



UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

**Analisi del *che* polivalente in commenti a post
di politici su Facebook**

Candidato: *Sara Dell'Orfanello*

Relatore: *Mirko Tavosanis*

Correlatore: *Maria Simi*

Anno Accademico 2018-2019

Indice

Introduzione	4
Il linguaggio del web	5
L'evoluzione del linguaggio politico	6
La comunicazione politica su Facebook	7
Metodologia della raccolta	9
Pulizia del corpus	10
Eliminazione di commenti non pertinenti all'analisi linguistica.....	11
Eliminazione dei tag	13
Normalizzazioni	13
Analisi con Read-IT	17
Read-IT.....	17
Caratteristiche estratte dal testo.....	18
Profilo di base	18
Profilo lessicale.....	19
Profilo sintattico.....	20
Analisi linguistica	23
Che polivalente.....	24
Tanl e UDPipe	28
Selezione dei commenti	30
Analisi manuale.....	31
Proposta tagset di classificazione	34
Conclusioni	37
Bibliografia	39
Sitografia	40
Appendice	41

Introduzione

L'enorme diffusione dei media telematici ha permesso di mettere in pratica una scrittura di massa, la quale ha contribuito a rivoluzionare il panorama linguistico così come lo conoscevamo fino a ora.

A questo ha certamente contribuito Facebook che è una delle reti sociali più diffuse in Italia e, come tale, è una risorsa linguistica importantissima.

Se si considera che per molte persone quella tramite il web è l'unica occasione di scrittura (Fresu, 2016, p. 340), risulta ben chiaro l'impatto che le nuove tecnologie hanno avuto sulle persone e sulla scrittura stessa. Tutto ciò ha infatti favorito un linguaggio più spontaneo, colloquiale e vicino al parlato con il conseguente aumento della soglia di tolleranza (Fresu, 2016, p. 340).

Le reti sociali hanno contribuito anche allo stravolgimento del modello di comunicazione utilizzato in politica, nel quale però resta centrale il ruolo svolto dal dialogo. La parola è infatti l'"arma" principale attraverso la quale un politico riesce a comunicare non solo con gli elettori, ma anche con i suoi avversari.

Il presente elaborato ha lo scopo di analizzare un corpus di 5000 commenti, appunto di tipo politico, raccolti sulle pagine Facebook ufficiali dell'ex Presidente del Consiglio Matteo Renzi e del Ministro dello Sviluppo Economico Luigi Di Maio.

Tale raccolta è stata interamente effettuata tra il 19 novembre 2018 e il 26 novembre 2018.

Il web si è rivelato una fonte importantissima per osservare e analizzare come e in quale misura i tratti neostandard si sono affermati nella scrittura in maniera sempre crescente.

Il termine neostandard indica infatti realizzazioni linguistiche normalmente utilizzate in situazioni di media formalità le quali, seppur non aderenti allo standard, sono ormai state accolte e utilizzate anche nello scritto giornalistico (Tavosanis, 2011, p. 109).

Ai fini della nostra analisi è risultato molto interessante esaminare il linguaggio tramite il quale i commenti raccolti sono stati realizzati, in particolar modo andando a ricercare la presenza di un tratto specifico tra quelli individuati come neostandard: il *che* polivalente.

Il linguaggio del web

L'avvento di internet e delle nuove tecnologie ha sancito un riavvicinamento alle pratiche di lettura e scrittura da parte di una fascia sempre più ampia di utenti.

Fino alla fine degli anni ottanta, il predominio dell'audiovisivo che i così detti media non alfabetici avevano portato con sé faceva presagire un'inesorabile perdita d'importanza da parte della scrittura, data per spacciata anche dagli studiosi più autorevoli. Ciò che nessuno aveva previsto era però l'inversione di marcia che avrebbe conferito la rete, la quale si presenta infatti con un atteggiamento fortemente "testocentrico" (Antonelli, 2017, p.12).

In generale risulta abbastanza complicato individuare caratteristiche adatte per tutti i generi di scrittura digitale, ma possiamo comunque individuare delle tendenze predominanti (Palermo, 2015, p.223-224):

- **Frammentarietà.** La scrittura digitale risulta frammentaria poiché l'utente tende a concepire il testo digitalizzato come modificabile e riassemblabile a proprio piacimento, molto più che nella scrittura tradizionale.
- **Brevità.** La principale motivazione della brevità dei testi digitali è legata alla facile caduta d'attenzione che la lettura su schermo comporta, perciò si predilige una scrittura chiara e concisa.
- **Iconicità.** A questo aspetto sono legati l'uso di emoticons, della punteggiatura espressiva e del maiuscolo come imitazione delle varie tonalità orali.
- **Dialogicità.** I testi digitali sono predisposti alla ricezione di commenti, interventi, "mi piace" e quant'altro, aumentando perciò la possibilità di comunicazione con altri utenti.

Dunque le nuove tecnologie hanno indubbiamente consentito un riavvicinamento alla scrittura, ma ne hanno anche profondamente cambiato la percezione. Infatti, la quotidianità e facilità con cui questa attività viene messa in pratica hanno favorito la desacralizzazione della scrittura (Antonelli, 2017, p.13).

Il concetto odierno di scrittura è lontano dalla solennità e dall'ufficialità con cui la stessa era concepita in passato. L'utente, contrariamente a quanto avveniva in precedenza, scrive in situazioni di scarsa concentrazione e senza una pianificazione del testo, ma soprattutto: si scrive dovunque per raggiungere chiunque e comunicare comunque (Antonelli, 2017, p.13).

Anche l'ansia da pagina bianca non ha più ragion d'essere. Lo schermo dei nostri dispositivi risulta più facile da riempire rispetto al tradizionale foglio di carta, forse perché già predisposto all'invio di una risposta o forse per la consapevolezza della potenziale presenza di un qualcuno ad attendere la nostra risposta dall'altra parte del dispositivo.

L'evoluzione del linguaggio politico

La politica è legata prevalentemente ad attività linguistiche: la parola risulta infatti il principale strumento utilizzato dai politici, soprattutto per cercare o rafforzare il consenso da parte degli elettori.

L'efficacia delle scelte linguistiche adoperate da un dato politico o movimento viene infatti valutata in base ai risultati ottenuti: più l'adesione al partito si rafforza, più le scelte si dimostrano adeguate.

Tutto ruota dunque intorno al linguaggio politico il quale, nel corso degli anni, si è modificato enormemente soprattutto grazie ai nuovi media. In particolare, l'11 ottobre 1960 è la data che sancisce l'ingresso della politica in televisione con la messa in onda della prima puntata di Tribuna elettorale: da quel giorno infatti, consapevoli del ruolo fondamentale svolto dal piccolo schermo nell'orientamento delle scelte elettorali, i leader politici hanno iniziato un lento ma graduale avvicinamento al nuovo mezzo di comunicazione fino a invadere tutti i generi televisivi (dai programmi destinati all'informazione politica a quelli d'intrattenimento) nelle fasce orarie più disparate.

Una delle principali caratteristiche del linguaggio politico televisivo è tuttavia l'oscurità: c'è la volontà di rendere complesso ciò che potrebbe essere espresso in modo chiaro, semplice e diretto.

Tra le motivazioni che spingono i politici ad adottare questo tipo di approccio, la principale è certamente quella di evitare prese di posizioni troppo nette che potrebbero scontentare, seppur parzialmente, il pubblico di elettori.

Al centro del discorso politico si colloca il ruolo del leader il quale, facendo leva sul proprio carisma, utilizza precise strategie linguistiche al fine di legittimare sé stesso e consolidare la propria figura politica. Lo scopo principale è conferire credibilità al leader convincendo chi ascolta del fatto che l'esponente politico ha tutte le carte in regola per realizzare concretamente quanto sta dicendo.

A partire dalle elezioni politiche del 2006 poi, a fianco dei tradizionali mezzi di comunicazione, il mondo della politica inizia a utilizzare i nuovi strumenti che il web mette a disposizione, specialmente blog e siti internet dedicati.

In particolare, con la nascita dei Social Network lo scopo principale diventa quello di instaurare una relazione “amichevole” con gli elettori parlando loro direttamente mediante l’utilizzo del “tu” al fine di farli sentire unici e speciali.

Si passa in questo modo dal *paradigma della superiorità* al *paradigma del rispecchiamento*, ovvero si utilizzano parole semplici e immediate in cui anche i più umili si possano rispecchiare (Antonelli, 2017).

L’impiego di un registro informale ha consentito perciò il passaggio dal *politichese*, tipico della Prima Repubblica, al così detto *gentese* (Antonelli, 2017, p. 222), caratterizzato da concetti elementari nonché dal desiderio di parlare alla sensibilità e ai sentimenti del popolo.

La volontà di comunicare alla “pancia degli italiani” ha portato con sé l’utilizzo di modi di dire popolari o dialettalismi, spingendosi addirittura fino al turpiloquio. È stata però proprio questa spontaneità ad avvicinare così tanto alla sfera politica il popolo, il quale poteva finalmente immedesimarsi in questa e pensare: “Quel politico parla come me e io parlo come lui: dunque mi piace” (Antonelli, 2017, p. 222).

La comunicazione politica su Facebook

Un altro decisivo stravolgimento è stato conferito al linguaggio politico dall’avvento dei Social Network.

La rapidissima diffusione e l’enorme importanza che gli stessi hanno acquisito all’interno della vita delle persone non sono passate inosservate agli occhi dei politici; così, tra il 2007 e il 2008, iniziano ad apparire i primi profili politici su Facebook.

Per la prima volta la politica ha la possibilità di rendersi autonoma rispetto ai tradizionali mezzi di comunicazione dialogando in prima persona con il proprio pubblico di elettori.

Si esce quindi dalla monodirezionalità della comunicazione televisiva abbandonando la centralità del leader che questo tipo di informazione prevedeva per dedicarsi a un’interazione diretta con i cittadini che mirasse a colpire i loro sentimenti e la loro interiorità.

Facebook è indubbiamente uno dei Social più utilizzati dal pubblico italiano: il 56% della popolazione italiana possiede un profilo Facebook (percentuale che sale a più del 70% nelle fasce più giovani di popolazione), mentre Twitter è fermo al 12,3% (Censis, 2018, p.13).

Nonostante l'enorme divario, le analisi linguistiche realizzate finora hanno coinvolto quasi esclusivamente Twitter mentre Facebook è rimasto pressoché privo di analisi. Anche la preferenza dei politici è ricaduta prevalentemente su Twitter; nonostante la maggior parte dei politici possieda un profilo Facebook, sono risultati molto più attivi sugli account Twitter.

Questa generale propensione da parte del panorama politico per Twitter è probabilmente proprio dovuta alla relativa élite di utenti da cui è popolato, la quale consente di instaurare dialoghi anche con singoli utenti. Facebook, al contrario, mantiene una comunicazione pressoché unidirezionale dal momento che i post vengono subito presi d'assalto dal grande numero di utenti, rendendo impossibile seguire qualsiasi conversazione, sia per i politici che per gli stessi utenti.

A rimarcare la differenza tra le due reti sociali contribuisce senza dubbio la grande differenza tra le interfacce, le quali vengono divise in due categorie: conversazioni¹ e broadcast² (Tavosanis, 2016, p. 679). Facebook appartiene alla seconda categoria: c'è un aggiornamento di stato come spinta iniziale alla quale fa seguito un certo numero di commenti.

Solitamente i politici iscritti su Facebook non possiedono un profilo privato bensì un account ufficiale il quale, nella maggior parte dei casi, viene gestito direttamente dagli uffici stampa per gli aggiornamenti di routine. I messaggi più importanti sono invece curati personalmente dai politici, o comunque concordati preventivamente con gli stessi.

Oltre al ruolo svolto dagli uffici stampa per conto dei parlamentari, bisogna anche tenere conto del fatto che normalmente i politici italiani non scrivono commenti su Facebook, perciò per l'osservazione e lo studio del linguaggio politico su Facebook sembra utile soffermarci più sull'analisi dei commenti lasciati dai loro elettori piuttosto che su quello dei parlamentari (Tavosanis, 2015).

¹ Le comunicazioni di questo tipo rappresentano le battute di un dialogo

² Le comunicazioni di questo tipo sono rivolte a un pubblico generico, possono essere atti di presenza in un dialogo o rimanere senza risposta

Riassumendo possiamo affermare che i messaggi politici su Facebook, realizzati spesso con un linguaggio colloquiale e molto vicino al parlato, possono godere di una enorme diffusione grazie alle caratteristiche intrinseche di questa rete sociale. Ciò comporta però un grandissimo numero di commenti, in molti casi anche estremamente volgari, che rendono impossibile l'instaurarsi di dialoghi, specialmente da parte del gestore della pagina.

Metodologia della raccolta

Per la realizzazione di questo corpus di commenti politici su Facebook sono stati presi in esame i post pubblicati da due delle identità politiche più in vista negli ultimi anni in Italia, Luigi Di Maio e Matteo Renzi.

Purtroppo non è stato possibile utilizzare le API di Facebook per lo scaricamento automatico del materiale, perciò per la raccolta dei commenti componenti il corpus si è rivelato necessario l'utilizzo di un semplice copia-incolla.

Per rendere il corpus sufficientemente vario, sono stati selezionati 5 differenti post per ciascun politico, ottenendo perciò un totale di 10 post distinti³.

La quantità di materiale disponibile sulle pagine dei due politici presi in oggetto era imponente, ma per l'analisi linguistica di nostro interesse sono stati selezionati in totale 5000 commenti, 2500 per ogni politico; per una corretta comparazione dei due corpora di commenti è stato poi deciso di raccogliere 500 commenti per ogni post scelto.

Avendo selezionato 5 post per ogni politico e avendo raccolto 500 commenti per ogni post, raggiungiamo quindi la quota dei 2500 commenti.

La raccolta del materiale è stata realizzata andando a prendere i commenti ricevuti da ogni post e le risposte ai commenti stessi.

Per maggiore chiarezza e a fini esemplificativi è stato riportato lo screenshot di uno dei post selezionati durante la raccolta dei dati (v. fig. 1). Appare chiaro come in questo caso siano state selezionate, oltre al commento del sig. Francesco S., anche le risposte sottostanti.

³ V. appendice

The image shows a Facebook post by Matteo Renzi, dated November 20 at 19:44. The post text reads: "Ho criticato spesso il sindaco di Roma Virginia Raggi per la mancanza di pulizia, per la qualità del manto stradale o per il servizio scadente degli autobus. Oggi sono felice di dare atto al sindaco e a tutta l'Amministrazione che l'operazione contro i Casamonica è davvero un'ottima cosa per Roma, per la legalità, per l'Italia. La serietà del confronto politico passa anche dal riconoscere i risultati degli avversari: con noi non lo hanno mai fatto ma anche su questo siamo diversi. Complimenti Sindaco. Sulla strada della legalità saremo sempre compagni di viaggio." The post has 11,983 reactions, 945 comments, and 1487 shares. Below the post, there are four comments: 1. Francesco S...: "Questo post non è fatto per fare i complimenti alla Raggi, serve per sottolineare il fatto che 'io Matteo Renzi sono diverso, sono bravo, faccio i complimenti ai miei avversari politici... poi ci butta in mezzo 'la mancanza di pulizia', 'la qualità del... Altro...' (6 likes). 2. Patrizia M...: "Già.. (4 likes). 3. Cinzia M...: "Francesco Ska Orrù 🤔🤔🤔 te avresti il 'sale in zucca'? (1 like). 4. Francesco S...: "Cinzia Melani, signora, capelli pochi, ma sale sicuro! Lei???" (2 likes).

Figura 1. Esempio esplicativo dell'estrazione dei commenti

Avendo dunque selezionato sia i commenti diretti al post che le risposte ricevute dai diversi commenti, si è rivelato necessario trovare un modo per riuscire a ricostruire la struttura dei post in esame. A questo scopo è stata realizzata una serie di codici per identificare in modo univoco ogni commento/risposta al commento del nostro corpus. Al fine di rendere più chiaro il procedimento utilizzato, ecco degli esempi:

- Il codice “MR1” indica il post numero uno di Matteo Renzi
- Il codice “MR2C3” indica il terzo commento al secondo post di Matteo Renzi,
- il codice “MR4C5R2” indica la seconda risposta al quinto commento del quarto post di Matteo Renzi.

Pulizia del corpus

Prima di effettuare l'analisi linguistica con Read-IT si è rivelato necessario ripulire i due corpus di commenti effettuando alcune normalizzazioni, necessarie per una corretta interpretazione dei dati durante la fase di analisi, ed eliminando contenuti privi di rilevanza a livello linguistico che avrebbero potuto alterarne il risultato. Di seguito sono riportate nel dettaglio le operazioni di pulizia operate sul materiale.

Eliminazione di commenti non pertinenti all'analisi linguistica

Prima di procedere all'analisi linguistica è stato necessario accertare che tutti i commenti fossero effettivamente rilevanti per il nostro scopo; molti commenti si sono infatti rivelati superflui poiché non portavano con sé informazioni linguistiche e perciò sono stati eliminati.

La non pertinenza dei commenti ha coinvolto due differenti casi:

- Commenti contenenti esclusivamente link (v. fig. 2).



Figura 2. Esempio di commento contenente link di collegamento

Casi come questo sono stati ignorati poiché, al di là del puro collegamento, non contenevano alcun dato linguistico che potesse in qualche modo arricchire la nostra analisi ma, al contrario, avrebbero potuto alterare alcuni risultati (es. numero di token).

Sono invece stati mantenuti link di collegamento inseriti all'interno di commenti più ampi poiché, eliminandoli, si sarebbe potuto perdere il significato originario del commento (v. fig. 3).

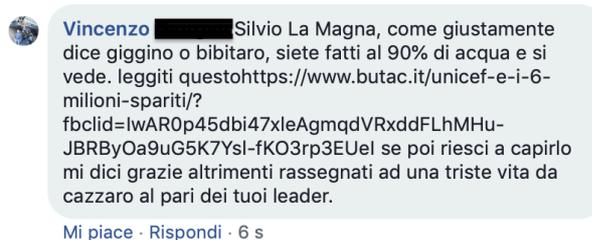


Figura 3. Esempio di commento contenente link di collegamento

- Commenti contenenti esclusivamente emoji o gif (v. fig. 4 e 5).



Figura 4. Esempio di commento contenente emoji



Figura 5. Esempio di commento contenente gif

Non è stato possibile mantenere commenti contenenti emoji o gif poiché non venivano interpretati correttamente dall'analizzatore linguistico.

Di sotto riportiamo una tabella riassuntiva riguardo il numero di link e di emoji/gif rinvenute nei commenti dei due diversi politici presi in esame.

Link					Emoji/Gif						
MR	MR1	2	DM	DM1	3	MR	MR1	0	DM	DM1	2
	MR2	1		DM2	4		MR2	0		DM2	2
	MR3	3		DM3	1		MR3	3		DM3	6
	MR4	3		DM4	6		MR4	0		DM4	9
	MR5	12		DM5	5		MR5	5		DM5	1
	TOT	21		TOT	19		TOT	8		TOT	20

Tabella 2. Numero di link e di emoji/gif

Appare chiaro che per i link le differenze tra i due politici sono minime: mentre per Matteo Renzi (MR) il totale dei commenti contenenti esclusivamente un collegamento link è di 21, per Luigi Di Maio (DM) il numero complessivo è 19, quindi inferiore di solo due unità.

La situazione cambia invece completamente nel caso delle emoji/gif. Come mostrato in tabella, per Matteo Renzi il numero di commenti con questo contenuto è pari solo a 8 elementi. Per Luigi Di Maio il numero complessivo è più che raddoppiato: si possono infatti contare 20 commenti appartenenti a questa specifica tipologia.

Eliminazione dei tag

Una delle principali funzionalità di Facebook consiste nella possibilità di *taggare* gli utenti iscritti al Social Network inserendo il nome della persona di nostro interesse all'interno di un commento. Ciò permette di condividere materiale e di sottoporre all'attenzione del destinatario i contenuti più disparati: post, foto, video e molto altro. Alcuni di questi tag sono parte integrante del commento che li contiene ed eliminandoli perderemmo materiale linguistico costitutivo della frase (v. fig. 6); sono perciò stati conservati tutti i casi appartenenti alla situazione sopra descritta.



Figura 6. Esempio di commento contenente tag mantenuto invariato

Si è invece deciso di procedere con la rimozione dei tag non dotati di valore sintattico (v. fig. 7) poiché assolutamente superflui ai fini dell'analisi linguistica.



Figura 7. Esempio di commento contenente tag oggetto di eliminazione

Nel corpus oggetto di studio sono stati eliminati un totale di 1634 tag privi di valenza a livello sintattico.

Normalizzazioni

Per una corretta interpretazione dei dati da parte dell'analizzatore linguistico è stato necessario ricorrere ad alcune normalizzazioni.

Molto importante si è rivelata la regola del "punto-spazio-maiuscola" (Lenci, Montemagni, Pirrelli, 2005, p.106), la quale consente di riconoscere il punto come indicatore di fine frase se è seguito da uno spazio e da una parola la cui lettera iniziale è indicata come maiuscola. La corretta identificazione dei confini frasali è infatti una condizione fondamentale e un punto di partenza indispensabile per una buona analisi linguistica. Proprio a questo scopo sono state adoperate modifiche differenti in

relazione alle situazioni incontrate, in particolare possiamo distinguere diverse categorie:

- Normalizzazione degli spazi interni e di confine frasale.

Molti dei commenti analizzati presentavano irregolarità rispetto agli spazi, sia quelli indicanti i confini frasali (v. fig. 8), sia quelli posti all'interno dello stesso periodo, specialmente dopo un segno di punteggiatura (v. fig. 13).

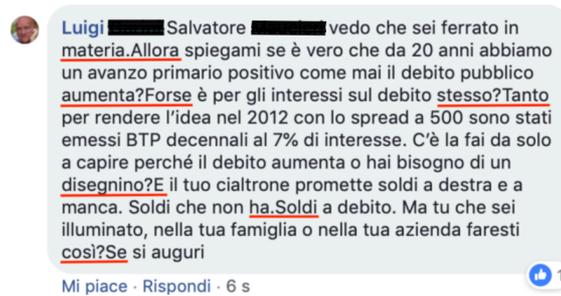


Figura 8. Esempio di commento normalizzato per gli spazi di confine frasale



Figura 9. Esempio di commento normalizzato per gli spazi interni

Casi come questo sono stati risolti semplicemente aggiungendo uno spazio dopo il segno di punteggiatura, così da rispettare la regola seguita dall'analizzatore.

- Normalizzazione delle maiuscole interne e di inizio frase.

Un altro caso molto frequente è quello del mancato uso della lettera maiuscola sia per nomi propri di persona e luogo (v. fig. 10), sia dopo un segno di punteggiatura forte (come il punto fermo, il punto interrogativo o il punto esclamativo) (v. fig. 11).

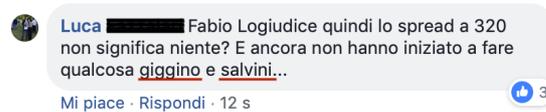


Figura 10. Esempio di commento normalizzato per le maiuscole interne alla frase



Figura 11. Esempio di commento normalizzato per le maiuscole di inizio frase

Anche in questo caso si è preferito procedere con la sostituzione dell'iniziale della parola con la lettera maiuscola corrispondente.

- Normalizzazione delle parole.

Forse a causa della velocità di scrittura, forse per l'utilizzo delle tastiere di smartphone o tablet, molto frequenti sono risultati i casi di parole scritte in modo errato (v. fig. 11).

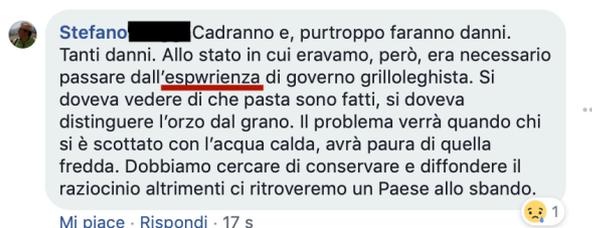


Figura 12. Esempio di commento contenente normalizzazione delle parole

Largamente diffusa sul web è anche l'espressività grafica, della quale sono stati trovati diversi esempi all'interno del corpus d'esame. Le devianze ortografiche più frequenti sono riconducibili a tre diverse tipologie: lettura endofasica (in cui i simboli vengono interpretati in base al loro nome, per esempio <6> in luogo di *sei*), abbreviazioni (per esempio *nn* in luogo di *non*) e variazioni grafiche (per esempio l'utilizzo di <k> al posto di <ch>).

In entrambi i casi, che si tratti di errori di battitura o di devianze ortografiche, per consentire all'elaboratore di individuare senza problemi il lemma di ogni singolo token, è stata effettuata una normalizzazione.

- Normalizzazione ortografica.

Nel corso della revisione dei commenti selezionati sono stati rilevati numerosi casi di anomalie ortografiche (come, ad esempio, l'utilizzo dell'apostrofo al posto dell'accento grave o acuto), le quali però non possono essere identificate come errori grammaticali a tutti gli effetti (v. fig. 12). Tali fenomeni possono infatti essere dovuti a diversi fattori quali, per esempio, il frequente utilizzo da parte degli utenti di tastiere non particolarmente agevoli (smartphone o tablet), o il non sapere come realizzare un determinato carattere; esemplare in questo caso è il grafema “È” il quale spesso viene realizzato come “E”.



Figura 13. Esempio di commento contenente normalizzazione grammaticale

Essendo impossibile distinguere gli errori “consapevoli” dagli errori effettivi, i casi di questo tipo non sono stati considerati errori grammaticali.

- Normalizzazione di errori grammaticali veri e propri.

Più rari, ma comunque significativi, i casi di errori ortografici veri e propri (v. fig. 15). Fra gli errori individuati, quelli più comuni sono senz'altro l'errato uso dell'h e la confusione nell'utilizzo dell'articolo indeterminativo maschile e femminile.



Figura 14. Esempio di commento contenente normalizzazione di errori grammaticali veri e propri

Di seguito è riportata una tabella riassuntiva che specifica il totale di normalizzazioni realizzate per ogni categoria, sia in numero che in percentuale. Poiché la lunghezza del commento potrebbe influire sulla presenza di errori, al fine di rendere i risultati direttamente confrontabili, la percentuale è stata calcolata rispetto al numero di token. A questo proposito è necessario specificare che il numero di token individuato per i due politici è differente, perciò la percentuale è calcolata su cifre diverse (come specificato nella sezione *Profilo di base*, pp. 18-19): 63068 per Renzi e 88090 per Di Maio. La percentuale totale è invece stata calcolata rispetto al numero complessivo di token di cui il corpus è composto, ovvero 151158.

	Norm. Spazi Inizio Frase	Norm. Maius. Inizio Frase	Norm. Parole	Norm. Ort.	Norm. Spazi Int.	Norm. Maius. Int.	Norm. Err. Gram.	Tot. Norm.
Renzi	122	121	213	301	247	177	33	1214
Renzi (%)	0,19	0,19	0,34	0,48	0,39	0,28	0,05	1,92
Di Maio	179	203	410	474	519	311	38	2134
Di Maio (%)	0,20	0,23	0,47	0,54	0,59	0,35	0,04	2,42
Totale	301	324	624	775	766	488	71	3349
Totale (%)	0,20	0,21	0,41	0,51	0,51	0,32	0,05	2,22

Tabella 2. Normalizzazioni effettuate

Complessivamente sono state realizzate 3349 normalizzazioni, pari al 2,22 % dei token totali del corpus.

Osservando le cifre riportate in tabella appare subito chiaro il netto stacco che separa i due politici: il corpus di commenti raccolto sulla pagina di Matteo Renzi ha subito il 1214 normalizzazioni, quello di Di Maio ben 2134.

Tranne che per le normalizzazioni degli errori grammaticali veri e propri, per i quali le differenze sono minime, il Ministro dello Sviluppo Economico ottiene punteggi nettamente più alti in tutte le categorie. Nonostante i casi di errori grammaticali riscontrati nel corpus di commenti di Luigi Di Maio siano 38 mentre quelli calcolati per Renzi solo 33, la percentuale di errore calcolata risulta essere più alta per l'ex Presidente del Consiglio.

La categoria nella quale è stata osservata la differenza più importante resta però la normalizzazione degli spazi interni alla frase: i commenti estratti dai post di Luigi Di Maio hanno subito infatti 272 normalizzazioni in più rispetto a quelli di Matteo Renzi.

Analisi con Read-IT

Read-IT

I dati raccolti all'interno del corpus sono stati analizzati mediante la versione demo di Read-IT⁴, strumento di analisi computazionale elaborato presso il CNR di Pisa. Read-IT consente, tra le altre cose, di ricavare informazioni riguardo il numero e la lunghezza media delle parole e dei periodi che compongono il corpus e altri dati utili. Dopo aver effettuato l'analisi linguistica, Read-IT consente di consultare una serie di sezioni specifiche contenenti i risultati estratti dal nostro corpus; in particolare possiamo trovare:

- Suddivisione in frasi, in cui è possibile individuare la suddivisione in periodi del corpus analizzato
- Suddivisione in token, in cui è riportata la suddivisione in token (o parole) del corpus analizzato
- Parti del discorso, in cui ritroviamo il lemma associato a ogni token del nostro corpus, con eventuali altre specifiche di tratti morfo-sintattici
- Annotazione, in cui, oltre a tratti morfo-sintattici, vengono individuati i rapporti di dipendenza fra gli elementi della frase

⁴ <http://www.italianlp.it/demo/read-it/>

- Analisi globale della leggibilità, che valuta il livello globale della leggibilità del corpus in esame stabilito sulla base di diversi fattori (informazioni generali sul documento, aspetti lessicali e sintattici ecc.)
- Proiezione della leggibilità sul testo, che individua l'indice di leggibilità per ogni frase all'interno del documento, indicando il grado di difficoltà lessicale e grammaticale per ognuna di queste.

Caratteristiche estratte dal testo

All'interno della sezione “Analisi globale della leggibilità” possiamo trovare la sottosezione “Caratteristiche estratte dal testo”, la quale contiene tutta una serie di informazioni riguardo il corpus analizzato. Queste informazioni sono suddivise all'interno di tre ulteriori sottosezioni che vedremo in dettaglio di seguito.

Profilo di base

Il profilo di base è la prima sottosezione oggetto di studio.

Al suo interno possiamo ritrovare le informazioni più generali rispetto al corpus che abbiamo deciso di analizzare. In particolare, vengono estratti dal testo i seguenti dati:

- Numero totale dei periodi, ossia il numero totale di frasi di cui il corpus è composto. I segni di punteggiatura forte (punto fermo, punto interrogativo e punto esclamativo) e i ritorni a capo vengono interpretati come separatori di frase.
- Numero totale delle parole, ovvero il numero complessivo di token che compongono il testo. Nel caso in esame va tenuto in considerazione che, insieme ai token, sono state conteggiate anche le emoticon, tipiche della comunicazione su Facebook, che hanno quindi incrementato il valore estratto dall'analisi.
- Lunghezza media dei periodi, ovvero la lunghezza media delle frasi, espressa in token, del corpus.
- Lunghezza media delle parole, che restituisce la lunghezza media dei token, espressa in caratteri, che si trovano all'interno del nostro testo.

La tabella di seguito riportata rappresenta i dati ottenuti dall'analisi dei due corpora di commenti.

	Numero tot. di periodi	Numero tot. di token	Lunghezza media dei periodi	Lunghezza media dei token
Matteo Renzi	4582	63068	13,8	4,7
Luigi Di Maio	5637	88090	15,6	4,7

Tabella 3. Analisi linguistica

Appare evidente che i commenti estratti dai post di Luigi Di Maio hanno restituito risultati sensibilmente più alti. Il numero dei periodi si distacca infatti per più di mille punti dal totale delle 4582 frasi di Matteo Renzi, così come il numero complessivo di token che, per quest'ultimo, è di soli 63068 mentre sale a 88090 per Di Maio.

Maggiore è anche la lunghezza media dei periodi che per Di Maio è uguale a 15,6, mentre per Renzi si ferma a solo 13,8.

Risulta invece identica la lunghezza media dei token che si stabilisce a 4,7 per entrambi i politici.

Profilo lessicale

La seconda sottosezione presa in analisi è il profilo lessicale.

Come suggerisce il nome stesso, questa sottosezione tratta dati più strettamente legati all'ambito lessicale.

Tra le varie informazioni messe a disposizione, quelle di nostro interesse sono:

- Rapporto tipo/unità, anche detto Type Token Ratio, utilizzato per valutare la varietà lessicale di un corpus. Viene calcolato sulle prime 100 parole del testo (al fine di rendere paragonabili i risultati) mediante il rapporto tra le parole tipo e le unità del vocabolario, ottenendo un valore compreso tra 0 e 1. Più il valore è vicino a 1, più siamo di fronte a un testo caratterizzato da una forte varianza lessicale e viceversa.
- Densità lessicale, che rappresenta il rapporto tra le parole semanticamente piene (nomi, aggettivi, avverbi e verbi) e le parole funzionali (articoli, preposizioni ecc.).

Viene calcolata mediante il rapporto tra le parole portatrici di significato (le parole piene, appunto) e il numero totale delle parole all'interno del testo.

	Rapporto tipo/unità	Densità lessicale
Matteo Renzi	0,69	0,575
Luigi Di Maio	0,69	0,568

Tabella 4. Analisi linguistica

Dalla tabella emerge una situazione interessante. Infatti, nonostante il numero dei token calcolato sia nettamente inferiore, il corpus di commenti di Matteo Renzi raggiunge lo stesso punteggio ottenuto da Di Maio in merito al rapporto tipo/unità . La densità lessicale risulta invece maggiore nei commenti dell'ex Presidente del Consiglio.

Profilo sintattico

Il profilo sintattico occupa la terza e ultima sottosezione.

Più precisamente tratta informazioni legate all'analisi morfo-sintattica del corpus, che si occupa di associare la corretta categoria grammaticale a cui appartiene un token, e alla struttura sintattica sottostante, il cui compito è quello di descrivere i periodi dal punto di vista delle dipendenze tra parole.

Per prima cosa l'analisi restituisce una "misura delle categorie grammaticali" espressa in percentuale, grazie alla quale è possibile prendere visione della distribuzione degli elementi grammaticali all'interno dei periodi.

Di sotto è riportata una tabella contenente i risultati ottenuti dall'analisi dei corpora in oggetto.

	Sost.	Nomi Pr.	Agg.	Verbi	Congiunzioni		
					Coord.	Sub.	
Renzi	15,60	5,70	4,90	17,50	5,20	63,10	36,90
Di Maio	16,60	5,20	4,90	17,40	5,50	63,20	36,80

Tabella 5. Analisi linguistica

Osservando la tabella appare chiaro che non vi sono grandi differenze tra i risultati ottenuti dai due politici, nella maggior parte dei casi si parla di uno 0,10/0,50% di differenza.

Il campo in cui osserviamo la differenza più importante è quello dei sostantivi, per i quali è stato registrato un 15,60% per Matteo Renzi e un 16,60% per Luigi Di Maio. Gli unici campi a non presentare differenze sono quelli legati alle congiunzioni coordinate e subordinate che presentano la stessa percentuale di distribuzione in entrambe le identità politiche.

È possibile trovare anche una parte dedicata all'articolazione interna del periodo il cui scopo è fornire maggiori informazioni riguardo il modo in cui le frasi sono organizzate.

Tra gli altri, possiamo trovare anche dati riguardanti:

- Numero medio di proposizioni per periodo, dato dal rapporto tra il numero di proposizioni e quello dei periodi. L'aumentare di questo valore è indice di una maggiore complessità del testo a livello sintattico.
- Percentuale di proposizioni principali e subordinate, che esamina la proporzione tra i due diversi tipi di proposizioni per iniziare a stabilire il grado di complessità grammaticale delle frasi che costituiscono il corpus; infatti, un alto numero di proposizioni subordinate contribuisce inevitabilmente ad innalzarne il livello.

	N. medio di proposizioni per periodo	Percentuale principali	Percentuale subordinate
Matteo Renzi	2,022	74,20	25,80
Luigi Di Maio	2,269	70,30	29,70

Tabella 6. Analisi linguistica

La tabella mostra i risultati ottenuti con la nostra analisi.

Il corpus estratto dalla pagina Facebook di Matteo Renzi ha una percentuale di proposizioni principali più alta rispetto a quella del suo concorrente, Luigi Di Maio. Ciò significa che i commenti di Renzi hanno un grado di complessità grammaticale inferiore rispetto a quelli di Di Maio.

Riguardo all'articolazione interna della proposizione, i parametri presi in esame sono stati il numero medio di parole per proposizione e il numero medio di dipendenti per testa verbale.

La tabella sottostante riporta i risultati ottenuti per questi parametri: le differenze risultanti sono minime.

	N. medio di parole per proposizione	N. medio di dipendenti per testa verbale
Matteo Renzi	6,808	2,021
Luigi Di Maio	6,886	2,019

Tabella 7. Analisi linguistica

I livelli di incassamento vengono valutati all'interno di un'apposita parte della tabella che prende il nome di "misura della profondità dell'albero sintattico".

Questa sezione risulta fondamentale quando ci troviamo di fronte a più di una proposizione subordinata perché consente di stabilire la complessità del testo mediante la ricostruzione dei rapporti di incassamento che intercorrono tra le diverse proposizioni.

Ciò che permette di risalire ai livelli di incassamento gerarchico è l'altezza massima dell'albero. Questa misura la distanza massima esistente tra la radice dell'albero e una foglia (ovvero una parola del testo che non ha dipendenti) mediante un numero di archi che rappresentano le relazioni di dipendenza esistenti.

Questo aspetto dell'analisi viene ulteriormente approfondito soffermandosi sulle strutture nominali complesse, composte da una testa nominale che viene modificata tramite aggettivi o complementi preposizionali, e sulla ricorrenza di proposizioni subordinate incassate ricorsivamente.

Come dimostra la tabella sottostante, anche in questo caso le differenze non sono particolarmente elevate sia per quanto riguarda la profondità media di strutture nominali che per la profondità media di catene di subordinazione. Più significativo invece lo scarto per la media delle altezze massime che per Di Maio raggiunge 4,342, mentre per Renzi si ferma a 3,913.

	Media delle altezze massime	Profondità media di strutture nominali complesse	Profondità media di "catene" di subordinazione
Matteo Renzi	3,913	1,152	1,237
Luigi Di Maio	4,342	1,17	1,248

Tabella 8. Analisi linguistica

Molto importante è anche la sezione dedicata alla “misura della lunghezza delle relazioni di dipendenza”, utile anch’essa per stabilire il livello di complessità del nostro testo. Com’è noto infatti, l’adiacenza di elementi semanticamente o sintatticamente simili consente di risalire più rapidamente ai rapporti esistenti tra le parole e di comprenderne appieno il significato.

La lunghezza delle relazioni di dipendenza corrisponde alla distanza, calcolata mediante il numero di token, tra la testa e il dipendente.

Due parametri importantissimi sotto questo punto di vista sono la media delle lunghezze di tutte le relazioni di dipendenza e la media delle lunghezze di dipendenza massime per ciascuna frase.

La tabella sotto riportata dimostra che non ci sono grandi differenze fra i due politici per quanto riguarda la lunghezza media, a differenza della media delle lunghezze massime per la quale troviamo un risultato di 5,720 per Matteo Renzi contro i 6,385 di Luigi Di Maio.

	Lunghezza media	Media delle lunghezze massime
Matteo Renzi	2,456	5,720
Luigi Di Maio	2,721	6,385

Tabella 9. Analisi linguistica

Analisi linguistica

Lo scopo principale di questo elaborato è un esame approfondito dei *che* presenti all’interno del corpus di analisi, con particolare attenzione alla variante del *che* polivalente.

La volontà è quella di andare a individuare i differenti valori assunti da questa specifica tipologia di *che* fornendo un’iniziale proposta di classificazione da utilizzare anche per l’analisi di corpora futuri.

Che polivalente

Il *che* polivalente, anche conosciuto come *che* subordinante generico (Serianni 2016, p. 569), è probabilmente il fenomeno linguistico più vistoso e complesso dell'italiano neostandard.

Contrariamente a quanto possiamo pensare, l'utilizzo del *che* polivalente ha origini antiche e ha lasciato le sue tracce come aferesi⁵ di *perché* (sotto forma di *ché*) già all'interno della Divina Commedia (Treccani alla voce "Che polivalente"), nei Promessi Sposi di Alessandro Manzoni ("ascoltatevi bene, *che* vedrò di farvela intendere" VI.31) (Serianni, 2016, p. 570) e nelle prose di Giovanni Verga ("intanto l'avvocato chiacchierava e chiacchierava *che* le parole andavano come la carrucola di un pozzo", I Malavoglia, 269) (Serianni, 2016, p. 570).

Nell'italiano standard la congiunzione *che* ha il compito di introdurre alcune tipologie di frasi subordinate: oggettive, soggettive e dichiarative (Treccani alla voce "Che polivalente"). Nell'italiano colloquiale e di uso medio si è però affermato un uso esteso del *che*, il quale viene utilizzato per introdurre subordinate le quali, secondo lo standard, necessiterebbero di congiunzioni più adatte a livello semantico.

In questi casi si parla quindi di *che* polivalente, il quale svolge una funzione di connettivo generico, ovvero mette in connessione due enunciati ma non ha una vera e propria funzione definita; infatti, come suggerisce il nome stesso, può raccogliere valori e funzioni differenti e può essere utilizzato per introdurre frasi temporali, causali, relative, finali, causali ecc.

Di seguito sono elencati i possibili valori che il *che* polivalente può assumere:

1. Uso **esplicativo-consecutivo**. Un esempio di tale tipologia può essere:

Vieni *che* ti pettino.

2. Uso **causale**. Il valore causale assunto dal *che* corrisponde alle subordinate di tipo causale: viene utilizzato al fine di specificare la causa, il motivo per cui ciò di cui stiamo parlando avviene (o dovrebbe avvenire). In questo caso il *che* può essere sostituito da: *perché*, *poiché*, *siccome*, *giacché*, *dal momento che*, *visto che*, *dato che* ecc. Esempi adatti per questa tipologia sono:

Copritevi, *che* fa freddo.

Vai a dormire *che* ne hai bisogno.

⁵ Fenomeno di tipo fonetico che comporta la caduta di uno o più foni iniziali di una parola

3. Uso **consecutivo-presentativo**. Il valore causale assunto dal *che* serve per esprimere la conseguenza o l'effetto di ciò di cui stiamo parlando, così come nelle subordinate di tipo consecutivo. In questo caso il *che* può essere sostituito da: *al punto che, a tal punto, in modo tale, così che, tanto che, in modo che* e simili. Esempi utili per questa tipologia sono:

È un funambolo, un equilibrista *che* quelli del circo di Pechino, al confronto, risultano dei dilettanti.

Io sono una donna tranquilla *che* sto in casa, lavoro.

4. Uso **relativo-temporale**. Il valore temporale assunto dal *che* specifica, come espresso dalle stesse subordinate temporali, il momento in cui si sta svolgendo, si è svolto o si svolgerà ciò di cui si sta parlando. In questo caso il *che* può essere sostituito da realizzazioni come: *mentre, quando, come, nel momento che/in cui, al tempo che* ecc. Esempi di tale tipologia possono essere:

Vado a lavorare *che* è ancora notte fonda.

Maledetto il giorno *che* ti ho incontrato.

5. Uso **finale**. Il valore finale assunto dal *che* serve, come espresso dalle stesse subordinate finali, per indicare il fine, lo scopo per il quale si sta compiendo l'azione di cui si sta parlando. In questo caso il *che* può essere sostituito da: *perché, in modo che, affinché, in modo da, al fine di, allo scopo di* e simili. Un esempio per questa tipologia è:

L'Emilia a volte mi chiamava dalle finestre, dal palazzo, *che* salissi, facessi, le portassi qualcosa.

6. Uso **enfaticamente-esclamativo**. Un altro caso particolare degno di essere preso in esame è indubbiamente il *che* enfaticamente-esclamativo (Treccani alla voce "Che polivalente"), il quale può essere osservato in frasi come:

Che sogno *che* ho fatto.

Che disordinati *che* siete!

Il *che* enfaticamente-esclamativo, anche individuato come aggettivo esclamativo in particolar modo nella sua variante intensiva (Serianni 2016, p. 324), è spesso utilizzato davanti a un aggettivo ed è tanto diffuso (sia nello scritto che nel parlato) quanto criticato dai grammatici.

7. Uso **pseudorelativo**. Un esempio per tale tipologia può essere:

Li vedo *che* scendono.

8. Uso **relativo**. In questo caso si individua l'uso generico di *che* in frasi relative nelle quali sarebbe più appropriato l'uso di realizzazioni quali *preposizione + cui* o (*preposizione*) + *art.* + *quale* (Palermo 2015, p. 208), come nelle frasi:

Il paese *che* [=in cui] sono stato domenica scorsa si chiama P.

L'amico *che* [=con cui] stavo parlando un attimo fa è una vecchia conoscenza.

Questo è un argomento *che* [= di cui/del quale] ne discutiamo spesso.

9. Uso con **clitico di ripresa**. Parlando degli usi non standard del *che* non possiamo non citare il caso del *che* con clitico di ripresa, riportato nell'esempio seguente:

È una cosa *che l'*ha detta il ministro.

Inizialmente si credeva che questo particolare utilizzo del *che* fosse per lo più presente nell'italiano popolare e colloquiale e che dunque la variazione più incidente fosse quella diastratica; in realtà è presente anche nel parlato di persone colte (Alfonzetti, 2002, p. 25).

La ripresa può essere messa in pratica in molte situazioni diverse, in particolar modo quando ci troviamo di fronte a oggetti o beneficiari (perlopiù animati). La ripresa di un oggetto diretto è favorita dall'impiego di verbi modali ("Difficoltà che dobbiamo affrontare e *che* dobbiamo cercare di risolverle") e dall'utilizzo di una terza persona indeterminata ("Io ho mangiato un pesce *che l'*hanno fatto buonissimo").

Nel caso di un oggetto indiretto ad influire molto sono, per esempio, la presenza di un soggetto animato ("È un tipo *che le* piace vivere avventurosamente"), l'uso di verbi come "farci caso", "tenerci" e "farci attenzione" che sono normalmente utilizzati con il clitico di ripresa ("È una cosa *che ci* faccio attenzione"), i quali sembrano portarlo con sé per una sorta di "solidarietà sintagmatica" (Alfonzetti, 2002, p.60) e molti altri esempi.

10. Uso **relativo-locativo**. In questo caso il *che* può essere sostituito da: *sul quale*, *dove* e simili. Un esempio rappresentativo di questa tipologia è:

Mi passi il libro *che c'*è scritto «Grammatica»?

11. Uso **concessivo**. Il valore concessivo assunto dal *che* viene utilizzato, così come le subordinate concessive, per esprimere una circostanza o un fatto parzialmente contrastante con ciò di cui stiamo parlando. In questo caso il *che*

può essere sostituito da: *anche se, benché, sebbene, nonostante, malgrado* e simili. Un esempio adatto a questa tipologia è:

Bevo ancora un po', *che* proprio non dovrei.

I primi nove possibili valori attribuibili al *che* sono stati ricavati dalla classificazione data da Giuliana Fiorentino (Treccani alla voce "Che polivalente"), il decimo e l'undicesimo caso sono invece stati tratti da un'altra classificazione (Panebianco, Varani, 2009, pp. 368-369).

È possibile individuare alcuni esempi estratti direttamente dal corpus oggetto di studio, e in particolare:

- **Uso causale:**
 - "(...) fate l'inceneritore a Pontassieve *che* in Campania ci sono già tanti bambini morti di tumore (...)"
 - "Antonio vai a dormire *che* non si può leggere come scrivi via ...scio"
 - "Sergio attenzione al fegato *che* schiatti"
 - "(...) Nonnetta va a dormire *che* è tardi..."
 - "(...) riguardati *che* l'inverno deve ancora arrivare (...)"
 - "(...) va a giocare a tennis *che* è meglio"
 - "Attenzione *che* è un modo di pensare che va per la maggiore (...)"
- **Uso consecutivo-presentativo:**
 - "Comunicazione falsa e bugiarda *che* se uno ci riflette fanno rabbrivire"
 - "(...) Inoltre le rammento, visto che avete la memoria corta, *che* spesso c'è bisogno di resettarla (...)"
- **Uso finale:**
 - "Fatemi sapere *che* provvedo."
 - "Porta argomenti *che* poi ti rispondiamo."
 - "Comunque vai a lavorare *che* fai crescere il PIL."
 - "Lei dove insegna *che* vengo a sentirla?"
 - "Completa la frase *che* faccio lo screenshot, sarai aggiunto alla lista di chi verrà denunciato per diffamazione."
 - "Bene ora voglio anche vedere il Daspo per qualcuno di FI e qualcuno del PD *che* ce li leviamo anche questi dai c*****i per sempre"
- **Uso enfaticamente-esclamativo:**
 - "*Che* brutta persona *che* sei..."

- “(...) *che* bella cosa *che* ha fatto.....”
- “*Che* brutte persone *che* siete (...)”
- “(...) lo vedi Renzi *che* brutta fine *che* ha fatto (...)”
- “Ammazza *che* TROLLACCIO *che* sei”
- “Ma *che* coraggiosi *che* siamo.”
- **Uso relativo:**
 - “Rispetto è una parola *che* gente come voi non conosce il significato”
- **Uso con clitico di ripresa:**
 - “(...) è una bufala QUELLA DI DARE LA COLPA AGLI ALTRI *che* la fate girare ad arte (...)”
 - “(...) le iene è un programma *che* non lo seguo.”
- **Uso concessivo:**
 - “(...) è importante comunicare nel linguaggio social...*che* ripeto per fortuna il mondo reale è altra cosa”

Tanl e UDPipe

La prima cosa da fare è individuare le caratteristiche principali dei due programmi utilizzati per l’analisi linguistica.

Tanl (Text Analytics and Natural Language) è una pipeline per l’analisi e l’elaborazione di testi⁶ ed è costituito da una serie di moduli che lo rendono in grado di effettuare una serie di operazioni, tra le quali:

- **Sentence Splitter:** realizza la divisione del testo in frasi
- **Word Tokenizer:** segmenta una frase in token
- **Word Aggregator:** aggrega espressioni polirematiche⁷ in un singolo token (es. “a meno che”)
- **POS Tagger:** arricchisce il token con attributi quali il POS⁸ e il lemma

⁶ http://www.lrec-conf.org/proceedings/lrec2016/pdf/992_Paper.pdf

⁷ Espressioni composte da due o più parole costituenti un unico sintagma non scomponibile e con significato autonomo rispetto ai singoli termini

⁸ Part-of-Speech (POS), categorie di parole che possiedono proprietà grammaticali simili e che mostrano comportamenti simili dal punto di vista sintattico (Fonte: https://en.wikipedia.org/wiki/Part_of_speech)

- **Morph Splitter:** si occupa sia della suddivisione del POS di ciascun token in diversi POS e attributi morfologici specifici, sia della suddivisione delle forme clitiche in due o più token (es. “averlo” > “aver – lo”)
- **Parser:** analizza le frasi e le dipendenze presenti tra le parole producendo alberi di dipendenza

Altri moduli fondamentali consentono a Tanl di trovare e distinguere nomi di persone, luoghi, organizzazioni, quantità, percentuali e simili (Named Entity Tagger) oppure di identificare ogni elemento all'interno di un testo e classificarlo entro categorie specifiche (Named Entity Recognition).

UDPipe⁹ a sua volta è una pipeline che effettua la tokenizzazione, l'etichettatura, la lemmatizzazione e l'analisi delle dipendenze all'interno di un testo. Nella sua versione come servizio online vi è sia la possibilità di inserire testo manualmente (selezionando l'opzione “Input Text”) oppure di caricare un file di testo direttamente dal proprio computer (utilizzando l'opzione “Input File”).

Il software gratuito è distribuito mediante Mozilla Public License 2.0¹⁰ ed esclusivamente sotto licenza CC BY-NC-SA¹¹.

In entrambi i casi l'aspetto di maggior interesse è quello relativo al tagger. Tanl include l'HMM Tagger il quale gestisce tagset di grandi dimensioni senza problemi; il tagset Tanl comprende infatti più di 300 tag contenenti specifiche morfologiche, per i quali non vengono registrati cali di prestazioni e/o accuratezza.

La definizione del tagset Tanl è stata possibile partendo da quello utilizzato da ISST¹² il quale, a sua volta, fa riferimento al tagset ILC / PAROLE¹³ conforme allo standard internazionale EAGLES¹⁴.

⁹ <http://lindat.mff.cuni.cz/services/udpipe/run.php>

¹⁰ <http://www.mozilla.org/en-US/MPL/2.0/>

¹¹ <https://creativecommons.org/licenses/by-nc-sa/4.0/>

¹² L'ISST (Italian Syntactic-Semantic Treebank) è un corpus italiano annotato a più livelli sviluppato tra il 1999 e il 2001. Informazioni più dettagliate sono consultabili all'indirizzo: <http://medialab.di.unipi.it/isst/ISST.html>

¹³ Informazioni più dettagliate riguardo il progetto ILC/PAROLE sono consultabili all'indirizzo: <http://webilc.ilc.cnr.it/viewpage.php/sez=ricerca/id=33/vers=ita>

¹⁴ Lo standard EAGLES (Expert Advisory Group on Language Engineering Standards) è un insieme di linee guida e raccomandazioni messe a punto da una vasta comunità

I POS tags messi a disposizione dal tagset TANL possono essere distinti in base alla loro granularità, si possono infatti individuare:

- Tag a grana “grossa” (o coarse-grain), come: A (Adjective), C (Conjunction), P (Pronoun) ecc.
- Tag a grana “fine” (o fine-grain), come: CC (Coordinative Conjunction), CS (Subordinative Conjunction), PR (Relative Pronoun) ecc.
- Morphed tags, i quali includono informazioni morfologiche come genere, numero, persona, modo, tempo e clitico.

Il tagger di UDPipe è invece composto da diversi modelli di MorphoDiTa¹⁵, ognuno dei quali riesce a risalire a POS tags e lemmi specifici.

Selezione dei commenti

Il punto di partenza fondamentale per l’analisi è stata la selezione manuale di 5000 commenti politici (descritta in dettaglio nella sezione *Metodologia della raccolta*) rilasciati dagli utenti ai post pubblicati sulle pagine Facebook ufficiali del Ministro dello Sviluppo Economico Luigi Di Maio e dell’ex Presidente del Consiglio Matteo Renzi.

Al fine di ottimizzare il processo di analisi, è stato utile andare a selezionare esclusivamente tutti i commenti in cui era presente un *che*. Tale operazione è stata realizzata tramite il comando *grep*, utilizzato dai sistemi Unix e Unix-like, il quale consente di ricercare all’interno di un file di testo, mediante l’utilizzo di espressioni regolari, la presenza di una determinata stringa di caratteri¹⁶ (v. fig. 17).

```
MacBook-Pro-di-Sara:desktop sahara$ grep -i che Commenti_Dell'Orfanello.txt > commentitot.txt
```

Figura 15. Utilizzo del comando *grep*

di ingegneri linguistici. Informazioni più dettagliate sono consultabili all’indirizzo: <http://www.ilc.cnr.it/EAGLES96/home.html>

¹⁵ MorphoDiTa (Morphological Dictionary e Tagger) è uno strumento open-source per l’analisi morfologica di testi in linguaggio naturale che esegue analisi morfologiche, tagging, tokenizzazione e altro. Informazioni più dettagliate sono consultabili all’indirizzo: <http://ufal.mff.cuni.cz/morphodita>

¹⁶ <https://it.wikipedia.org/wiki/Grep>

È stata utilizzata anche l'opzione *-i* di *grep*, la quale ha permesso di selezionare tutti i *che* presenti all'interno del file ignorando le differenze tra lettere maiuscole e minuscole.

Fatto ciò è stato possibile proseguire analizzando il nuovo file mediante programmi di analisi automatica del linguaggio e, successivamente, procedere a un'analisi manuale del risultato ottenuto per verificare la correttezza dell'analisi effettuata dal programma. Per la realizzazione del Gold Standard di riferimento per questo elaborato sono stati rivisti e corretti un totale di 2076 commenti contenenti complessivamente 3690 *che*.

Analisi manuale

Prima di andare avanti con l'analisi è stato fondamentale appurare che l'interpretazione effettuata dagli analizzatori fosse corretta o, se il caso lo richiedeva, procedere alla correzione degli eventuali errori. È stato perciò necessario mettere a punto un Gold Standard test set, ovvero l'output corretto di riferimento ottenuto mediante l'annotazione manuale del test corpus.

Il Gold Standard è stato realizzato valutando l'output ottenuto mediante l'analisi di UDPipe ed è stato poi utilizzato anche per valutare il risultato restituito da Tanl.

Sia Tanl che UDPipe si attengono alle specifiche definite dal progetto Universal Dependencies¹⁷, per il quale lo scopo principale è quello di facilitare l'annotazione di costruzioni simili in lingue differenti, seppur permettendo estensioni specifiche per ogni lingua quando necessario, fornendo linee guida specifiche e un inventario di categorie universale.

Per fare ciò viene utilizzato il formato CoNLL-U¹⁸, una versione rivisitata del formato CoNLL-X, codificato in file di testo (UTF-8).

I campi specificati nell'annotazione sono dieci, tra i quali possiamo distinguere:

1. **ID**: numero intero che rappresenta l'indice di parola. Alla prima parola di ogni frase viene affidato il numero 1, poi la numerazione prosegue in modo crescente
2. **FORM**: forma della parola o segno di punteggiatura
3. **LEMMA**: lemma o radice della parola

¹⁷ <http://universaldependencies.org>

¹⁸ <https://universaldependencies.org/format.html>

4. **UPOS**: POS tags universali
5. **XPOS**: POS tags specifici per ogni lingua. Se il campo non viene compilato viene inserito un underscore (_) per segnalare un valore non specificato
6. **FEATS**: contiene un elenco di caratteristiche morfologiche; se tali attributi non vengono specificati viene inserito un underscore (_) per segnalarne la mancanza, se invece vengono inseriti più tratti morfologici viene inserita una barra verticale (|) per segnalarne la separazione. Tutti i tratti morfologici vengono indicati come coppie valore = attributo
7. **HEAD**: testa del token corrente, la quale può essere rappresentata dal valore dell'ID o dal numero zero (0)
8. **DEPREL**: indica la relazione di dipendenza tra l'HEAD e il token corrente o uno specifico sottotipo della lingua in analisi
9. **DEPS**: elenco delle dipendenze secondarie
10. **MISC**: qualsiasi altra annotazione.

Il Gold Standard di riferimento, composto complessivamente da 3690 *che*, distribuiti su un totale di 2076 commenti, è stato realizzato mediante la revisione e correzione manuale dell'interpretazione fornita dal programma. I restanti 2924 commenti non sono stati presi in considerazione poiché non utili ai fini del presente elaborato.

Di seguito è riportata una tabella riassuntiva contenente il numero di interpretazioni corrette ed errate elaborate dal programma, con le relative percentuali calcolate rispetto al totale di *che* analizzati.

Totale	Corretto	%	Sbagliato	%
3690	3224	87,37%	466	12,63%

Tabella 10. Tabella riassuntiva analisi generica del *che*

Su un totale di 3690 *che*, l'87,37% del materiale analizzato è stato interpretato in modo corretto mentre il 12,63% ha richiesto una correzione.

Si è rivelato utile calcolare i valori di *precision*¹⁹, per valutare la correttezza delle risposte fornite dal sistema, e *recall*²⁰, al fine di verificare la copertura del sistema stesso.

Viene riportato anche il valore di *F-score*, il quale rappresenta la media armonica tra i valori di *precision* e *recall*.

	Precision	Recall	F-score
Tanl	0,8333	0,8241	0,8262
UDPipe	0,8841	0,8737	0,8748

Tabella 11. Tabella riportante i valori di Precision e Recall

Come possiamo osservare sia la *precision* che la *recall* ottengono valori leggermente più alti nei risultati forniti da UDPipe, indice non solo di una copertura maggiore del sistema, ma anche di una maggiore precisione e correttezza delle risposte fornite.

Osservando i risultati dell'analisi da parte di entrambi i programmi è stato possibile individuare le categorie di POS maggiormente presenti all'interno del nostro corpus di commenti, e precisamente:

- PRON (pronome), che conta 1895 occorrenze di cui 1809 PR (pronomi relativi) e 86 PQ (pronomi interrogativi)
- CONJ (coniunzione subordinativa), che conta 1619 occorrenze

Per il nostro studio i casi di maggior interesse restano però quelli del *che* polivalente, di cui sono stati individuati complessivamente 85 casi all'interno del corpus preso in esame.

¹⁹ Indica il rapporto fra il numero di elementi la cui etichettatura si è rivelata corretta e il numero complessivo di elementi etichettati nel corso dell'analisi (Fonte: https://it.wikipedia.org/wiki/Precisione_e_recupero).

²⁰ Ovvero il rapporto fra la quantità di elementi etichettati in modo corretto e il numero complessivo di elementi che appartengono a quella classe (compresi gli oggetti che non sono stati etichettati come appartenenti alla classe ma dovrebbero esserlo) (Fonte: https://it.wikipedia.org/wiki/Precisione_e_recupero).

Proposta tagset di classificazione

Come dimostrato nei paragrafi precedenti, è possibile individuare differenti tipologie di *che* polivalente. I programmi di analisi automatica del linguaggio allo stato attuale non sono in grado di individuare e distinguere queste tipologie, perciò si è dimostrato interessante predisporre risorse in funzione di questo obiettivo.

A questo scopo sono stati recuperati i corpora utilizzati dal dottor Burgarella e dal dottor Nasuto e sono stati uniti al corpus di commenti di riferimento per il presente elaborato, integrando a quest'ultimo i casi di *che* polivalente segnalati all'interno dei due nuovi corpora e raggiungendo così un totale di 183 casi di *che* polivalente. Fatto ciò è stato possibile procedere con l'annotazione dei casi di *che* polivalente riscontrati. Tale annotazione è stata inserita come feature morfologica nel modo seguente:

Poli=<etichetta>

Ovviamente le etichette variano a seconda della tipologia di *che* polivalente incontrata, per questo motivo è stato realizzato un tagset più specifico di riferimento:

Tipologia	Etichetta
Esplicativo-Consecutivo	ESC
Causale	CAU
Consecutivo-Presentativo	COP
Relativo-Temporale	RET
Finale	FIN
Enfatizzante Esclamativo	ENF
Pseudorelativo	PSE
Relativo	REL
Clitico di ripresa	CLI
Relativo-Locativo	RLO
Concessivo	CON

Tabella 12. Tagset *che* polivalente

Il risultato di tale operazione è quindi una risorsa Gold nella quale sono individuati i casi di *che* polivalente, distinti per tipologia e annotati secondo il tagset sopra riportato. Lo scopo è quello di fornire una risorsa iniziale utile in prospettiva di elaborazioni successive specifiche che mirino all'apprendimento, nonché alla distinzione delle differenti tipologie, di questo specifico tratto anche da parte di programmi di analisi automatica del linguaggio.

Il tagset propone un'etichetta per ognuna delle 11 categorie di *che* polivalente individuate durante lo studio approfondito del fenomeno (vedasi la sezione *Che polivalente*, pp. 24-28); di fatto però le etichette effettivamente utilizzate all'interno del corpus di analisi sono state solamente 9, e precisamente: ESC, CAU, COP, RET, FIN, ENF, REL, CLI, CON.

Complessivamente il numero di casi riscontrati all'interno del corpus è così distribuito:

Tipologia	Etichetta	%
Esplicativo-Consecutivo	5	2,73
Causale	122	66,67
Consecutivo-Presentativo	7	3,83
Relativo-Temporale	3	1,64
Finale	5	2,73
Enfatizzante Esclamativo	10	5,46
Relativo	21	11,47
Clitico di ripresa	7	3,83
Concessivo	3	1,64

Tabella 13. Distribuzione tipologie

Le tipologie esplicativo-consecutivo e relativo-temporale sono state riscontrate esclusivamente nel corpus del dottor Burgarella.

La percentuale riportata in tabella è calcolata rispetto al totale dei 183 casi di *che* polivalente analizzati.

Al termine dell'analisi la categoria maggiormente presente risulta essere la tipologia causale, per la quale si contano complessivamente 122 casi.

Di seguito viene riportato un esempio dell'applicazione della classificazione per ogni tipologia di *che*.

Tipo	Commento	Proposta
Esplicativo-Consecutivo	Vieni con i pullman pagliaccio <i>che</i> ti inforchiamo anche a te dentro e ti buttiamo a mare	Poli=Esc
Causale	Non ti fare consigliare da loro <i>che</i> perdi la pensione e figli e nipoti restano con un debito fino al secolo prossimo	Poli=Cau
Consecutivo-Presentativo	È l'invasione degli i*****i Antonio. Comunicazione falsa e bugiarda <i>che</i> se uno ci riflette fanno rabbrivire, loro fanno la politica con i tweet.	Poli=Cop
Relativo-Temporale	Come rosicheranno <i>che</i> il Pil schizzerà quest'anno.	Poli=Ret
Finale	Lei dove insegna <i>che</i> vengo a sentirla?	Poli=Fin
Enfatizzante Esclamativo	Ma <i>che</i> coraggiosi <i>che</i> siamo.	Poli=Enf
Relativo	Rispetto è una parola <i>che</i> gente come voi non conosce il significato.	Poli=Rel

Clitico di ripresa	Vecchia, è una bufala QUELLA DI DARE LA COLPA AGLI ALTRI <i>che</i> la fate girare ad arte , PERCHÈ SIETE INCAPACI DI GOVERNARE , NON AVETE FATTO NULLA IN 9 MESI.	Poli=Cli
Concessivo	..il tutto sta nell'uso <i>che</i> se ne fa, se si comunica sui social a mio avviso, è importante comunicare nel linguaggio social... <i>che</i> ripeto per fortuna il mondo reale è altra cosa.	Poli=Con

Tabella 14. Esempio proposta classificazione *che* polivalente

Conclusioni

L'analisi, svolta con un approccio linguistico-computazionale, è stata realizzata grazie a strumenti per l'annotazione linguistica.

Dopo aver selezionato tutti i commenti contenenti il “*che*” si è potuto procedere con l'analisi linguistica: a partire dal risultato fornito da UDPipe è stato messo a punto il Gold Standard di riferimento, utilizzato poi anche per la valutazione delle prestazioni di Tanl (Text Analytics and Natural Language).

Per la realizzazione del Gold Standard sono stati rivisti e corretti manualmente 3690 *che*, con particolare attenzione ai casi di *che* polivalente e, nello specifico, delle sue diverse tipologie. Allo scopo di rendere i programmi per l'analisi automatica del linguaggio capaci di individuare e distinguere le differenti tipologie di *che* polivalente, è stato realizzato un primo passo mettendo a punto un tagset più specifico riportante etichette distinte per ognuna delle categorie individuate per questo fenomeno. Tale tagset è stato applicato sia al corpus di commenti raccolti per il presente elaborato, sia

ai corpora utilizzati dal dottor Burgarella e dal dottor Nasuto nel corso delle loro analisi, fornendo perciò un'iniziale risorsa Gold utile per future analisi in questa direzione. A questo proposito, potrebbe essere interessante ampliare il corpus con materiale ulteriore, magari estendendo l'analisi al di là del linguaggio politico o includendo materiale proveniente da risorse linguistiche diverse (per esempio altri Social Network).

Sarebbe utile anche andare a verificare la correttezza del Part-of-Speech associato a ogni token del corpus realizzando così un Gold Standard valido per l'intero documento, il quale consentirebbe di stimare l'accuratezza degli strumenti in modo più preciso. Questa operazione risulterebbe utile anche al fine del riconoscimento dei casi di *che* polivalente dal momento che gli strumenti utilizzano il contesto in fase di addestramento.

L'analisi potrebbe poi essere portata avanti utilizzando le etichette individuate nel tentativo di addestrare gli strumenti al riconoscimento automatico dei casi di *che* polivalente e delle sue differenti tipologie.

È necessario spendere qualche parola anche in merito alle percentuali di distribuzione delle varie tipologie di *che* polivalente. Come osservato, la tipologia più frequente risulta essere quella causale la quale, proprio a causa della grande diffusione di cui gode, tende a essere avvertita come non troppo distante dalle realizzazioni più vicine allo standard; infatti, espressioni come “Vai a giocare a tennis *che* è meglio” non sembrano avere alcunché di strano. Questo però non si verifica con altre categorie: espressioni come “Rispetto è una parola *che* gente come voi non conosce il significato”, annoverata nella categoria di valore relativo, continuano difatti a essere percepite molto distanti dallo standard.

In conclusione, è possibile affermare che le sfumature linguistiche sono innumerevoli e la strada da fare per permettere agli strumenti di coglierle e analizzarle nel modo corretto è ancora lunga. Inoltre, essendo quelli utilizzati strumenti statistici, sarebbe necessario lavorare su un numero di casi molto maggiore poiché, soprattutto per alcuni tag, gli esempi a disposizione sono pochissimi. Le potenzialità a disposizione degli strumenti stessi restano comunque elevate e ulteriori studi condotti in questo campo non potranno che migliorarne le prestazioni.

Bibliografia

Alfonzetti, Giovanna (2002), *La relativa non-standard. Italiano popolare o italiano parlato?*, Palermo, Centro Studi Filologici.

Antonelli, Giuseppe (2017), *L'italiano nella società della comunicazione 2.0*, Bologna, il Mulino

Burgarella, Gabriele (2018), *Analisi linguistica di commenti estratti da post di politici italiani Facebook*, elaborato di laurea triennale, corso di laurea in Informatica Umanistica, Università di Pisa, anno accademico 2018-2019.

Cassarà, Giuseppe (2017), *Analisi linguistica di commenti a post Facebook di personaggi del mondo dello spettacolo*, elaborato di laurea triennale, corso di laurea in Informatica Umanistica, Università di Pisa, anno accademico 2017-2018.

Fondazione Censis (2018), *Quindicesimo rapporto sulla comunicazione. I media digitali e la fine dello star system*, Roma, Franco Angeli Editore

Dell'Orletta F., Montemagni S., Venturi G. "READ-IT: assessing readability of Italian texts with a view to text simplification". In: SLPAT '11 – SLPAT '11 Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (Edimburgo, UK, 30 Luglio 2011). Proceedings, pp. 73 – 83. Association for Computational Linguistics Stroudsburg, PA, USA, 2011.

Fresu, Rita (2016), *L'italiano dei semicolti*, in: Lubello, Sergio, *Manuale di linguistica italiana*, Berlino, Walter de Gruyter GmbH & Co KG, pp. 328-350.

Lenci, Alessandro, Montemagni, Simonetta, Pirrelli, Vito (2005), *Testo e computer. Elementi di linguistica computazionale*, Roma, Carocci editore, pp. 101-110.

Nasuto, Paolo (2018), *Applicazione di strumenti di analisi linguistica a un corpus di commenti provenienti da Facebook*, elaborato di laurea triennale, corso di laurea in Informatica Umanistica, Università di Pisa, anno accademico 2017-2018.

Palermo, Massimo (2015), *Linguistica italiana*, Bologna, Il Mulino.

Panebianco, Beatrice, Varani, Antonella (2009), *Metodi e fantasia. Narrativa*, Bologna, Zanichelli editore, pp. 368-369.

Serianni, Luca (2016), *Grammatica italiana. Italiano comune e lingua letteraria*, Novara, De Agostini Scuola SpA.

Serianni, Luca, Della Valle, Valeria, Patota, Giuseppe (2015), *Il bello dell'italiano. Comprendere, ragionare, comunicare. La grammatica*, Milano-Torino, Scolastiche Bruno Mondadori.

Spina, Stefania (2012), *Openpolitica. Il discorso dei politici italiani nell'era di Twitter*, Milano, Franco Angeli Editore

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

Tavosanis, Mirko (2016), *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.

Tortorelli, Maria Cristina (2017), *Analisi linguistica di commenti ai post delle pagine Facebook dei Ministri della Repubblica Italiana*, elaborato di laurea magistrale, corso di laurea in Informatica umanistica, Università di Pisa, anno accademico 2016-2017.

Sitografia

Documentazione di accompagnamento alla demo di Read-IT, consultabile all'indirizzo: <http://www.italianlp.it/wp-content/uploads/2016/01/Documentazione-READ-IT.pdf>

Tanl: <http://tanl.di.unipi.it/it/overview.html> e tutte le pagine riguardanti il progetto disponibili sul sito

Tavosanis, Mirko, «...e tu pensi che un premier abbia il tempo di rispondere ai post?». *La politica su Facebook*, in *Treccani Magazine*: http://www.treccani.it/magazine/lingua_italiana/speciali/leader/Tavosanis.html, 2015.

Treccani alla voce che polivalente: [http://www.treccani.it/enciclopedia/che-polivalente_\(Enciclopedia-dell%27Italiano\)/](http://www.treccani.it/enciclopedia/che-polivalente_(Enciclopedia-dell%27Italiano)/)

UDPipe: <http://ufal.mff.cuni.cz/udpipe> e tutte le pagine riguardanti il progetto disponibili sul sito

Universal Dependencies: <http://universaldependencies.org> e tutte le pagine riguardanti il progetto disponibili sul sito

Appendice

Post politici raccolti per l'analisi linguistica

In questa sezione sono riportati i post politici estratti dalle pagine Facebook ufficiali dei politici selezionati per l'analisi linguistica. I post sono stati trascritti in modo fedele; non sono infatti state apportate modifiche, interventi o normalizzazioni. I post in esame vengono riportati all'interno di una tabella così che sia possibile risalire all'autore del post, alla data della sua creazione e al codice associato al post stesso, per poter eventualmente prendere visione dei commenti ricevuti.

Autore	Data	Codice	Post
Matteo Renzi	19/11/18	MR1	<p>Dice Di Maio che gli impianti per i rifiuti sono "vintage". Certo: non facciamo niente, neanche un impianto di compostaggio! Diciamo NO a tutto, è così bello e facile dire NO. Per Di Maio la modernità, l'innovazione, il futuro sono i rifiuti che vengono portati in Germania, ovviamente su gomma, perché le infrastrutture ferroviarie vanno bloccate. Per Di Maio la modernità, l'innovazione, il futuro sono i condoni edilizi come quello approvato per Ischia, sono le norme per aumentare i fanghi nei terreni agricoli, sono le chiusure domenicali, sono i sussidi per chi sta a casa mentre si tagliano gli incentivi a chi crea lavoro. I cinque stelle giocano a fare gli ambientalisti ma non si rendono conto che saranno per sempre quelli dei condoni e del fango, in agricoltura e sulla rete.</p>
Matteo Renzi	20/11/18	MR2	<p>Salvini e Di Maio hanno scelto una linea di politica economica suicida per l'Italia. E i mercati stanno presentando il conto, ma lo paghiamo noi cittadini. Lo Spread a 320 significa che il prossimo anno avremo 6 miliardi di euro in più di interessi sul debito: tre volte il Canone Rai, due volte l'Imu sulla prima casa, quanto la cifra aggiuntiva stanziata per il finto reddito di cittadinanza. Lo spread così alto farà danno alle famiglie, alle imprese, ai risparmiatori delle banche, agli artigiani.</p> <p>Ricordate le polemiche sulle quattro banche fallite? Allora i soldi persi coi subordinati dai risparmiatori furono 300 milioni di euro. Da quando è in carica questo governo i risparmiatori hanno perso 300 MILIARDI di euro.</p> <p>I numeri sono aridi ma spiegano la realtà.</p> <p>E i numeri ci trasmettono altre immagini negative: chi ha i soldi li porta all'estero, i BTP faticano ad essere</p>

			<p>collocati, gli investitori stranieri hanno paura a intervenire in Italia.</p> <p>Mi immagino la reazione</p> <p>Basta con le critiche, Matteo. Fai una controproposta. Sappiamo che loro sono incapaci, ma tu che proponi? Semplice.</p> <p>Prendete la proposta lanciata in Leopolda a ottobre con Padoan. Abbassate subito al 2.1% il deficit. Fate le tre misure che abbiamo lanciato per famiglie, lavoro, immobiliare. Gli effetti saranno immediati: lo spread si dimezza, le tasse scendono.</p> <p>E a chi dice: perché non lo avete fatto voi? Perché quando eravamo al Governo noi lo spread era a 100. Il pasticcio è nato quando Salvini e Di Maio hanno iniziato a fare a pugni con l'Europa. E soprattutto a fare a pugni con la logica.</p> <p>Noi non siamo solo capaci di criticare. Noi facciamo proposte alternative. Quello che è certo è che con questa strategia Salvini/Di Maio l'Italia va a sbattere contro il muro. E chi paga sono sempre i più poveri. Cari VicePremier, fermatevi! E ragionate, se potete. L'Italia vale più del vostro orgoglio.</p> <p>Ora è il momento della saggezza.</p>
Matteo Renzi	20/11/18	MR3	<p>Sapete da cosa capisco che il clima sta cambiando? Dalle risposte che danno i profili finti, i troll di Lega e Cinque Stelle. Prima per ogni argomento contestavano il nostro operato, ora si limitano a insultare e chiedere di sparire.</p> <p>Fateci caso, amici.</p> <p>Da quando Travaglio è stato condannato due volte, nessuno ci attacca sui finti scandali.</p> <p>Da quando lo spread vola oltre 300, nessuno mette più in discussione i risultati economici del nostro governo.</p>

			<p>Da quando i grillini hanno votato il condono edilizio, nessuno parla più di onestà e di ambiente.</p> <p>Al massimo ti dicono: Sparisci, Ritirati, Hai perso il referendum, Non parlare, Vattene.</p> <p>Su questo ancora non vediamo passi in avanti. Perché il fatto che la democrazia sia confronto, dialogo, civiltà proprio non riescono a capirlo. E l'idea che io sia senatore perché eletto dai cittadini del mio collegio non riescono a digerirla.</p> <p>Noi continueremo a criticare il Governo ma a offrire per ogni critica un'alternativa. Ogni posto noi rispondiamo con proposte in positivo: cosa fare sulla Tav, sul ponte di Genova, sullo spread, sulle FakeNews.</p> <p>Loro spargono fango e odio, noi replichiamo con idee e progetti. Si stancheranno prima loro.</p>
Matteo Renzi	20/11/18	MR4	<p>Ho criticato spesso il sindaco di Roma Virginia Raggi per la mancanza di pulizia, per la qualità del manto stradale o per il servizio scadente degli autobus. Oggi sono felice di dare atto al sindaco e a tutta l'Amministrazione che l'operazione contro i Casamonica è davvero un'ottima cosa per Roma, per la legalità, per l'Italia. La serietà del confronto politico passa anche dal riconoscere i risultati degli avversari: con noi non lo hanno mai fatto ma anche su questo siamo diversi. Complimenti Sindaco. Sulla strada della legalità saremo sempre compagni di viaggio.</p>
Matteo Renzi	21/11/18	MR5	<p>Salvini e Di Maio stanno mettendo a rischio i soldi degli italiani. Sono due irresponsabili che pensano di ottenere più consenso sfasciando i conti. La bocciatura europea e dei mercati è un fatto molto grave. E questo Governo si conferma il miglior amico</p>

			<p>degli speculatori che stanno facendo soldi contro l'Italia.</p> <p>Mi appello al Presidente del Consiglio e al Ministro dell'Economia: FERMATEVI! Abbassate il deficit, aprite un dialogo con la Commissione, cambiate le misure.</p> <p>Noi abbiamo formulato con Padoan una controproposta che dimezzerebbe lo spread (che oggi è oltre 300 e che il nostro Governo aveva portato sotto quota 100) e abbasserebbe le tasse.</p> <p>Siamo pronti a collaborare per aiutare l'Italia e gli italiani. Ma voi fermatevi perché altrimenti il prezzo della vostra follia sarà pagato caro dai cittadini.</p>
Luigi Di Maio	21/11/18	DM1	<p>Sia noi che l'Europa vogliamo la stessa cosa: ridurre il debito. E l'Unione europea si convincerà che, per raggiungere l'obiettivo, abbiamo scelto l'unica strada che funziona: aiutare le famiglie e le imprese, creare nuove opportunità di lavoro per i giovani. È così che l'Italia finalmente può crescere.</p>
Luigi Di Maio	20/11/18	DM2	<p>Questa mattina ho fatto visita alla CNA. Gli artigiani e le piccole medie imprese sono stati le forze a cui l'Italia si è aggrappata nel momento di crisi più nera per rimanere in piedi e questo governo non se lo scorderà mai. Ho ricordato ai rappresentanti seduti al tavolo che questo governo vuole bene all'Italia e alle imprese italiane.</p> <p>Voi ci avete fatto grandi nel mondo negli anni più difficili e adesso è arrivato il momento di restituirvi un po' di quello che avete dato.</p>
Luigi Di Maio	23/11/18	DM3	<p>Ho chiesto ai probiviri di avviare il procedimento disciplinare per Maurizio Pascucci, indicando che secondo me, vista la gravità, merita il massimo della sanzione cioè l'espulsione dal Movimento 5 Stelle.</p> <p>Sulla mafia non è concesso neppure peccare</p>

			d'ingenuità da parte di chi si candida a ricoprire cariche pubbliche. Ci aspettavamo scuse, non arroganza. Questo non è un comportamento da MoVimento 5 Stelle e come tale deve essere sanzionato immediatamente. Qualora qualcuno della lista fosse eletto, gli verrà subito ritirato il simbolo.
Luigi Di Maio	24/11/18	DM4	<p>Gli italiani sono passati negli ultimi 10 anni per riforme del lavoro, aumento dell'Iva, riforma delle pensioni, aumento delle tassazioni per le imprese... e tutti i politici ci dicevano che erano misure che ci chiedeva l'Europa.</p> <p>Io credo in una Unione Europea solidale che porta avanti i suoi principi, però la macelleria sociale che ci stanno chiedendo non è possibile. Dialogheremo senza rinunciare ai diritti degli italiani.</p>
Luigi Di Maio	25/11/18	DM5	<p>Avrete visto il servizio delle Iene. E avrete visto anche la mia intervista.</p> <p>Come sapete, in tutti questi anni, alle Iene abbiamo sempre dato il massimo della disponibilità, non abbiamo chiesto di non mandare in onda servizi, a differenza di altri; non abbiamo mai chiesto alcun trattamento di favore e quando ci hanno rivelato qualcosa di importante li abbiamo ringraziati.</p> <p>Il caso di stasera riguarda un lavoratore che 8 anni fa ha lavorato in nero per mio padre. Sono contento che Salvatore - l'operaio - abbia trovato il coraggio di denunciare pubblicamente dopo 8 anni. Ho letto dei commenti che lo attaccano per averlo detto pubblicamente solo ora, personalmente non credo lo si debba aggredire, inoltre credo che Salvatore Pizzo abbia anche votato il Movimento alle ultime elezioni, visto che ha aderito alla nostra campagna di maggio #ilmiovotoconta (goo.gl/GyebKD).</p>

		<p>Salvatore Pizzo all'epoca dei fatti si è rivolto al Sindacato CGIL che gli consigliò di trovare un accordo con mio padre per farsi assumere, e infatti poi ha ottenuto un contratto regolare. Successivamente gli fu corrisposto anche un indennizzo.</p> <p>8 anni fa, come avrete visto dal servizio io non ero né socio dell'azienda, né mai mi sono occupato delle questioni di mio padre. Mio padre ha fatto degli errori nella sua vita, e da questo comportamento prendo le distanze, ma resta sempre mio padre. E capirete anche che sia improbabile che un padre racconti al figlio 24enne un accaduto del genere. A maggior ragione se, come ho detto nel servizio, abbiamo anche avuto un rapporto difficile, che sono contento sia migliorato negli ultimi anni.</p> <p>Come sempre, manterrò gli impegni presi e domani consegnerò a Filippo Roma i documenti su questa vicenda in particolare, che intanto ho chiesto di procurare a mio padre, e faremo tutte le verifiche che servono su quanto raccontato da Salvatore nel servizio.</p> <p>Buona serata a tutti.</p>
--	--	---

Abbreviazioni di uso frequente

Di seguito è riportato un elenco delle abbreviazioni più frequentemente utilizzate all'interno di questo elaborato, in particolare:

- Aggettivi: agg.
- Congiunzioni: cong.
- Coordinanti: coord.
- Figura: fig.
- Nomi propri: nomi pr.
- Numero: n.

- Pagina: p.
- Pagine: pp.
- Sostantivi: sost.
- Subordinanti: sub.
- Vedi: v.