riguardavano la sfera personale e venivano fornite informazioni su cosa si era fatto durante la giornata. Col tempo poi, e con la diffusione del servizio, gli utenti hanno iniziato a raccontare e a condividere tutto quello che ritenevano importante, non solo aspetti della propria vita ma anche notizie, aggiornamenti, consigli su prodotti, richieste di supporto, foto, video e tanti altri generi di informazione. Per questo, nel 2009, Twitter ha sostituito la domanda originaria *What are you doing* con quella *What's happening?*²⁴, che poi, a partire dall'11 dicembre 2009, è stata tradotta insieme a tutto il resto dell'interfaccia anche in francese, italiano, tedesco e spagnolo, oltre alle lingue inglese e giapponese che erano già presenti.

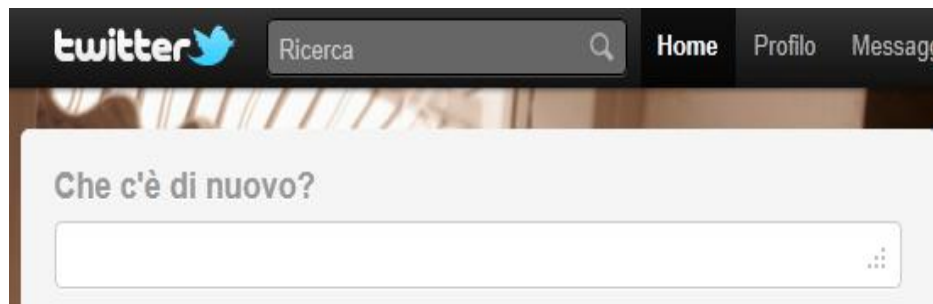
²¹ Per citarne alcuni, è il caso di www.bit.ly, www.tinyurl.com o www.OW.ly.

²² L' URL, *Uniform Resource Locator*, è l' indirizzo univoco di un sito web che inizia in genere con la stringa <http://>.

²³ *Che cosa stai facendo?*

²⁴ *Che c'è di nuovo?*

FIGURA 2.4 La nuova interfaccia di Twitter in italiano



Luca Conti nel suo libro *“Comunicare con Twitter. Creare relazioni, informarsi, lavorare.”* suddivide i diversi tipi di tweet in:

- Tweet con link

È il modo di pubblicazione di tweet più comune. Consiste nel pubblicare un messaggio aggiungendo un link dove verificare o approfondire quello che abbiamo appena scritto, questo perché è impossibile trattare un argomento in maniera estesa nei 140 caratteri a disposizione.

- Tweet tradizionale: che cosa sto facendo adesso

È il tipo di tweet originariamente usato nella prima versione di Twitter, che vedeva la pubblicazione di messaggi che raccontavano aneddoti della vita quotidiana. Ancora oggi è molto frequente imbattersi in questo tipo di tweet, anche se è consigliato non raccontare troppo della propria vita personale.

- Tweet d’opinione: che cosa sto pensando

È il tipo di tweet che riguarda l’espressione di una propria opinione su un determinato argomento, sia personale, sia d’attualità, sia inerente ciò che altri utenti stanno discutendo su Twitter.

- Tweet “missione compiuta”: ciò che ho appena fatto

È il tweet che esprime pubblicamente quello che abbiamo appena fatto. Può risultare particolarmente utile in quanto solitamente genera da parte dei lettori una risposta.

- Tweet di intrattenimento

Vi appartengono quei tweet piacevoli o divertenti che generano interesse nel web sociale.

- Tweet interrogativo

È quel tweet che prevede la formulazione di una domanda allo scopo di attivare una discussione su un determinato argomento, o più semplicemente ricevere una risposta per sciogliere un proprio dubbio.

- Tweet multimediale

Vi appartengono quei tweet che contengono immagini o video da condividere con tutti i propri follower. La condivisione può essere effettuata con numerose applicazioni apposite, per esempio Twitpic o Yfrog, anche se nell'agosto 2011 Twitter ha inserito un apposito pulsante integrato nel campo dove viene composto il messaggio che permette l'inserimento diretto dell'immagine.

2.3.2 Retweet

Il retweet è stato introdotto da Twitter nel 2009 e consiste nel rilanciare ai propri follower un tweet pubblicato da un altro utente. Ovviamente si *ritwetta* qualcosa che riteniamo interessante, qualcosa che condividiamo, poiché questo messaggio viene letto da tutti i nostri follower, visto che viene pubblicato sulla nostra bacheca e nella cronologia degli utenti che ci seguono.

Ritwittare un messaggio significa soprattutto aumentarne la diffusione, dargli una maggiore popolarità facendolo conoscere un maggior numero di persone. I Retweet mostrano, oltre al messaggio originale, anche il nome utente dell'autore che lo ha scritto, la sua immagine del profilo e il numero di retweet.

È possibile ritwittare qualsiasi messaggio ed è possibile farlo in due modi, o dalla homepage, semplicemente cliccando sul link *Retweet* contenuto all'interno del box contenente il tweet, o aggiungendo l'abbreviazione *RT* più *@nomeutente* e facendole seguire il testo del messaggio che si intende ritwittare.

Il metodo più diffuso, anche perché più semplice e immediato, è il primo, mentre il secondo, oltre che caratteristico della prima versione di Twitter, viene oggi utilizzato per i retweet effettuati dal cellulare.

FIGURA 2.5 Un esempio di Retweet ottenuto col primo metodo. In questo caso il tweet è stato scritto dall'utente *AGIRE* ed è stato ritwittato dall'utente *pandemia*.

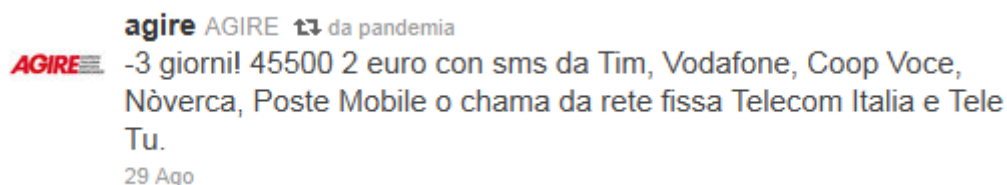



FIGURA 2.6 Un esempio di Retweet ottenuto col secondo metodo. In questo caso il tweet è stato scritto dall'utente *Maumol* ed è stato ritwittato dall'utente *Francesca Oliva*.



I retweet effettuati tramite il link integrato in twitter sono caratterizzati dall'icona  e hanno una loro scheda specifica nel menù della Home che permette di visionarli.

Il menù Retweet contiene tre diversi elenchi di retweet:

- Ritwittati da altri

In questo elenco compaiono i retweet effettuati dalle persone che seguiamo;

- Ritwittati da te

Qui compaiono i tweet altrui che abbiamo ritwittato. Se anche altri utenti hanno ritwittato questi messaggi vedremo il loro avatar all'interno della lista;

- I tuoi tweet, ritwittati

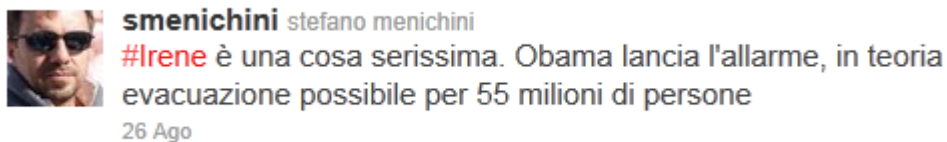
Qui troviamo, eventualmente, i nostri messaggi che sono stati ritwittati da altri.

2.3.3 Hashtag

Un hashtag è una parola chiave preceduta dal simbolo # e viene utilizzato per enfatizzare una parola all'interno di un tweet.

Quando si inserisce un hashtag, e quindi si antepone il simbolo # a una parola, questa si trasforma in un link cliccabile, rimandando ad una apposita pagina di ricerca contenente soltanto i tweet con quella parola chiave. Cliccando sul link è come se effettuassimo una ricerca su tutti i tweet filtrandoli per un argomento specifico perché l'hashtag ha proprio la funzione di rendere ricercabile e tracciabile un tweet attraverso il motore di ricerca di Twitter.

FIGURA 2.7 Un esempio di hashtag, dove è stata scelta come parola chiave la parola #Irene.



Ultimamente l'hashtag viene utilizzato spesso durante eventi, conferenze o incontri, durante i quali gli organizzatori scelgono e comunicano la parola chiave ufficiale da utilizzare come riferimento per partecipare all'evento. Ricercando quindi quell'hashtag dal motore di ricerca di Twitter è possibile instaurare conversazioni tra gli utenti, che potranno così commentare in tempo reale quello che si sta dicendo all'evento o che potranno partecipare anche se non fisicamente presenti.

2.3.4 Liste

Quando il numero degli utenti che seguiamo diventa elevato risulta difficile riuscire a seguire tutti gli interventi degli utenti perché ci troviamo di fronte ad un traffico di messaggi molto intenso. Per ovviare a questo problema ci vengono in aiuto le liste, costituite da un insieme di profili raggruppati per tema o per qualche altra caratteristica in maniera tale da facilitare la lettura dei tweet.

Ogni utente può creare al massimo 20 liste, contenenti non più di 500 utenti cadauna. Una volta creata una lista, cliccandoci sopra, potremo vedere soltanto i messaggi inseriti dagli utenti facenti parte di quella lista, col risultato quindi di una

potente azione di filtraggio. È inoltre possibile inserire nelle liste utenti che non abbiamo deciso di seguire, magari perché poco interessanti, o magari perché pubblicano troppi tweet; in questa maniera i loro messaggi non compariranno nella nostra cronologia ma soltanto nella pagina della lista in cui li abbiamo inseriti.

3. Modalità di raccolta del corpus

3.1 Criteri per la selezione degli utenti

Questo studio è stato realizzato per osservare il tipo di linguaggio e le abitudini linguistiche che gli utenti italiani assumono all'interno di Twitter.

Per raccogliere il corpus di tweet utile alla nostra analisi il primo passo è stato quello di ricercare su Twitter centosessanta utenti italiani aventi profilo pubblico, nello specifico ottanta utenti professionisti²⁵ e ottanta utenti non professionisti²⁶.

Per la selezione degli utenti professionisti si è cercato di selezionare i personaggi più affermati del momento in svariati campi di occupazione. Tra i campi di occupazione presi in esame sono stati scelti, dall'apposita pagina di Twitter *Sfoglia Interessi*²⁷, quelli più popolari, ovvero sport, politica, musica, tecnologia e spettacolo, ai quali è stata volutamente aggiunta anche quella dei giornalisti-scrittori, che non era contenuta nella pagina.

FIGURA 3.1 La pagina Sfoglia Interessi di Twitter



Si è deciso di includere nel campione analizzato anche questa categoria sia perché oggi sono sempre più numerosi gli scrittori e i giornalisti che utilizzano blog e i microblog per diffondere i loro articoli, sia perché, trattandosi di un'analisi

²⁵ Fanno parte degli utenti professionisti coloro che utilizzano il proprio profilo per divulgare informazioni affini la propria attività e che sono conosciuti a livello nazionale.

²⁶ Tutti coloro che non rientrano nella categoria dei professionisti.

²⁷ È una pagina che raggruppa i maggiori argomenti presenti su Twitter. All'interno di ogni argomento sono contenute le persone più popolari del momento che Twitter consiglia di seguire.

linguistica, è interessante vedere se quelli che sono considerati veri e propri professionisti della scrittura utilizzano un metodo di scrittura diverso all'interno della rete, e in particolare se sono condizionati dai 140 caratteri messi a disposizione dal servizio.

Dopo aver stabilito i campi di occupazione da prendere in esame, il passo successivo è stato quello di ricercare all'interno di ogni campo un numero casuale di professionisti italiani con profilo pubblico. Cercando di tenere conto del loro grado di notorietà, e quindi selezionando quelli più conosciuti al grande pubblico, si è proceduto alla selezione, sia tra quelli suggeriti automaticamente da Twitter nella pagina sopracitata, sia attraverso una ricerca libera, realizzata sempre all'interno della pagina *Sfoglia Interessi* tramite l'apposito motore di ricerca *cerca per nome o argomento*. In questo caso, per completare la lista degli utenti mancanti, sono stati inseriti nel motore di ricerca i campi di occupazione presi in esame, usando nello specifico le seguenti chiavi di ricerca: *sportivi*, *politici*, *cantanti*, *spettacolo*, *giornalisti*, *scrittori* e *attori*, e selezionando poi, dai risultati, sempre gli utenti più noti al grande pubblico. Soltanto due utenti²⁸ sono stati cercati manualmente dall'autore, ed entrambi appartengono al genere dei *blogger informatici*, in quanto i risultati ottenuti con la chiave di ricerca *blogger tecnologia* o *blogger informatica* indirizzavano verso profili Twitter di noti siti web e portali del settore, ma non verso profili di singole persone. La ricerca si è conclusa con il raggiungimento complessivo di ottanta²⁹ utenti.

²⁸ Salvatore Aranzulla, Giorgio Tave.

²⁹ Lista degli utenti professionisti: Salvatore Aranzulla, Lorenzo De Silvestri, Emiliano Viviano, Fabio Volo, Andrea Bargnani, Niccolò Savino, Luca Bizzarri, Danilo Gallinari, Antonio Di Pietro, Frankie Hi-NRG, Silvio Muccino, Gianni Soldini, Alessandro Del Piero, Negramaro, Luca Sofri, Lorenzo Jovanotti, Ilario D'Amico, Dario Di Vico, Stefano Menichini, Mario Pappagallo, Gianni Riotta, Francesco Costa, Luca Conti, Sandro Ruotolo, Massimo Donadi, Vasco Rossi, Mario Adinolfi, Rudy Bandiera, Pierferdinando Casini, Luca Ward, Luca Telese, Alberto Angela, Daniele Bossari, Marco Bardazzi, Antonio Cantoro, Francesca Olica, Gianluca Di Marzio, Trio Medusa, Andrea Sarubbi, Alessandro Cattelan, Laura Pausini, Sabina Guzzanti, Marco Melandri, Alessia Marcuzzi, Sandrone Dazieri, Beppe Severgnini, Salvo Sottile, Giorgio Tave, Luca Toni, Ivan Basso, Federica Pellegrini, Paolo De Ceglie, Cesare Cremonini, Francesco Facchinetti, Alberto Brandi, Fiorello, Marco Baldini, Debora Serracchiani, Giuliano Pisapia, Andrea Marcucci, Francesco Rutelli, Sonia Alfano, Paolo Attivissimo, Riccardo Luna, Federico Rampini, Andrea D'Ambrosio, Marco Alfieri, Dario Olivero, Antonio Sofi, Maurizio Sacconi, Bruno Arena, Lia Capizzi, Federico Taddia, Ferruccio De Bortoli, Pierfrancesco Favino, Lorella Cuccarini, Diego Bianchi, Alberto Gilardino.

Va segnalato che 2³⁰ degli 80 utenti sono costituiti da più persone, in quanto trattasi di gruppi musicali e dello spettacolo.

Di seguito è possibile osservare la ripartizione degli utenti professionisti analizzati in base al loro campo di occupazione:

TABELLA 3.1

Il campo di occupazione degli utenti professionisti analizzati.

Campo di occupazione	Numero di utenti
Giornalisti-Scrittori	28
Attori-Conduuttori	18
Sportivi	12
Politici	11
Cantanti	7
Blogger Informatici	4
Totale	80

Il numero maggiore di giornalisti-scrittori, attori-conduttori e sportivi rispetto ai blogger informatici, è dovuto al fatto che i primi sono più noti al grande pubblico, trattandosi di campi meno specialistici. Per quanto riguarda invece il basso numero di politici e cantanti questo è dipeso dal fatto che la maggior parte di loro scrive in terza persona, ciò fa quindi pensare che non siano loro a scrivere ma abbiano dato in gestione il profilo a specialisti del social media marketing³¹. In questo caso si è cercato di selezionare politici e cantanti che scrivono in 1^a persona e che apparentemente sembrano gestire personalmente il loro profilo Twitter.

Gli utenti non professionisti sono stati invece selezionati tra i follower³² degli utenti professionisti, in particolare 17 sono stati prelevati dai follower dei giornalisti-scrittori, 21 dagli attori-conduttori, 15 dagli sportivi, 12 dai politici, 11 dai cantanti e

³⁰ Negramaro, Trio Medusa.

³¹ Il Social Media Marketing è quella parte del Marketing che si occupa di generare visibilità sui Social media.

³² I follower sono gli utenti che seguono il profilo di una persona scegliendo di ricevere i suoi aggiornamenti.

4 dai blogger informatici, per un raggiungimento totale di ottanta utenti³³. Questa ripartizione è stata casuale; durante la selezione di questi 80 utenti si è soltanto prestato attenzione al fatto che l'utente avesse compilato il proprio profilo personale con informazioni sulla propria età e sul proprio status di studente o lavoratore. Là dove lo status di studente o lavoratore non fosse stato specificato vi è stato possibile risalirvi, ad eccezione di 6 persone, attraverso una rapida lettura dei tweet del singolo utente, o attraverso appositi link (rimandanti a profili facebook o a blog personali) inseriti sempre all'interno della descrizione del profilo.

Grazie alle informazioni del profilo è quindi stato possibile provvedere ad una suddivisione analoga per campo di occupazione.

TABELLA 3.2

Il campo di occupazione degli utenti non professionisti analizzati

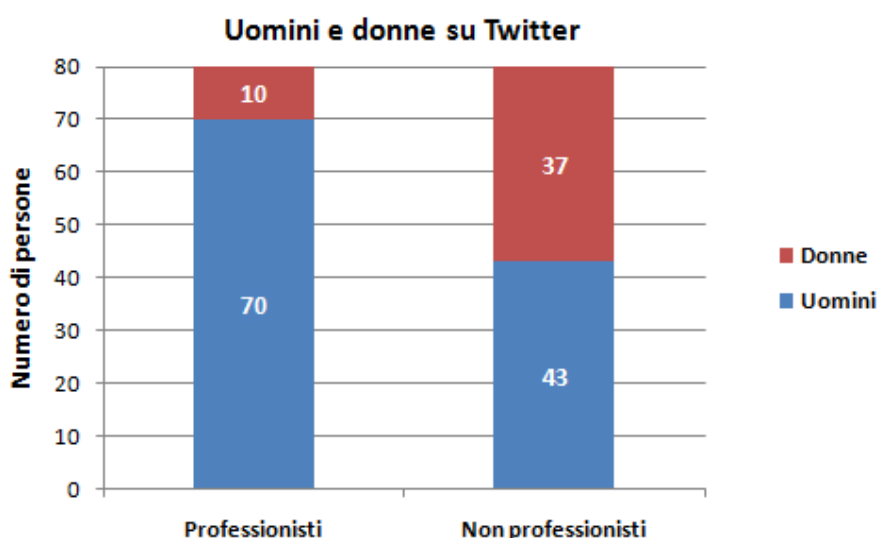
Campo occupazione	Numero di utenti
Studenti Universitari	31
Lavoratori	30
Teenagers³⁴	12
Pensionati	1
Non specificato	6
Totale utenti	80

³³ Daniela Frig, Giorgia Zuccalà, Lorenzo Roversi, Enri Paf Padili, Daniele De Iaco, Michele Scibè, Sara Ceccherini, Lorenzo Di Carlo, Emanuele Voci, Silvia Papi, Emilio Titi, Dennis Magri, Giorgio Favaretto, Matteo Plazzi, Francesco Tusi, Maurilio Anese, Marianna Kafka Fiori, Giovanni Piscopo, Giuly Cogoni, Arianna Giannatempo, Mariangela Mazza, Claudia Romagnoli, Claudia Branciaroli, Eleonora Visconti, Simona Rossi, B di Bina, Camilla Smirulz, Francesca Art, Martino Carbotti, Emilia Piermartini, Anna Zignale, Tiziano Sebastiani, GianMatteo, Stefano Paolucci, Luca G., Maria Luciana, Giulia Zontini, Francesca Boboli, Rocco Sorrentino, Massimo Pinto, Marco Rubertelli, Gloria Fabiano, Francesca Arpaia, Elena Quadra, Danilo G. Cacucciolo, Gianluca Scudieri, Simone Santoleri, Elvira Vaglio, Giulia Sciannella, Marta Boschi, Stefania Esposito, Aldo Di Muccio, Anita Sessa, Dario Botticini, Sara Petralia, Massimo Lupi, Lorenzo Alessandri, Cecilia, Antonella Calandrucchi, Luca Mancino, Raffaele, Pasquale Gionti, Ennio Comini, Domenico Asprone, Jessica Di Francesco, Francesco, Armando Di Lorenzo, Veronica Antinucci, Luca Longo, Laletta, Betti Sguaita, Tiziano Castelli, Martina, Giangiacomo Ceresara, Leone Francesco, Stefano Villa, Domenico Schillaci, Laura Zucchini, Stefano, Adolfo Speranza, Mino Paradiso.

³⁴ Per teenagers si intendono le persone dai 13 ai 18 anni.

Nello specifico tra gli 80 utenti professionisti sono presenti 70 uomini e 10 donne, mentre tra gli 80 utenti non professionisti troviamo 43 uomini e 37 donne. Questa forte differenza può essere dipesa dal fatto di aver selezionato i non professionisti tra i follower di attori, cantanti e sportivi di sesso maschile. Va comunque segnalato che tutti gli sportivi suggeriti da Twitter nella pagina *Sfoglia Interessi - Sport* appartenevano al genere maschile.

FIGURA 3.1 La distribuzione degli uomini e delle donne all'interno del nostro campione di utenti



3.2 Criteri per la raccolta dei tweet

Per facilitare la raccolta dei tweet gli utenti sono stati inseriti in due liste³⁵, Professionisti e NonProfessionisti, create appositamente su Twitter mediante l'opzione *Crea una Lista*.

Inizialmente il campione di utenti preso in esame era composto da 100 utenti (50 utenti per ciascuna categoria), poi si è deciso di ampliarlo a 160 utenti (80 utenti per ciascuna categoria).

Per ogni singolo utente sono stati prelevati gli ultimi 20 tweet inseriti in ordine cronologico inverso a partire da domenica 8 maggio 2011, per i primi 100 utenti, e da domenica 11 settembre 2011 per gli altri 60 utenti.

Dato che il criterio di ricerca è stato casuale, è capitato di incorrere in utenti con meno di 20 tweet inseriti, 4 tra gli utenti professionisti³⁶ e 22 tra i non professionisti³⁷; in questi casi sono stati presi in considerazione soltanto i tweet effettivamente pubblicati. Inoltre, per far sì che fossero selezionati soltanto i tweet scritti dai singoli utenti, sono stati tolti tutti i retweet trovati, in modo da non condizionare l'analisi linguistica del corpus con messaggi scritti da altri.

Una volta raccolti tutti i tweet questi sono stati trasportati, senza interventi sugli errori di ortografia o di battitura, in due distinti file di testo pronti per essere analizzati, per un corpus totale di 2.895 tweet

3.3 L'analisi del corpus

3.3.1 Uno sguardo approfondito ai tweet

L'analisi linguistica del corpus si è basata sullo studio di 2.895 tweet inseriti dalle due categorie di utenti presi in esame.

Dei 2.895 tweet a nostra disposizione gli utenti professionisti ne hanno inseriti 1.563, formati da 20.116 parole e 120.314 caratteri (spazi inclusi 138.814), mentre i non professionisti ne hanno inseriti 1.332, formati da 13.897 parole e 80.782 (spazi inclusi 93.336) caratteri.

Sulla base di questi dati è stata calcolata la lunghezza media dei tweet inseriti, col risultato che la media di parole per tweet di un utente professionista è di 12,87 parole e la media dei caratteri è 76,98, mentre quella di un non professionista è di circa 10,43 parole e 60,65 caratteri.

³⁶ Emiliano Viviano, Alessia Marcuzzi, Dario Olivero, Alberto Gilardino.

³⁷ Michele Scibè, Emanuele Voci, Emilio Titi, Denis Magrì, Marianna Fiori, Giovanni Piscopo, Giuly Cogoni, Mariangela Mazza, Claudia Romagnoli, Simona Rossi, Marta Boschi, Emilia Piermartini, Massimo Pinto, Arpaia Francesca, Aldo Di Muccio, Sara Petralia, Massimo Lupi, Raffaele, Ennio Comini, Domenico Asprone, Jessica Di Francesco, Betti Sguaita, Tiziano Castelli.

Come detto in precedenza i tweet sono stati raccolti in ordine cronologico inverso a partire da domenica 8 maggio 2011 e da domenica 11 settembre 2011.

Nella giornata di domenica 8 maggio, sul campione iniziale di 50 utenti, i professionisti che hanno utilizzato il servizio e hanno inserito almeno un tweet sono stati 30, mentre sono stati 20 i non professionisti.

Nella settimana 01/05/2011 - 08/05/2011 è emerso invece che 47 professionisti hanno scritto almeno una volta nell'arco della settimana, contro i 38 non professionisti.

Nella giornata di domenica 11 settembre, sul campione successivo di 30 utenti, i professionisti che hanno inserito almeno un tweet sono stati 19, mentre sono stati 14 i non professionisti.

Nella settimana 04/09/2011 - 11/09/2011 è emerso invece che 25 professionisti hanno utilizzato il servizio almeno una volta nell'arco della settimana, contro i 20 non professionisti.

Da questi risultati è evidente un maggiore utilizzo di Twitter da parte dei professionisti, che oltre a inserire interventi più lunghi scrivono anche un numero più ampio di tweet quotidiani, dimostrando una maggiore continuità nell'uso del servizio. Su questo maggior utilizzo influisce sicuramente il fatto che i professionisti hanno un interesse diretto all'uso del servizio, che probabilmente li porta ad utilizzare Twitter quotidianamente.

Un altro dato interessante analizzato è stato quello della lingua maggiormente utilizzata per la scrittura dei tweet.

Trattandosi, come abbiamo visto all'inizio, dello studio di un campione di 100 utenti italiani la lingua maggiormente utilizzata è stata l'italiano, seguita da inglese, spagnolo, tedesco e francese. Nello specifico:

- i professionisti hanno inserito 1.417 tweet in italiano (90,66 %), 104 tweet in inglese (6,65%), 3 in tedesco (0,19%) e 2 in francese (0,13%);
- i non professionisti hanno inserito 1.044 tweet in italiano (78,38%), 237 tweet in inglese (17,79%) e 4 in spagnolo (0,30%);

Tra gli utenti professionisti coloro che fanno maggiormente uso della lingua inglese appartengono al genere degli sportivi e dei cantanti, questo probabilmente a causa della loro notorietà internazionale e al conseguente numero di fan di lingua non italiana da cui sono seguiti. Infatti, dei totali 104 tweet in lingua inglese, 68 (il 65,38%) sono stati scritti proprio da utenti appartenenti al genere degli sportivi e dei cantanti. Nello specifico un utente³⁸ professionista ha scritto soltanto tweet in lingua inglese.

Di seguito alcuni tweet in lingua inglese degli utenti professionisti:

1. am I gonna see the sun before leaving TO? ³⁹
2. @NuggetsAreBoss I do like denver..thanks for the love..I will do my best for you guys..⁴⁰
3. Thank you for being close and for all your affection, it is really very important for me.⁴¹

I non professionisti invece utilizzano la lingua inglese per comunicare con gli utenti di lingua straniera da loro seguiti, che coincidono sempre con personaggi di successo. Il maggior numero di tweet in inglese che si ritrova nei loro post (237 tweet contro 104) è dovuto molto probabilmente a questo motivo.

Si riportano alcuni tweet in inglese dei non professionisti indirizzati a celebrità:

1. @ParisHilton wow... life it's hard!!!
2. @katyperry good luck!!!
3. @wes_sneijder10 we are all waiting you to score in Serie A. Sunday will be your day man, go and win!

³⁸ Alessandro Del Piero

³⁹ Andrea Bargnani, cestista italiano dei Toronto Raptors, NBA, Stati Uniti.

⁴⁰ Danilo Gallinari, cestista italiano dei Denver Nuggets, NBA, Stati Uniti.

⁴¹ Alessandro Del Piero, calciatore italiano della Juventus.

TABELLA 3.3
Tabella riassuntiva

Tipologia utenti	Professionisti	Non professionisti
Numero tweet	1.563	1.332
Numero parole	20.116	13.897
Numero caratteri	120.314	80.782
Lunghezza media tweet in parole	12,87	10,43
Lunghezza media tweet in caratteri	76,98	60,65
Tweet in italiano	1417	1044
Tweet in inglese	104	237
Tweet in francese	2	0
Tweet in spagnolo	0	4
Tweet in tedesco	3	0

Riepilogo analisi del corpus in oggetto

3.3.2 L'espressività grafica del corpus

Per consentire l'aggiornamento delle proprie pagine anche attraverso un SMS Twitter impone ai messaggi un limite di 140 caratteri. Questo limite sicuramente influenza gli scriventi, che spesso fanno ricorso all'uso di emoticon, di abbreviazioni, di acronimi e di altre variazioni dell'italiano standard. Tutti questi fenomeni sono presenti, anche se in maniera diversa, all'interno del nostro corpus.

3.3.3 Le Emoticon

Le emoticon sono delle faccine, in inglese chiamate anche *smiley*⁴², utilizzate nella comunicazione scritta di Internet e degli SMS per esprimere stati d'animo e

⁴² Sono dette anche *smiley* (faccine sorridenti) proprio perché quella più comune ed utilizzata è lo smile (sorriso), formata dai due punti, trattino e parentesi chiusa :-)

emozioni, compensando così la mancanza del linguaggio non verbale all'interno della rete.

La parola emoticon è la fusione delle parole inglesi *emotional* e *icon*, proprio per indicare un'icona emotiva volta a rappresentare delle emozioni.

L'emoticon è caratterizzata da una serie continua di caratteri rappresentati da numeri, lettere e punteggiatura, che ricordano un'espressione facciale specifica.

1. Si studia :(
2. sto mangiando troppi dolci :O :(
3. Mormora, la gente mormora..falla tacere praticando l'allegria ^_^

L'origine delle emoticon risale al 19 settembre 1982, data in cui Scott Fahlman, un informatico statunitense, propose per la prima volta di inserire nei messaggi elettronici le sequenze di caratteri :-) e :-(per distinguere i messaggi scherzosi da quelli seri. Da quella data si è assistito ad un grandissimo incremento del numero di emoticon, nate dalle più svariate combinazioni di caratteri. Una delle caratteristiche più interessanti delle emoticon è che, essendo formate da segnali grafici, sono indipendenti dalla lingua di chi scrive e anche continuamente in evoluzione, tanto da arrivare a fare una suddivisione tra emoticon occidentali ed emoticon orientali (dette anche emoticon giapponesi).

La differenza principale tra le emoticon occidentali e quelle orientali è che quelle occidentali vengono lette orizzontalmente (da sinistra verso destra), richiedendo al lettore di girare la testa per comprenderle, mentre quelle orientali vengono lette verticalmente, avendo gli occhi in alto e la bocca in basso.

Per mostrare le differenze riportiamo di seguito alcune emoticon appartenenti ai due generi.

TABELLA 3.4
Tabella riassuntiva

Significato	Occidentali	Orientali
Sorriso	:-) :) :o) 8)	^_^ ^-^ ^.^ ' _'
Sorpresa	>:O :-O :O	O_o @_@ O_O
Rabbia	>:C >: (>_< >.<
Pianto	: ' (; * (; _; T _T

Differenze tra emoticon occidentali ed orientali

Nel campione dei tweet analizzato troviamo un totale di 392 emoticon, un numero piuttosto ridotto considerando il numero di parole da cui è composto il corpus (34.013). Nello specifico troviamo 97 emoticon nel corpus degli utenti professionisti (pari allo 0,005%) e 295 emoticon nel corpus dei non professionisti (pari al 0,02%).

In base alla suddivisione delle emoticon occidentali e orientali fatta in precedenza vediamo che delle 97 emoticon degli utenti professionisti, 95 appartengono al repertorio di quelle occidentali e 2 soltanto al repertorio di quelle orientali. Diversa, anche se di poco, è la situazione degli utenti non professionisti. Su 295 emoticon hanno utilizzato 258 emoticon appartenenti al repertorio occidentale (87,46%) e 37 appartenenti al repertorio orientale (12,54%).

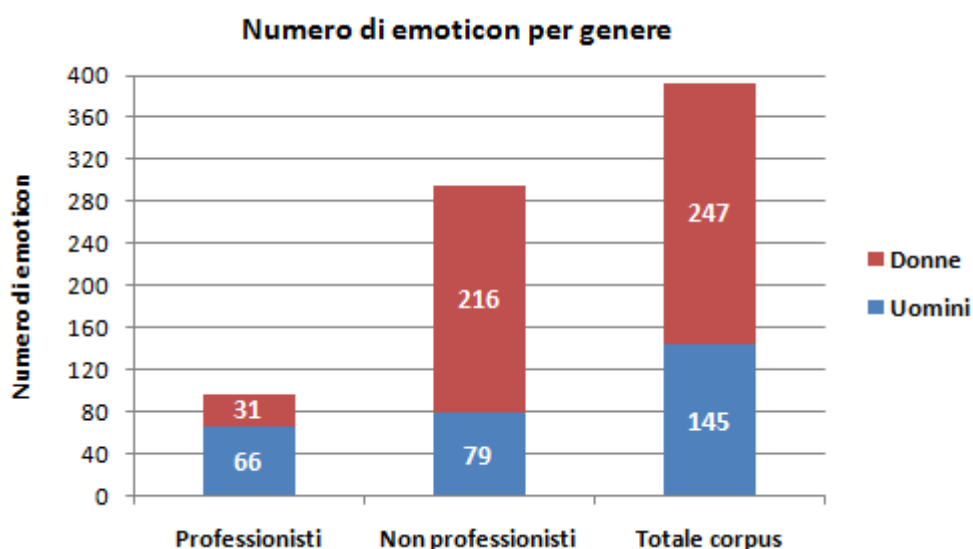
Degli 80 utenti professionisti sono 31 coloro che utilizzano le emoticon nei loro tweet (il 38,75%), nello specifico 23 sono maschi (con 66 emoticon inserite su 97 totali) e 8 sono femmine (con 31 emoticon inserite su 97 totali). Considerando che degli 80 utenti presi in considerazione 70 sono maschi e 10 sono femmine, la percentuale di persone di sesso femminile che utilizza le emoticon è del 80% (8 su 10), contro il 32,86% di quella degli utenti di sesso maschile (23 su 70) .

Tra gli 80 utenti non professionisti sono invece 48 coloro che utilizzano le emoticon nei loro tweet (il 60%), nello specifico 29 femmine (con 216 emoticon inserite su 295 totali) e 19 maschi (con 79 emoticon inserite su 295 totali). Tenendo

in considerazione che nel campione degli 80 utenti 37 sono le femmine e 43 i maschi, la percentuale maggiore di persone che utilizza le emoticon è sempre quella di sesso femminile, il 78,38% (29 su 37), contro il 44,18% di quella degli utenti appartenenti al genere maschile (19 su 43).

Se si raggruppano questi dati il maggior utilizzo delle emoticon da parte del sesso femminile è ancora più evidente. Si nota infatti che delle 392 occorrenze verificate nel corpus ben 247 appartengono alle donne (il 63%), mentre le restanti 145 (il 37%) appartengono agli uomini.

FIGURA 3.2 Grafico riepilogativo del numero di emoticon inserite dagli uomini e dalle donne all'interno del corpus analizzato



Vediamo di seguito tutte le emoticon presenti all'interno del corpus.

TABELLA 3.5

Tabella riassuntiva delle emoticon occidentali presenti nel corpus

Occidentali	Professionisti	Non professionisti
:)	23	75
: -)	24	8
: (4	15

: ((0	2
:- (3	2
=)	0	9
=B	0	1
=D	0	1
=P	0	3
:P	0	9
:D	13	29
:DD	0	1
:-D	2	4
;D	0	3
D:	0	1
;))	9	40
;-)	7	1
:O	0	2
XD	0	11
xD	0	4
:	0	1
:S	0	4
:/	0	1
:')	0	1
;))))	0	1
;)))	1	1

;))	2	1
:))	2	0
:)))))	0	1
;~))	1	0
:~))	1	0
:~)))))	1	0
:@	0	1
(:	1	2
:*	0	1
:**	0	1
:~/	1	0
<3	0	21
Totale	95	258

TABELLA 3.6

Tabella riassuntiva delle emoticon orientali presenti nel corpus

Orientali	Professionisti	Non professionisti
ㄨ_ㄨ	0	1
_\'	0	1
^_^	0	2
_	0	3
-.-	0	3
-.-\`	0	2
-.-"	0	3

U.U	0	2
u.u	0	2
w	0	4
.	0	3
o_o	0	2
o_o	0	1
o_o	1	0
o.o	0	1
o.o	0	2
T.T	0	2
.-.	0	1
^^	0	2
@@	1	0
Totale	2	37

Nel 2010 Maurizio Maltese ha cercato di classificare le emoticon in tre macrocategorie: emoticon di sostituzione morfo-sintattica, emoticon espressive e emoticon di intrattenimento.

La prima, quella delle emoticon di sostituzione morfo-sintattica, si riferisce a le emoticon che sostituiscono parti del discorso o anche interi discorsi (infatti si contano in questa categoria anche quei messaggi composti solamente da un'emoticon).

La seconda categoria, quella delle emoticon espressive, si riferisce a quelle emoticon utilizzate per aggiungere informazione grafica al testo e dare informazioni

sugli stati d'animo dello scrivente. Non sostituiscono il testo ma aggiungono informazioni ad esso.

Fanno parte dell'ultima categorie le emoticon di intrattenimento, usate con ruolo puramente estetico e con lo scopo di divertire.

Grazie ad excel è stato poi possibile calcolare la distribuzione delle emoticon all'interno delle due categorie di utenti, vedendo quanti sono gli utenti che hanno inserito un certo numero di emoticon.

Tra i professionisti, come si evince dal grafico, la distribuzione è piuttosto scarsa, e il massimo numero di emoticon utilizzate da un utente ammonta a 11.

FIGURA 3.3 Grafico di distribuzione delle emoticon nel corpus degli utenti professionisti



Per quanto riguarda i non professionisti anche in questo caso possiamo vedere dal grafico che il massimo numero di emoticon utilizzate da un utente è 22. Le restanti, comunque, sono ben distribuite, e l'utilizzo che ne viene fatto è decisamente maggiore.

FIGURA 3.4 Grafico di distribuzione delle emoticon nel corpus degli utenti professionisti



Dai dati in nostro possesso notiamo anche che le emoticon vengono usate in prevalenza alla fine della frase, quasi a sostituire il saluto finale.

Su 392 emoticon infatti 272 (il 69,39%) sono usate al termine della frase, mentre le restanti 69 (il 30,61%) vengono utilizzate o all'inizio di essa o al suo interno.

Nello specifico è emerso che su 97 emoticon i professionisti ne hanno inserite al termine della frase 59 (70,09%), mentre i non professionisti, su 295, ne hanno inserite alla fine 213 (72,20%). Una percentuale pressoché identica.

Il corpus inoltre presenta 58 tipi di emoticon differenti. Quella più utilizzata è il classico smiley :), che conta 98 occorrenze (equivalente al 25% del totale), seguita dalla faccina ;) , che conta invece 49 occorrenze (il 12,5% del totale).

A fronte di tutti questi dati è quindi possibile affermare che gli utenti non professionisti fanno un utilizzo maggiore delle emoticon, 295 occorrenze contro 97.

Come si è visto questa grossa differenza risulta essere dovuta, in primis, al maggior numero di donne presenti all'interno del campione dei non professionisti che risultano essere le maggiori utilizzatrici delle emoticon. In aggiunta si nota che il campione degli utenti non professionisti è costituito prevalentemente da adolescenti e giovani under 30 e che quindi, questi, siano più propensi all'utilizzo del linguaggio giovanile. A dimostrazione di ciò basta osservare che sui 58 tipi diversi di emoticon presenti nel corpus i professionisti ne utilizzano soltanto 18, di cui 16 appartenenti

alla categoria occidentali, le più semplici e conosciute, mentre i non professionisti utilizzano 51 tipi di emoticon, 33 appartenenti alla tipologia occidentale e 18 a quella delle più recenti emoticon orientali, che contano ben 37 occorrenze non professioniste contro le 2 professioniste. Inoltre, un altro fattore di cui tener conto, è sicuramente quello della categoria di appartenenza. Gli utenti professionisti, in quanto tali, tendono sicuramente ad utilizzare un linguaggio più formale nei loro messaggi, essendo probabilmente anche influenzati dal numero maggiore dei loro follower e dal loro grado di personaggi pubblici.

3.3.4 La punteggiatura

Un elemento fondamentale di ogni testo è la punteggiatura, usata da un lato per riprodurre le pause, l'espressività, l'intonazione della lingua parlata; dall'altro per evidenziare le componenti grammaticali e sintattiche di una frase o di un discorso, separandone o segnalandone le varie parti.

All'interno del nostro corpus di testo l'interpunzione è piuttosto presente; questo probabilmente è dovuto al fatto che spesso i tweet vengono scritti da una tastiera standard di un computer, piuttosto che da quella di un cellulare, quindi è più facile inserire e revisionare la stessa. Va anche evidenziato che nel corpus, sia nei tweet dei professionisti che in quelli dei non professionisti, viene fatto un uso massiccio della punteggiatura (puntini di sospensione, punti interrogativi, punti esclamativi). In particolare questa è di tipo ritmico piuttosto che di tipo sintattico, a indicare la volontà di riprodurre nello scritto le pause e i fenomeni utilizzati nel parlato. Si trovano infatti numerosi puntini di sospensione, sia intermedi che finali, utilizzati per riprodurre le pause del parlato e anche punti esclamativi e interrogativi ripetuti, per enfatizzare parti del discorso che nella comunicazione verbale avrebbero una determinata intonazione.

1. @Ale_Marra tutti all'Arredomania!!!!!!
2. @roberta_nav Ciao Grande Roby!!!come stai??
3. Non riesco a dormire....troppi pensieri mi attanagliano....

3.3.5 Il lessico

Il lessico utilizzato dalle due tipologie di utenti presenta meno tratti in comune e differenze più accentuate rispetto a quelle appena viste nell'interpunzione.

Il tratto comune ad ambedue le categorie di utenti è sicuramente l'utilizzo di un linguaggio informale, dove non esiste un'alternativa all'utilizzo del tu, che viene sempre impiegato per rivolgersi a qualunque tipo di persona. Si registra talvolta anche l'utilizzo di un linguaggio volgare, con fenomeni di turpiloquio presenti soprattutto nel corpus degli utenti non professionisti.

Si contano infatti 37 occorrenze di espressioni volgari dei non professionisti contro 13 dei professionisti.

1. Si inizia! Oggi primo ciak! **Merda! Merda! Merda!**# fb
2. L' uomo ha inventato la sedia elettrica il topo non inventerebbe mai una trappola per topi... Einstein Maledette armi MONDO DI **MERDA!!!**
3. @acmilan dai **cazzo**, sabato si gioca per vincere e non per pareggiare!! Forza ragazzi!

Si ritrovano inoltre anche 66 espressioni dialettali (v. par. 3.3.6), di cui 22 appartenenti ai professionisti e 44 ai non professionisti.

3.3.6 Abbreviazioni e variazioni

La comunicazione scritta di Internet e degli SMS ha portato, sia per esigenze di limiti di caratteri e sia per velocizzare la scrittura, all'utilizzo di un linguaggio più sintetico, caratterizzato da abbreviazioni, sostituzioni e variazioni grafiche di vario tipo applicate al testo.

In particolare Tavosanis, in *L'italiano del Web* (ivi, pp. 76-77), descrive diversi meccanismi di alterazione:

1. lettura endofasica: è il caso della sostituzione grafica di una parola con un simbolo (per esempio <I> per *uno*, <6> per *sei*, <X> per *per* e, a volte, la sequenza di fonemi /per/ all'interno di una parola più ampia (per esempio *xkè*);
2. abbreviazioni non realizzate abitualmente nel parlato (come *nn* al posto di *non*, *qst* al posto di *questo*, *cmq* al posto di *comunque* e così via;
3. variazioni grafiche vere e proprie (*k* per sostituire *ch*, *c* o *q*);
4. <x> al posto di <ss> o <zz> in sostituzione di parolacce o parole non adatte a tutti i tipi d'età (*caxxo*, *sexxo* ecc.);
5. <j> al posto di <gl> nel dialetto romanesco (*mejo* al posto di *meglio* ecc.)

Questi meccanismi di alterazione sono rilevabili, in maniera differente, anche all'interno del nostro corpus.

Si conta un totale di 41 occorrenze per le variazioni endofasiche, 73 occorrenze per le abbreviazioni, 23 occorrenze per le variazioni grafiche, 66 per le espressioni dialettali, 51 tweet maiuscoli e 62 tweet alternanti l'uso delle maiuscole e delle minuscole.

Entrando nei dettagli è possibile notare che all'interno dei tweet dei professionisti il numero di questi fenomeni è piuttosto ridotto.

Si contano infatti le seguenti occorrenze:

- Scrittura endofasica: 13 occorrenze, di cui 8 *x* al posto di *per*, 3 + per *più*, 1 *I* per *uno* e 1 *8* al posto di *ott*^a;

1. @andreaganduglia nella realtà la serie 24 si è fermata all'**8va**.

2. cin palazzo ore 18 assemblea con cittadinanza e rappr di comune, prov, e regione. **x** no definitivo a bisca + tavolo di trattativa **x** futuro

3. I veri #mattidalegare faranno 'a nuttata? Tra #UsOpen (ora Nadal vs Murray + semifi femmin) e #RugbyWorldCup h 5.30 #Italia vs Australia...
- Abbreviazioni: 32 occorrenze, di cui 6 *w* per *viva*, 5 *nn* per *non*, 2 *app* per *applicazione*, 2 *cmq* per *comunque*, 1 *1^a* per *prima*, 1 *2^a* per *seconda*, 1 *ape* per *aperitivo*, 1 *ca.* per *circa*, 1 *scengg* per *sceneggiatura*, 1 *diff* per *differenza*, 1 *pres* per *presidente*, 1 *rappr* per *rappresentanti*, 1 *prov* per *provincia*, 1 *picc* per *piccione*, 1 *puniz* per *punizione*, 1 *settim* per *settimana*, 1 *prox* per *prossima*, 1 *semifin* per *semifinale*, e 3 abbreviazioni di città, *mima* per *Milano Marittima*, *Bo* per *Bologna* e *AL* per *Alessandria*. Tra le abbreviazioni abbiamo incluso anche l'abbreviazione tradizionale *viva*, rappresentata con *w* e un tempo realizzata scrivendo due *W* parzialmente sovrapposte. Inoltre troviamo anche abbreviazioni utilizzate nel parlato come *app*, *ape* e *prox*.
 1. **nn** so se conoscete la serie 24, ma direi unica *diff* realtà/ finzione è che lì il **pres** buono è nelle prime 3, nella realtà è nelle ultime
 2. @vivaelbarca : non ho l'ansia da share, **cmq** il risultato è buono soprattutto alla luce del #clasico
 3. L'unica cosa positiva e' ... Che i tre punti vanno in Sicilia ! **W** S ROSALIA!buonanotte ps domani l'edicolante Cesare mi massacrerà ! Notte
 - Variazioni grafiche: 4 occorrenze, 2 *minkia* per *minchia*, 1 *sbrokko* per *sbrocco*, 1 *makkè* per *macché*. Da notare che appartengono tutte allo stesso utente⁴³.
 1. @braccinocorto hehehe **makkè**... Qualche magrebino ma niente brasilere ;)

⁴³ Rudy Bandiera.

2. @AlessiaSelene no no, era ottimo e molto abbondante. I ragazzi gentili e un po di cibarie, ma cinque euro... **Minkia.**
- Espressioni dialettali: 22 occorrenze;
1. L'ultimo FF della serata per @guglielmoscilla . Ho sentito ora (better late than never!) #atupergu. Bravo! **Daje!**
 2. @frankiehinrgmc paoletta **nun se batte....!** nel lavoro e nella vita...
 3. @alsim71 ma tu sei pazzo, stai sempre a **tuittà**, se t'aggiungo poi leggo solo te. L'importante è sapere che **ce sei, nte preoccupà ;)**

I tweet interamente composti in maiuscolo sono 38; di cui 20 (e cioè tutti quelli inseriti) appartengono allo stesso utente⁴⁴, che probabilmente ha l'abitudine costante di scrivere tweet in maiuscolo. Gli altri 18 tweet scritti in maiuscolo appartengono invece a cinque utenti, che hanno utilizzato questo metodo di scrittura in alcuni dei loro tweet o per mettere in risalto un argomento di particolare importanza o per evidenziare il titolo di un articolo apparso su un sito web (a cui poi hanno fatto seguire un link specifico).

1. QUALE SARA' LO SCOOP DI MERCATO DI GIANLUCA? SU SKY SPORT 1 PUNTATONA QUESTA SERA A "SPECIALE
2. PER VINCERE... <http://ff.im/-CMZkm>
3. PER IL MIO COMPLEANNO UN GRANDE REGALO PER VOI. albertogilardino.it/ita/news/21/pe...

⁴⁴ Massimo Donadi.

Diversa è la situazione che riguarda i tweet che alternano la scrittura in maiuscolo/minuscolo. In questo caso i tweet interessati sono 24 e vengono utilizzati o per mettere in risalto una parte del messaggio considerata di maggiore importanza o per evidenziare un titolo di un articolo sul web.

1. IN VISTA DEL MILAN: La sfida con il Milan alle ultime battute di campionato è più impegnativa che mai. Ci faremo... <http://bit.ly/m3pgYM>
2. IL NUOVO ALBUM ESCE il 29 MARZO E NEL 2011 TORNA NEGLI STADI Trapelano le prime indiscrezioni sulle date itunes/iFb9s4
3. con GASperini all' inter piu ALLEGRI

Per quanto concerne i non professionisti i meccanismi di alterazione della scrittura sono presenti in maniera maggiore. Nello specifico si contano le seguenti occorrenze:

- Scrittura endofasica: 28 occorrenze, di cui 13 *x* al posto di *per*, 6 *d* per *di*, 1 *xk'* al posto di *perché*, 1 *xk* al posto di *perché*, 1 *xchè* al posto di *perché*, 1 + per *più*, 1 *sxo* per *spero*, 1 <1> al posto di *una*, 1 <1> al posto di *uno*, 1 *sxiamo* al posto di *speriamo*, 1 <6> al posto di *sei*, voce del verbo essere, presente, 2ª persona singolare;

1. @chiellini Ah porta qst mex alla squadra:"SIETE **1** GRANDE SQUADRA E VI AMO!!!! ps. DISTRUGGETE IL GENOAA!!" TI PREGO FALLO **x** ME!!!
2. Ma **xk'** dai pronostici del #giro levano sempre #Menchov e #Sastre e si limitano a #Contador #Nibali e #Scarponi
3. Zarate è uscito xchè l'arbitro si è accorto Che in realtà era pandev travestito @interistiorg #amala

- Abbreviazioni: 32 occorrenze, di cui 15 *nn* per *non*, 4 *qst* per *questo*, 4 *cmq* per *comunque*, 2 *sn* per *sono*, 2 *cn* per *con*, 1 *tnt* per *tanto*, 1 *pers* per *persona*, 1 *mex* per *messaggio*, 1 *qnt* per *quanti*, 1 *stt* per *stato*, 1 *dv* per *dove*, 1 *prpr* per *proprio*, 1 *ftt* per *fatto*, 1 *tvb* per *ti voglio bene*, 1 *tto* per *tutto*, 1 *stas* per *stasera*, 1 *comm* per *commerciale*, 1 *qll* per *quello*, 1 *fanta* per *fantacalcio*, 1 *pome* per *pomeriggio*, 1 *Berlu* per *Berlusconi*.

1. @mariop89 prima, sempre e comunque. **nn** credo cambierà mai **qst** tipo **d** cultura. purtroppo.
2. @Mythra__ **cmq** sono tutte su @JaredLetoOnline io **nn** ho nessun merito!!!
3. @Thefrututa c'è tanta ansia.... :) anke se si sa **qll** ke contiene il pacco atteso... :)

- Variazioni grafiche: 19 occorrenze, 10 *ke* per *che*, 3 *k* per *ke*, 3 *anke* per *anche*, 1 *ki* per *chi*, 1 *qlke* per *qualche*, 1 *kiederti* per *chiederti*;

1. **Ke** belle le immagini di Torino dall'alto fra tricolori e rosa inizia il #giro...finalmente!!
2. @VeroJonasLover ciao ho sempre dimenticato di **kiederti...d dv** 6??

- Espressioni dialettali: 44 occorrenze;

1. **Belin**, oggi son riuscito a risolvere 4 rebus della settimana enigmistica...
2. @3v1n0 non ti preoccupare siamo donne serie!!:) **vi si fa piglia** anke la sbornia:))!!
3. @astro_paolo "**u signuruzzu t'u paga**" (il signore te lo rende)

I tweet scritti interamente in maiuscolo sono 13 e sono stati inseriti da otto utenti differenti. L'uso del maiuscolo è stato fatto per mettere in evidenza un argomento di particolare importanza o per evidenziare il titolo di un articolo apparso su un sito web (a cui poi hanno fatto seguire un link specifico). Nessuno di questi sette utenti ha però scritto tutti i propri tweet in maiuscolo, al contrario di quello che avevamo notato tra i professionisti.

1. APPROFITTIAMONE FINCHÈ QUALCUNO NN DECIDERA' CHE È ILLEGALE...
2. PENSA GIULIA, PENSA!
3. ALLE ORE 10:52 USO PER LA PRIMA VOLTA TWITTER..

Sono 38 invece i tweet che alternano parole in maiuscolo a parole in minuscolo, utilizzando appunto le maiuscole per mettere in risalto una parte del messaggio che probabilmente ritenevano più importante.

1. Buongiorno!! ...e BUONA FESTA DELLA MAMMA!!!
2. Oooooooooooooooooooooooooooooo la!!!!!! FINALMENTE!!!!xD
3. Cercasi STATISTI disperatamente

Di seguito la tabella riassuntiva delle variazioni dell'italiano standard appena descritte.

TABELLA 3.7
Tabella riassuntiva dei fenomeni

Tipo di variazione	Professionisti	Non professionisti	Totale corpus*
Particolarità endofasiche	13 (0,06%)	28 (0,2%)	41 (0,12%)
Abbreviazioni	32 (0,16%)	43 (0,3%)	75 (0,22%)
Variazioni grafiche	4 (0,01%)	19 (0,13%)	23 (0,06%)
Espressioni dialettali	22 (0,1%)	44 (0,31%)	66 (0,19%)

Tweet maiuscoli	38(2,43%)	13 (0,97%)	51 (1,76%)
Tweet maiuscoli/minuscoli	24 (1,53%)	38 (2,85%)	62 (2,14%)
Totale fenomeni rilevati	133	185	318

* Percentuali calcolate sul totale delle parole dei professionisti e dei non professionisti.

3.3.7 Gli errori

La presenza di errori all'interno del corpus può essere suddivisa in tre gruppi:

1. presenza di accenti non standard, in cui i caratteri accentati vengono rappresentati in modo sbagliato e sostituiti dalla sequenza lettera + apice. Questo avviene o perché il supporto di scrittura è inadatto, tale quindi da non consentire la corretta rappresentazione della parola (è il caso delle tastiere inglesi che non hanno i caratteri accentati italiani) o perché viene ignorata la differenza tra accenti acuti e accenti gravi;
2. errori di tipo meccanico, detti in inglese anche *typo*. In questi casi lo scrivente sa come scrivere la parola ma non riesce a farlo per un errore involontario, come la pressione accidentale di un tasto;
3. errori di grammatica, tutti quegli errori che evidenziano dei limiti e delle lacune personali nell'utilizzo dell'italiano standard.

Questi errori sono presenti all'interno del nostro corpus in maniera differente.

Per quanto riguarda gli utenti professionisti si trovano all'interno del primo gruppo:

- 10 *perché* corretti contro 3 *perchè* errati;
 1. Il governo dovrebbe ri-proclamare lo stato di emergenza per i rifiuti in Campania e non lo fa **perchè** significherebbe il fallimento di B.

2. Credo che qualcuno debba delle scuse a @danipedrosa .. Sono davvero contento per lui **perchè** si è sempre rialzato! grande.
- 66 e', 21 a', 9 o' 7 di u' e 2 i', invece della corrispondente lettera accentata; anche se questo potrebbe essere dipeso dal fatto della mancanza di una tastiera italiana a disposizione, considerando poi che tra il campione di utenti professionisti si ritrovano utenti che trascorrono, a causa della loro professione, alcuni periodi della loro vita fuori dall'Italia.
1. @lucadegennaro cerchero' di essere il cuoco **più'** "poser" della storia!!
 2. La scena della pallavolo tra cardinali di "HabemusPapam" (Nanni Moretti) **e'** troppo lunga. #nonsonomereghetti
 3. Consiglio facolta' a princeton university: difendere lingua italiana da rischio estinzione

Nel secondo, quello degli errori di tipo *typo*, troviamo:

- 31 occorrenze, ripartite in: 5 *po* per *po'*, 1 *c'e* per *c'è*, 1 *citta* per *città*, 1 *l anno* per *l'anno*, 1 *anhe* per *anche*, 1 *e* per *è*, 1 *lòa* per *là*, 1 *li* per *lì*, 1 *npoco* per *poco*, 1 *unza* per *una*, 1 *apprezzoh* per *apprezzo*, 1 *sensatodi* per *sensato di*, 1 *madridi* per *madri di*, 1 *saluta* per *saluto*, 1 *scuste* per *scusate*, 1 *cin* per *in*, 1 *citadinanza* per *cittadinanza*, 1 *ripondo* per *rispondo*, 1 *pero* per *però*, 2 *perche* per *perché*, 1 *purttroppo* per *purtroppo*, 1 *fiduci* per *fiducia*, 1 *dell0intervento* per *dell'intervento*, 1 *natalei* per *nata lei*, 1 *nagli* per *negli*, 1 *spoglitoi* per *spogliatoi*;
1. hai capito come? Non c'è gara!! Ma complimenti **an he** a te ?
 2. visto david donatello, **scuste** nn mi ero accorta di quanto siamo orrendi. devo rivedere tutti i piani. scola eccezionale nel contesto

3. con GASperini all' inter piu' calore **nagli**
spoglitoi

Infine, per quanto riguarda gli errori di competenza, troviamo:

- 7 occorrenze, 1 bho per boh, 2 spriz per spritz, 1 sta sera per stasera, 3 pò per po'.

1. Il treno, alle sette di sera dopo uno **spriz**, é la cosa più soporifera del mondo. Specie in primavera.

2. **sta sera** cinema palazzo roma vergassola poi concerto. a chi chiede positività ripondo, venite qua. è un'isola felice

Nel corpus dei non professionisti troviamo per il primo gruppo:

- 6 *perché* corretti contro 13 *perchè* errati;

1. **Perchè** non valgo la polvere calpestata dai piedi degli impiccati. Ti ho dato la mia anima, lasciatemi il nome.

2. @riccardowired ovviamente dormono. Anche **perchè** scusa, Riccardo, cosa vuoi che sia qualche spicciolo di evasione fiscale?

- l'uso di 15 *e'*, di 3 *a'*, di 2 *o'*, di 1 *i'* e 1 di *u'*, invece della corrispondente lettera accentata; e anche qui potrebbe essere dipeso dal fatto della mancanza di una tastiera italiana a disposizione.

1. La vita a volte ti lascia interdetto, stordito, ti **puo'** dare una profonda tristezza o una gioia immensa. Ora **e'** tristezza.....Forza Rosanna....

2. **e'** brutto portarsi dietro problemi del 2010.. Anno nuovo vita nuova? Solo se si vuole davvero..e a volte non basta.

Per il secondo gruppo, quello degli errori di tipo *typo*, troviamo:

- 37 occorrenze così distribuite, 2 *piu* per *più*, 1 *tho* per *ho*, 1 *raramente* per *raramente*, 1 *Irlanda_* per *Irlanda*, 1 *fedteggia* per *festeggia*, 1 *sisi* per *sì*, 2 *semblato* per *sembrato*, 2 *vedele* per *vedere*, 1 *l esame* per *l'esame*, 1 *lse* per *se*, 1 *mabbuongiorno* per *ma buongiorno*, 1 *ance* per *anche*, 1 *lavoroche* per *lavoro che*, 1 *barbache* per *barba che*, 1 *comè* per *com'è*, 1 *beativoi* per *beati voi*, 1 *vabbe* per *vabbè*, 1 *frote* per *fronte*, 1 *preferiscr* per *preferisce*, 1 *Gesu* per *Gesù*, 1 *messuno* per *nessuno*, 1 *magara* per *magari*, 1 *po* per *po'*, 1 *rasdio* per *radio*, 1 *pura* per *paura*, 1 *acui* per *a cui*, 1 *ommioddio* per *o mio dio*, 1 *iamo* per *siamo*, 1 *die* per *dei*, 1 *l invidia* per *l'invidia*, 1 *difenerà* per *difenderà*, 1 *profiloo* per *profilo*, 1 *po* per *po'*.

1. Fuochi d'artificio a Milano, chissa' cosa si
fedteggia!
2. @Mars_Rock quella su @JaredLetoOnline....a **rasdio**
104.5.....una delle più' esilaranti mai viste!!!
3. Un costosissimo #Abbiati a 60 fantamilioni
difenerà la porta dell'Internacional! #fantacalcio

In questo caso *semblato* e *vedele* e *mabbuongiorno* e *ommioddio*, sono errori voluti, coi primi due che appartengono ad una celebre espressione utilizzata all'interno di un cartone animato⁴⁵ e i secondi due utilizzati per riprodurre un'intonazione del parlato nello scritto.

Allo stesso modo possiamo considerare un errore volontario anche l'utilizzo di *sallo*, che troviamo nella frase:

1. @Punk14bz Mi hai fatto venire voglia di cornetto e
succo anche a me -.-" sallo che ti odio!

⁴⁵ Titti e Silvestro, Looney Tunes, Warner Bros.

Per quanto riguarda invece gli errori grammaticali, troviamo:

- 15 occorrenze, 2 *menu* per *menù*, 2 *po* per *po'*, 1 *pò* per 1 *po'*, 1 *qual'è* per *qual è*, 1 *sto'* per *sto*, 1 *mha* per *mah*, 1 *l'ho giudico* per *lo giudico*, 2 *freccie* per *frecce* e 1 *c'ho* per *ci ho*, 1 *iniziero* per *inizierò*, 1 *melenzane* per *melanzane*, 1 *melenzana* per *melanzana*.

1. Ke belle le nostre freccie tricolore!! #giro

2. @mariop89 ma nn **l'ho giudico** per le sue cazzate...nn lo faccio mai cn nessuno.

3. @fracchinetti a casa mia tramezzini di **melenzana**.... :)

Dai dati presi in esame si nota come il corpus degli utenti non professionisti sia molto più ricco di variazioni appartenenti all'italiano non standard. Si contano infatti 28 fenomeni endofasici, 43 occorrenze di abbreviazioni, 19 variazioni grafiche e 44 espressioni dialettali per un totale di 134 fenomeni, contro i 13 fenomeni endofasici, le 32 occorrenze di abbreviazioni, le 4 variazioni grafiche e le 22 espressioni dialettali inserite dai professionisti, per un totale di 71 fenomeni rilevati.

Tutto questo porta a evidenziare come, almeno all'interno del nostro corpus, tali fenomeni siano maggiormente utilizzati dagli adolescenti e dai non professionisti. Considerando anche che, degli 80 utenti non professionisti 31 sono studenti universitari, 7 di scuola superiore e gli altri 42 dimostrano comunque di saper utilizzare l'italiano standard (infatti i 15 errori grammaticali trovati nel corpus appartengono soltanto a 8 utenti⁴⁶) il fatto di essere personaggi pubblici professionisti incide, probabilmente, sull'uso di queste variazioni, portandoci a pensare che, anche in porzioni così brevi di testo (140 caratteri) si abbia la presenza di quella che Tavosanis, in *L'Italiano del web* (*ivi*, p. 46) chiama scrittura rapida e revisionata, ovvero "(...) testi scritti rapidamente, senza progettazione preliminare, ma con qualche grado di revisione (per esempio, correzioni di errori di battitura)".

⁴⁶ Sara Ceccherini, Giorgia Zuccalà, Marianna Fiori, Arianan Giannatempo, Francesca Art, Gianluca Scudieri, Antonella Caldandrucci, Adolfo Speranza.

4. L' analisi dei dati con Excel

4.1 Criteri di classificazione

Tutti i dati raccolti durante questo lavoro di tesi sono stati inseriti nel file excel denominato *StatisticheTweet.xls* e prodotto con Microsoft® Excel 2007.

Il file è composto da quattro fogli di lavoro, *Utenti professionisti*, *Utenti non professionisti*, *Riepilogo*, *Distribuzione emoticon*, *Correlazioni professionisti e Correlazioni non professionisti*.

Nel primo e nel secondo foglio sono presenti tutti i dati appartenenti agli utenti professionisti e non professionisti suddivisi nelle seguenti colonne:

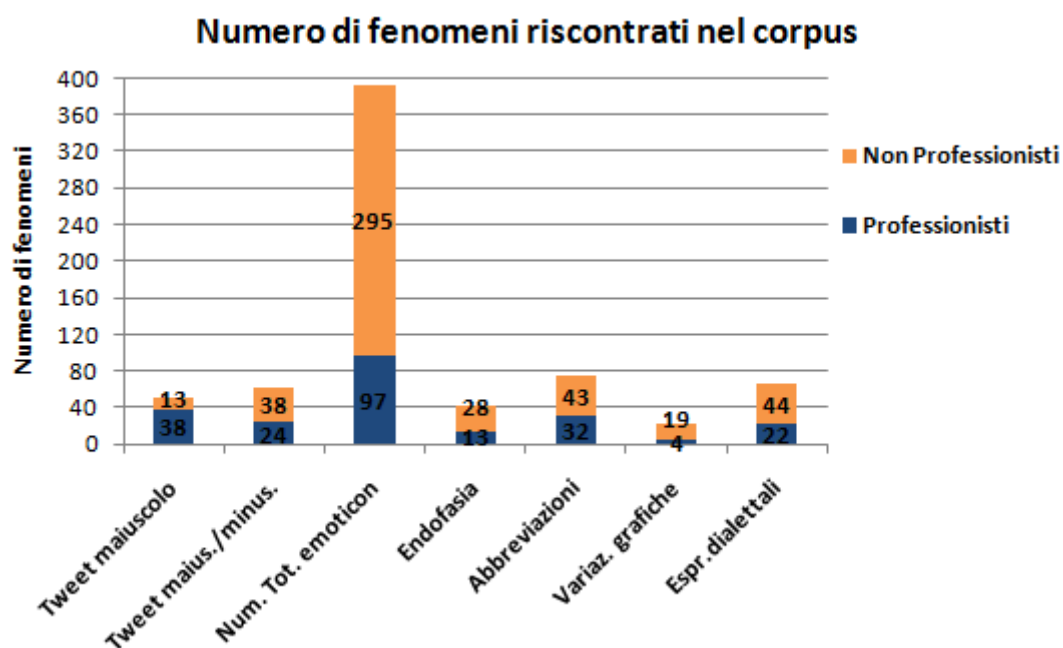
- *Nome*: vi sono specificati, se disponibili, il nome e il cognome dell'utente;
- *Url*: contiene l'indirizzo internet della pagina del profilo dell'utente di Twitter;
- *Num. Tot. Tweet*: contiene il numero totale dei tweet di ogni utente. Il valore massimo è 20;
- *Tweet italiano*: contiene il numero dei tweet in italiano scritti dall'utente;
- *Tweet inglese*: contiene il numero dei tweet in inglese scritti dall'utente;
- *Tweet tedesco*: contiene il numero dei tweet in tedesco scritti dall'utente;
- *Tweet francese*: contiene il numero dei tweet in francese scritti dall'utente;
- *Tweet spagnolo*: contiene il numero dei tweet in spagnolo scritti dall'utente;
- *Tweet solo link*: vi è specificato il numero dei tweet in cui è presente soltanto un link, senza ulteriori caratteri o parole;

- *Tweet ibridi*: contiene il numero dei tweet multilingua scritti da ogni utente;
- *Tweet solo emoticon*: contiene il numero di tweet che hanno al loro interno soltanto emoticon;
- *Tweet solo hashtag*: contenente il numero di tweet aventi come testo soltanto un hastag;
- *Tweet Maiusc.*: vi è specificato il numero dei tweet scritti interamente in maiuscolo;
- *Tweet Maiusc./Minusc.*: vi è specificato il numero dei tweet scritti in parte in maiuscolo e in parte in minuscolo;
- *Num. Tot. Emoticon*: contiene il numero delle emoticon inserite dall'utente;
- *% Emoticon*: contiene la percentuale di emoticon utilizzate dai singoli utenti in rapporto al totale delle emoticon usate dalla categoria di appartenenza;
- *Endofasia*: contiene il numero di varietà endofasiche impiegate da ogni utente;
- *Abbreviazioni*: contiene il numero di abbreviazioni utilizzate da ogni utente;
- *Variaz. Grafiche*: contiene il numero di variazioni grafiche rilevate per ogni utente;
- *Espress. Dialettali*: contiene il numero di espressioni dialettali utilizzate da ogni utente;

Nel foglio di lavoro *Riepilogo* si trovano invece alcuni dati utilizzati per la costruzione dei grafici contenuti all'interno dello stesso foglio. In particolare nel foglio *Riepilogo* si trovano il numero totale dei tweet inseriti dalle due categorie, il

confronto tra le tipologie di sesso - maschile o femminile - riscontrate nei due campioni di utenti, il numero dei tweet nelle diverse lingue individuate e il riepilogo di tutte le varietà dell'italiano standard riscontrate nel corpus. Sulla base di tutti questi dati sono stati poi costruiti i relativi grafici.

FIGURA 4.1 Uno dei grafici presenti all'interno del file excel *StatisticheTweet.xls* e raffigurante il numero di fenomeni riscontrati all'interno di tutto il corpus



Il foglio *Distribuzione emoticon* contiene una tabella pivot che mostra la distribuzione delle emoticon all'interno delle rispettive categorie di utenti. Sui dati di questa tabella pivot sono stati poi costruiti i relativi grafici sia per i professionisti che per i non professionisti (v. FIGURA 3.3, 3.4).

Gli ultimi due fogli, *Correlazioni professionisti* e *Correlazioni non professionisti*, illustrano le correlazioni tra i diversi fenomeni contenuti nel corpus e sono accompagnati dai rispettivi grafici.

4.2 Le correlazioni tra i diversi fenomeni

La correlazione è la misura della relazione fra due o più variabili e indica la tendenza che hanno due variabili, X e Y, a variare insieme (covarianza).

In statistica si usa definire il grado di correlazione fra due variabili attraverso un indice di correlazione, detto anche indice di correlazione di Pearson e indicato con ρ_{xy} , che può andare da -1 a 1 .

$$-1 \leq \rho_{xy} \leq 1$$

Se:

- $\rho_{xy} > 0$, le variabili x e y sono direttamente correlate, o correlate positivamente;
- $\rho_{xy} = 0$, le variabili x e y non sono correlate;
- $\rho_{xy} < 0$, le variabili x e y sono inversamente correlate, o correlate negativamente.

Per la correlazione diretta si distingue inoltre:

- $0 < \rho_{xy} < 0,3$: si ha una correlazione debole;
- $0,3 < \rho_{xy} < 0,7$: si ha una correlazione moderata;
- $\rho_{xy} > 0,7$: si ha una correlazione forte.

Inoltre, quando si parla di statistica e di correlazione, è utile introdurre anche il concetto di coefficiente di determinazione, meglio noto come R^2 .

R^2 esprime in quale misura il modello è in grado di spiegare la relazione tra due o più variabili e può assumere un valore compreso tra 0 e 1 . Più è elevato il valore di R^2 , tanto meglio la variabile dipendente (Y) è spiegata dai fattori considerati nel modello, e tanto più affidabile è il risultato ottenuto.

Non è possibile definire in modo certo e assoluto la qualità di R^2 , ma si possono definire gli ordini di grandezza che aiutano a capire e a interpretare meglio il suo risultato.

Se:

- $R^2 \geq 0,8$: il risultato ottenuto è molto attendibile. I fattori considerati nel modello contribuiscono a spiegare bene l'entità di Y.
- $R^2 \geq 0,6 < 0,8$: il risultato ottenuto è abbastanza attendibile. I fattori considerati nel modello contribuiscono a spiegare in modo soddisfacente Y.
- $R^2 \geq 0,4 < 0,6$: il risultato ottenuto è poco attendibile.
- $R^2 < 0,4$: il risultato è troppo poco attendibile o potrebbe essere basato su dati errati.

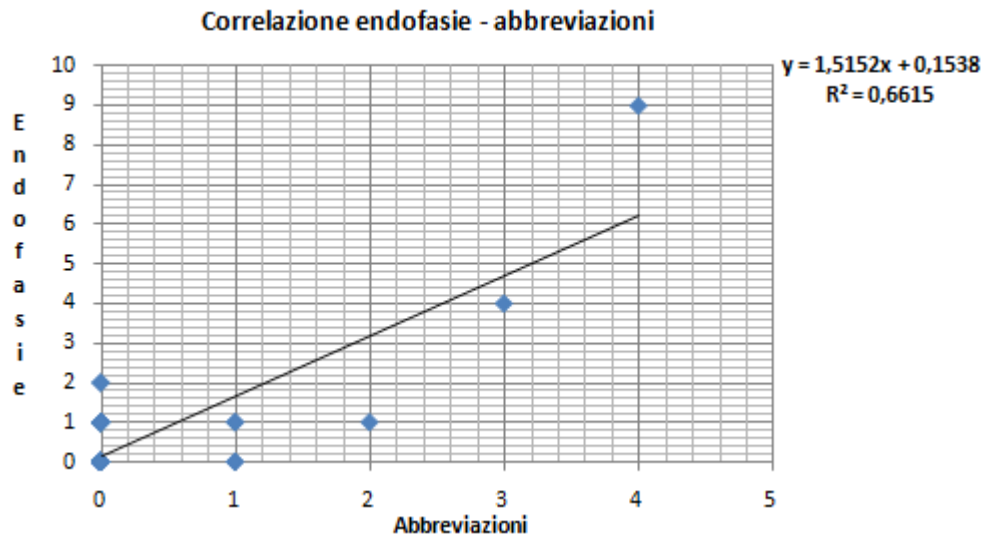
Sulla base di tutte queste informazioni si sono calcolate quindi le varie correlazioni tra i vari fenomeni contenuti all'interno del corpus.

Nello specifico, per quanto riguarda il corpus dei professionisti, sono state calcolate le seguenti correlazioni:

- Correlazione tra il numero di tweet in inglese e il numero delle emoticon;
- Correlazione tra il numero di endofasie e il numero di abbreviazioni;
- Correlazioni tra il numero di abbreviazioni e il numero di espressioni dialettali;
- Correlazioni tra il numero di emoticon e il numero di abbreviazioni;

col risultato che soltanto una, quella tra endofasie e abbreviazioni, mostra una correlazione forte (0,81) e un R^2 abbastanza attendibile (0,66), come si evince dal grafico sottostante.

FIGURA 4.2 Il grafico della correlazione tra le endofasie e le abbreviazioni riferito al corpus dei professionisti.



Per quanto riguarda invece il corpus dei non professionisti sono state calcolate le seguenti correlazioni:

- Correlazione tra il numero di tweet in inglese e il numero delle emoticon;
- Correlazione tra il numero di endofasie e il numero di abbreviazioni;
- Correlazione tra il numero di variazioni grafiche e il numero di espressioni dialettali;
- Correlazioni tra il numero di abbreviazioni e il numero di espressioni dialettali;
- Correlazione tra il numero di variazioni grafiche e il numero di endofasie;
- Correlazioni tra il numero di emoticon e il numero di abbreviazioni;

col risultato che due, quella tra numero di endofasie e numero di abbreviazioni e quella tra numero di variazioni grafiche e numero di endofasie, mostrano correlazioni abbastanza soddisfacenti. In particolare la correlazione tra i fenomeni di endofasia e il numero di abbreviazioni è di 0,74, ma con un coefficiente di determinazione R^2 di 0,54 che, nonostante la correlazione positiva, ci dice che i risultati non sono

completamente attendibili. Allo stesso modo la correlazione tra il numero di variazioni grafiche e il numero di endofasie ha una correlazione di 0,66, ma ha pur sempre un indice R^2 di 0,44, che mostra un risultato poco attendibile.

FIGURA 4.3 Il grafico della correlazione tra le endofasie e le abbreviazioni riferito al corpus dei non professionisti.

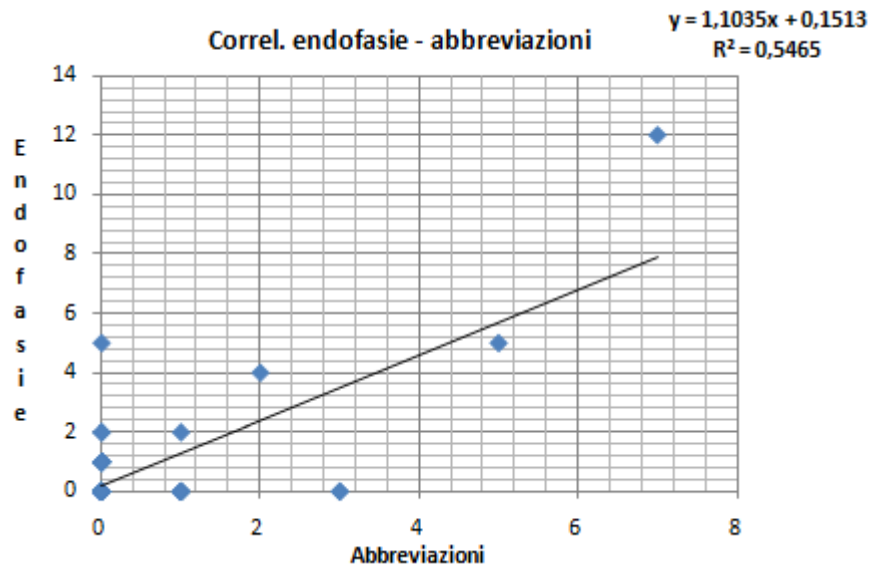
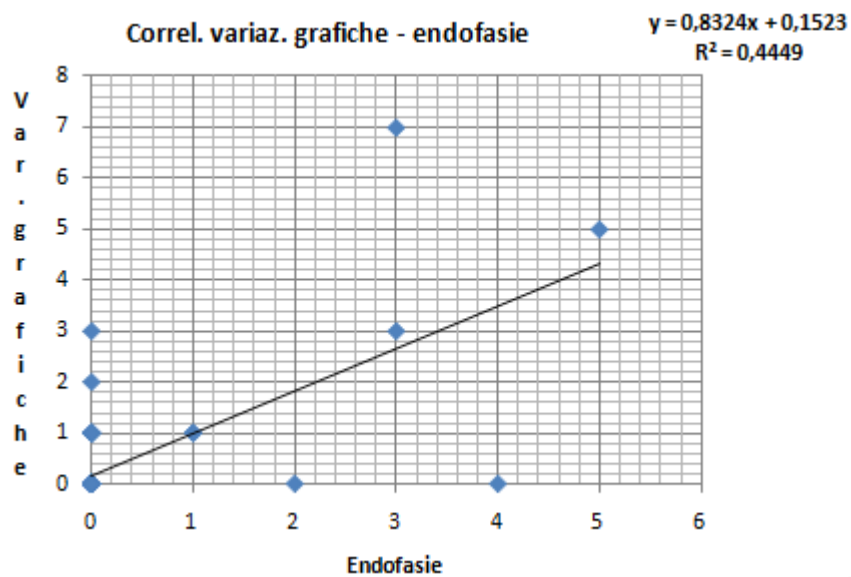


FIGURA 4.4 Il grafico della correlazione tra il numero di variazioni grafiche e il numero di endofasie riferito al corpus dei non professionisti.



Conclusioni

Nelle pagine precedenti si è visto come Twitter oggi sia diventato in pochissimo tempo uno strumento di comunicazione di massa e di giornalismo partecipativo utilizzato da circa 200 milioni di persone appartenenti a diverse tipologie di utenti: adolescenti, giovani, adulti, studenti, celebrità e professionisti di vario genere.

Nello specifico, in questo lavoro di tesi, si sono analizzati linguisticamente 2.895 tweet inseriti da 160 utenti italiani.

Dai dati in nostro possesso è emerso che il linguaggio utilizzato su Twitter presenta molte caratteristiche che lo mettono in analogia con il linguaggio degli SMS, con quello della Rete e con quello giovanile. Tra queste caratteristiche ci sono senz'altro l'utilizzo delle emoticon, dei fenomeni endofasici, delle abbreviazioni e il numero ridotto di caratteri - i 140 di Twitter contro i 160 degli SMS che, anche se non sfruttati interamente, hanno comunque degli effetti sulla pianificazione del testo e sull'utilizzo di numerose scelte grafiche -.

Tutte queste varietà sono in larga parte contenute all'interno del corpus degli utenti non professionisti, e questo probabilmente perché costituito in prevalenza da giovani, donne e adolescenti.

Inoltre, sempre all'interno del corpus, sono presenti anche termini specifici di Twitter stesso, come tweet, retweet, hashtag, follower e così via. Tuttavia non sarebbe giusto parlare di un gergo di Twitter, in quanto sono poche le parole di questo genere impiegate dagli utenti.

Il linguaggio di Twitter è inoltre di tipo informale. La deissi non si allontana mai dal tu e gli utenti, pur non conoscendosi, non si danno mai del Lei. L'interpunzione come abbiamo visto è abbastanza presente, anche se spesso segue l'intonazione del parlato piuttosto che la forma logico-sintattica dello scritto.

La scarsa pianificazione dei testi può spiegare la presenza degli errori typo, mentre l'appartenere all'una o all'altra categoria può giustificare la presenza di errori grammaticali, che comunque sono meno di quello che ci si aspettava di trovare, perlomeno nel corpus degli utenti non professionisti.

Certamente questa analisi non intende generalizzare i tratti riscontrati all'interno dei dati presi in esame; definiamo quindi i suoi relativi risultati come rappresentativi del corpus analizzato e non come aspetti specifici di tutte le centinaia di migliaia di tweet in lingua italiana disponibili su Twitter.

È quindi prevedibile che in un campione maggiore di tweet sia possibile riscontrare che qualche tratto qui analizzato risulti più o meno presente o del tutto assente; nonostante ciò questi dati restano comunque validi a illustrare le particolarità e le varietà che l'italiano può assumere in corrispondenza di determinate tipologie di utenti e di ambienti sociali.

Potrebbe quindi essere un'idea quella di ampliare questo lavoro di tesi, nello specifico analizzare un altro corpus di tweet inserito da un altro numero di utenti e verificare se i risultati ottenuti si allontanano o si avvicinano, e in tal caso di quanto, a quelli rilevati all'interno di queste pagine.

Bibliografia

Huberman, Bernardo A., Romero, Daniel M. e Wu Fang. 2008. *Social networks that matter: Twitter under the microscope*. pp.1-12. Social Computing Lab, Hp Laboratories, Palo Alto and Cornell University, Ithaca, U.S.A.

Boyd, Danah, Golder, Scott e Lotan Gilad. 2009. *Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter*. 42nd Hawaii International Conference on System Sciences, pp.1-10.

Boyd, Danah e Ellison N.B. 2007. *Social Network Sites: Definition, History and Scholarship*. Journal of Computer-Mediated Communication, vol. 13.

Collins, Tim cur. Incorvaia, Antonio. 2009. *Il piccolo libro di Twitter*. Sperling & Kupfer.

Conti, Luca. 2010. *Comunicare con Twitter. Creare relazioni, informarsi, lavorare*. Milano, Hoepli.

Dardi, Federica. 2011. *Twitter, guida per il professionista della chiacchiere digitali*. Milano, Apogeo.

Herring, Susan e Honeycutt Courtenay. 2009. *Beyond Microblogging: Conversation and Collaboration via Twitter*. 42nd Hawaii International Conference on System Sciences, pp.1-10.

Lavrenko, Victor, Osborne Miles e Petrovic Sasa. 2010. *The Twitter Edinburgh Corpus*. NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media, pp.25-26. Los Angeles, U.S.A.

Maltese, Maurizio. 2010. *Due punti parentesi chiusa: un profilo linguistico dell'emoticon nei forum italiani*. Elaborato di laurea triennale, Università di Pisa.

Masilla, Rosanna. 2011. *Raccolta e analisi di un corpus di testi da Twitter*. Tirocinio formativo, Università di Pisa.

Nelson, Stephen. 2002. *Excel Data Analysis*. Wiley Publishing, cap. 4 e 5.

Pak, Alexander, Paroubek, Patrick. 2010. *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Proceedings of the Seventh conference on International Language Resources and Evaluation, European Language Resources Association, pp. 1320-1326. Valletta, Malta.

Pistolesi, Elena. 2009. *Il parlar spedito. L'italiano di chat, e-mail e SMS*. Padova, Esedra editrice.

Riva, Giuseppe. 2010. *I Social Network*. Bologna, Società editrice il Mulino.

Serianni, Luca. 1997. *Italiano*. Torino, Garzanti Editore.

Tavosanis, Mirko. 2011. *L'italiano del web*. Roma, Carocci Editore.