



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

**Analisi linguistica di commenti estratti da post
di politici italiani su Facebook**

Candidato: *Gabriele Burgarella*

Relatore: *Mirko Tavosanis*

Correlatore: *Maria Simi*

Anno Accademico 2018-2019

Indice

1. Introduzione	p. 5
2. Perché Facebook?	p. 6
2.1 Dove si informano gli italiani?	p. 9
2.2 Pagine ufficiali e pagine unofficial	p. 10
3. Estrazione dei dati	p. 12
3.1 Api Graph, Scraper e Netvizz v1.6.....	p. 12
3.2 Facepager	p. 15
3.3 Scelta del campione	p. 20
3.3.1 Prima rilevazione: 13 -15 marzo	p. 20
3.3.2 Seconda rilevazione: 30 marzo	p. 26
4. Analisi	p. 29
4.1 Il “che polivalente”	p. 29
4.2 Strumenti per l’analisi	p. 30
4.2.1 Universal Dependencies	p. 31
4.2.2 UDPipe	p. 33
4.3 Il “che” polivalente nel corpus di commenti analizzati	p. 37
4.3.1 Principali usi non standard del “che”	p. 38
4.3.2 Altri usi non standard del “che”	p. 39
4.3.3 “Che” di difficile interpretazione	p. 41
4.3.4 “Che” non classificabili	p. 41
4.3.5 “Che” standard atipici	p. 41
4.3.6 Costruzioni ricorrenti con il “che”	p. 43
4.3.7 Uso volutamente erroneo del “che”	p. 44
5. Conclusioni	p. 45
Bibliografia	p. 46

Sitografia p. 47

1. Introduzione

Si dice che Kennedy abbia vinto grazie alla televisione. Quaranta anni più tardi Barak Obama ha vinto le elezioni grazie alla rete. In occasione delle presidenziali del 2008 è stato infatti il primo politico a sfruttare i social per la propria campagna elettorale. Da quel famoso *Yes We Can*, entrato nell'immaginario collettivo come sinonimo di rinnovamento e impegno civile, ad oggi, i social hanno cambiato radicalmente la politica. Oggi, dopo 11 anni da quelle elezioni, la politica sembra non poter più fare a meno di internet. I leader di ogni paese si sfidano infatti ogni giorno a colpi di tweet, post e selfie instaurando un contatto diretto con i propri seguaci, i quali a loro volta commentano le attività dei propri leader preferiti generando una quantità incredibile di dati linguistici. Mentre però il linguaggio dei politici non è particolarmente interessante dal punto di vista linguistico perché sembra poco innovativo rispetto a quello della comunicazione tradizionale broadcast, affidata alla carta stampata o alla televisione, al contrario quello dei commenti, scritti da non professionisti, rappresenta un campione di comunicazione politica dal basso (Tavosanis, 2015).

Nella nostra analisi sono stati considerati quindi i commenti degli utenti sotto ai post dei maggiori leader politici italiani su Facebook ed è stata effettuata l'analisi linguistica sui dati ricavati. Ciò è stato possibile grazie a nuovi strumenti per l'analisi linguistica che hanno rivoluzionato la ricerca. Uno di questi, chiamato UDPipe, è stato utilizzato per annotare morfo-sintatticamente il corpus e scoprire in quanti casi gli utenti italiani utilizzano il "che" in modo improprio.

Il corpus su cui è stato effettuato lo studio contiene commenti estratti dai post di 4 politici (Matteo Salvini, Luigi Di Maio, Giorgia Meloni e Matteo Renzi) e un giornalista (Enrico Mentana).

2. Perché Facebook?

Facebook nasce nel 2004 ad Harvard e dopo solo 13 anni raggiunge quota 2,23 miliardi di utenti attivi mensilmente (Scarcella, 2017), classificandosi come primo servizio di rete sociale per numero di utenti attivi. Oggi è il social network con più utenti sul web e da anni ormai costituisce terreno di confronto politico. Secondo il report 2019 di *wearesocial.com* in Italia si registrano circa 35 milioni di utenti (pari al 60% della popolazione totale). La crescita costante di Facebook in termini di utenti in Italia appare chiara: a fronte di una piccola diminuzione della popolazione nel nostro paese, gli utenti italiani sul social network sono cresciuti di 1 milione solo nell'ultimo anno.

Fin da subito molti politici hanno intravisto la possibilità di usare il social per avere un ulteriore mezzo di contatto con i cittadini, per ascoltare le loro richieste e per capire le esigenze dell'elettorato. Il primo che capì le straordinarie possibilità di Facebook fu Barak Obama, ex Presidente degli Stati Uniti che sfruttò la rete per la campagna per le presidenziali del 2008 contro John McCain. Da quel momento la pagina dell'ex presidente ha continuato a crescere e oggi conta addirittura 54 milioni di "Mi Piace". Molti sono i leader che negli anni si sono approcciati al social: da Narendra Modi (Primo Ministro indiano) a Erdogan (presidente turco) passando per Angela Merkel (Cancelliere tedesco) e Donald Trump (attuale Presidente degli Stati Uniti). In Italia l'esponente con maggiore seguito è attualmente Matteo Salvini, Ministro degli Interni e leader della Lega (3,5 milioni di "Mi Piace").

In Italia fino a 8 anni fa il quadro era decisamente diverso. Il politico più seguito era Nicola Vendola (514 mila "Mi Piace" - Presidente della Regione Puglia) seguito da Silvio Berlusconi e Antonio Di Pietro. Faceva già parlare di sé Matteo Renzi, allora Sindaco di Firenze, per il suo dialogo costante con i cittadini sulla piattaforma. Secondo uno studio realizzato nel 2011 da Stefano Epifani, docente alla facoltà di Scienze della Comunicazione presso la Sapienza di Roma, e condotto per conto dell'Istituto Superiore di Studi Politici San Pio V, solo il 29% di chi sedeva alla Camera o al Senato aveva aperto un profilo su Facebook. Oggi ci

stupiremmo se non trovassimo sui social media la pagina di un politico. Può suonare davvero strano a confronto con i numeri in cui ci imbattiamo ogni giorno ma il superamento da parte di “Nichi Vendola” del mezzo milione di seguaci sui social ebbe grande risonanza mediatica. Quei numeri dopo 8 anni non basterebbero per trovarsi tra i 10 politici più seguiti su Facebook. Questo perché con l’incremento vertiginoso del numero degli utenti, il social network creato da Mark Zuckerberg è diventato uno strumento imprescindibile per ogni personaggio politico italiano. Infatti su Facebook ogni giorno migliaia di utenti trattano dei più svariati temi. Avere a disposizione uno spazio per intercettare bisogni e ricevere feedback diventa importante ogni giorno di più soprattutto in un campo sociologico come la politica. Da quando la scena politica ha iniziato a comunicare tramite il social media i cittadini si sono sempre più sentiti partecipi nelle discussioni politiche tramite l’illusione che il network potesse creare una sorta di canale comunicativo diretto con il leader politico di riferimento. Ciò ha portato a un numero elevato di interazioni virtuali su temi di interesse comune. Tutta questa grande quantità di dati sarà oggetto di esami e di studio per capire “dove” e “come” i cittadini si approcciano alla politica su Facebook. Sarà molto interessante comprendere i cambiamenti linguistici e il modo di comunicare degli utenti in situazioni di dibattito su canali non convenzionali. In ultimo luogo, prima di effettuare l’analisi, è necessario inoltre capire quale sarà il campione d’indagine. Facebook è infatti una piattaforma aperta a tutti gli utenti senza distinzioni di sesso, età, religione o titolo di studio.

Secondo il quindicesimo rapporto Censis nel 2018 circa il 72,5% degli italiani utilizza i social network per comunicare. Le percentuali legate alla differenza di genere sono pressoché irrilevanti. Infatti sono presenti il 71,6% degli uomini e il 73,6% delle donne. Invece è particolarmente interessante il dato riguardante le età degli utenti. È presente sui social l’85% dei nati negli anni successivi al 1990 (in leggero calo) e l’88% di chi ha tra i 30 e i 44 anni. Sono in vertiginosa crescita gli utenti nati tra il 1960 e il 1974 che rispetto al 2017 aumentano di 11 punti raggiungendo il 75%. In aumento anche gli over 65 che si approcciano a questo modo di comunicare: poco più di un terzo utilizza social network. È possibile notare inoltre una grossa differenza nella presenza sui social tra chi possiede la licenza

elementare o media e chi possiede un diploma o laurea. Sulle varie reti sociali difatti è presente l'85% di diplomati o laureati, mentre si registra soltanto il 62% di chi possiede la licenza elementare o media. Dopo aver inquadrato l'utenza dei social network è necessario esaminare più in dettaglio la situazione di Facebook e per farlo ci serviremo di nuovo del quindicesimo rapporto dell'Istituto Censis. A Facebook è iscritto il 56% della popolazione italiana pari a circa 34 milioni di persone (dato in calo del 7% rispetto al 2017). Nel 2018 si registra il 74% degli utenti di età compresa tra i 30 e i 44 anni e il 70% dei nati dopo il 1990.

Campione	Social Network (%)	Facebook (%)
Totale popolazione	72,5	56,0
Maschi	71,6	54,8
Femmine	73,2	57,1
Licenza elementare/media	62,0	46,4
Diploma/Laurea	85,4	67,9
14 - 29 anni	85,1	70,7
30 - 44 anni	88,5	74,0
45 - 64 anni	75,2	55,2
65 - 80 anni	34,3	20,9

Tabella 1 - Gli italiani su Facebook (Fonte: CENSIS 2018)

I numeri sono molto interessanti anche se è necessario specificare che i giovanissimi su Facebook risultano in netto calo rispetto ai dati del 2016 e 2017. Oltre a Instagram, che in un anno ha visto crescere i suoi iscritti tra i 14 e i 29 anni del 7% circa e sicuramente ha influito molto su questo calo di popolarità, anche la maggiore frequentazione da parte degli adulti di questi spazi ha contribuito al calo dei giovanissimi su Facebook (Censis, 2018, p. 41). I numeri restano comunque molto interessanti e tutti, come da previsione, a favore dei più giovani. Nella Tab.1 è possibile notare questa tendenza. Non sono riportate le analisi l'uso di Facebook nelle varie aree geografiche perché la variazione è minima: i picchi sono costituiti dall'Italia centrale e dal Nord-est in cui è iscritto il 61,5% e il 60% della popolazione, i punteggi più bassi si registrano nel Sud/Isole e nel Nord-Est con il 53% e il 52,9%.

2.1 Dove si informano gli italiani?

L'avvento dei social network ha senz'altro cambiato le regole dell'informazione. Le nuove reti sociali costituiscono però soltanto la punta dell'iceberg della trasformazione dei mezzi di informazione iniziata più di dieci anni fa. Secondo il quindicesimo rapporto del Censis, dal 2007 a oggi l'utenza che utilizza media tradizionali come tv, radio, libri e quotidiani è diminuita (e in alcuni casi in modo drastico: quotidiani in generale -30,6%) a scapito di nuovi media come siti web di informazione, internet e smartphone. Questo processo di trasformazione è avvenuto prevalentemente negli ultimi 7-8 anni. Infatti dal 2001 al 2011 i media tradizionali non hanno subito eccessivamente l'avvento di siti web e blog (la tv ha perso solo l'1,4% della propria utenza) e anzi è aumentato l'uso della radio e dei libri. La carta stampata era l'unica ad aver registrato una diminuzione dell'utenza. Fino al 2011 dunque i media tradizionali non sono stati impensieriti dall'avvento di internet. Dal 2011 al 2018 invece si sono registrati dei cambiamenti significativi. La tv (-4,5%), i libri (-14,3%) e i quotidiani (-10,4%) hanno visto una grande perdita in termini di utenza negli ultimi 7 anni. Al contrario Internet (+25,3%) e i telefoni cellulari (+8,9%) hanno visto crescere il loro utilizzo. Un dato che può aiutare a capire questo processo riguarda senz'altro le spese effettuate dalle famiglie italiane. Negli ultimi 10 anni abbiamo assistito a un calo del 37,2% degli acquisti di giornali e libri a fronte di aumenti vertiginosi di spesa per telefoni (+187%) e computer (+47%).

Il Censis definisce questo quadro "transmedialità", ovvero l'ulteriore sviluppo della multimedialità. "Con quest'ultima si indicava un contesto all'interno del quale operavano diversi mezzi di comunicazione" e nel quale "tutti i mezzi dialogavano con tutti mantenendo ognuno la propria specificità" (Censis, 2018). Oggi, con l'avvento di Internet e dei social media, questa specificità è venuta meno. Non si parte più dai media veicolando il messaggio in base al mezzo che dovrà eseguire l'azione bensì dai contenuti per passare solo in un secondo momento alle piattaforme attraverso le quali diffonderli. I nuovi media, potendo veicolare contenuti diversi, possono svolgere un maggior numero di funzioni rispetto ai media tradizionali.

Questa trasformazione ha influito non poco anche sugli strumenti che gli italiani utilizzano per informarsi. Nel 2017 infatti l'uso dei siti web di informazione ha scavalcato quello dei quotidiani cartacei. Facebook rappresenta per moltissimi una fonte di informazione giornaliera. Infatti grazie a strumenti come la condivisione di link e articoli, permette a molte persone di rimanere aggiornate. La percentuale degli italiani che dichiara di informarsi su Facebook è del 25,9 %, cifra che, anche se in netto calo, supera quella dei quotidiani cartacei, dei notiziari non-stop, dei giornali radio e dei siti web di informazione.

2.2 Pagine ufficiali e pagine Unofficial

Per l'analisi dei "che polivalenti" saranno usate esclusivamente pagine verificate che possiedono il *badge* blu. Esistono però pagine su Facebook chiamate dagli addetti ai lavori "unofficial". Si tratta di spazi che non rispondono a nessuna fazione politica, né tantomeno a leader. Inoltre essendo entità senza nome e cognome risulta difficile risalire all'autore del contenuto. In realtà queste pagine, neutre e apolitiche di facciata, si occupano, tra le altre cose, di propaganda vera e propria. Tutti i post pubblicati contengono una foto e una scritta breve che esprime derisione o indignazione. Alcuni esempi di pagine non ufficiali sono: "Roby", "Buona E Sincera,ma Quando Serve Bastarda", "Contro il Potere" ecc. La discussione attorno alle pagine "unofficial" si è riaccesa nel settembre 2018 quando Alessio De Giorgi, social media strategist di Matteo Renzi e consulente del Partito Democratico, ha riferito i risultati della sua analisi riguardante le pagine cosiddette non ufficiali ai media. Dal suo report (ottobre 2017 – settembre 2018) emerge che circa 9 pagine non ufficiali su 10 supportano le forze di governo. Secondo De Giorgi nel periodo considerato le pagine con il maggior tasso di crescita sono state quelle legate alla Lega passate da 9 a 15 milioni di interazioni mentre l'area vicina al Movimento 5 Stelle e quella vicina al centrosinistra sono rimaste in sostanza inalterate. Il dato più interessante è però quello legato al numero delle interazioni delle pagine unofficial: nel giro di dieci mesi hanno raddoppiato il numero di interazioni (da 19 milioni a 32,7 milioni) superando le principali testate giornalistiche italiane e alcuni dei principali leader politici. Federico Fubini, giornalista del Corriere, per verificare i dati di De Giorgi ha chiesto di esaminare i dati forniti dal social media strategist a

Giovanni Zagni, giornalista di Pagella Politica (Fubini, 2018b). Dall'analisi incrociata di Zagni emerge che la crescita del mondo unofficial è stata addirittura superiore rispetto a quanto registrato da De Giorgi. Viene sottolineato però che, anche se sono prevalentemente filogovernative, alcune di queste risultano orientate verso il centrosinistra. Nella Tab. 2 è riportata la diffusione delle pagine a seconda della loro tipologia.

Tipologia	%
Unofficial	33,5
Media	34,1
Esponenti Politici	27,4
Partiti	5,0

Tabella 2 - Suddivisione per tipologia delle pagine che trattano di politica (Fonte: Osservatorio Social di Alessio De Giorgi)

In un articolo pubblicato sul Corriere il 17 settembre 2018 Federico Fubini ha cercato di fare chiarezza su questo universo unofficial intervistando Francesco Gangemi, muratore disoccupato di Taurianova (RC) e admin di una delle pagine non ufficiali (“sputtaniamotutti.com”). La sua pagina ha generato quasi un milione e mezzo di interazioni su Facebook ed è arrivata a raggiungere anche dieci milioni di lettori con un solo post (Fubini, 2018a). Nell'intervista Gangemi ha spiegato che la sua attività non ha alcun intento politico. Infatti il suo unico scopo è quello di generare contenuti forti in modo che i lettori dei suoi post clicchino sul link condiviso così da essere reindirizzati al suo sito. Questo fenomeno, chiamato clickbaiting¹, riesce a fargli guadagnare fino a 600 euro al mese grazie alle pubblicità fornite dal servizio pubblicitario di Google (AdSense). Dalla versione di Gangemi emerge che difficilmente su queste pagine vi saranno contenuti impopolari ma non è detto che il governo rimanga popolare per tutta la legislatura. Ecco che le pagine potrebbero spostarsi quindi su altri leader e altri temi. La

¹ È un termine che indica un contenuto web la cui principale funzione è di attirare il maggior numero possibile d'internauti, per generare rendite pubblicitarie online.

versione di Gangemi non convince molti esponenti sia perché considerano impossibile generare tale indotto sia perché, nonostante la situazione precaria del disoccupato, non reputano accettabili i toni usati da queste pagine. Attorno a questo fenomeno sono ancora in corso studi e approfondimenti. Oggi, nonostante la costante attenzione mediatica su questi temi, non disponiamo di strumenti e dati a sufficienza per capire e classificare le pagine che contribuiscono a creare questo universo. Per questo motivo, nonostante l'elevata portata del fenomeno, nella nostra analisi non sono state considerate le pagine "unofficial" ma soltanto quelle ufficiali.

3. Estrazione dei dati

3.1 Api Graph, Scraper e Netvizz v1.6

L'estrazione dei commenti fino a non molto tempo fa era consentita dalle API Graph di Facebook.² Questo strumento, creato da Facebook stesso per gli sviluppatori, permetteva di estrarre tramite richieste GET dettagli pubblici relativi a pagine. Il suo funzionamento era semplice e intuitivo: una volta collegati al sito ufficiale bastava accedere con il proprio profilo e creare una app.



Figura 1 – L'interfaccia del tool di esplorazione per la API Graph

² <https://developers.facebook.com/>

Grazie a un tool di esplorazione, dopo aver ricevuto un access token valido (una *key* segreta) era possibile elaborare query e ricevere risposte dall'API Graph. Nella fig. 1 è mostrata l'interfaccia del tool di esplorazione.

L'Application Programming Interface Graph di Facebook, dal 2018, in seguito ad alcuni eventi è diventato uno strumento di difficile utilizzo per l'estrazione dei dati dal social network.

- 1) Il 17 marzo 2018 The New York Times e The Observer hanno riferito che la società di consulenza britannica *Cambridge Analytica*, fondata da Steve Bannon e Robert Mercer, aveva estratto informazioni personali su 50 milioni di utenti senza autorizzazione per scopi di tracciamento psicologico tesi a influenzare il comportamento dell'individuo (Rosenberg, 2018). Secondo le stime di Facebook sono stati acquisiti dati su 87 milioni di persone (Australian Broadcasting Corporation News, 2018).
- 2) Il 25 maggio dello stesso anno è stato attuato nell'Unione Europea il regolamento generale per l'utilizzo dei dati personali n. 2016/679 (General Data Protection Regulation) che tutela la circolazione dei dati personali.³

In seguito a questi eventi il social network ha deciso di innalzare le barriere di sicurezza e per questo motivo i dati non sono più facilmente accessibili dall'API Graph Explorer fornito. Come già spiegato nell'elaborato del dott. Nasuto (2018) lo *scraper* di Max Woolf⁴ utilizzato dalla dott.ssa Tortorelli (2017) non è più utilizzabile.

Sono stati comunque effettuati dei tentativi con altri scraper. In primo luogo, è stato testato "*Ultimate-Facebook-Scraper*" di Hassaan Elahi distribuito da GitHub⁵ con licenza MIT che, come da descrizione, non utilizza le API Graph di Facebook. Lo scraper utilizza uno script in python per estrarre i dati. È necessaria l'installazione

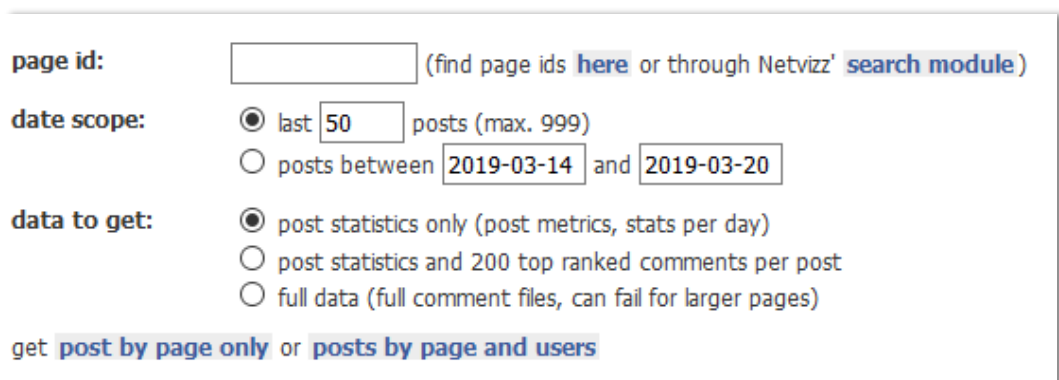
³ Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016)

⁴ Scraper di Max Woolf: <https://github.com/minimaxir/facebook-page-post-scraper>

⁵ Per testare lo scraper: <https://github.com/harismuneer/Ultimate-Facebook-Scraper>

del modulo “selenium” per effettuare la ricerca. Una volta lanciato lo script su un documento .txt all’interno del quale dovremo inserire l’indirizzo URL della pagina Facebook da “scrapare”, ci viene richiesta l’immissione di username e password del nostro profilo utente sul social network. Lo scraper però, molti tentativi effettuati, non solo non raccoglie le informazioni relative alla dopo pagina pubblica ma non estrae nemmeno i dati relativi ai commenti, oggetto di studio della tesi. Per questo motivo è stato abbandonato.

Successivamente è stata testata “Netvizz v1.6”⁶, un’applicazione presente su Facebook. Netvizz è uno strumento che aiuta ad analizzare diverse sezioni della piattaforma Facebook - principalmente pagine - a scopo di ricerca. L’applicazione è stata sviluppata nel 2009 da Bernhard Rieder, docente universitario di Lettere e Filosofia presso la University of Amsterdam. L’app riesce ad analizzare network di pagine connesse tra loro tramite “Mi Piace”, esaminare i post di una pagina, analizzare immagini di una pagina e generare statistiche per i link condivisi su Facebook. Per l’analisi dei commenti ai post, una volta effettuato l’accesso all’app, basterà cliccare su “page post”. Dopodiché si caricherà una nuova pagina dove, una volta inserito l’ID numerico della pagina che si intende prendere in considerazione, sarà possibile selezionare nel campo “data scope” il numero degli ultimi post da esaminare o in alternativa i post compresi in un certo periodo di tempo. Netvizz fornisce inoltre la possibilità di scegliere se estrarre statistiche sui post selezionati, statistiche sui post e i 200 migliori commenti per ognuno di essi oppure se esaminare tutto il materiale (compresi i commenti).



The screenshot shows the Netvizz v1.6 interface with the following fields and options:

- page id:** A text input field followed by the text "(find page ids [here](#) or through Netvizz' [search module](#))".
- date scope:** A radio button selected for "last posts (max. 999)". Below it, an unselected radio button for "posts between and ".
- data to get:** A radio button selected for "post statistics only (post metrics, stats per day)". Below it, two unselected radio buttons: "post statistics and 200 top ranked comments per post" and "full data (full comment files, can fail for larger pages)".
- At the bottom, the text "get [post by page only](#) or [posts by page and users](#)".

Figura 2 – L’interfaccia grafica di Netvizz v.1.6

⁶ Per maggiori informazioni su Netvizz v1.6: <https://apps.facebook.com/107036545989762/>

In quest'ultima opzione ("full data") viene anticipato tra parentesi che per i file di grandi dimensioni l'applicazione potrebbe fallire. In Fig. 2 è mostrata l'interfaccia grafica di Netvizz. Al termine della compilazione dei campi della ricerca avverrà l'estrazione e il download dei dati. L'app esporta tre file in formato .tsv (Tab Separated Values) all'interno dei quali troveremo il testo dei commenti (ovviamente senza autore), reazioni al post, statistiche giornaliere per il periodo coperto dal post e una serie di analisi sui post analizzati.

Il creatore però, in una lettera aperta sul sito *thepoliticsofsystems.net* dell'11 agosto 2018, afferma che l'app non verrà purtroppo aggiornata e non è possibile valutare per quanto potrà ancora essere utilizzata visto l'elevato numero di "salti mortali" necessari per tenerla al passo con le modifiche attuate sulle API di Facebook. (Rieder, 2018) Nonostante Netvizz costituisca oggi un ottimo strumento per la raccolta e l'analisi dei dati, per la ricerca che dovrà essere svolta non verrà utilizzata questa applicazione, rinunciando all'ampia gamma di soluzioni che offre. Questo perché la replicabilità dell'esperimento costituisce un requisito fondamentale per la scelta dello strumento per l'estrazione. Purtroppo la situazione di incertezza che oggi offre Netvizz ci obbliga a prendere in considerazione strumenti più affidabili.

3.2 Facepager

Per procedere all'estrazione dei dati, è stata scelta l'applicazione "Facepager" rilasciata da Jakob Jünger e Till Keyling (rispettivamente Ricercatore presso l'Università di Greifswald e Assistente Scientifico presso la Ludwig-Maximilians-Universität di Monaco di Baviera) su GitHub con licenza MIT⁷. Facepager, come si legge in descrizione sulla pagina dell'applicazione e sul blog di Keyling, "è stata creata per recuperare dati pubblici disponibili da Facebook, Twitter e altre API basate su JSON senza l'uso di linguaggi di programmazione o script predefiniti, lasciando ampi gradi di libertà all'utente". I dati sono memorizzati in database SQL e possono essere esportati in formato .csv (comma-separated values). I pacchetti di installazione sono disponibili sulla releases page, dove vengono inserite le versioni

⁷ Link a Facepager: <https://github.com/strohne/Facepager>

aggiornate.⁸ Per la nostra analisi utilizzeremo l'ultima versione rilasciata, la 3.10.2 pubblicata il 6 ottobre 2018. Una volta eseguito il download del file "Facepager_Setup_3_10.exe" è possibile iniziare l'installazione e, una volta terminata, cominciare ad usare l'app. Innanzitutto è necessario effettuare il login a Facebook dall'app inserendo e-mail e password. Dopo l'accesso viene rilasciato un Access Token valido per poter analizzare liberamente le varie pagine. Successivamente è possibile selezionare in alto "Add Nodes" per aggiungere il nodo che intendiamo visitare. Questo nodo corrisponderà alla pagina (o alle pagine) da analizzare. Basterà dunque aggiungere, nel campo di testo che comparirà, la *vanity url* (una o più) che si trova sia nell'indirizzo URL della pagina Facebook, sia sotto all'immagine del profilo. In figura 3 è mostrato il campo di inserimento.

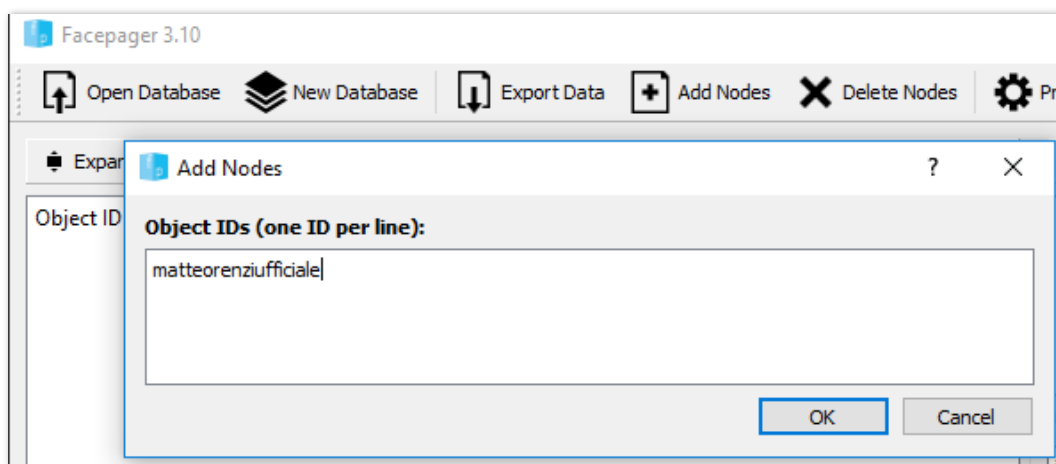


Figura 3 - Inseriamo come esempio il vanity url della pagina di Matteo Renzi

Una volta aggiunta la vanity url della pagina da analizzare sarà possibile notare come nell'interfaccia dell'applicazione è comparsa accanto a questo indirizzo, sotto al campo "Object Type" la notazione "seed". Questo sarà il nodo di partenza per tutte le analisi che verranno fatte su quella pagina. In fig. 4 è mostrata l'interfaccia principale.

Object ID	Object Type	Query Status	Query Time	Query Type	id
matteoreenziufficiale	seed				

Figura 4- Come appare il nodo padre nell'interfaccia di estrazione

⁸ Releases: <https://github.com/strohne/Facepager/releases>

Individuata la pagina occorre definire quali informazioni vogliamo ottenere dall'estrazione effettuata da Facepager. Sotto all'interfaccia mostrata precedentemente è programmata una finestra. All'interno di questa è possibile dare indicazioni all'applicazione riguardo alle operazioni che devono essere svolte. Per l'estrazione dei commenti degli utenti dai post di pagine su Facebook è necessario selezionare il campo "Facebook". Successivamente, per il corretto funzionamento, occorre mantenere aggiornato il percorso⁹. Nella voce "Resource" sarà necessario inserire quale campo della pagina individuata deve essere analizzato. Per estrarre i commenti selezioneremo il campo `<page>/post` e come parametro passeremo `<page>`. In questo modo saranno ottenuti gli ultimi post pubblicati da quella pagina. In fondo alla finestra verrà richiesta anche la quantità di post, espressa col numero di "pagine" da compilare con i post o i commenti. Quest'ultima sezione è probabilmente la più controversa e merita un breve approfondimento. Il procedimento di estrazione, classificando il numero dei commenti in pagine, fin da subito è risultato di difficile comprensione perciò è stato contattato direttamente Jacob Jünger il quale ha riferito che Facebook imposta il limite a 500 commenti estraibili per sessione. Se vogliamo estrarre per esempio 2000 commenti dovremo dunque dividerli, ed è esattamente ciò che fa Facepager: divide i commenti da estrarre in blocchi da 100 che chiama "pagine". Quindi per estrarre 2000 commenti dovremo selezionare nel campo "Maximum page" il numero 20 e il software estrarrà ciò che è stato richiesto. In fig. 5 i vari campi.

Figura 5 - Campi di ricerca: esempio di estrazione di 100 post da una pagina Facebook

⁹ Il percorso base solitamente corrisponde alla versione corrente dell'API di Facebook. La guida fornita consiglia agli utenti di controllare sempre l'ultima versione rilasciata. Nel caso in cui l'API venga aggiornata è dunque necessario cambiare il percorso, ad esempio da "graph.facebook.com/v2.0" a "graph.facebook.com/v2.10"

Dopo aver compilato i campi soprastanti è necessario controllare tutte le impostazioni prima di passare alla prima estrazione dei dati. Tra le impostazioni è possibile selezionare un numero massimo di errori consecutivi per il quale non viene invalidata l'estrazione, il numero dei *thread* da analizzare parallelamente (più il numero è alto più velocemente avverrà l'estrazione ma ci saranno anche più errori), il limite di richieste massimo per minuto e il livello del nodo su cui verranno effettuate le operazioni. Inoltre tra le impostazioni è possibile notare il campo “Log all request”, che è consigliabile selezionare per tenere memoria nello “Status Log” di tutte le richieste effettuate a Facepager e controllare gli errori. Una volta impostata correttamente la sezione “Settings” è possibile iniziare a estrarre i dati cliccando su “Fetch Data”. L'operazione durerà qualche secondo e, nella sezione citata in precedenza, sarà possibile visualizzare un report di almeno tre righe sull'analisi appena effettuata. Nella riga iniziale spunterà un messaggio che certifica l'inizio dell'analisi e il numero dei nodi esaminati. Nelle righe centrali (tante quante le pagine di post da esaminare) verrà riportato il messaggio di avvenuta identificazione del profilo da osservare, l'inizio dell'estrazione e l'Access Token utilizzato per accedere. Nell'ultima riga troveremo il numero di nuovi nodi creati seguito da un riepilogo delle risposte che può avere come codice 200 (success) o 400 (error). In Fig. 6 un esempio dell'estrazione avvenuta con successo.

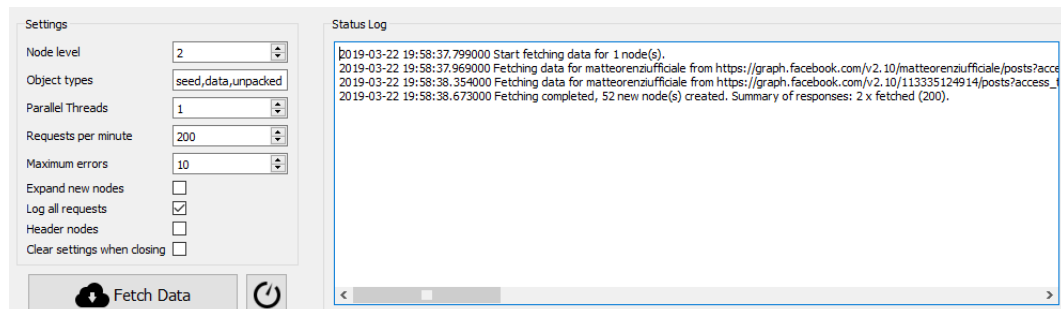


Figura 6 - Un esempio di estrazione avvenuta con successo nel quale sono stati estratti 52 post

Una volta terminata l'estrazione sarà possibile notare che nell'interfaccia principale della pagina, sotto al seme iniziale (“matteoreenziufficiale” di Fig. 4), compariranno tanti nodo-post quanti indicati nello Status Log. Ognuno di questi nodo-post ha: un ID, un tipo (che per questi nodi sarà “data” e non “seed”), un “Query Status” (che segnala la riuscita o meno dell'operazione), un “Query Time” (l'orario in cui è stata effettuata l'operazione), un “Query Type” che nel mio caso sarà

“Facebook:<page>/post” e il campo “message” all’interno del quale è presente il testo del post.

Per estrarre i commenti il procedimento è analogo ma anziché che selezionare il nodo principale sarà necessario selezionare i nodo-post, figli di primo grado del nodo-pagina. Una volta effettuata la selezione dei nodi sui quali si intende operare, non resta che modificare i campi di ricerca e i parametri. In “Resource” sceglieremo l’opzione <post>/comments mentre in “Parameters” <post>. Nelle impostazioni sarà necessario modificare il livello del nodo da “1” a “2”. Completate tutte queste operazioni sarà possibile lanciare il comando per l’estrazione dei dati. Una volta riscontrato dallo Status Log che l’operazione è terminata, sarà necessario verificare la riuscita. Come si può notare dalla figura 7, sono stati estratti i commenti sotto al post con ID 113335124914_10156591105899915 pubblicato dalla pagina “matteorenziufficiale”. È opportuno effettuare una controprova per dimostrare la veridicità e l’affidabilità dell’estrazione. Come dimostrano la fig. 8 e la fig. 9, Facepiger ha estratto perfettamente i commenti.

Object ID	Obj	Query Status	Query Time	Query Type	message
matteorenziufficiale	seed				
> 113335124914_10156591266159915	data	fetchd (200)	2019-03-22 ...	Facebook:<pag...	Noi non siamo come loro. Noi non mostr...
113335124914_10156591105899915	data	fetchd (200)	2019-03-22 ...	Facebook:<pag...	Ecco la mia intervista a Bruno Ventavoli p...
10156591105899915_10156591116179915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Sei e rimani L unico Grande Leader, Aspet...
10156591105899915_10156591130129915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	che cosa devo commentare ? sei un gra...
10156591105899915_10156591120499915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Abbiamo in comune lo stesso modo di le...
10156591105899915_10156591145094915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Siamo arrivati a un bivio, aspetto con fidu...
10156591105899915_10156591177069915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	MAT TEO RENZI un uomo pieno di cultur...
10156591105899915_10156591110009915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	"Un libro è un giardino che puoi custodir...
10156591105899915_10156591114604915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Scrivete fin da bambino e lo si capisce!!! L...
10156591105899915_10156591139644915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Anche a me piace molto leggere, ora per...
10156591105899915_10156591111999915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Senatore, sono orgoglioso di averlo come...
10156591105899915_10156591169169915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Wow, mi è piaciuta tanto questa intervist...
10156591105899915_10156591148304915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Tu a Di Maio, come persona, non sei deg...
10156591105899915_10156591140544915	data	fetchd (200)	2019-03-22 ...	Facebook:<post...	Mi è molto piaciuta la citazione dell'inter...

Figura 7 – Esempio di estrazione avvenuto con successo: nella casella “message” il testo dei commenti



Figura 8 - Il post su cui nell’esempio è stata effettuata l’estrazione

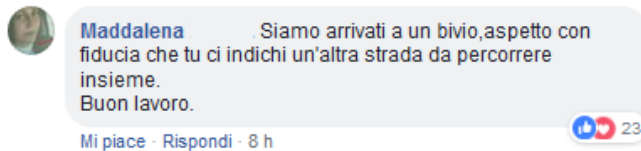


Figura 9 - Il commento evidenziato in fig.7

Una volta verificati i dati non resta che esportarli da Facepager. Prima di procedere è necessario selezionare le colonne funzionali alla ricerca nella sezione “Custom Table Columns”. Per l’analisi dei commenti saranno utilizzati i dati nelle colonne “id” e “message”. Verranno indicati (in due righe diverse) questi due campi. Conclusa la digitazione basta cliccare “Apply Columns” per far sì che Facepager selezioni soltanto i dati richiesti. Successivamente, una volta selezionati con il mouse i dati che si intendono esportare, è possibile completare l’operazione con “Export Data”. I file ottenuti dalle query effettuate dall’applicazione saranno in formato .csv.

3.3 Scelta del campione

L’analisi riguarderà i commenti lasciati dagli utenti sotto ai post di natura politica. Sarà necessario per questo motivo individuare gli spazi politici più frequentati dagli italiani su Facebook. È verosimile pensare che gli italiani parlino di politica sotto ai post di politici italiani, giornalisti e pagine di informazione politica. Così in primo luogo sono stati analizzate le 10 migliori pagine per ognuna di queste tre categorie in termini di “Mi Piace” totali. La ricerca è stata effettuata con l’ausilio del sito www.baroncelli.eu che monitora, in termini di “Mi Piace” le pagine Facebook di politici, giornalisti e di informazione. La rilevazione è stata effettuata tra il 13 e il 15 marzo 2019.

3.3.1 Prima rilevazione: 13 -15 marzo 2019

Per quanto riguarda l’analisi delle pagine dei leader politici, è stato interessante vedere come tra le prime 10 più seguite su Facebook ci siano soltanto quelle di tre esponenti dell’attuale governo (Matteo Salvini, 1^a con 3,5 milioni di “Mi Piace”, Luigi Di Maio, 2^a con 2,1 milioni di “Mi Piace” e Giuseppe Conte, 10^a con 858 mila “Mi Piace”). Ancora più singolare è vedere quella del Presidente del Consiglio

dei Ministri, Giuseppe Conte, dietro non solo a quella di Giorgia Meloni (8^a con 1 milione di “Mi Piace”), leader di Fratelli D'Italia (FDI), partito che alle elezioni politiche del 2018 ha ricevuto il 4,3% dei consensi, ma anche alle spalle di quella del Sindaco di Roma Virginia Raggi (9^a con 912 mila “Mi Piace”). È inoltre doveroso sottolineare come delle 10 pagine prese in esame ben 4 siano di esponenti del Movimento 5 Stelle. È anche interessante notare come gli esponenti più in vista dei partiti “storici” italiani (Matteo Renzi e Silvio Berlusconi¹⁰) non siano tra i 5 politici più seguiti su Facebook.

Successivamente, per le 5 più seguite nelle rispettive categorie, è stata effettuata un’analisi più dettagliata. Una volta individuati gli ultimi 10 post pubblicati da ogni pagina a partire dalle ore 23 del giorno 7 marzo 2019, è stata calcolata la media di reazioni e commenti per post. La scelta del giorno da cui iniziare la rilevazione non è stata casuale. I post presi in esame risalgono infatti a una settimana prima rispetto all’inizio dell’analisi. Questo per evitare che subiscano grandi variazioni in termini di reazioni e commenti in futuro. Anche il giorno della settimana (giovedì) non è casuale. Nei weekend si registra infatti un minor numero di utenti connessi al social network.

Per la categoria dei politici sono state effettuate analisi sulle pagine di “Matteo Salvini” (3,5 mln di “Mi Piace”), “Luigi Di Maio” (2,1 mln di “Mi Piace”), “Beppe Grillo” (2 mln di “Mi Piace”), “Vittorio Sgarbi” (1,7 mln di “Mi Piace”) e “Alessandro Di Battista” (1,5 mln di “Mi Piace”). È necessario fare una precisazione sulla classificazione. Vittorio Sgarbi viene annoverato tra i politici, nonostante la fama nazionale di critico d’arte, in quanto eletto alla Camera Dei Deputati tra le fila del partito “Forza Italia”. In seguito è passato al “Gruppo Misto” e tuttora svolge attività parlamentare.

Per la categoria dei giornalisti sono state effettuate analisi sulle pagine di “Roberto Saviano” (2,5 mln di “Mi Piace”), “Marco Travaglio” (1,7 mln di “Mi Piace”),

¹⁰ La pagina di Matteo Renzi, senatore del Partito Democratico (PD) vanta 1,1 milioni di “Mi Piace”. Quella di Silvio Berlusconi, leader di Forza Italia (FI) conta invece 1 milione di “Mi Piace”.

“Enrico Mentana” (1,1 mln di “Mi Piace”), “Massimo Gramellini” (471 mila “Mi Piace”) e “Andrea Scanzi” (450 mila “Mi Piace”). Per la rilevazione abbiamo preso in considerazione esclusivamente le pagine di giornalisti che hanno pubblicato almeno un aggiornamento di stato di carattere politico negli ultimi 10 post (sempre partendo dalle 23.00 del 7 marzo 2019).

Per la categoria delle pagine di informazione sono state effettuate analisi sulle pagine di “Fanpage” (8 mln di “Mi Piace”), “La Repubblica” (3,7 mln di “Mi Piace”), “The Social Post” (3,6 mln di “Mi Piace”), “Today.it” (3 mln di “Mi Piace”) e “Corriere della Sera” (2,5 mln di “Mi Piace”). Per la rilevazione abbiamo preso in considerazione esclusivamente le pagine di informazione che hanno pubblicato almeno una notizia di carattere politico negli ultimi 10 post (sempre partendo dalle 23.00 del 7 marzo 2019).

Una volta conclusa l’analisi, esaminando il numero medio di reazioni e commenti per post, è possibile affermare che gli italiani preferiscono parlare di politica sulle pagine Facebook dei leader politici piuttosto che su quelle di giornalisti o di informazione. Ci sono però alcune eccezioni importanti. Beppe Grillo per esempio vanta 2 milioni di “Mi Piace” alla sua pagina Facebook ma in realtà il numero medio di reazioni (1419,7) e commenti (84,3) ai suoi post è addirittura più basso di quello di giornalisti politici come Enrico Mentana (R: 3380.4, C: 753.6) e Andrea Scanzi (R: 2815.9, C: 328.5) che vantano rispettivamente 1,1 milione e 450 mila “Like”. Lo stesso vale per la pagina di Vittorio Sgarbi dove a fronte di un numero elevatissimo di “Mi Piace” alla pagina (1,7 mln), il livello di interazioni complessive è inferiore a quello di tutti i giornalisti politici esaminati.

Non esiste dunque una chiara correlazione tra il numero complessivo di “Mi Piace” alla pagina e tasso di *engagement*¹¹ ai post. Infatti se per i casi di “Matteo Salvini” e “Roberto Saviano” esiste un rapporto (rispetto alle altre pagine del loro campo), per “Marco Travaglio” e “Fanpage.it” non possiamo dire lo stesso. La pagina di

¹¹ Per Engagement Rate si intende il tasso di interazione generato da un contenuto, una pagina, un’attività che può dar luogo ad interazioni (Cosenza, 2018)

Enrico Mentana, che vanta “solo” 1,1 milioni di “Mi Piace”, ben 600.000 in meno di quella di Travaglio, ottiene risultati nettamente migliori (tra 3 e 7 volte in più). Ricapitolando, l'analisi effettuata dimostra come gli italiani preferiscano parlare di politica sulle pagine dei leader politici, motivo per cui avremo necessariamente bisogno di più dati a riguardo visto che quelli relativi a Beppe Grillo e Vittorio Sgarbi non sono soddisfacenti. Per effettuare un'analisi sincrona dovremo inoltre mettere da parte anche i dati raccolti sui post della pagina di Alessandro Di Battista considerato che sono precedenti di oltre due settimane rispetto a quelle dei suoi colleghi. Accantonati i dati raccolti su queste tre pagine, sono state esaminate quelle dei loro diretti inseguitori nella classifica iniziale basata sul numero di “Mi Piace” totali alla pagina, vale a dire “Matteo Renzi” (1,1 milioni di “Mi Piace”), Silvio Berlusconi (“1 milione di “Mi Piace”) e “Giorgia Meloni” (1 milione di “Mi Piace”). Una volta analizzate le tre pagine secondo lo stesso metro utilizzato per quelle dei loro colleghi emerge un quadro che conferma quanto affermato in precedenza: non vi è alcuna correlazione tra il numero “Mi Piace” totali alla pagina e tasso di engagement. Infatti se analizziamo i numeri, è possibile notare come quella di Silvio Berlusconi ottenga numeri decisamente inferiori ai suoi equivalenti. Un altro dato interessante è costituito dal rapporto tra reazioni e commenti. Come è possibile vedere in Tab.3, il dato non varia di molto da pagina a pagina ma si riscontra una maggiore propensione dell’utente a lasciare commenti sotto a pagine di informazione (anche se ciò non implica che vi siano sotto i post di quella pagina un maggior numero di commenti). Ecco i risultati complessivi dell’analisi dei 180 post:¹²

Nome pagina	Like	MRP	MCP	R/C	NM
Matteo Salvini	3,5 mln	27446,1	5165,0	5,31	15
Luigi Di Maio	2,1 mln	12359,4	2034,0	6,07	4
Beppe Grillo	2,0 mln	1419,7	84,3	16,84	9

¹² **LEGENDA:** **Like** = Numero di “Mi Piace” totali alla pagina, **MCP** = Media di commenti per post, **MRP** = Media di reazioni per post, **R/C** = Rapporto tra media di reazioni e media di commenti (ogni quante reazioni c’è un commento), **N** = numero medio di post pubblicati al giorno

Vittorio Sgarbi	1,7 mln	852,5	87,0	9,79	2
Alessandro Di Battista	1,5 mln	9967,1	2299,1	4,33	1
Matteo Renzi	1,1 mln	5443,6	820,7	6,63	1
Silvio Berlusconi	1,0 mln	2101,2	423,7	4,95	3
Giorgia Meloni	1,0 mln	8131,9	805,2	10,09	6
Roberto Saviano	2,5 mln	5815,8	647,3	8,98	1
Marco Travaglio	1,7 mln	1073,9	175,0	6,13	7
Enrico Mentana	1,1 mln	3380,4	753,6	4,48	6
Massimo Gramellini	471mila	1379	79,1	17,43	1
Andrea Scanzi	450mila	2815,9	328,5	8,58	3
Fanpage.it	8,0 mln	453,2	169,7	2,67	5
La Repubblica	3,7 mln	1500,5	290,8	5,15	12
The Social Post	3,6 mln	343,9	531,3	0,65	4
Today.it	3,0 mln	438,3	347,8	1,26	9
Corriere della Sera	2,5 mln	671,9	430,1	1,56	5

Tabella 3 – Le migliori pagine in termini di reazioni e commenti su Facebook

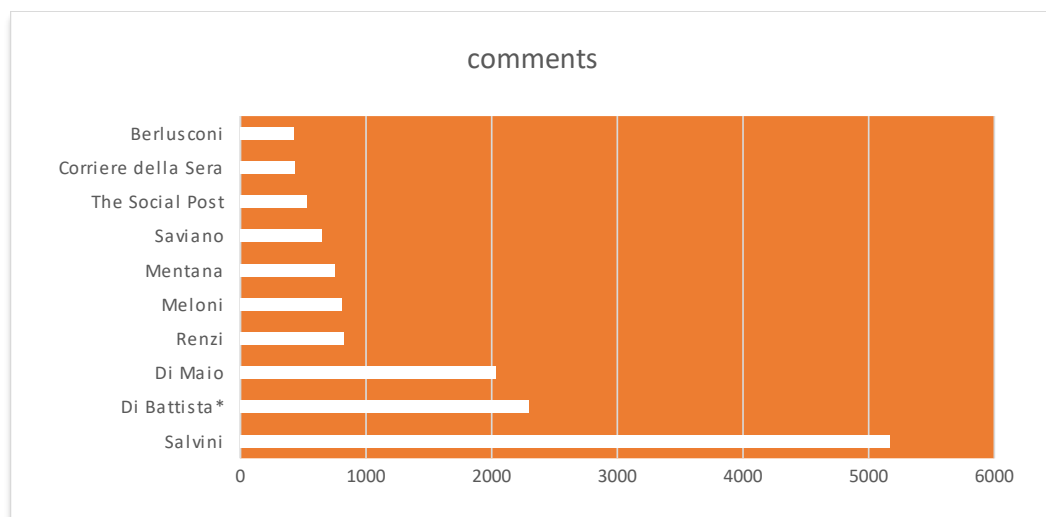


Figura 10 - Sotto ai post di quale pagina gli italiani si confrontano? (* La rilevazione su Alessandro Di Battista è asincrona. La nostra rilevazione riguarda i post precedenti alle 23.00 del 7 marzo 2019. L'ultimo post di Alessandro Di Battista risale invece al 13 febbraio 2019)

Per individuare gli spazi su Facebook dove gli italiani parlano di politica sarà necessario analizzare questi dati e prestare maggiore attenzione al campo della media dei commenti per post (MCP). Come è possibile vedere dalla Fig. 10, Matteo Salvini è la pagina che su Facebook riscuote maggior successo in termini di commenti seguito da Luigi Di Maio, Matteo Renzi, Giorgia Meloni ed Enrico Mentana.

Prima di iniziare l'estrazione dei commenti da queste pagine sarà opportuno anche tenere in considerazione il numero dei post che vengono pubblicati in una giornata da una singola pagina. È vero che l'elevato numero di post pubblicati può influire negativamente sui risultati che l'aggiornamento postato raggiunge? No. Questo almeno per quanto riguarda l'algoritmo di Facebook. Da diversi anni infatti il social network ha infatti introdotto lo *Story Bumping* (Vahl, 2013) che ha affiancato il più conosciuto *Time Decay*. Ci si basa essenzialmente non più solo sulla "freschezza" dei contenuti, ma sulla "pesantezza" in termini di interazioni. L'interazione più pesante e che permette al post di tornare ad essere virale anche a distanza di tempo è il click sul post (ogni volta che un post viene visualizzato in una tab tramite il click su di esso), seguito nell'ordine da: share & comment (condividi e commenta), instant share (condividi ora), commenti, reazioni ("Love", "Haha", "Sad", "Wow", "Angry"), like ("Mi Piace") e visualizzazioni. A queste interazioni se ne aggiunge un'altra: il tempo di permanenza davanti a un post senza compiere azioni. Grazie a questo nuovo algoritmo che non valuta più solo quanto una notizia sia recente ma anche quanto sia importante, verrebbe da pensare che non oggi non sia più così fondamentale pubblicare costantemente. Analizzando il numero dei commenti lasciati dagli utenti nelle pagine precedentemente esaminate, nell'arco di una settimana, sarà possibile verificare se esista un nesso tra l'algoritmo e la realtà. Considerati gli scarsi risultati che ottengono le pagine di informazione in termini di reazioni e soprattutto commenti, questa seconda analisi non le riguarda. I post sono stati esaminati il 30 marzo 2019 e risalgono alla settimana compresa tra le 23.00 del 7 marzo e le 23.00 del 28 febbraio.

3.3.2 Seconda rilevazione: 30 marzo

Dall'analisi della quantità dei commenti totali che riceve una determinata pagina in una settimana, emerge che "Matteo Salvini" è la pagina che raccoglie il maggior numero di opinioni (160mila commenti) sotto i suoi post da parte degli utenti Facebook. Al secondo posto, a sorpresa per la verità, troviamo la pagina "Enrico Mentana" con 74mila risposte dagli utenti. Quest'ultima è l'unica della categoria "giornalisti" tra le prime 5. Dopo quella del direttore del Tg La7 troviamo la pagina di Giorgia Meloni con 53mila commenti ricevuti in una settimana. A seguire troviamo quelle di Luigi Di Maio e Matteo Renzi, rispettivamente con 43mila e 25mila commenti in una settimana.

Adesso è possibile valutare anche se esista un nesso tra il numero dei post pubblicati e il numero dei commenti ricevuti. È molto difficile rispondere a questa domanda ma dall'analisi il quadro che emerge è abbastanza chiaro. Prima di confrontare i dati è necessaria una precisazione. Durante la rilevazione sono sempre stati fatti controlli incrociati con lo studio effettuato tra il 13 e il 15 marzo. Nella Tab. 3 è possibile vedere che "Matteo Salvini" pubblica mediamente circa 15 post al giorno. Nel nostro studio attuale (30 marzo) invece risulta che la pagina del Vicepremier abbia pubblicato soltanto 32 post nella settimana esaminata, ovvero soltanto 4-5 post al giorno. Quest'ultimo è un dato che senza dubbio va in aperto conflitto con il dato di Tab. 3 che invece faceva pensare a quantità di post (e di commenti) decisamente superiore (almeno 90-100 post a settimana). Per capire il perché di questa discordanza numerica è stato fondamentale riprendere in mano i dati sui post pubblicati da "Matteo Salvini" analizzati a metà marzo. Da un rapido controllo risulta che alcuni post pubblicati dal Ministro dell'Interno, sono stati nascosti o cancellati e per questo motivo non è possibile trovarli nella home della pagina. In particolare, dei 10 post analizzati in Tab.3, ben 4 sono stati nascosti o cancellati. Dovrebbe essere adesso più chiaro il motivo della discordanza numerica. Questa cancellazione comporta soprattutto la perdita di una grande quantità di informazione. Nei casi di "Giorgia Meloni" e "Enrico Mentana" possiamo affermare che esiste un nesso tra il numero dei post pubblicati e il numero dei commenti ricevuti. Infatti la pagina della politica romana pubblica tra i 7 e gli 8

post al giorno ottenendo ogni volta buoni risultati in termini di commenti (per quanto inferiori rispetto ad altre pagine). Per la leader di Fratelli d'Italia è decisiva la quantità di contenuti pubblicati. La pagina "Enrico Mentana" invece pubblica leggermente meno (6 post al giorno) ma ha anch'essa straordinari risultati probabilmente dovuti anche alla grande frequenza con cui il giornalista risponde ai commenti dei fan scherzando con loro o offrendo spunti di riflessione. Diametralmente opposta è la comunicazione che propongono le pagine di Luigi Di Maio e Matteo Renzi. Il Vicepremier e l'ex Presidente del Consiglio pubblicano pochissimo ma grazie all'algoritmo di Facebook "Story Bumping" e alla loro popolarità riescono a ottenere discreti risultati in termini di commenti. Ogni volta che una di queste due pagine scrive un post riceve molte interazioni che le permettono di apparire in alto nella home degli utenti anche a distanza di tempo. Lo Story Bumping dunque funziona, anche se la strategia introdotta dalle pagine di "Matteo Salvini" e "Giorgia Meloni" è decisamente più efficace. Pubblicare un alto numero di contenuti originali ha un effetto concreto sul numero dei commenti che una pagina riceve. La Tab.4 riassume quanto detto.

Nome Pagina	NPS	TCS 28/02-7/03	MCP
Matteo Salvini	32	160464	5014,5
Enrico Mentana	44	74638	1696,3
Giorgia Meloni	53	53115	1002,1
Luigi Di Maio	19	43120	2269,4
Matteo Renzi	17	25661	1509,4
Beppe Grillo	62	9457	152,5
Marco Travaglio	56	9434	168,4
Silvio Berlusconi	10	8293	829,3
Andrea Scanzi	17	6648	391,0

Roberto Saviano	6	6235	1039,1
Vittorio Sgarbi	15	2199	146,6
Massimo Gramellini	7	947	135,2

Tabella 4 – Commenti totali ai post pubblicati dalle pagine nella settimana compresa tra le 23.00 del 28/02/19 e le 23.00 del 7/03/19. **LEGENDA** → **NPS** = Numero dei Post pubblicati nella Settimana; **TCS**: Totale dei Commenti sui post pubblicati dalla pagina nella Settimana; **MCP** = Media dei commenti per post

Durante lo studio è stato interessante notare come a distanza di oltre 10 giorni dalla pubblicazione un post continui ad incrementare le interazioni su di esso. Questo è lo Story Bumping che permette anche a chi non pubblica molti contenuti di apparire anche a distanza di giorni. Emblematico è il caso delle pagine di Silvio Berlusconi, Roberto Saviano e Massimo Gramellini. Tutti e tre hanno scritto 10 o meno post nella settimana presa in considerazione (28 febbraio - 7 marzo). Quei post sono esattamente gli stessi che erano stati analizzati il 15 marzo. È stato così semplice verificare a distanza di tempo l'andamento di quei contenuti. Il caso di "Roberto Saviano" è molto interessante. Ha pubblicato 6 post nella settimana che abbiamo esaminato. L'incremento di commenti è stato vertiginoso. Da una media di 452 commenti per post (15 marzo) è passato a una media di 1039 commenti (30 marzo). La variazione è stata del + 129%. Non è stato possibile misurare la variazione del numero di commenti su tutte le altre pagine analizzate poiché la rilevazione del 15 marzo fa riferimento a soli 10 post precedenti alle 23.00 del 7 marzo. Molte di queste pagine, come possiamo vedere da Tab. 4, pubblicano oltre 30 post a settimana quindi i 10 post che abbiamo analizzato a metà marzo costituiscono un dato interessante e rappresentativo, ma incompleto. Invece il confronto sui post di Saviano, Gramellini e Berlusconi può essere considerato "puro" e mediamente il numero dei commenti ha registrato un incremento dell'86%. Emerge dalla tabella che gli algoritmi tendono a "spingere" i post più pesanti (in termini di interazioni) e funzionano indiscriminatamente non solo sulle pagine che pubblicano molto poco ma anche su quelle che pubblicano molti post al giorno. Come già detto, il numero dei commenti sulle pagine Facebook aumenta col passare del tempo. Sarà quindi fondamentale prendere in considerazione per l'estrazione dei commenti soltanto le

pagine che oggi, 1° aprile 2019, si trovano ai primi posti per numero di commenti: Matteo Salvini, Enrico Mentana, Giorgia Meloni, Luigi Di Maio e Matteo Renzi.

4. Analisi

Dopo aver effettuato l'estrazione dei dati con Facepager è stato possibile analizzarli con strumenti linguistico-computazionale. Prima di continuare è utile segnalare che su un campione di 207.699 commenti estratti dai post dei 5 leader politici presi in esame, solo 56.804 contenevano il pronome/congiunzione “che”. Per effettuare l'analisi sul “che polivalente” sarà necessario dunque prendere in considerazione soltanto quelli che presentano il suddetto *token*. Il corpus a nostra disposizione è formato da 23.823 commenti provenienti dai post della pagina “Matteo Salvini”, 14.713 commenti provenienti dalla pagina “Enrico Mentana”, 8.622 commenti provenienti dalla pagina “Luigi Di Maio”, 7.736 commenti provenienti dalla pagina “Giorgia Meloni” e 2.102 provenienti dalla pagina “Matteo Renzi”.

4.1 Il “che” polivalente

Prima di addentrarci nell'analisi sarà necessario individuare tutte le specificità del “che polivalente”. Mentre in italiano standard la congiunzione “che” è spesso utilizzata per introdurre alcune frasi subordinate come le oggettive, le soggettive e le dichiarative, nel parlato e nel colloquiale assistiamo alla tendenza a estendere l'uso del “che” anche per introdurre subordinate che nell'italiano standard presenterebbero congiunzioni subordinanti semanticamente più precise (Fiorentino, 2010). Questo uso del “che” in parte si sovrappone a quello come pronome relativo. Questa tendenza è considerata da Berruto uno dei tratti di ristrutturazione e ristandardizzazione della lingua contemporanea (1998). Si utilizza il “che polivalente” per introdurre frasi di significato esplicativo-consecutivo, causali, consecutivo-presentative, relative-temporali, finali, pseudorelative e frasi in cui il “che” ha valore enfaticamente esclamativo.

Questo uso del “che” ha in realtà radici ben profonde nella lingua italiana. Basti pensare al celebre passo della Divina Commedia: “*Nel mezzo del cammin di nostra*

vita // mi ritrovai per una selva oscura // ché la dritta via era smarrita” (Dante, Inf. I, 1-3) nel quale il “che” è trattato come *aferesi*¹³ di perché. Già nel XVI secolo Pietro Bembo si interrogava riguardo alla possibilità di accettare il che polivalente anche nell’italiano standard. Nelle sue “Prose della volgar lingua” (1525, III, LXIV), analizzando gli scritti di Boccaccio e Petrarca, condannava apertamente quest’uso della congiunzione a cui preferiva alternative più ricche di informazione sintattica come pronomi relativi e congiunzioni subordinanti appropriate. Come sostiene Fiorentino (2010), l’intervento di Bembo segna un punto di svolta nella lingua italiana: la sua presa di posizione infatti sancisce l’abbandono di quest’uso del “che” dalla norma dell’italiano. Tuttavia le attestazioni del “che polivalente” persistono nella storia. In epoche più recenti è stato utilizzato infatti anche da Italo Calvino, Cesare Pavese e molti altri per marcare una scelta stilistica di tipo popolare.

Oggi il “che polivalente” appare non solo in varietà diastraticamente basse ma anche in varietà di parlato o scritto poco controllate (come ad esempio Facebook). In molti casi l’uso del “che polivalente” può essere considerato accettabile (specie se viene trattato come una congiunzione relativa-temporale, causale, finale o consecutiva) nell’italiano medio colloquiale. In altri casi l’uso del “che polivalente” costituisce un vero e proprio errore sintattico e caratterizza l’enunciato come popolare. In registri bassi è possibile trovare inoltre anche il “che” usato per rafforzare un’altra congiunzione. Inoltre in alcuni casi è difficile stabilire il valore sintattico del “che polivalente” sia in relazione alla semantica propria, ma anche rispetto alla possibilità che si tratti di coordinazione e non subordinazione (Fiorentino, 2007).

4.1 Strumenti per l’analisi

Una volta individuato l’ambito è possibile effettuare la vera e propria analisi linguistica dei commenti. Per fare ciò ci serviremo di UDPipe, una pipeline realizzata da Milan Straka, docente della Univerzita Karlova di Praga, per

¹³ Fenomeno che comporta la caduta di uno o più foni iniziali di una parola

tokenizzare, etichettare, lemmatizzare e analizzare dipendenze dei file CoNLL-U. UDPipe nasce dal progetto internazionale Universal Dependencies.

4.2.1 Universal Dependencies

Il progetto Universal Dependencies (UD)¹⁴ nasce ufficialmente nel 2014 ed è il risultato di varie iniziative tese a combinare le Stanford Dependencies e i tag di Google in uno schema di annotazione universale. Le Stanford Dependencies sono state sviluppate nel 2005 come riconoscitori di sistemi testuali e successivamente sono diventate uno standard nell'analisi delle dipendenze dell'inglese. Il tagset di Google nasce nel 2007 per il POS tagging non supervisionato e ben presto diventa uno standard in questo campo. Ci sono stati nel corso degli anni vari tentativi di unificare i due strumenti. Il primo di questi è stato il progetto Universal Dependency Treebank (UDT) che rilasciò *treebanks* per 6 lingue nel 2013 e per 11 lingue nel 2014. Poco dopo, sempre nel 2014, la seconda versione di HamleDT ha fornito l'annotazione Stanford/Google per 30 lingue. Successivamente sono state sviluppate le Universal Stanford Dependencies (USD). La nuova Universal Dependencies racchiude tutte queste iniziative basandosi sulle Universal Stanford Dependencies, una versione estesa del tagset universale Google e una versione riveduta del formato CoNLL-X (che si chiama CoNLL-U). A 5 anni dalla sua nascita ufficiale il progetto sta proseguendo con lo scopo di “fornire un inventario universale di categorie e linee guida per facilitare l'annotazione coerente di costruzioni simili in tutte le lingue, consentendo estensioni specifiche della lingua quando necessario.”¹⁵ In altre parole è finalizzato alla definizione di uno schema di annotazione sintattica a dipendenze applicabile a tutte le lingue. Per fare ciò, Universal Dependencies ha una open community con oltre 200 collaboratori che producono più di 100 treebanks in oltre 70 lingue. L'obiettivo è quello di sostituire tutti gli schemi finora esistenti e, per farlo, si propone di fornire un tagset che permetta di enfatizzare le similarità fra le lingue senza però appiattirne le differenze

¹⁴ <https://universaldependencies.org/>

¹⁵ <https://universaldependencies.org/introduction.html>

quando necessario (Nivre, 2016). Il progetto è coordinato da Joakim Nivre. I rilasci (compresa la validazione e la documentazione) sono gestiti da Filip Ginter, Sampo Pyysalo e Dan Zeman.

Nell'annotazione UD le parole sono l'unità minima dell'analisi grammaticale e le caratteristiche morfologiche sono codificate come proprietà delle parole. UD mette a disposizione risorse linguistiche costruite attraverso tokenizzazione, annotazione morfologica e annotazione sintattica. Esistono tre diversi livelli di rappresentazione di una parola: lemma, tag PoS e Features. Il lemma riguarda il contenuto semantico e il tag Part of Speech (tag della parte del discorso) indica la categoria lessicale a cui appartiene una parola e si esprime con una sigla relativa categoria sintattica. Le features riguardano le proprietà di una determinata parola (categoria grammaticale e proprietà morfosintattiche). Di seguito in Tab.5 e Tab.6 i tag assegnabili da UD.

Open class words	Closed class words	Other
ADJ	ADP	PUNCT
ADV	AUX	SYM
INTJ	CCONJ	X
NOUN	DET	
PROPN	NUM	
VERB	PART	
	PRON	
	SCONJ	

Tabella 5 - Part of Speech Tag

Lexical features	Inflectional features	
	Nominal	Verbal
PronType	Gender	VerbForm
NumType	Animacy	Mood
Poss	NounClass	Tense
Reflex	Number	Aspect
Foreign	Case	Voice
Abbr	Definite	Evident
	Degree	Polarity
		Person
		Polite
		Clusivity

Tabella 6 - Features Tag

4.2.3 UDPipe

UDPipe è un'applicazione gratuita distribuita con Mozilla Public License 2.0 e i modelli linguistici sono gratuiti per uso non commerciale e distribuiti sotto licenza CC BY-NC-SA¹⁶. Per questa analisi sarà utilizzato il servizio web ma è disponibile un software per Linux / Windows / OS X, come libreria per C ++, Python, Perl, Java, C#. La caratteristica che rende UDPipe un'applicazione molto utile, è quella di poter svolgere le proprie funzioni indipendentemente dalla lingua utilizzata e di poter essere addestrato con dati annotati nel formato CoNLL-U.¹⁷ Il servizio web UDPipe presenta un blocco all'interno del quale è possibile scegliere i modelli e le azioni che deve svolgere. Nel primo campo è possibile scegliere la versione con cui effettuare l'analisi. Per questa analisi sceglieremo la versione 2.3, l'ultima rilasciata. Inoltre, sempre nel primo campo c'è la possibilità di scegliere tra modelli di oltre 50 lingue. Per l'italiano sono presenti ISTD, ParTUT e PoSTWITA. Per l'analisi verrà utilizzato "italian-isdt-ud". Più tardi analizzeremo con più precisione il modello ISTD. Nella Fig. 11 è mostrata l'interfaccia del servizio web di UDPipe.

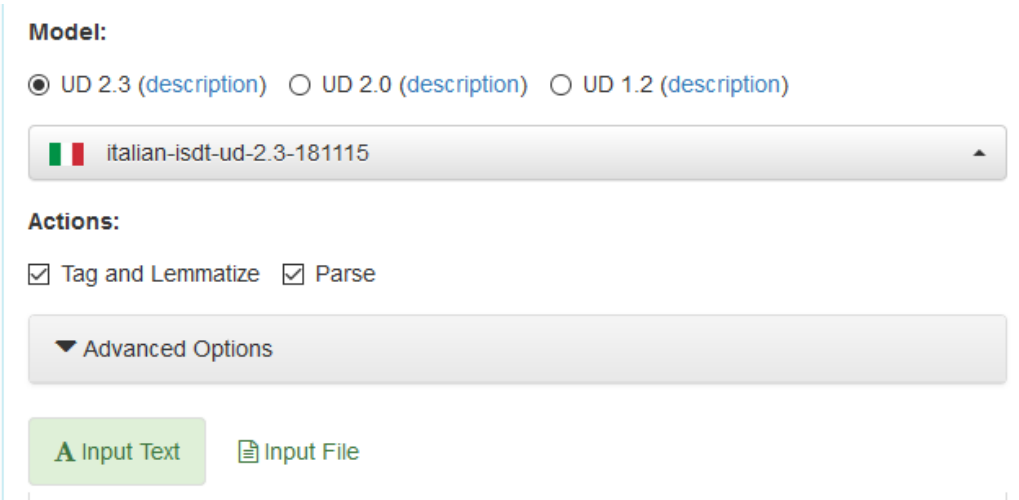


Figura 11 – L'interfaccia del servizio web UDPipe: è stato selezionato l'Italian Stanford Dependency Treebank

¹⁶ La licenza permette di distribuire, modificare, creare opere derivate dall'originale, ma non a scopi commerciali, a condizione che venga: riconosciuta una menzione di paternità adeguata, fornito un link alla licenza e indicato se sono state effettuate delle modifiche; e che alla nuova opera venga attribuita la stessa licenza dell'originale (quindi a ogni opera derivata non verrà consentito l'uso commerciale)

¹⁷ <http://lindat.mff.cuni.cz/services/udpipe/>

Nel secondo campo è possibile scegliere il tipo di azioni da compiere. Le opzioni sono 2 ed è possibile taggare e lemmatizzare e/o parsare il testo. Inoltre tra le opzioni avanzate c'è la possibilità di tokenizzare interamente il testo, scegliere il formato CoNLL-U¹⁸ e scegliere se visualizzare l'analisi orizzontalmente o verticalmente. Se viene spuntata la casella relativa alla tokenizzazione del testo si può addirittura optare per normalizzare gli spazi, contare il numero di caratteri/spazi totali e indicare se il testo è già tokenizzato manualmente così da ottimizzare il lavoro. Per ogni dubbio il servizio web fornisce un manuale semplice ed intuitivo. Nel riquadro sottostante è possibile inserire righe di testo oppure caricare direttamente un file sulla piattaforma.

ISDT

ISDT è l'acronimo di Italian Stanford Dependency Treebank. Lo schema di annotazione di Stanford è stato adattato alla specificità della lingua italiana. È stato ottenuto attraverso un processo di conversione semiautomatica a partire da MIDT (Merged Italian Dependency Treebank), primo progetto finalizzato a fondere due treebanks italiani esistenti (TUT e ISST-TANL). La composizione del corpus di ISDT è la stessa del MIDT, per un totale di circa 200.500 token. Gli antenati di MIDT sono, come già detto, TUT (Turin University Treebank) e ISST-TANL (Italian Syntactic Semantic Treebank) (Bosco, 2013). TUT, sviluppato da un'equipe di ricerca sul NLP¹⁹ dell'Università di Torino, rappresenta cinque diversi generi di testo (giornali, codice civile italiano, JRC-Acquis Corpus5, Wikipedia e la Costituzione Italiana) per un totale di circa 3500 frasi. ISST-TANL è invece stato sviluppato dall'Istituto di Linguistica Computazionale (ILC-CNR) in collaborazione con l'Università di Pisa e raccoglie articoli di giornali e periodici, selezionati per coprire una grande varietà di argomenti (politica, economia, cultura, scienza, salute, sport, tempo libero, ecc.) per oltre 4000 frasi.

¹⁸ Si tratta di una versione riveduta del formato CoNLL-X. Le annotazioni sono codificate in file di testo (UTF-8). Le linee annotate iniziano con il #, le frasi sono delimitate da righe vuote.

¹⁹ Natural Language Processing

La versione di ISDT 2.0 è stata rilasciata in occasione del dependency parsing shared task di Evalita²⁰ del 2014. Il formato di MIDT (e di conseguenza anche di ISDT) è CoNLL, uno standard de facto per quanto riguarda il NLP. Come si legge sul sito MediaLab, nel formato ISDT i dati rispettano alcune regole: 1) i file contengono una o più frasi separate da una riga vuota; 2) una frase è composta da uno o più token, ognuno dei quali inizia su una nuova riga; 3) un token è costituito da dieci campi (ID, FORM, LEMMA, CPOSTAG, POSTAG, FEATS, HEAD, DEPREL, PHEAD e PDEPREL); 4) I campi sono separati da un carattere di tabulazione 5) tutti i dati contengono questi dieci campi anche se ne vengono visualizzati meno; 6) i file di dati sono codificati UTF-8 (unicode). Ecco un quadro descrittivo dei 10 campi.

Field Number	Field	Description
1	ID	Token counter, starting at 1 for each new sentence.
2	FORM	Word form or punctuation symbol
3	LEMMA	Lemma of word form
4	CPOSTAG	Coarse-grained part-of-speech tag.
5	POSTAG	Fine-grained part-of-speech tag.
6	FEATS	Morpho-syntactic features depend on the POS, as detailed in the linked file.
7	HEAD	Non-projective head of current token, which is either a value of ID or zero ('0')
8	DEPREL	Dependency relation to the non-projective-head, which is 'ROOT' when the value of HEAD is zero.
9	PHEAD	Projective head of current token, which is always an underscore because it is not available for the Italian treebank.
10	PDEPREL	Dependency relation to projective head, which is always an underscore, because it is not available for the Italian treebank.

ISDT raccoglie oggi 200.516 token e grazie all'impegno e allo sforzo profuso dai collaboratori per specializzare la Stanford Dependencies sulle peculiarità della lingua italiana si ritiene che questo treebank migliori significativamente l'usabilità

²⁰ [EVALuation of NLP and Speech Tools for ITALian](#)

della risorsa (Bosco, 2013). Per la nostra analisi utilizzeremo dunque questa risorsa su UDPipe.

ParTUT

ParTUT (Parallel Turin University Treebank) è una raccolta di frasi annotate morfo-sintatticamente parallelamente per l'italiano, il francese e l'inglese. Comprende testi provenienti da fonti diverse e vari generi e domini pubblicati in diversi formati di annotazione. È il risultato di un lavoro più che ventennale iniziato con TUT.²¹ (Bosco, 2000).

PoST**wITA**

POSTWITA (Part of Speech Twitter Italia) è una raccolta di tweet italiani annotati che possono essere utilizzati per l'addestramento di sistemi di *Natural Language Processing* per migliorare le loro performance sulle analisi dei testi provenienti dai social media.²² Con lo sviluppo dei social media (in questo caso Twitter) molti linguisti hanno reputato interessante analizzare i testi che gli italiani producono sul web ogni giorno. Il corpus originale è composto da oltre 6.500 tweet. Come si legge sul sito di Universal Dependencies, “il processo di conversione e annotazione sintattica è stato effettuato attraverso fasi alternate di scripting automatico e revisione manuale”. Dal maggio del 2018, con l'introduzione del nuovo regolamento europeo sul trattamento dei dati personali (GDPR), sono stati sostituiti gli ID originali e i nomi utente. Hanno contribuito alla creazione di questa *treebank* Fabio Tamburini e Oronzo Antonelli dell'Università di Bologna, Alessandro Mazzei e Cristina Bosco dell'Università di Torino e Andrea Bolioli dell'azienda “Celi” e Alberto Lavelli della Fondazione Bruno Kessler di Trento.

Prima di continuare è necessario precisare in che modo sono stati utilizzati i risultati del lavoro svolto. Dopo aver esaminato i singoli “che” polivalenti trovati all'interno

²¹ <http://www.di.unito.it/~tutreeb/treebanks.html>

²² https://universaldependencies.org/treebanks/it_postwita/index.html

del corpus di commenti, ad ognuno di essi è stato assegnato un tag. Successivamente sono state sostituite le etichette assegnate ai “che” polivalenti da UDPipe nel file in formato CoNLL-U con quelle ricavate dalla nostra analisi in modo che questi dati possano essere utilizzati per addestrare il sistema in futuro.

4.3 I “che” polivalenti presenti nel corpus di commenti analizzato

Analizzati “che” contenuti nei commenti con UDPipe è possibile esaminare quali sono stati usati in modo improprio. È necessario specificare che sono stati esaminati 1495 commenti contenenti il “che” bilanciando l’analisi in base al numero di commenti riscontrati sotto i post dei politici. Per avere un campione rappresentativo sono stati quindi esaminati 625 commenti estratti dalla pagina “Matteo Salvini”, 386 estratti dalla pagina “Enrico Mentana”, 226 estratti dalla pagina “Luigi Di Maio”, 203 estratti dalla pagina “Giorgia Meloni” e 55 estratti dalla pagina “Matteo Renzi”. I commenti sono stati scelti casualmente. Nella Tab.7 sono riportati i risultati dell’analisi.

Risultati dell’analisi dei “che polivalenti”

Tipologia	N°	%
<i>Che</i> complessivi	1971	100,00
Valore relativo	18	0,91
Valore causale	58	2,94
Valore consecutivo	5	0,25
Valore relativo- temporale	4	0,20
Valore concessivo	2	0,10
Valore incerto	20	1,01
Uso intensivo	11	0,55
Con clitico di ripresa	6	0,30
<i>Che</i> polivalenti totali	108	5,47

Tabella 7 – Risultati dell’analisi per tipologia

Nell’analisi sono stati esaminati 1971 “che” provenienti da commenti su Facebook. È emerso che gli italiani spesso usano la congiunzione con valore causale (58 casi

corrispondenti al 3% circa): “*Stai zitto **che** è meglio*”. Risultano pochissimi casi in cui viene assegnato valore relativo-temporale o concessivo mentre sembra un fenomeno diffuso l’uso del *che* in sostituzione di un pronome relativo più adatto (art. o prep. + *cui* ecc.). Vediamo in dettaglio gli usi del “*che*” da parte degli italiani nel corpus analizzato.

4.3.1 Principali usi non standard del “*che*”

Ecco i principali valori del “*che* polivalente” riscontrati nel nostro corpus:

- Relativo (uso del “*che*” in frasi relative nelle quali sarebbe più appropriato l’utilizzo di “preposizione + *cui*” oppure “preposizione+ articolo + *quale*”) (Palermo, 2015, p. 208)
 - “*Ecco che mi affido a te MATTEO , quel carnevale **che** lanciano arance , non mi sembra idoneo come divertimento*”
 - “*xke nn creare una gamma METANO al di fuori del twin air; con la distribuzione di metano **che** l'Italia dispone sarebbe una grossa alternativa economica ed efficiente!!!*”
- Causale (uso del “*che*” per introdurre subordinate di tipo causale o per esprimere la causa di ciò di cui si parla. In sostituzione di: “*perché*”, “*poiché*”, “*dato che*” ecc.)
 - “*si, molto bene... ma non parlare di vaccini **che** fai solo autogol*”
 - “*Occhio **che** in fondo al corridoio ti trovi Renzi e Martina*”
 - “*Ma ritirati dalle scene **che** di danni ne hai fatti a sufficienza!!!!*”
- Consecutivo-presentativo (uso del “*che*” per esprimere la conseguenza o l’effetto di ciò di cui si parla. In sostituzione di: “*al punto che*”, “*così che*”, “*in modo che*” ecc.)
 - “*E a Metaponto quando vieni con i pullman pagliaccio **che** ti inforchiamo anche a te dentro e ti buttiamo a mare*”
 - “*Forza guerriero.. aiutati **che** DIO ti aiuta..non mollare*”
- Relativo-temporale (uso del “*che*” per indicare il momento in cui si svolge un’azione. In sostituzione di: “*mentre*”, “*come*”, “*quando*”, “*nel momento in cui*” ecc.)

- *“Ormai **che** sono sul pulman vai verso il porto e carica le navi!!!!”*
- *“Come rosicheranno **che** il Pil schizzerà quest'anno. Già me Io vedo in coda alle farmacie per acquistare Maloox 😊”*
- Concessivo (uso del “che” per introdurre subordinate concessive o per indicare una circostanza in contrasto con quanto si è detto in precedenza. In sostituzione di: “ma”, “benché”, “sebbene”, “nonostante”, “anche se” ecc.)
 - *“Sulla TAV non si sono mai messi d'accordo prima perché le mazzette o le buste chiamatelo come volete erano poche volevano maturare di più **che** nel frattempo anno comunque mangiato”*

4.3.2 Altri usi non standard del “che”

Tra gli usi del “che” non standard vale la pena segnalare quello in cui è utilizzato prima di un clitico che riprende l’oggetto diretto o indiretto. Nei commenti analizzati il “che” con il clitico di ripresa ricorre 6 volte. Di seguito sono stati inseriti due esempi dal nostro corpus. Nel due esempi vediamo due tipici esempi di clitico di ripresa dell’oggetto diretto (“dolore”).

*“Quei genitori sono già immensamente immersi nel dolore **che** non lo potranno mai colmare”*

*“La perdita di un figlio e un dolore troppo grande **che** lo può capire solo chi ci passa”*

Durante l’analisi dei commenti è emerso inoltre l’uso del “che” aggettivo nella sua variante intensiva (Serianni, VII, 125). Questo utilizzo del “che”, volto a enfatizzare il sostantivo o verbo che lo precede, rafforza l’enunciato. Nel nostro corpus è stato registrato 11 volte. Ecco un esempio estratto dai commenti analizzati:

*“**Che** grinta **che** hai, sei veramente un grande Manuel”*

Tra gli usi non standard del “che” troviamo anche quello rafforzativo del pronome neutro in espressioni ironicamente affermative. Nel corpus sono presenti cinque casi. Questo tipo di costrutto è costituito solitamente da una preposizione articolata seguita da un sostantivo a cui seguono il “che” e una frase dichiarativa.

A queste espressioni interiettive costruite con il pronome, Serianni aggiunge anche quelle di tipo interrogativo (es. “Che cavolo vuoi?”). Ecco due esempi dal corpus di riferimento:

*“Finalmente col cavolo **che** li accolgono se non rubacchiano soldi”.*

*“(...) Col c... **che** ti voto !”*

Singolare è invece l’uso del “che” in una delle espressioni analizzate. Nella frase “(...) infatti ho notato che sono meno, da noi erano fissi **che** entravano nelle case a rubare (...)” notiamo subito che si tratta di un uso non standard del “che” usato in sostituzione di una più pertinente preposizione. Sia se nel caso in cui *fissi* venga considerato aggettivo sia che a “fissi” venga attribuita funzione di avverbio per indicare la frequenza, facendo riferimento alla persistenza e all’intensità, “essere fissi” non può essere seguito da una congiunzione. La costruzione verbale avrebbe dovuto reggere un complemento indiretto introdotto dalla preposizione semplice *in* (“essere fissi in un proposito, in una determinazione”). Probabilmente una resa molto più vicina allo standard sarebbe stata “*erano fissi a entrare nelle case*” anche se probabilmente sarebbe stato opportuno rivedere tutta la costruzione sintattica (“*da noi entravano fissi nelle case*”). Giuliana Fiorentino inserisce casi come questo in varietà diafasicamente marcate come basse e popolari in cui il valore sintattico non può essere stabilito.²³ Un altro caso in cui è impossibile indicare quale valore assuma il “che” è contenuto nel commento seguente: “*Uguale a Renzi quando parlava in inglese **che** ridevano tutti, brava Giorgia*”. Come segnalato anche nel cap. 4.1, esiste infatti la possibilità che si tratti di valore coordinante e non subordinante (“...quando parlava in inglese *e* ridevano tutti”).

4.3.3 “Che” di difficile interpretazione

Un “che” di difficile interpretazione è quello contenuto nella frase “*Meno male grazie Salvini **che** ai messo fine a sto schifo*”. Il dubbio in questo caso riguarda il

²³ Fiorentino Gabriella, voce “che polivalente”, Treccani

valore assunto. Nel caso in cui l'intenzione dell'autore fosse quella di ringraziare il Vicepremier in quanto autore dell'"aver messo fine", il "che" sarebbe considerato un pronome relativo. Se invece i ringraziamenti fossero stati inviati per l'"aver messo fine", allora probabilmente si tratterebbe di aferesi di "poiché" e il "che" assumerebbe valore causale. Altre casi simili nei quali è impossibile determinare il valore del pronome sono i seguenti: "(...) *adesso devi fare piazza pulita di tutta questa mondezza **che** non hanno diritto di stare in questo paese*", "*acquistate auto italiane **che** sono superiori*" ecc. Tutti questi casi sono stati classificati tra quelli con valore incerto.

4.3.4 "Che" non classificabili

Nei commenti analizzati è presente anche un commento impossibile da classificare. Il commento recita: "*Aspetta **che** allora magne con i volontari*". È molto difficile considerare questo "che" come polivalente perché non è ben chiaro il messaggio che ha voluto mandare l'autore. Dal contesto purtroppo non si evince se intendesse dire "*Aspetta (per)che allora mangio con i volontari*" oppure "*Aspetta che mangino con i volontari*". Nel primo caso si sarebbe trattato sicuramente di aferesi e quindi di un "che" con valore causale mentre nel secondo non ci sarebbero stati problemi a classificare l'utilizzo della congiunzione come standard.

4.3.5 "Che" standard atipici

Nel corpus analizzato è necessario segnalare il "che" in principio di frase col senso di "inoltre" o "alla fin fine". È il caso di frasi come "*Che poi avete notato?*" o "*Che poi, seriamente, come si può dare credito a quello che dice Adinolfi?*". Sono stati registrati solo due casi di questo tipo nel nostro corpus e sono stati annoverati tra i "che" standard.

Sono stati registrati 18 casi in cui il "che" viene utilizzato come congiunzione per rafforzare un'esortazione o un invito. Questo tipo di costruzione, molto simile a quella del "che" esplicativo-consecutivo, è composta da un verbo all'imperativo (in alcuni casi al posto del verbo si trova un sostantivo) seguito dal "che". In questo caso è stato assegnato valore standard poiché l'intento dell'autore non è tanto quello

di servirsene per indicare la causa o il motivo (“*Forza perché sei forte*”) di quanto espresso nella prima parte della frase, bensì di incoraggiare o prendersi gioco ironicamente del destinatario del messaggio. Ecco alcuni esempi:

*“Forza **che** sei forte complimenti”*

*“Dai **che** ci riesci, col tuo coraggio...”*

Un caso controverso e di difficile interpretazione è quello contenuto nella seguente frase: “*Lui è razzista, voi del PD che avete tolto la casa alla signora Peppina che cosa siete? **Che** grazie al razzista le è stata ridata.*” In questo caso il “che”, classificato tra quelli con valore causale (afèresi di “poiché”) in prima battuta, può avere un valore diverso. Infatti l’irregolarità sintattica della costruzione inganna. In questo caso il “che” ha valore di pronome neutro in funzione di “la qual cosa” ed è riferito a un concetto precedentemente espresso. Questo utilizzo, spesso usato dagli scrittori antichi all’inizio di un periodo è solitamente (ma non in questo caso) determinato dall’articolo. Dopo una breve analisi l’interpretazione diventa dunque più semplice e considerando anche la presenza del clitico “*le*” riferito sempre alla cosa che “*è stata ridata*” possiamo classificare tranquillamente il “che” come pronome relativo standard.

Un commento nel quale è stato complicato individuare il valore della congiunzione è il seguente: “*Dovete cambiare la legge elettorale regionale, dovete fare in tal modo **che** si vota per tutti nello stesso giorno, (...)*”. Oltre all’errata concordanza temporale del verbo “votare” che avrebbe dovuto presentare la sua forma al congiuntivo, è stato difficile comprendere se il “che” avesse valore consecutivo standard e introducesse una subordinata modale (“*in modo che si voti per tutti...*”) oppure se dovesse essere considerato afèresi di “perché” in una costruzione che presentasse, oltre alla principale (“*Dovete...regionale*), una coordinata contenente un pronome relativo (“*dovete fare in tal modo*”) nella quale si ribadiva il comando espresso nella frase principale, a cui sarebbe seguita una proposizione finale standard contenente il “che” (“*affinché si voti per tutti lo stesso giorno*”). La presenza di questo tipo locuzione avverbiale (“*in tal modo*”) ha complicato molto l’interpretazione del “che”. Dopo aver esaminato il contesto è stata verificata la

correttezza della costruzione sul vocabolario Treccani. Alla voce “tale” è possibile trovare la seguente definizione: “In tal modo *si utilizza in testi antichi e/o letterari e può essere utilizzato come antecedente di una proposizione consecutiva introdotta da che o da*”. Tutti i dubbi sono stati dunque fugati e al “che” è stato assegnato un valore consecutivo standard (“...*dovete fare in modo che si voti...*”). Per quanto riguarda l’intera frase invece, sarebbe stato più opportuno evitare il rafforzamento del messaggio, tramite l’introduzione di una coordinata alla principale con un secondo comando, e inserire direttamente la proposizione subordinata consecutiva (“*Dovete cambiare la legge elettorale regionale, ~~dovete fare~~ in tal modo **che** si vota per tutti nello stesso giorno...*”).

Anche la frase “(...) *Quella sì **che** è colpa nostra!*” è stata annoverata tra le costruzioni regolari. In una prima analisi questo uso, non essendo molto ricorrente nell’italiano scritto, era stato segnalato come incerto e solo in un secondo momento è stato classificato come standard. Il “sì”, utilizzato in collaborazione con il sostantivo espresso dal pronome dimostrativo “quella”, agisce in sostituzione di “certo”, “sicuro” (“*Quella certo che è colpa nostra*”). Nel corpus è stata trovata solo una costruzione di questo tipo.

4.3.6 Costruzioni ricorrenti con il “che”

È necessario segnalare alcune costruzioni standard ricorrenti nel nostro corpus. Sono state registrate 48 occorrenze nel corpus del costrutto “che + sostantivo + (clitico) + (articolo) + verbo al congiuntivo” con il “che” utilizzato con funzione ottativa²⁴, 34 occorrenze del costrutto “(verbo) + locuzione temporale + che” con il quale si introduce una contestazione, una frase di rammarico o di rimpianto (“*ora **che** sei sposato*” e simili). Un altro uso, registrato 23 volte, è quello del “che” preceduto da “altro” utilizzato quando la locuzione indica esclusione o preferenza rispetto a un altro elemento (Treccani, 2018). Un uso standard diffusissimo è quello del “che” esclamativo. Viene usato in frasi esclamative o davanti a un aggettivo che

²⁴ La frase ottativa esprime un desiderio, segnalato dal modo verbale (per lo più il congiuntivo) o da particolari marche lessicali (avverbi come magari, almeno)

la norma vorrebbe seguito da nome²⁵. Durante l'analisi sono state riscontrate 66 occorrenze di questo tipo.

*“(...) e **che** dio ti guidi nella tua vita un abbraccio”*

*“Era una vita **che** un popolo non amava più un premier”*

*“Altro **che** Renzi!”*

*“**che** tristezza!!!”*

4.3.7 Uso volutamente erroneo del “che”

Infine si segnala un caso di costruzione volutamente errata da parte dell'autore. Nel commento *“Ciao mio eroe mateo salvini. finalmente ai fato la legie giusta cosi se mi entrano li zingari a casa di don tonio li poso pichiare co la zironia che cio sempre con me. **Che** prima co i comunisti **che** se li tocavi a li stra comunitari ti a restavano a te”* è infatti possibile notare come alcuni errori morfo-sintattici vengano accentuati per dar vita a un commento sgrammaticato e ironico. L'uso del “che” è totalmente arbitrario fatta eccezione per il primo caso in cui ricorre (*“co la zironia **che** cio sempre con me”*) in funzione di pronome relativo. Infatti se anche la seconda volta in cui si registra potesse lasciare dei dubbi riguardo a un possibile errore involontario di sostituzione della più corretta congiunzione causale (*“**Che** prima co i comunisti..”*), nel terzo caso non ci sono dubbi nel considerarlo un errore voluto (*“prima co i comunisti **che** se li tocavi a li stracomunitari”*). Il contesto ci aiuta sicuramente a classificarlo come un errore cercato dall'autore (*“mateo salvini”* e *“stracomunitari”*). Se si fosse trattato di un errore involontario non ci sarebbero stati problemi nel classificare alcuni tipi di errore nella frase, prendendo in prestito la categoria dall'inglese, come *“misspelling”* (Tavosanis, 2011, p.75). In questo tipo di costruzioni il “che” è esasperato e non è possibile confrontarlo con gli altri casi in cui, seppur non sia stato usato in modo opportuno, emerge l'involontarietà dell'errore. Quest'utilizzo volutamente erroneo si è registrato solo una volta nel

²⁵ Vocabolario Treccani, voce “che”

nostro corpus ma è necessario precisare che spesso è difficile individuare questa tipologia d'uso.

6. Conclusioni

L'analisi è stata svolta con un approccio linguistico-computazionale grazie a strumenti per l'annotazione linguistica. Dopo una prima fase impiegata nella ricerca del campione d'indagine sono state individuate le 5 pagine che trattano di politica più commentate su Facebook. Grazie al software Facepager è stato possibile estrarre i commenti degli utenti dalle pagine in questione. Successivamente sono stati selezionati i commenti contenenti il "che" e su 1495 di essi è stata effettuata l'analisi linguistica prima con l'ausilio di UDPipe, pipeline addestrata sull'Italian Stanford Dependency Treebank, e soltanto in un secondo momento manualmente. Sono stati esaminati 1953 "che" ed è emerso un largo uso da parte degli italiani del "che" con valori non standard. Tra questi è da segnalare l'utilizzo della congiunzione esaminata in sostituzione della più opportuna causale. Questo tratto, sicuramente molto più vicino al parlato che allo scritto, si inserisce nel grande dibattito sul rapporto tra lingua parlata e reti virtuali su cui anche Crystal si espresse. L'analisi degli usi del "che" da parte degli italiani rimane comunque un aspetto molto interessante e variegato che si inserisce all'interno della più generale tendenza della lingua. Sarebbe interessante proseguire gli studi in questo ambito, estendendo l'analisi anche all'universo di pagine unofficial su Facebook che giorno dopo giorno stanno crescendo e, secondo molti esponenti, superando quelle ufficiali. Potrebbe essere utile cercare di capire se l'uso improprio del "che" si manifesti anche su altri social network come Youtube o Whatsapp. Inoltre, dal punto di vista prettamente linguistico sarebbe sicuramente funzionale all'analisi avere a disposizione una libreria di "che" polivalenti per individuare e monitorare i cambiamenti linguistici in atto che anche su questo fronte sono in corso e che sono stati evidenziati nell'elaborato. L'avvento dei social network ha portato ad una delle più veloci e prorompenti rivoluzioni della lingua ma di pari passo ha fornito dati linguistici che fino a qualche anno fa era impensabile ricavare (Tavosanis, 2011, p.

24). Grazie a software per l'estrazione di questi dati e a strumenti per l'analisi linguistico-computazionale oggi è possibile analizzare una vastissima varietà di costruzioni a addentrarsi nella "pancia" dell'italiano per capire in quale misura i social hanno cambiato e stanno cambiando questa lingua.

Bibliografia

Berruto, Gaetano. 1998. *Sociolinguistica dell'Italiano contemporaneo*. Roma, Carocci (1° ed. 1987).

Bosco, Cristina, e Simonetta Montemagni, Maria Simi. 2013. *Converting Italian treebanks: Towards an Italian Stanford dependency treebank*. In "7th Linguistic Annotation Workshop and Interoperability with Discourse". pp. 4-5.

Bosco, Cristina, e Vincenzo Lombardo, Leonardo Lesmo, Daniela Vassallo. 2000. *Building a treebank for italian: a data-driven annotation schema*. In "Proceedings of the LREC'00".

Fiorentino, Giuliana. 2011. *Che polivalente*. In: Raffaele Simone, Gaetano Berruto, Paolo D'Achille. "Enciclopedia dell'Italiano". vol. I. Roma, Istituto Della Enciclopedia Italiana.

Fiorentino, Giuliana. 2007. *Relative 'pragmatiche' in italiano*. In Federica Venier (a cura di). *Relative e pseudorelative tra grammatica e testo*. Alessandria, Ed. dell'Orso, pp. 53-71.

Fondazione Censis. 2018. *Quindicesimo Rapporto sulla comunicazione - I media digitali e la fine dello Star System*. Roma, Franco Angeli.

Nasuto, Paolo. 2018. *Applicazioni di strumenti di analisi linguistica a un corpus di commenti provenienti da Facebook*. Tesi di laurea, Università di Pisa.

Nivre, Joakim, e Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, Daniel Zeman. 2016. *Universal Dependencies v1: A Multilingual Treebank Collection*. In "Proceedings of LREC". pp.1659-1666.

Palermo, Massimo. 2015. *Linguistica Italiana*. Bologna, Il Mulino.

Serianni, Luca. 2005. *Grammatica Italiana*. Torino, UTET Università.

Tavosanis, Mirko. 2015. *Il linguaggio della comunicazione politica su Facebook*. In *L'italiano della politica e la politica per l'italiano, Atti del XI Convegno ASLI Associazione per la storia della Lingua Italiana (Napoli 20-22 novembre 2014)* a cura di Rita Librandi e Rosa Piro, Firenze, Cesati Editore.

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

Tortorelli, Cristina. 2017. *Analisi linguistica di commenti ai post delle pagine dei Ministri della Repubblica Italiana*. Tesi di laurea, Università di Pisa.

Treccani. 2012. *La grammatica italiana*, Istituto della Enciclopedia Italiana fondata da Giovanni Treccani.

Sitografia

Australian Broadcasting Corporation News. 5/04/18. *Facebook says up to 87m people affected in Cambridge Analytica data-mining scandal*. ABC News. <https://www.abc.net.au/news/2018-04-05/facebook-raises-cambridge-analytica-estimates/9620652>. Visitato il 21/03/2019.

Cosenza, Vincenzo. 9/10/18. *Come si calcola l'Engagement Rate?*. vincos.it. <https://vincos.it/2018/10/09/come-si-calcola-lengagement-rate/>. Visitato il 29/03/19.

Fubini, Federico. 16/09/2018. a. *Le interazioni delle pagine Facebook unofficial vicine al governo*. Corriere della Sera. www.corriere.it/cronache/18_settembre_16/interazioni-pagine-facebook-unofficial-vicine-governo-b7341b54-b9f1-11e8-a205-6445d272b52d.shtml. Visitato il 27/03/19.

Fubini, Federico. 16/09/18. b. *Chi sono i leader (politici) sul web: Post brutali? La gente vuole solo quelli*. Corriere della Sera. www.corriere.it/cronache/18_settembre_16/leader-politico-piu-ascoltato-web-muratore-taurianova-pro-salvini-gente-vuole-quello-1bc6eac6-b9e6-11e8-a205-6445d272b52d.shtml. Visitato il 27/03/19.

Rieder, Bernhard. 11/08/2018. *"Facebook app and how independent research just got a lot harder"*. Thepoliticsofssystem. <http://thepoliticsofssystem.net/2018/08/facebooks-app-review-and-how-independent-research-just-got-a-lot-harder/>. Visitato il 29/03/2019.

Rosenberg, Matthew, Nicholas Confessore, Carole Cadwalladr. 17/03/2018. *How Trump Consultants Exploited the Facebook Data of Millions*. New York Times. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>. Visitato il 21/03/2019.

Scarcella, Luca. 28/06/2017. *Facebook raggiunge 2 miliardi di utenti attivi al mese*. La Stampa. <https://www.lastampa.it/2017/06/28/tecnologia/due-miliardi-di-utenti-usano-facebook-One6S3f8L1Pu6ebMRsB8RK/pagina.html>. Visitato il 26/03/19.

Vahl, Andrea. 3/09/2013. *"Facebook News Feed Updates: How Marketers Should Respond to Story Bump"*. Socialmediaexaminer.com. <https://www.socialmediaexaminer.com/story-bump/>. Visitato il 29/03/2019.

Baroncelli: www.baroncelli.eu

Facebook: sezione per gli sviluppatori: <https://developers.facebook.com/>

Facepager: <https://github.com/strohne/Facepager>

Osservatorio Social di Alessio De Giorgi: www.osservatoriosocial.it

Turin University Treebank: <http://www.di.unito.it/~tutreeb/treebanks.html>

UDPipe: <http://lindat.mff.cuni.cz/services/udpipe/>

Universal Dependencies: <https://universaldependencies.org/>

We Are Social: www.wearesocial.it