UNIVERSITÀ DI PISA

Corso di Laurea in Informatica umanistica

RELAZIONE

Analisi del *che polivalente* all'interno dei commenti di Facebook

Candidato: *Matteo Mannini*

Relatore: *Mirko Tavosanis*

Correlatore: *Felice Dell'Orletta*

Anno Accademico 2020-2021

Sommario

Introduzione

L'obiettivo di questo elaborato di laurea è analizzare il *che polivalente* in un contesto particolare, quello dei social network, nello specifico all'interno dei commenti di Facebook.

Inizialmente è stato presentato lo sviluppo della scrittura e del web negli ultimi anni, fino ad arrivare ai social network, in particolar modo a Facebook. Successivamente, dopo aver inquadrato il *che polivalente* dal punto di vista linguistico, sono stati estratti 408 commenti (dal corpus di Fanpage e Repubblica su Facebook) e analizzati dal punto di vista grammaticale, lessicale e sintattico, utilizzando il software Read-IT. Per l'analisi del *che polivalente*, dopo una procedura di normalizzazione, è stato utilizzato il software UD-Pipe per ottenere un riscontro su come interpretasse questo fenomeno linguistico. Dopodiché, per verificare se l'uso del *che polivalente* fosse influente, dal punto di vista statistico nel campo linguistico, è stato effettuato il Test di Wilcoxon. Infine, per ottenere la distribuzione del *che polivalente* all'interno dei commenti, è stata eseguita un'analisi manuale, volta al riconoscimento dei vari usi.

1 Il Web e l'evoluzione della scrittura

L'avvento della televisione, della radio e del telefono permise al canale audio-visivo di prevalere per anni sulla scrittura. Con la nascita di Internet questa seconda oralità, definita così per la sua rinnovata dominanza del parlato, fu sorpassata dal ritorno della scrittura. [1]

Negli anni '90 l'infrastruttura Internet permise al web di configurarsi come uno spazio in cui esiste l'informazione e dove poterla cercare, uno spazio in cui ogni singolo elemento possiede un indirizzo. Tale circostanza consentì la creazione di una rete di informazioni, dove tutto e tutti risultano collegati. [2]

Grazie allo sviluppo del web, identificato come *web 2.0*, anche la scrittura ha subito un processo di modernizzazione. Questo processo portò alla creazione di testi digitali che permise alla scrittura di riassumere un ruolo centrale nella nuova società moderna. Nello "scrivere" d'oggi risulta difficile determinare caratteristiche comuni, sia per la quantità di testi digitali, sia per la differenza di pubblico coinvolto con essi. Tuttavia, è possibile individuare alcune tendenze dominanti:

- **Frammentarietà**: rispetto alla scrittura tradizionale, gli autori sono indotti a interpretare il testo come spostabile, assemblabile e modificabile, facilitandone la sua destrutturazione;
- **Brevità:** per far sì che l'utente legga il testo nel minor tempo possibile e focalizzi la sua attenzione su di esso, il web necessita di testi brevi e concisi. Tutto ciò ha anche facilitato la creazione di termini come *tvb* (ti voglio bene), *xchè* (perché), *cm* (come) e ecc.;
- **Iconicità**: l'utilizzo di una punteggiatura espressiva tipica dei *social network* (ma davvero!?!?) e degli *emoticons* ha determinato la riproduzione di atteggiamenti e gesti tipici dell'oralità in forma scritta;
- **Dialogicità**: il testo digitale viene trasmesso tra due o più persone e prevede di solito più risposte da parte dei destinatari, accentuando così la caratteristica propria del web, ovvero la condivisibilità;
- **Ridefinizione del rapporto tra testo, cotesto e contesto**: nella scrittura tradizionale i confini tra testo, cotesto, e contesto, risultano sostanzialmente definiti. Nella scrittura digitale non è così, diventano continuamente variabili. È possibile spostare

[1] Durant, 2009, p.5.
[2] Treccani alla voce "Web".

un testo inserendolo in un contesto differente da quello originario, entrare e uscire a piacimento da un testo all'altro e modificare in breve tempo le parti che lo compongono; [3]

- **Persistenza**: le pagine di un testo a stampa tradizionale sono statiche e permanenti. Invece, un testo digitale, non ha nulla di permanente. Questo confronto permette di esaltare la sua caratteristica dinamicità. Un utente può visionare la pagina di un sito web e il giorno dopo scoprire che quella determinata pagina non esiste più;
- **Link ipertestuali**: nel web qualsiasi unità di informazione su un supporto digitale può avere o meno un collegamento con un'altra. Quando tale caratteristica è presente, l'utente può passare da un'unità all'altra semplicemente con un "click". Un link ipertestuale, di solito, può essere riconosciuto facilmente attraverso un cambio di colore nel testo. [4]

1.1 I social network

Grazie al *web 2.0*, l'utente non è più identificato come uno "spettatore" di fronte ai servizi proposti dal web, ma diventa partecipe e interessato alla creazione di nuovi contenuti. In aggiunta a questo interesse, l'utente del web sente la necessità di poter condividere i propri pensieri e le proprie idee. Nascono così i *social network*, servizi informatici online che consentono di realizzare reti sociali virtuali. Consentono agli utenti di interagire tra di loro, di condividere contenuti testuali e contenuti multimediali (immagini, audio e video). Di solito, per poter essere iscritti, è necessario creare un profilo (protetto da password), fornendo i propri dati personali. Una volta terminata la fase di registrazione è possibile individuare gli altri utenti iscritti.

Nel 2002 fu lanciato Friendster, il precursore dei moderni social network. Esso implementava le funzioni di chat ed e-mail, con la possibilità di organizzare i contatti in gruppi di amici. Nonostante il successo, ben presto verrà soppiantato da LinkedIn, MySpace e soprattutto da Facebook, che diventerà un vero colosso del settore.[5]

Oggi i social più utilizzati sono: Facebook, YouTube, Whatsapp, Instagram e Wechat.[6]

[3] Palermo, 2015, pp. 223-224.
[4] Crystal, 2011, pp. 28-29.
[5] Treccani alla voce "Social network".
[6] Digital 2021, i dati globali:
https://wearesocial.com/it/blog/2021/01/digital-2021-i-dati-globali

1.2 Facebook

Facebook è stato fondato nel febbraio 2004 da Mark Zuckerberg e Dustin Moskowitz. Inizialmente è nato come rete sociale tra gli studenti universitari statunitensi, con il passare del tempo ha esteso la propria utenza a tutto il mondo. [7] Oggi è arrivato a contare 2,7 miliardi di utenti nel mondo [8], ed è il social network più utilizzato.

Gli utenti di Facebook possono iniziare a relazionarsi attraverso la funzionalità dell'amicizia. Le sezioni che consentono a Facebook di mantenere sempre "aggiornato" il legame tra amici sono principalmente due: il diario e i post.

Il diario (oppure timeline) permette agli utenti di modificare il proprio profilo, attraverso il caricamento di foto, video e testo. Nel tempo è possibile scorrere e visionare quanto è stato archiviato dall'utente sul sistema. Un ulteriore funzionalità è quella di aggiungere un avvenimento importante alla propria timeline, utile per ricordare momenti come lauree, matrimoni e feste particolari.

I post permettono agli utenti di pubblicare le proprie idee e opinioni utilizzando contenuti testuali con l'aggiunta o meno di foto e video. Con i post è possibile relazionarsi attraverso tre funzionalità: condivisione, commento e reazione (in sostituzione al tradizionale "mi piace"). La condivisione permette di divulgare ai propri contatti quanto pubblicato da un utente, dando risalto alla fonte originaria, il commento permette di lasciare un messaggio sotto al post in questione e la reazione consente di inserire un emoji (rappresentante una faccia divertita, una soddisfatta, una stupita, una triste e una arrabbiata) per esprimere apprezzamenti diversi sul contenuto condiviso.

[7] Treccani alla voce "Facebook".
[8] Digital 2021, i dati globali:
https://wearesocial.com/it/blog/2021/01/digital-2021-i-dati-globali

2 Italiano neo-standard

Il concetto di *neo-standard* è stato introdotto da Gaetano Berruto nel 1987. Nel repertorio dell'italiano, indica la varietà della lingua più vicina all'uso del parlato in contrapposizione allo standard normativo proposto dalle grammatiche.

Le principali aree della lingua interessate in questa distinzione sono:

- **Il sistema pronominale:** la norma tradizionale prevede l'utilizzo di forme differenziate dei pronomi personali a seconda della funzione logico-sintattica (soggetto o complemento). Un esempio che evidenzia la contrapposizione è la semplificazione della seconda/terza persona singolare e la terza persona plurale (l'utilizzo di *te* al posto di *tu*, di *lui/lei* al posto di *egli/ella* e di *loro* al posto di *essi/esse*). Fa parte di questa categoria anche il *che polivalente* (analizzato nel paragrafo 3.2);
- **Il sistema verbale:** esistono forme verbali che svolgono anche i compiti di quelle che vengono usate raramente; ciò le sottopone a **sovraccarico funzionale.** Il 90% delle forme verbali utilizzate nelle frasi principali appartengono a tre soli tempi dell'indicativo: il presente (72,6%), il passato prossimo (9,5%) e l'imperfetto (5,3%);
- **La sintassi:** i rapporti subordinativi vengono ridotti e maggiormente limitati al primo o al secondo grado, evidenziandone la semplificazione. Inoltre, ha anche acquisito un ruolo importante nella lingua scritta la **dislocazione a sinistra**, in particolare quella dell'oggetto, che comporta l'**accusativo preposizionale** ("a me la sua idea non mi convince").[9]

2.1 La relativa non-standard

Nella distinzione tra italiano standard e neo-standard un fenomeno interessante riguarda la trattazione della frase relativa, in particolare il ruolo del *che* utilizzato come introduttore di frasi relative. Esistono tre differenti casi:

- La costruzione analitica, formata da *che*, con la funzione di subordinatore generico, e dalla ripresa per mezzo di un personale clitico;

[9] Palermo, 2015, pp. 205-210.

- *Che* privo di qualsiasi mezzo di ripresa per tutte le posizioni sintattiche, dove la funzione sintattica dell'elemento relativizzato non viene espressa morfologicamente;
- L'uso del *che* con pronome di ripresa per l'oggetto diretto e raramente per il soggetto.[10]

2.2 Che polivalente

Nell'italiano standard il *che* è utilizzato per introdurre alcune frasi subordinate: soggettive ("è possibile che io sia in ritardo domani"), oggettive ("vedo che ti stai comportando bene"), dichiarative ("su questo siamo d'accordo, che la crisi sta mettendo in ginocchio il paese"), e le relative quando l'antecedente del pronome relativo è un soggetto o un complemento oggetto ("l'uomo che ha preso la parola è suo fratello").

Invece, nell'italiano colloquiale, è diffusa la tendenza a estendere l'uso del *che* come introduttore di subordinate che potrebbero essere introdotte da congiunzioni semanticamente e sintatticamente più precise. In tutti questi casi viene utilizzata l'espressione *che polivalente*.

L'uso del *che polivalente* è riscontrato già nell'italiano antico, spesso trattato come *aferesi*[11] di *perché*,[12] ad esempio nei primi versi della Divina Commedia: «Nel mezzo del cammin di nostra vita / mi ritrovai per una selva oscura / ché la dritta via era smarrita».[13]

Si parla di *che polivalente* nel caso in cui il *che* è utilizzato in particolari contesti:

- **Enfatico** ("Ma che bellini che sono!");
- **Consecutivo-presentativo:** il *che* sostituisce espressioni come *in modo tale, a tal punto, ecc.* ("beh, caro Franco…io sono una **che** ci crede!");
- **Subordinatore generico avverbiale:**
  - **Causale:** il *che* sostituisce espressioni come: *poiché, giacché, perché, ecc.* ("non voglio dire nulla va…**che** è meglio!");
  - **Concessivo:** il *che* sostituisce espressioni come *anche se, benché, nonostante, ecc.* ("Bevo ancora un po', **che** proprio non dovrei");

[10] Alfonzetti, 2002, p. 22.
[11] Soppressione di una vocale o sillaba iniziale.
[12] Treccani alla voce "Che polivalente".
[13] Dante Alighieri, *Inf.* I, 1-3.

    - **Finale:** il *che* sostituisce espressioni come *in modo che, affinché, al fine di, ecc.* ("sono senza dignità ma stanno con il rosario in mano pregando i santi **che** li aiutino");
- **Consecutivo-esplicative:** il *che* sostituisce espressioni come *così, cosicché, sicché, ecc.* ("dormi **che** è meglio");
- **Relativo per i casi indiretti:** il *che* sostituisce l'uso di espressioni quali *preposizione + cui* oppure *preposizione + articolo + quale* ("Perché la Rai **che** pago il canone a detto poco di Gela e Bagheria **che** ha vinto il centro sinistra […]");
- **Relativo con clitico di ripresa:** in questo caso si accompagna a un pronome clitico che sostituisce l'elemento relativizzato;
- **Pseudo relative** ("Li sento **che** ringhiano e piagnucolano perché non mi trovano");
- **Con funzione di avverbio perché causale** ("Non capisco **che** la nostra aviazione abbiano a lasciarli entrare così") o di **congiunzione ad inizio frase**, solitamente seguito da *poi* ("**che** poi vi lamentate che vi legge sempre meno gente…");[14]
- **Relativo-locativo:** il *che* sostituisce espressioni come *sul quale, dove, su cui, ecc*. ("passami il libro che c'è scritto Grammatica");[15]
- **Relativo-temporale:** il *che* sostituisce espressioni come *quando, mentre, nel momento in cui, ecc.* ("maledetto il giorno **che** ti ho incontrato");[16]

[14] Classificazione ripresa da Bagaglini Veronica in "La relativa non-standard su Facebook".
[15] Classificazione ripresa da Filosa Matteo, "Analisi del *che polivalente all'interno dei commenti di Facebook".*
[16] Treccani alla voce "Che polivalente".

3 Raccolta dei dati

L'utilizzo del *che polivalente* è stato analizzato prendendo in considerazione un sotto-campione del corpus CoCIF, realizzato da Veronica Bagaglini. Esso contiene i commenti degli utenti delle pagine Facebook di Fanpage e Repubblica, per un periodo di tempo che va dal 2017 al 2019.

I commenti sono stati estratti utilizzando la versione 3.10.2 di *Facepager*[17], un software sviluppato da Jakob Jünger (Università di Greifswald) e Till Keyling (Ludwig-Maximilians-Universität di Monaco di Baviera) che permette di estrarre i dati provenienti dai vari social media, tra cui Facebook. *Facepager* è disponibile gratuitamente, può essere scaricato da *github.com* e installato sul proprio computer.

Per poter esaminare solo i commenti contenenti il *che* è stato utilizzato *AntConc* [18] un software sviluppato da Laurence Anthony per l'analisi del testo e concordanze. Il software può essere scaricato gratuitamente dal sito web di Laurence Anthony.

Sia per Fanpage che per Repubblica sono stati esaminati nove file testuali, tre per ogni annata, per un totale di 70.093 commenti: 41.870 per quanto riguarda Fanpage, 28.223 per quanto riguarda Repubblica. In particolare, per poter analizzare l'uso non-standard del *che,* sono stati considerati i seguenti casi:

- relativa appositiva oggetto con strategia di ripresa (tag analitico RA o + ripresa);
- subordinata relativa appositiva oggetto diretto (tag analitico RA oi);
- subordinata relativa appositiva genitivo (possessore e complemento di specificazione, tag analitico RA gen.);
- subordinata relativa appositiva altri complementi (argomento, causa, limitazione, mezzo, modo, etc. tag analitico RA ind.);
- subordinata relativa appositiva complemento di luogo (tag analitico RA loc.);
- subordinata relativa appositiva complemento di tempo (tag analitico RA temp.);
- relativa restrittiva oggetto con strategia di ripresa (tag analitico RR o + ripresa);
- subordinata relativa restrittiva oggetto indiretto (tag analitico RR oi);
- subordinata relativa restrittiva genitivo (possessore e complemento di specificazione, tag analitico RR gen.);

[17] https://github.com/strohne/Facepager/releases
[18] https://www.laurenceanthony.net/software/antconc/

- subordinata relativa restrittiva altri complementi (argomento, limitazione, causa, mezzo, modo, etc. tag analitico RR ind.);
- subordinata relativa restrittiva complemento di luogo (tag analitico RR loc.);
- subordinata relativa restrittiva complemento di tempo (tag analitico RR temp.).

3.1 Estrazione e creazione del sotto-campione

Per poter estrarre i casi descritti sopra è stato creato uno *script*[19] nella versione 2.7.16 di *Python*.[20] Tale *script* utilizza le *espressioni regolari*, una notazione algebrica che permette di determinare particolari schemi di stringa. Questi schemi sono chiamati *pattern* e permettono di descrivere le regole che devono possedere le stringhe per poter essere individuate nel testo. Un'espressione regolare caratterizza tutte le stringhe che corrispondono a un particolare schema.[21]

Ogni file testuale, proveniente dal corpus CoCIF, è stato analizzato singolarmente utilizzando la classe *codecs* ed il metodo *open()* di *Python*,[22] collocati nel *main* dello *script*. Il metodo *read()* è utilizzato per leggere il file ed assegnare tutto il suo contenuto ad una variabile di tipo *String*[23]. La funzione *CalcolaRe* è utilizzata per calcolare l'espressione regolare utile all'estrazione dei commenti che contengono l'uso non-standard del *che*, considerando i casi descritti nel paragrafo precedente. La funzione sfrutta il modulo *re* (fornisce le operazioni necessarie per il trattamento delle espressioni regolari) e il metodo *re.findall()*. [24] Quest'ultimo, se non trova corrispondenze, restituisce una lista vuota. In caso contrario restituisce la lista di tutte le sequenze di caratteri che soddisfano l'espressione regolare all'interno della stringa.

L'utilizzo dello *script* ha permesso di ottenere 408 commenti di uso non-standard del *che:* 274 per quanto riguarda Fanpage, 134 per quanto riguarda Repubblica.

3.2 Formazione e caratteristiche del campione

[19] Script 1 in appendice.
[20] https://www.python.org/downloads/release/python-2716/
[21] Lenci, 2016, p.111.
[22] Manuale di Python, https://docs.python.org/2.7/library/codecs.html
[23] Manuale di Python, https://docs.python.org/2.7/library/string.html
[24] Manuale di Python, https://docs.python.org/2.7/library/re.html

Per poter inserire i commenti estratti in un contesto linguistico di riferimento, è stato utilizzato Read-IT.[25]

Read-IT è un software elaborato presso il CNR di Pisa che consente di eseguire analisi linguistiche (informazioni lessicali, morfosintattiche e sintattiche). Disponibile tramite l'interfaccia web, consente di inserire il testo da analizzare in un apposito spazio, per poi far partire l'analisi mediante il tasto dedicato. Una volta conclusa l'analisi, i risultati compaiono divisi in diverse sezioni: *suddivisione in frasi, suddivisione in token, parti del discorso, annotazione, analisi globale della leggibilità e proiezione della leggibilità sul testo.*

Per questo elaborato di laurea è stata presa in esame la sezione *analisi globale della leggibilità,* in particolare sono state analizzate le sue sottosezioni: *profilo di base, profilo lessicale, profilo sintattico.*

Profilo di base

In questa sezione sono presenti le informazioni generali sul testo di analisi:

- numero totale di periodi, ovvero il numero totale di frasi da cui il testo è formato;
- numero totale di parole;
- lunghezza media dei periodi;
- lunghezza media delle parole.

Di seguito sono stati riportati i risultati del campione di Fanpage e di Repubblica:

FANPAGE			
NUMERO TOT. DI PERIODI	NUMERO TOT. DI PAROLE (TOKEN)	LUNGHEZZA MEDIA DEI PERIODI	LUNGHEZZA MEDIA DELLE PAROLE (TOKEN)
274	5985	20,1	4,5

REPUBBLICA

[25] http://www.italianlp.it/demo/read-it/

NUMERO TOT. DI PERIODI	NUMERO TOT. DI PAROLE (TOKEN)	LUNGHEZZA MEDIA DEI PERIODI	LUNGHEZZA MEDIA DELLE PAROLE (TOKEN)
134	3437	24,6	4,6

Profilo lessicale

In questa sezione sono stati analizzati i valori della *densità lessicale* e della *type token ratio*.

La densità lessicale esprime il rapporto tra parole semanticamente piene (unità lessicali, lessemi) e parole funzionali (unità grammaticali). La *type token ratio* invece può essere interpretata come l'indice di ricchezza lessicale di un corpus ed è il risultato del rapporto tra le parole tipo e le unità del vocabolario. Il risultato può variare nel range tra 0 (varietà lessicale nulla) e 1 (alta varietà lessicale). [26]

FANPAGE	
TYPE TOKEN RATIO	DENSITÀ LESSICALE
0,730	0,536

REPUBBLICA	
TYPE TOKEN RATIO	DENSITÀ LESSICALE
0,710	0,559

Inoltre, è stata analizzata la percentuale di lemmi appartenenti al Vocabolario di Base (VdB) e la ripartizione dei lemmi riconducibili al VdB rispetto ai repertori di utilizzo:

FANPAGE
VOCABOLARIO DI BASE
73,2%

[26] Lenci, 2016, p.133.

FONDAMENTALE	ALTO USO	ALTA DISPONIBILITÀ
79,5%	14,3%	6,2%

REPUBBLICA		
VOCABOLARIO DI BASE		
68,1%		
FONDAMENTALE	ALTO USO	ALTA DISPONIBILITÀ
77,5%	15,9%	6,6%

La percentuale dei lemmi appartenenti al *lessico fondamentale*, 79.5% in Fanpage e 77,5% in Repubblica, è indicativa di un testo pensato per essere leggibile da un ampio pubblico di lettori.

Profilo sintattico

La sezione del profilo sintattico è divisa in due parti: la *misura delle categorie morfo-sintattiche,* in cui sono contenute le informazioni sulla quantità di categorie grammaticali presenti nel testo (sostantivi, aggettivi, verbi, ecc.) e la *struttura sintattica a dipendenze.* La *struttura a dipendenze* è suddivisa a sua volta in quattro sottosezioni: *articolazione interna del periodo, articolazione interna della preposizione, misura della profondità dell'albero sintattico e misura della lunghezza delle relazioni di dipendenza.*

Di seguito sono riportati i dati relativi all'*articolazione interna del periodo*:

- numero medio di proposizioni per periodo, ossia il rapporto tra il numero di preposizioni e il numero di periodi, un valore direttamente proporzionale alla complessità del testo a livello sintattico;
- percentuale di proposizioni principali e subordinate, dove un valore più alto di subordinate significa avere un testo più complesso.

I risultati ottenuti sono stati seguenti:

FANPAGE

NUM. MEDIO DI PROPOSIZIONI PER PERIODO	PERCENTUALE PROPOSIZIONI PRINCIPALI	PERCENTUALE PROPOSIZIONI SUBORDINATE
3,285	80,8%	19,2%

REPUBBLICA		
NUM. MEDIO DI PROPOSIZIONI PER PERIODO	PERCENTUALE PROPOSIZIONI PRINCIPALI	PERCENTUALE PROPOSIZIONI SUBORDINATE
3,743	71,7%	28,3%

Per quanto riguarda l'*articolazione interna della preposizione*, di seguito sono riportati i dati relativi al numero medio di parole per proposizione e al numero medio di dipendenti per testa verbale.

FANPAGE	
NUM. MEDIO DI PAROLE PER PROPOSIZIONE	NUM. MEDIO DI DIPENDENTI PER TESTA VERBALE
6,113	2,408

REPUBBLICA	
NUM. MEDIO DI PAROLE PER PROPOSIZIONE	NUM. MEDIO DI DIPENDENTI PER TESTA VERBALE
6,559	2,246

Riguardo alla *misura della profondità dell'albero sintattico* sono riportati i dati relativi alla *media delle altezze massime*, alla *profondità media di strutture nominali complesse* e alla *profondità media di catene di subordinazione.* Queste sezioni forniscono le informazioni relative alla complessità del testo, mediante la ricostruzione dei rapporti di incassamento tra le diverse

proposizioni. La distanza massima tra la radice e una foglia (altezza massima dell'albero) aiuta a capire i livelli di incassamento gerarchico.

FANPAGE		
MEDIA DELLE ALTEZZE MASSIME	PROFONDITÀ MEDIA DI STRUTTURE NOMINALI COMPLESSE	PROFONDITÀ MEDIA DI CATENE DI SUBORDINAZIONE
5,117	1,084	1,140

REPUBBLICA		
MEDIA DELLE ALTEZZE MASSIME	PROFONDITÀ MEDIA DI STRUTTURE NOMINALI COMPLESSE	PROFONDITÀ MEDIA DI CATENE DI SUBORDINAZIONE
6,143	1,117	1,233

I valori della *profondità media di catene di subordinazione,* 1,140 in Fanpage e 1,233 in Repubblica, sono indicativi di un testo di non facile lettura, dal punto di vista sintattico.

Infine, riguardo alla *misura della lunghezza delle relazioni di dipendenza* (calcolata come distanza in parole tra testa e dipendente), sono riportati i dati relativi alla lunghezza media e la media delle lunghezze massime. Questa sezione permette di capire ulteriormente il grado di complessità del testo.

FANPAGE	
LUNGHEZZA MEDIA	MEDIA DELLE LUNGHEZZE MASSIME
3,014	10,040

REPUBBLICA	
LUNGHEZZA MEDIA	MEDIA DELLE LUNGHEZZE MASSIME

3,081	12,321

I valori della *media delle lunghezze massime*, 10,040 in Fanpage e 12,321, confermano i dati della *profondità media di catene di subordinazioni*, indicando un testo non semplice dal punto di vista sintattico.

4 Analisi linguistica: UD-Pipe e Profiling-UD

L'analisi linguistica del *che polivalente* è stata svolta utilizzando due software, UD-Pipe[27] e Profiling-Ud[28], sviluppati presso l'Istituto di linguistica computazionale Zampolli.[29]

In un primo momento è stata utilizzato UD-Pipe per verificare come interpretasse il *che polivalente* dal punto di vista grammaticale e se fosse riconosciuto o meno. Successivamente è stato utilizzato Profiling-UD per ottenere i dati utili allo svolgimento del test di Wilcoxon, impiegato per cercare di rilevare, tramite valori statistici, se la presenza o meno del *che polivalente* comportasse variazioni a livello linguistico, rispetto a frasi contenenti l'uso standard del *che*.

4.1 UD-Pipe

UD-Pipe è una pipeline che permette di effettuare la tokenizzazione, l'etichettatura, la lemmatizzazione e l'analisi delle dipendenze all'interno di un testo. Il software è disponibile online, oppure può essere scaricato su Windows, Linux e OS. È distribuito mediante Mozilla Public License 2.0[30] e sotto licenza CC BY-NC-SA.[31]

Per l'analisi del *che polivalente* è stata utilizzata la versione online *italian-isdt-ud-2.6-200830*, che ha permesso il caricamento dei file di testo mediante l'apposita casella (File di input). È possibile specificare le caratteristiche del File di input mediante tre opzioni:

- *Tokenize plain text*: testo non tokenizzato;
- *CONLL-U*: testo in formato CONLL-U;
- *Horizontal*: testo con una frase per ogni riga (come in questo caso);
- *Vertical*: testo con una parola per ogni riga.

Per quanto riguarda l'output, UD-Pipe permette di visualizzare e salvare sul proprio computer:

[27] http://lindat.mff.cuni.cz/services/udpipe/run.php
[28] http://linguistic-profiling.italianlp.it/
[29] http://www.italianlp.it/
[30] https://www.mozilla.org/en-US/MPL/2.0/
[31] https://creativecommons.org/licenses/by-nc-sa/4.0/

- l'output testuale dell'analisi svolta;
- una tabella in cui ogni token viene analizzato a livello grammaticale;
- un albero di derivazione sintattica per ogni frase presente.

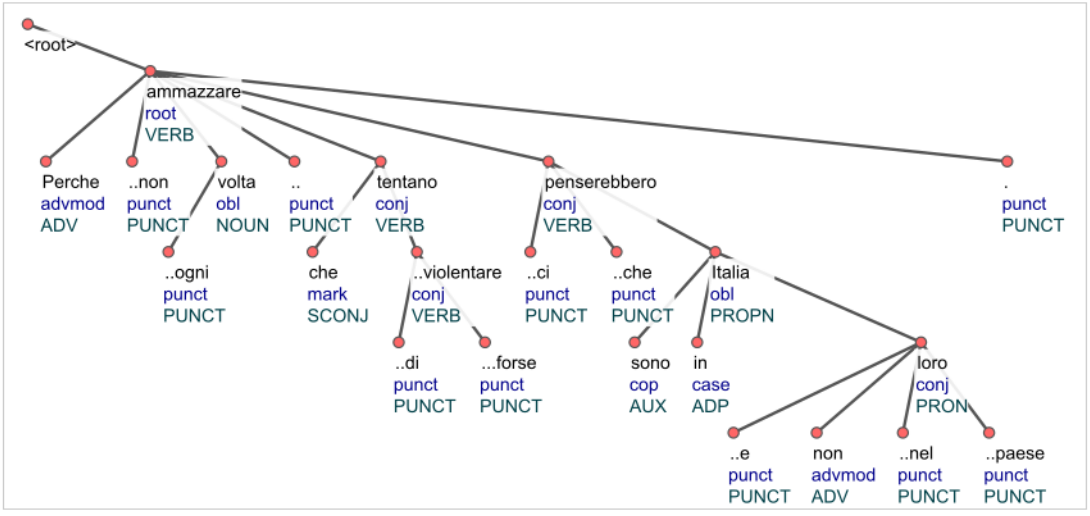
4.2 Prima analisi con UD-Pipe

Come primo passo, dopo aver visto che il software non ha riconosciuto il *che polivalente*, è stato deciso di confrontare due campioni di testo: un campione *normalizzato*[32] e un campione contenente i commenti originali provenienti da Facebook, contenenti 25 frasi ciascuno.

La *normalizzazione* ha interessato: l'eliminazione dei puntini prima del *che*, l'inserimento della punteggiatura dove ritenuto necessario, il ripristino del corretto ordine tra *maiuscolo* e *minusculo* e la correzione di errori evidenti all'interno della frase.

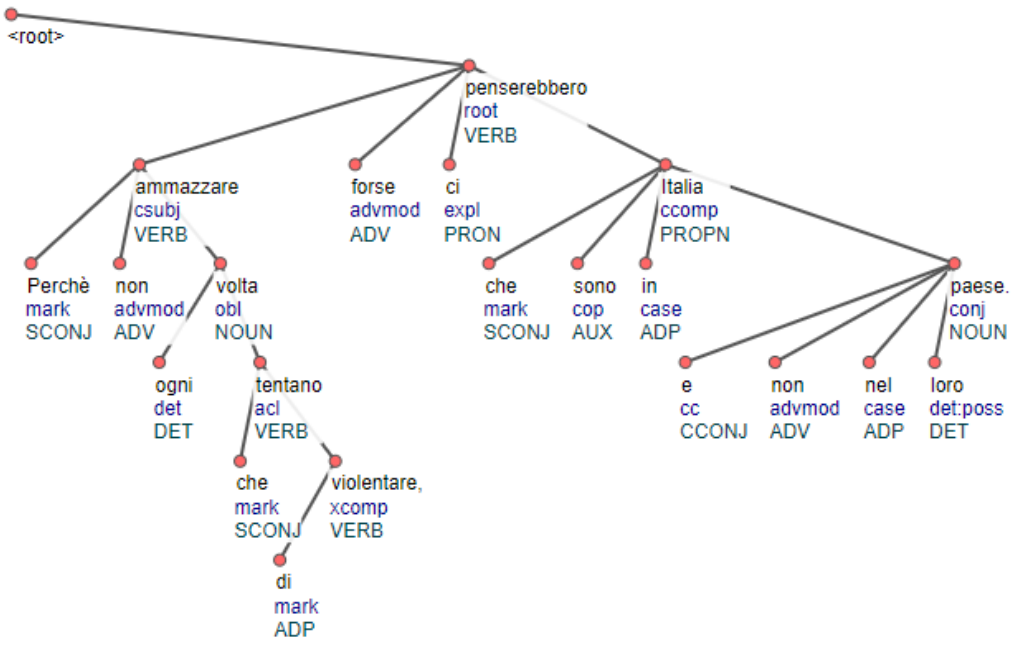
Per ottenere un riscontro sono state prese in esame le 6 frasi, dove il ruolo del *che* a livello grammaticale ha subito un cambiamento, da *pronome* a *congiunzione* o viceversa e l'albero sintattico ha ottenuto una forma diversa:

a) Campione originale: "Perche ..non ammazzare ..ogni volta .. che tentano ..di ..violentare ...forse ..ci penserebbero ..che sono in Italia ..e non ..nel loro ..paese ."



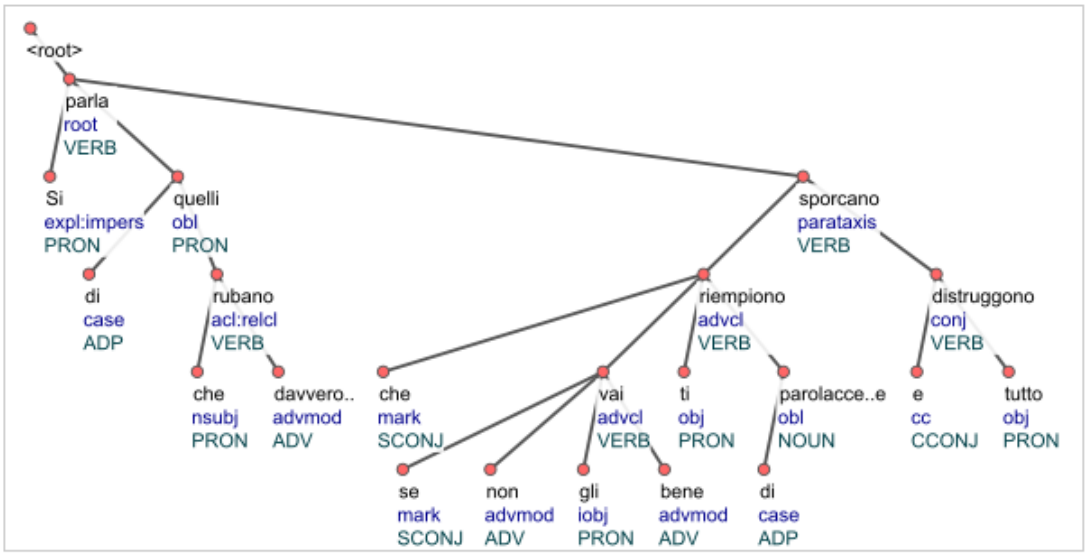
Campione normalizzato: "Perchè non ammazzare ogni volta che tentano di violentare, forse ci penserebbero che sono in Italia e non nel loro paese."

[32] Procedimento volto all'eliminazione della ridondanza informativa e di rischio di incoerenza, al fine di evitare anomalie.



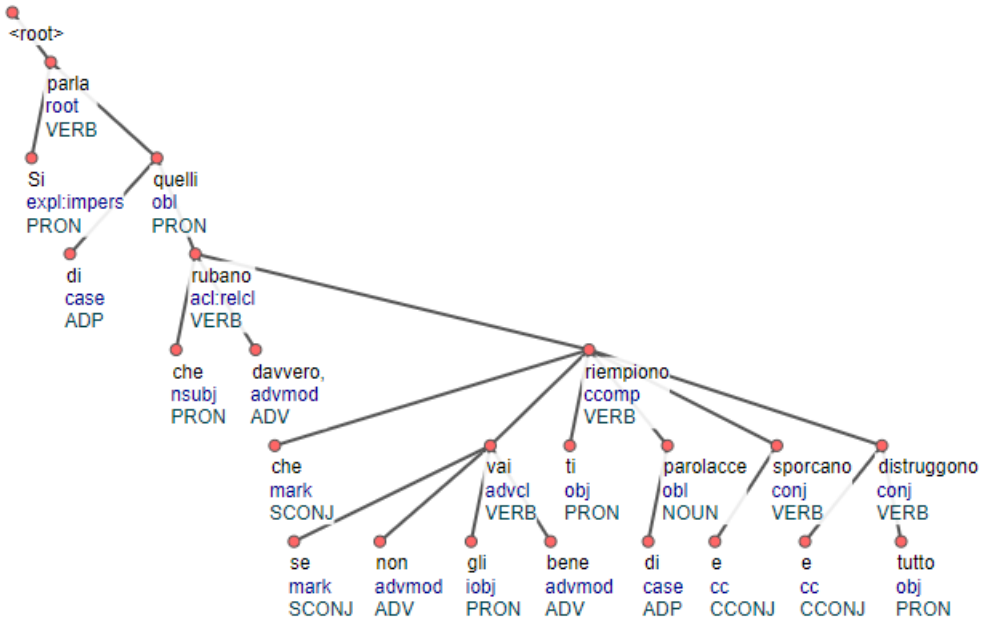
Dopo aver eliminato i puntini, in entrambi casi il *che* viene identificato come congiunzione, cambia completamente anche l'albero sintattico.

b) Campione originale: "Si parla di quelli che rubano davvero.. che se non gli vai bene ti riempiono di parolacce..e sporcano e distruggono tutto"



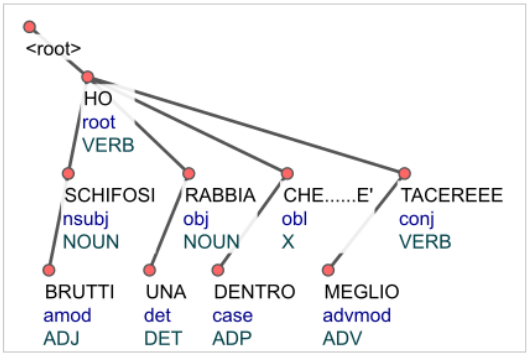
Campione normalizzato: "Si parla di quelli che rubano davvero, che se non gli vai bene ti riempiono di parolacce e sporcano e distruggono tutto"

Aggiunta virgola

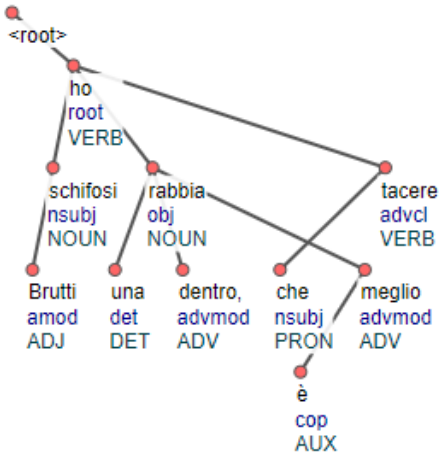


Dopo aver eliminato i puntini, in entrambi i casi il *che* viene identificato come congiunzione e cambia l'albero sintattico.

c) Campione originale: "BRUTTI SCHIFOSI HO UNA RABBIA DENTRO CHE.....E' MEGLIO TACEREEE"

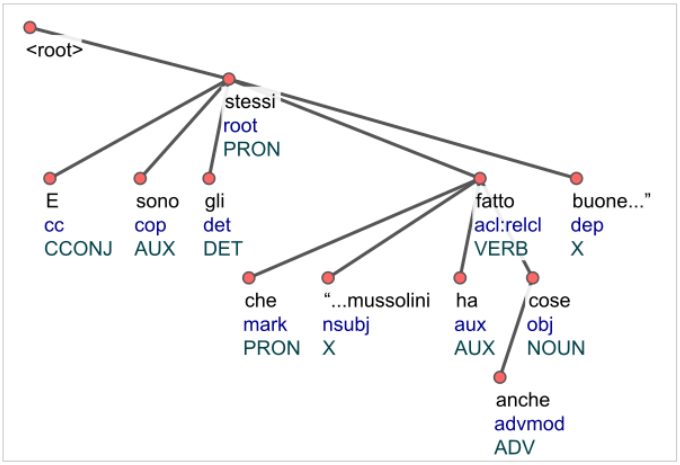


Campione normalizzato: "Brutti schifosi ho una rabbia dentro, che è meglio tacere"

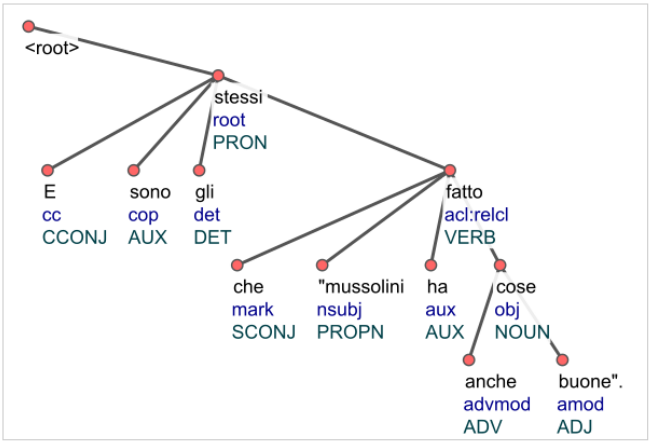


Eliminati i puntini il *che* viene identificato come pronome.

d) Campione originale: “E sono gli stessi che “...mussolini ha fatto anche cose buone...””

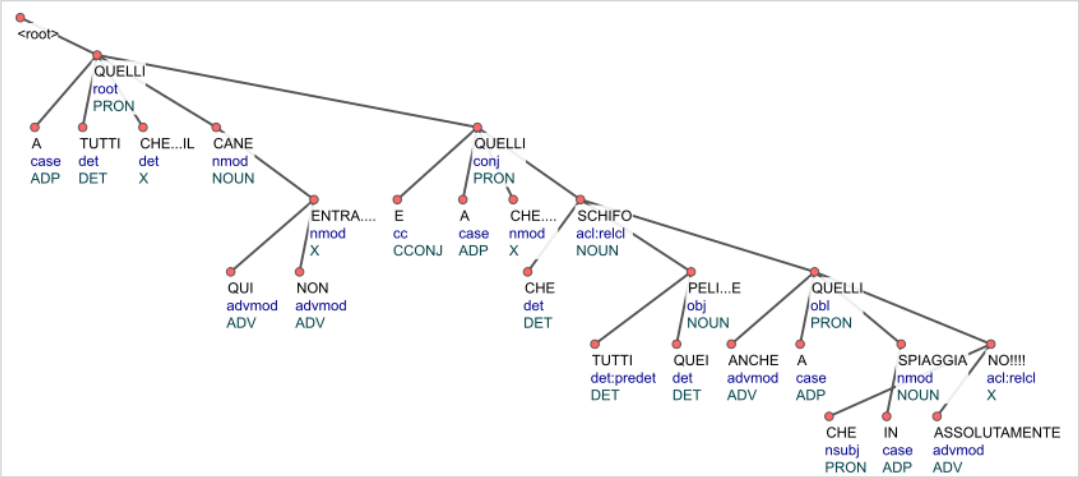


Campione normalizzato: “E sono gli stessi che "mussolini ha fatto anche cose buone".”

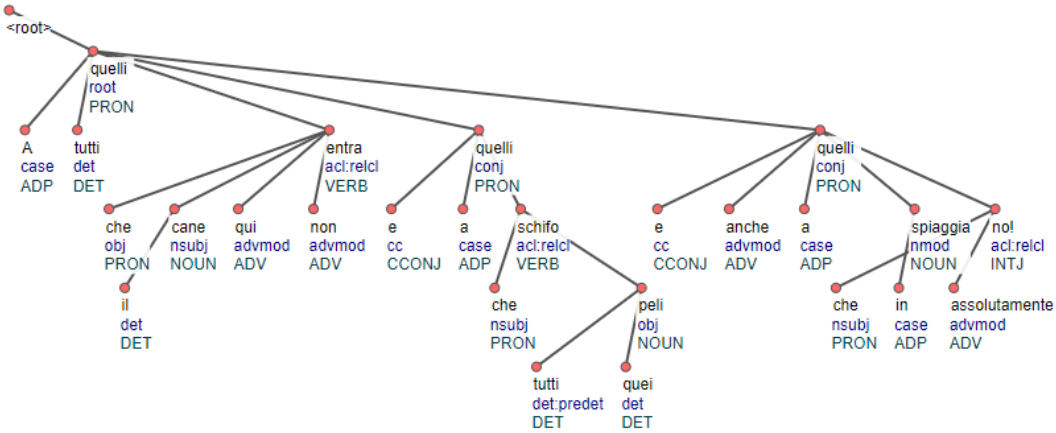


Eliminando i puntini cambia l'albero sintattico e il *che* viene identificato come congiunzione.

e) Campione originale: "A TUTTI QUELLI CHE...IL CANE QUI NON ENTRA.... E A QUELLI CHE.... CHE SCHIFO TUTTI QUEI PELI...E ANCHE A QUELLI CHE IN SPIAGGIA ASSOLUTAMENTE NO!!!!"

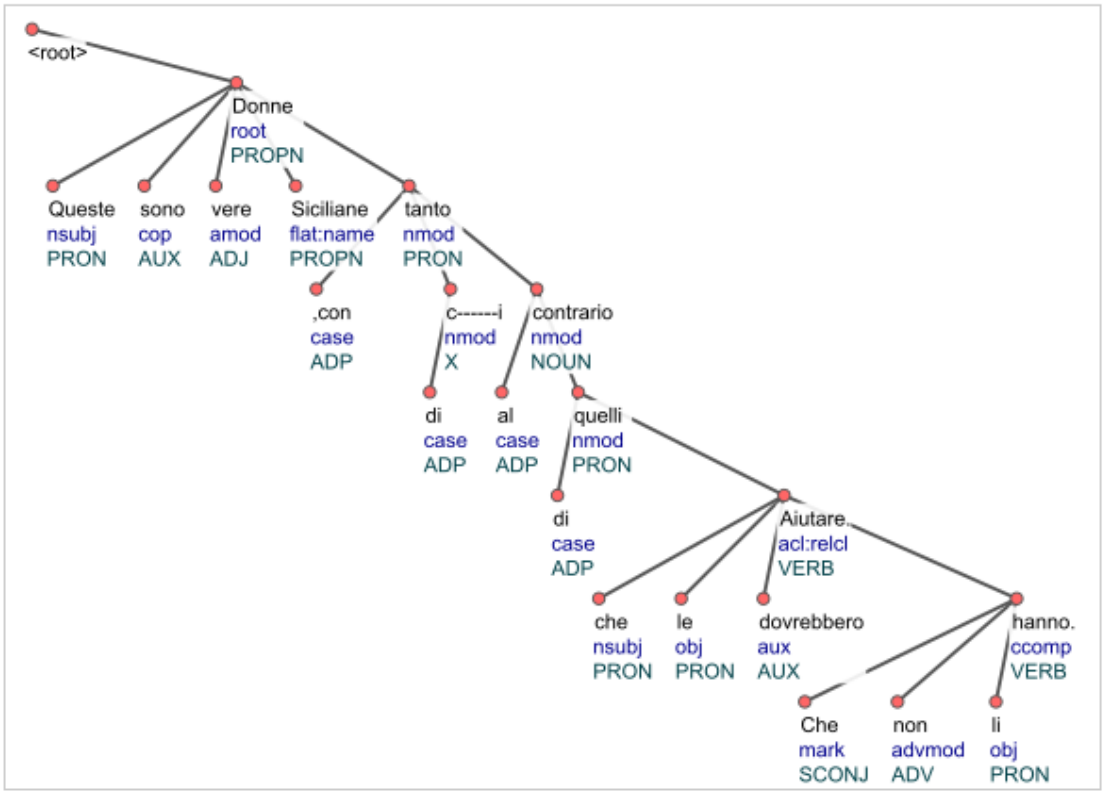


Campione normalizzato: "A tutti quelli che il cane qui non entra e a quelli che schifo tutti quei peli e anche a quelli che in spiaggia assolutamente no!"

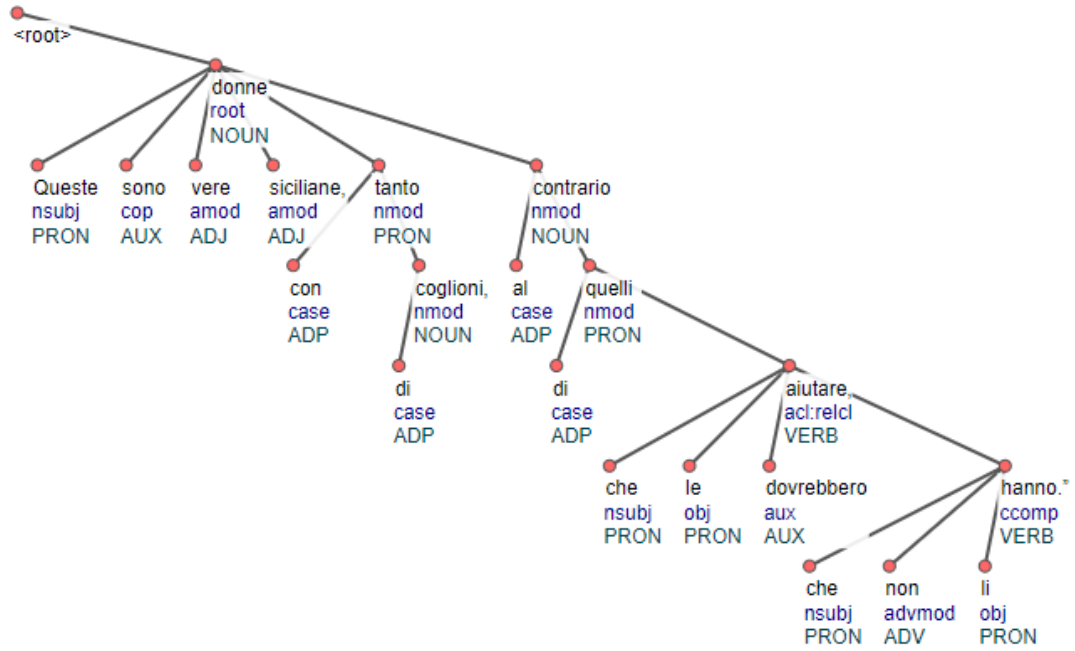


La frase non risulta chiarissima; eliminando i puntini e la ripetizione il *che* viene identificato come pronome.

f) Campione originale: "Queste sono vere Donne Siciliane ,con tanto di c------i al contrario di quelli che le dovrebbero Aiutare. Che non li hanno."



Campione normalizzato: “Queste sono vere donne siciliane, con tanto di coglioni, al contrario di quelli che le dovrebbero aiutare che non li hanno.”



Dopo aver eliminato il punto e aggiunto le virgole, il *che* viene identificato come pronome.

Alla luce dei risultati ottenuti, avendo potuto constatare la diversa identificazione del *che* (pronome o congiunzione) e la diversa struttura dell'albero sintattico, è stato deciso di adottare la normalizzazione su tutto il corpus.

4.3 Seconda analisi con UD-Pipe

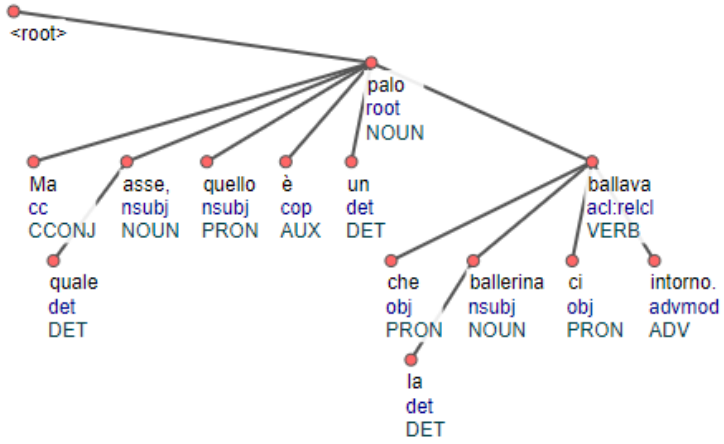
Dopo aver normalizzato il corpus, è stato eseguito un nuovo studio per verificare se vi fossero delle regolarità o meno nel processo di analisi, osservando l'*antecedente* del *che.*

A seguito dell'analisi, il *che* è stato riconosciuto come *pronome relativo, congiunzione subordinante* e *determinante.* Nelle tabelle che seguono sono stati riportati i dati relativi degli antecedenti presenti all'interno del corpus, suddivisi per Fanpage e per Repubblica:

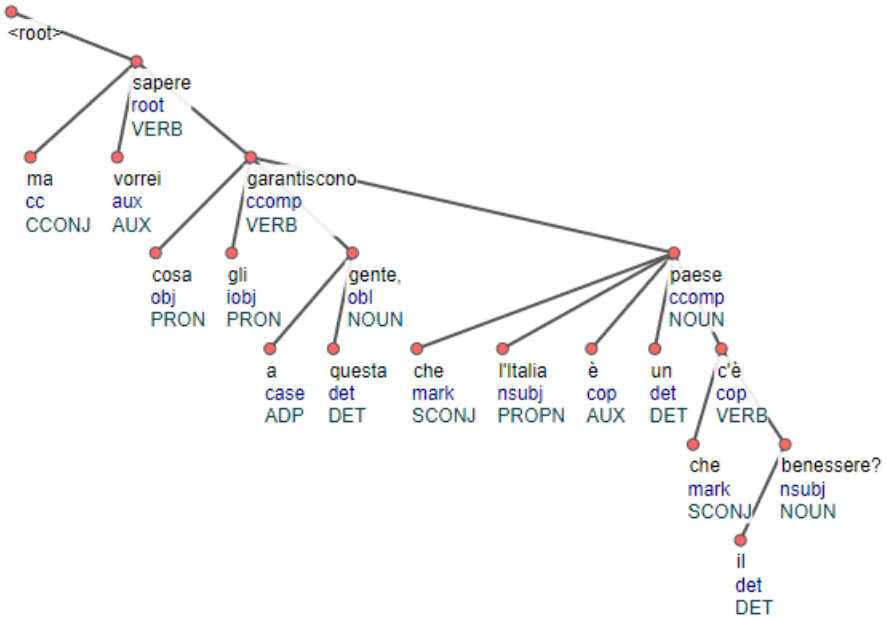
FANPAGE: 274 commenti		
ANTECEDENTE	OCCORRENZE	PERCENTUALE
Nome	211	77%
Aggettivo	21	7,66%
Pronome	21	7,66%
Verbo	4	1,46%
Congiunzione	5	1,82%
Avverbio	9	3,28%
Preposizione	1	0,36%
Numero	1	0,36%

Nei 211 casi in cui l'antecedente era nome il software ha etichettato il *che* come pronome (144 casi) e come congiunzione (67 casi). Di seguito alcuni esempi:

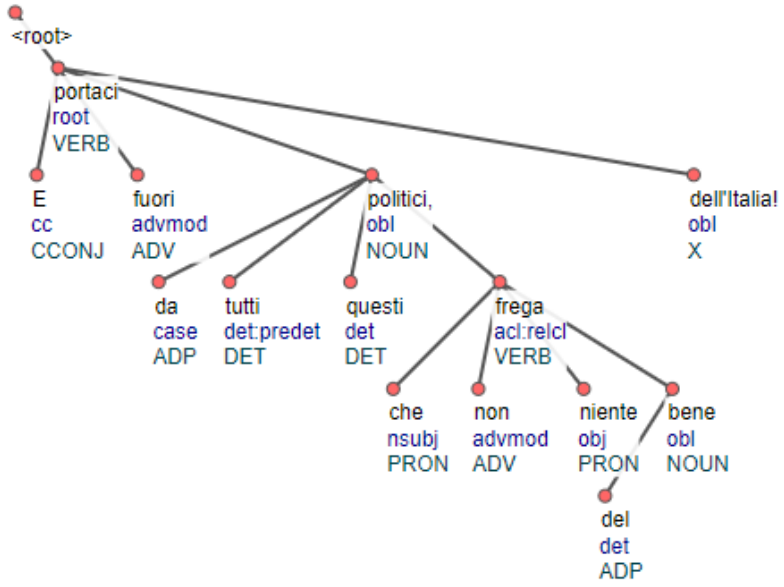
- "Ma quale asse, quello è un palo che la ballerina ci ballava intorno."



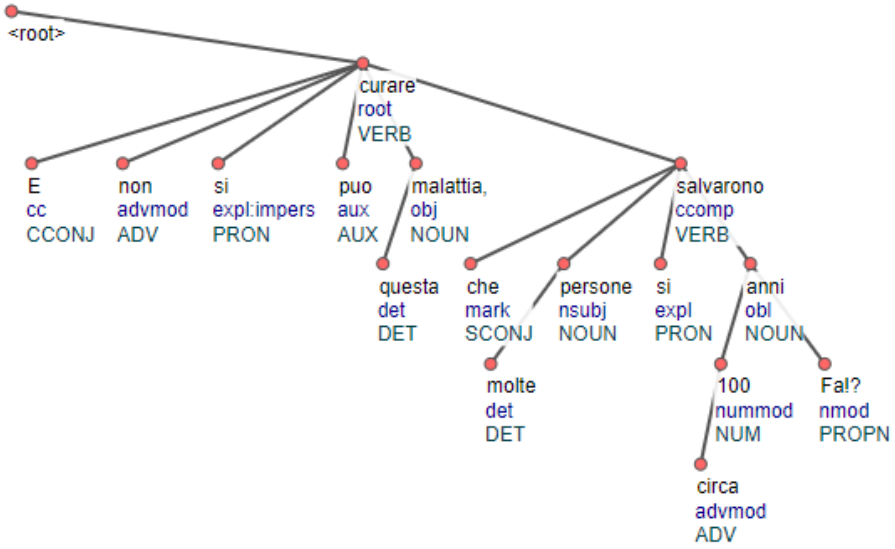
- “ma vorrei sapere cosa gli garantiscono a questa gente, che l'Italia è un paese che c'è il benessere?”.



- “E portaci fuori da tutti questi politici, che non frega niente del bene dell'Italia!”

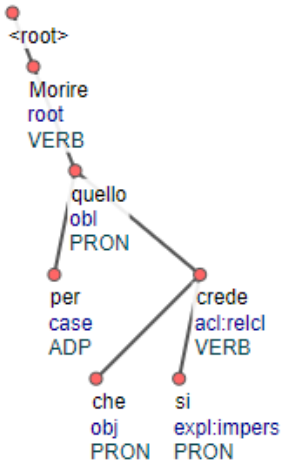


- “E non si puo curare questa malattia, che molte persone si salvarono circa 100 anni Fa!?”

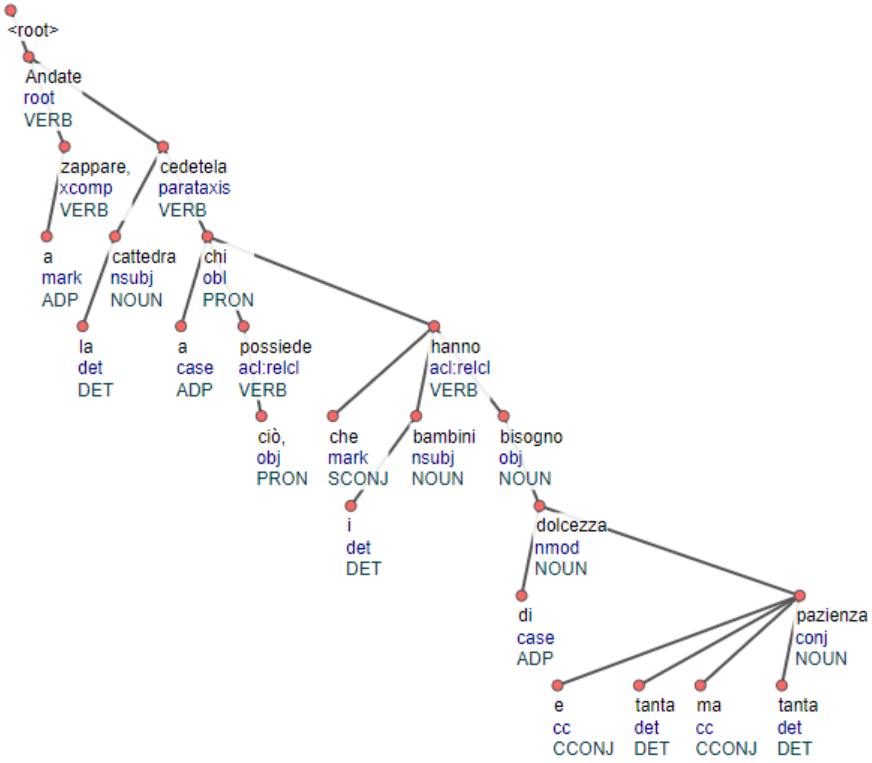


Nei 21 casi in cui l’antecedente era un pronome il *che* è stato identificato come pronome (18 casi) e come congiunzione (3 casi). Di seguito alcuni esempi:

- “Morire per quello che si crede”

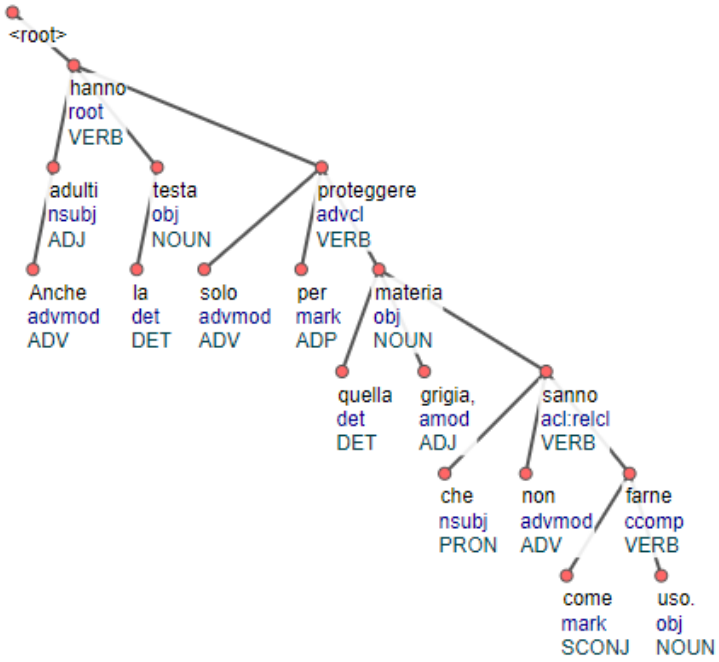


- “Andate a zappare, la cattedra cedetela a chi possiede ciò, che i bambini hanno bisogno di dolcezza e tanta ma tanta pazienza”

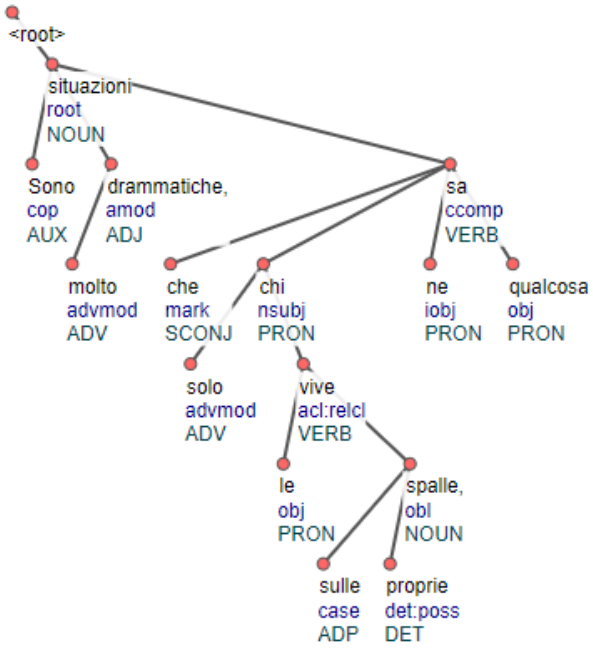


Nei 21 casi in cui l’antecedente era un aggettivo il software ha identificato il *che* come pronome (12 casi) e come congiunzione (9 casi). Di seguito alcuni esempi:

- “Anche adulti hanno la testa solo per proteggere quella materia grigia, che non sanno come farne uso”

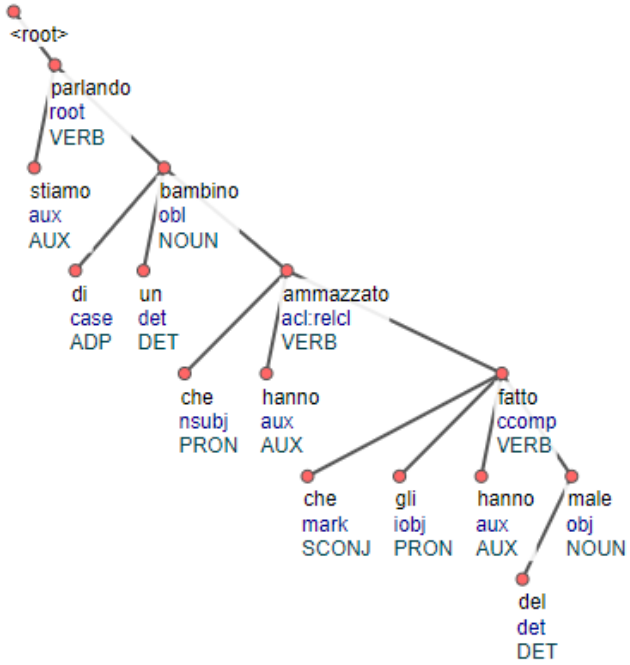


- “Sono situazioni molto drammatiche, che solo chi le vive sulle proprie spalle, ne sa qualcosa”

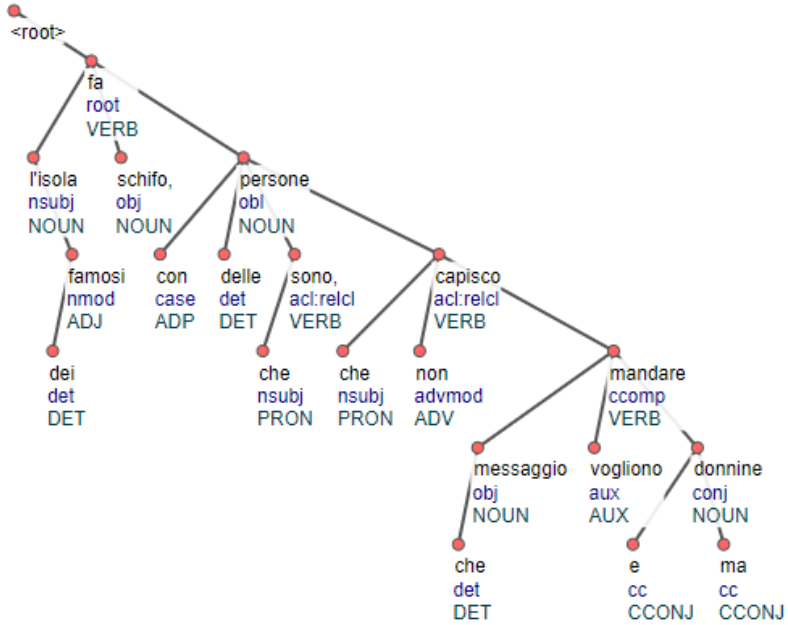


Nei 4 casi in cui l’antecedente era un vero il *che* è stato identificato come congiunzione (3 casi) e come pronome (1 caso). Di seguito alcuni esempi:

- “stiamo parlando di un bambino che hanno ammazzato che gli hanno fatto del male”

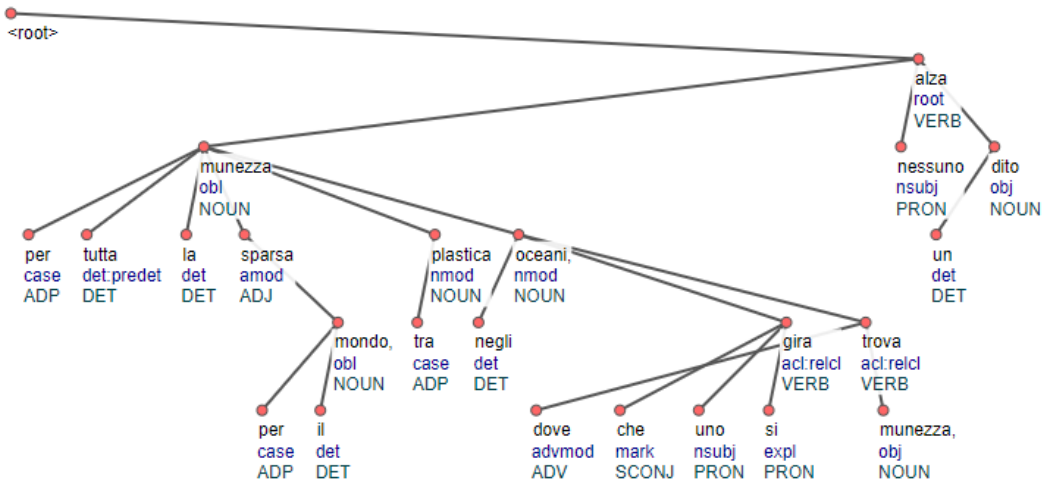


- “l'isola dei famosi fa schifo, con delle persone che sono, che non capisco che messaggio vogliono mandare e donnine ma”

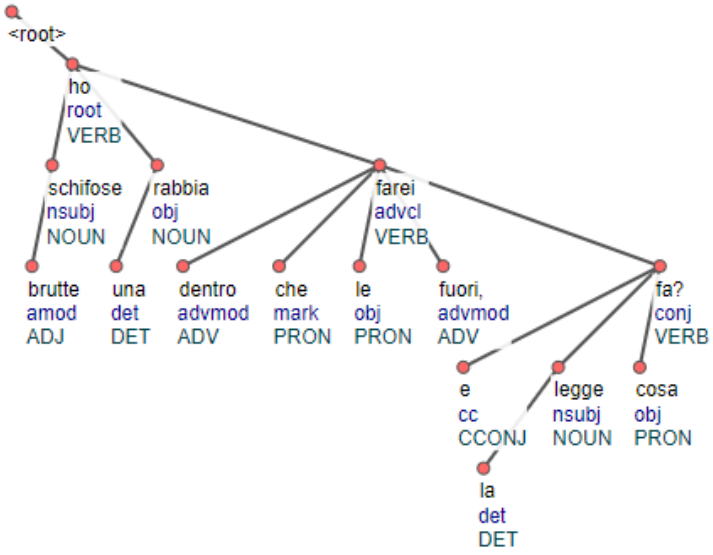


Nei 9 casi in cui l’antecedente era un avverbio il *che* è stato identificato come congiunzione (7 casi) e come pronome (2 casi). Di seguito alcuni esempi:

- “per tutta la munezza sparsa per il mondo, tra plastica negli oceani, dove che uno si gira trova munezza, nessuno alza un dito”

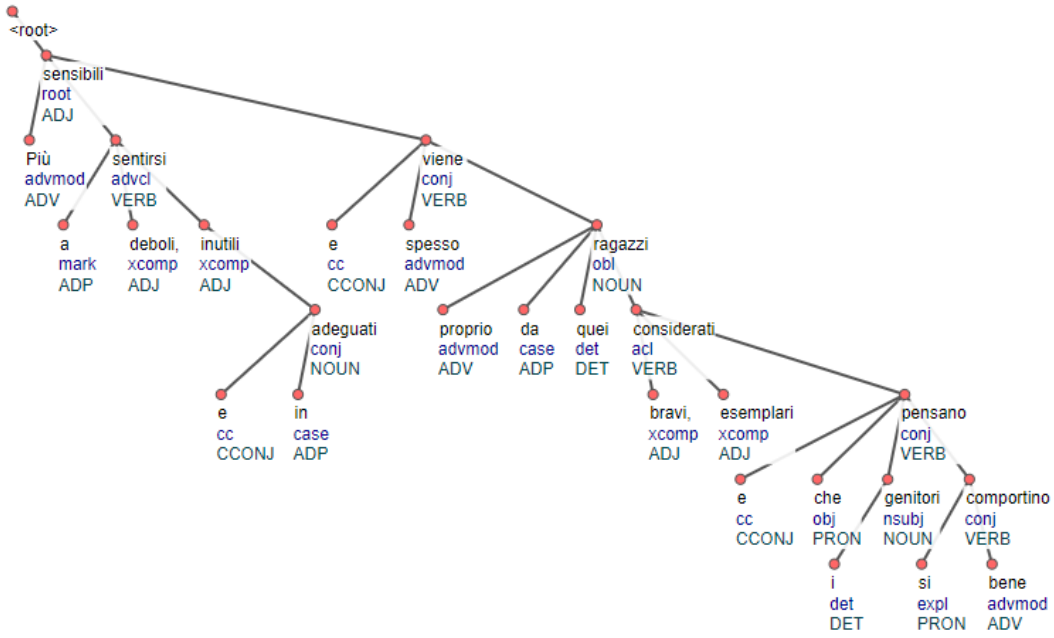


- “brutte schifose ho una rabbia dentro che le farei fuori, e la legge cosa fa?”



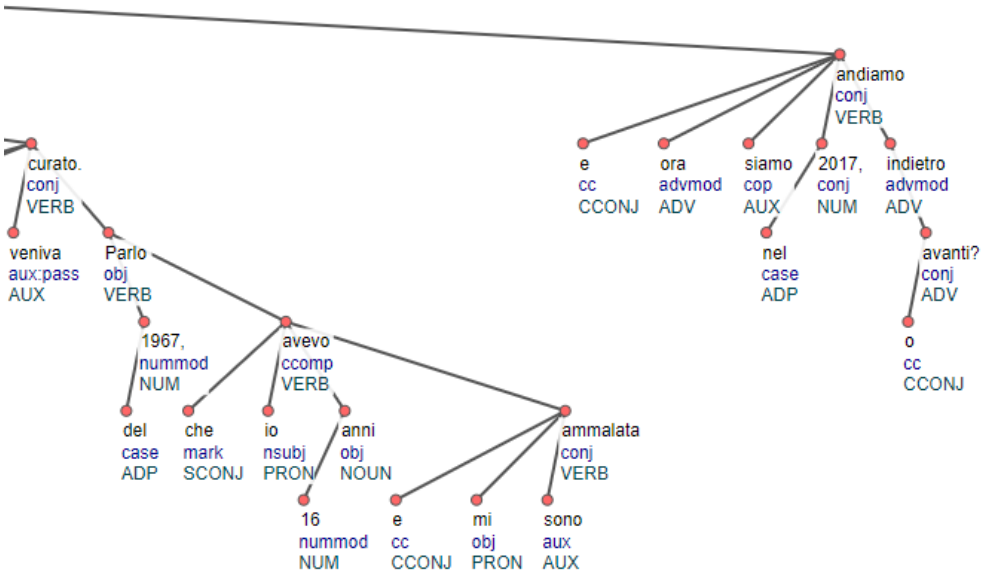
Nei 5 casi in cui l’antecedente era una congiunzione il *che* è stato identificato come pronome. Di seguito un esempio:

- “Più sensibili a sentirsi deboli, inutili e in adeguati e spesso viene proprio da quei ragazzi considerati bravi, esemplari e che i genitori pensano si comportino bene”



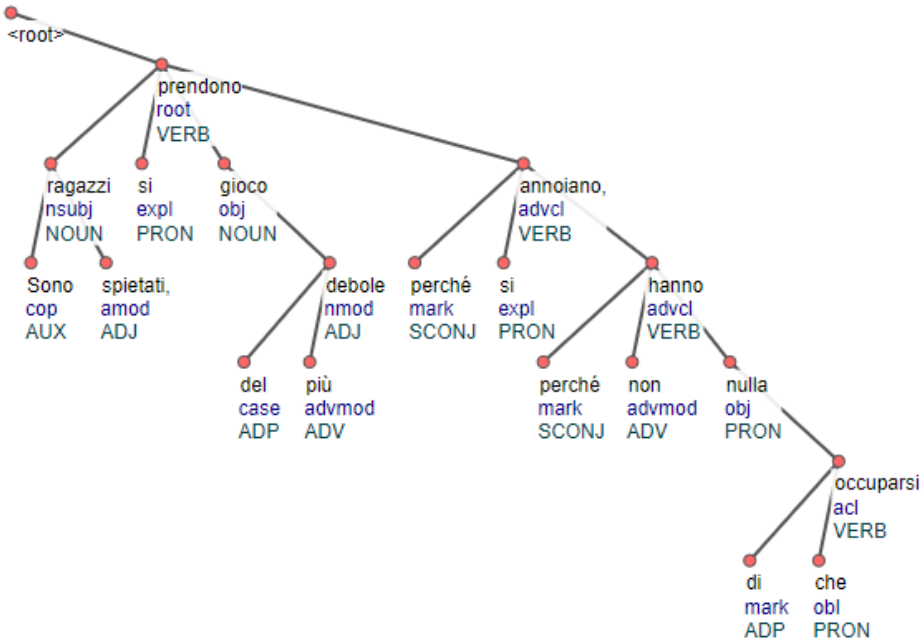
Nel caso in cui l'antecedente era un numero il *che* è stato identificato come congiunzione:

- "Si, io vorrei sapere se mi spiegano perché ai miei tempi non esistevano i vaccini e chi si ammalava veniva curato. Parlo del 1967, che io avevo 16 anni e mi sono ammalata e ora siamo nel 2017, andiamo indietro o avanti?"



Nel caso in cui l'antecedente era una preposizione il *che* è stato identificato come pronome:

- "Sono ragazzi spietati, si prendono gioco del più debole perché si annoiano, perché non hanno nulla di che occuparsi"



REPUBBLICA: 134 commenti		
ANTECEDENTE	OCCORRENZE	PERCENTUALE
Nome	104	77,61%
Aggettivo	7	5,22%
Pronome	14	10,44%
Verbo	2	1,49%
Congiunzione	3	2,23%
Avverbio	3	2,23%
Che a inizio frase	1	0,74%

In Repubblica sono stati riscontrati i seguenti risultati:

- Nei 104 casi in cui l'antecedente era un nome il *che* è stato identificato come pronome (52 casi), congiunzione (50 casi) e determinante (2 casi);
- Nei 7 casi in cui l'antecedente era un aggettivo è stato identificato come pronome (6 casi) e come congiunzione (1 caso);
- Nei 14 casi in cui l'antecedente era un pronome è stato identificato come pronome (10 casi) e come congiunzione (4 casi);
- Nei 3 casi in cui l'antecedente era una congiunzione il che è stato identificato come pronome (2 casi) e come congiunzione (1 caso);

- Nei 2 casi in cui l'antecedente era un verbo è stato identificato come congiunzione (1 caso) e come pronome (1 caso);
- Nei 3 casi in cui l'antecedente era un avverbio è stato identificato come congiunzione (2 casi) e come pronome (1 caso)
- Nel caso in cui il *che* si trovava a inizio frase è stato identificato come congiunzione.

Alla luce dei risultati emersi l'antecedente più ricorrente è il nome, a seguire il pronome e l'aggettivo, in questi casi il *che* viene identificato maggiormente come pronome.

4.4 Profiling-UD

Profiling-UD è un'applicazione web pensata per effettuare la profilazione linguistica di un testo per più lingue. Lo strumento lavora seguendo un processo a due fasi: l'annotazione del testo viene eseguita da UD Pipe, successivamente i testi annotati vengono utilizzati come input per la componente di profilazione linguistica.

Nell'interfaccia web è possibile caricare un file di testo oppure inserire il testo direttamente nello spazio apposito. Prima di eseguire l'analisi è possibile scegliere la lingua del testo di input, l'unità di analisi su cui deve essere eseguito il profilo linguistico (documento o frase) e la volontà di mantenere o meno la segmentazione delle frasi (se presente).

Dopo aver eseguito l'analisi, Profiling UD permette di scaricare tre file diversi:

- Un file in formato *CoNLLU* contenente i risultati della fase di annotazione automatica;
- Un file in formato *csv* contenente i risultati della profilazione linguistica;
- Un file in formato *txt* contenente la legenda delle caratteristiche della profilazione linguistica.

4.5 Analisi con Profiling-UD e test di Wilcoxon

Per verificare se vi fossero delle caratteristiche linguistiche, che variavano in maniera statisticamente significative nell'uso del *che polivalente* all'interno dei commenti, è stato utilizzato Profiling-UD come base di partenza nella raccolta dei dati, successivamente utilizzati per effettuare il test di Wilcoxon.

Le caratteristiche linguistiche estratte da Profiling-UD sono state:

- **N_tokens**: numero di token;
- **Upon_dist_Adj:** distribuzione degli aggettivi;
- **Upon_dist_Noun:** distribuzione dei nomi;
- **Upon_dist_Verb:** distribuzione dei verbi;
- **Lexical_density:** densità lessicale;
- **Avg_max_depth:** valore medio della profondità massima dell'albero estratto da ogni frase del documento;
- **Avg_links_len:** valore medio di parole che occorrono linearmente tra ciascuna testa sintattica e il suo dipendente;
- **Avg_prepositional_chain_len:** valore medio delle catene preposizionali;
- **Obj_post:** distribuzione dei complementi oggetto situati dopo il verbo;
- **Subj_post:** percentuale del soggetto situato dopo il verbo;
- **Avg_subordinate_chain_len:** la lunghezza media delle catene subordinate;
- **Principal_proposition_dist:** distribuzione delle proposizioni principali;
- **Subordinate_proposition_dist:** distribuzione delle proposizioni subordinate.

Il test di Wilcoxon è un test non parametrico utilizzato per verificare se due campioni statistici (non indipendenti) provengono dalla stessa popolazione.[33] In questo elaborato è stato utilizzato per verificare una possibile variazione delle caratteristiche linguistiche tra un campione di 30 commenti contenente l'uso del *che standard* e un campione di 30 commenti contenente l'uso del *che polivalente*.

I campioni utilizzati sono stati i seguenti:

- *Che standard:*

[33] Test di Wilcoxon-Mann-Whitney, in *Wikipedia,* consultato il 28 giugno 2021, https://it.wikipedia.org/wiki/Test_di_Wilcoxon-Mann-Whitney

1. “se voleva suicidarsi non c’era bisogno di fare spettacolo, penso che ci sono anche cose dietro”;
2. “Dice che ci sono meno arrivi di migranti e che sono diminuiti i morti”;
3. “Il bello è che mentre sbarcavano viene a dire che ci sono REGOLE e leggi che devono essere rispettate!”;
4. “Per 65 volte FN è stato processato per ricostituzione del Partito Nazionale Fascista e per 65 volte i giudici hanno stabilito che ciò non era vero!”;
5. “Un certo Giulietto Chiesa avvalendosi di prove, suppone che ciò possa avvenire intorno al 2024, abbiamo bruciato ulteriormente i tempi!?”;
6. “Attaccare le persone di successo sui social, è segnale di sfigatezza, strano che coincida con i sostenitori del Capitone.”;
7. “Credo che combattere questa politica nevrotica e razzista, venga spontaneo agli italiani, avendone già respinta un'altra a calci nel culo, qualche decennio fa”;
8. “Ciò non toglie che proprio perché è Madonna, una che non deve dimostrare nulla a nessuno, e che comunque era una performance vocalmente al di sopra dei suoi standard recenti, non capisco questa mossa.”
9. “Certo è che con questi titoli fuorvianti la Repubblica non fa bella figura ma nemmeno chi commenta solo leggendo il titolo Dell articolo”;
10. “Vedo che con questo governo i disordini sono diminuiti”.
11. “Ognuno della propria vita ne fa ciò che vuole, il mio consiglio è che continui con lo sciopero.”
12. “Credo che dai numeri ce ne eravamo resi conto da tempo.”
13. “ma l'aspetto veramente grave della faccenda è che da parte sua non vengano prese in nessuna considerazione le leggi razziali (naziste e fasciste, non comuniste)”
14. “E pensare che da piccolo lo sceriffo Padano aveva subito un "trauma" perché gli rubarono o perse il suo giocattolo di Zorro”
15. “Ovvio che, da un punto di vista puramente economico, sarebbe una perdita totale visto che l'ascoltano in 4.
16. “io non ho mai fatto il giornalista ma credo che da voi sarei uno dei migliori?”

17. “Pensare che definiamo gli stranieri trogloditi incivili, quando noi siamo retrogradi e ottusi.”
18. “Ma gli americani da Vietnam e Afghanistan non hanno imparato che devono stare zitti e fermi a casa loro eh?”
19. “Non si può negare che di anno in anno il partito Comunista cinese divenga sempre più aggressivo ed ultimamente si sono visti comportamenti da "bullo" per quanto riguarda i tratti di mare contesi e le isole artificiali”
20. “guarda poi che fine ha fatto il Messico.”
21. “vedrete che di meio e Conte si metteranno d'accordo con il buon Salvini per non perdere la poltrona!”
22. “Quello che mi inquieta è che di quel 90% una buona fetta non ha capito nemmeno il titolo?”
23. “Questo mi fa pensare che di solito canti in playback.”
24. “ma vedo che vogliono parlare solo loro e che di tolleranza non capiscono nulla”
25. “Gio Cant qualcuno glielo dice a questa che Diana e Sarah Ferguson erano grandi amiche?”
26. “Ma come è possibile che la gente non capisce che dice solo bugie e di esempi ne abbiamo già diversi”
27. “Un giornale da un titolo fuorviante il 90% della gente abbocca e nonostante un 10% faccia notare che dietro al fatto c'è dell'altro rispetto al titolo la gente continua a prenderlo per vero.”
28. “Ester, vedi che dobbiamo fare”
29. “Quelli che fanno commenti idioti sul fatto che non è una notizia, che non è rilevante, che dobbiamo pensare ai barconi”
30. “Martina Minardi mi sa che dobbiamo tornarci”

- *Che polivalente:*
    1. “Invece siamo contornati da ministri politici, che dirgli incompetenti è fargli un grosso complimento (Lezzi, Taverna, Castelli etc.etc)”;
    2. “A tutti quelli che gli andava bene l'assistenzialismo degli anni scorsi, fatevene una ragione siete i terroni oramai in via d'estinzione”;
    3. “Ma quale asse, quello è un palo che la ballerina ci ballava intorno”;

4. “Dite a Salvini che si dimezzi lo stipendio e faccia restituire i 49 milioni di euro che noi tutti ce li abbiamo messi di tasse”;
5. “L'unica volta che non vedo grande fratello, leggo i commenti qua”;
6. “Ecco a voi la gente che piace applaudire e poi lamentare in continuazione”;
7. “in tanti anni che seguo ballando non avevo mai visto un allievo così bravo, è forte da far volare la maestra”;
8. “Ah già, basta attaccare Salvini che si può dire qualsiasi cosa!”;
9. “Perché non ammazzare ogni volta che tentano di violentare, forse ci penserebbero che sono in Italia e non nel loro paese”;
10. “Dovevano tirare fuori gli attributi la prima volta che ti hanno tirato per la giacca”.
11. “arriverà la resa dei conti anche per voi, che vi definisco dei boia.”
12. “Una vita che abbiamo votato per essere ingannati”
13. “Ci sono le sgallettate che basta che sposano un qualunque con qualche soldo.”
14. “Viviamo in un paese che chi fa queste cose con poco esce.”
15. “Poi un italiano che ci scappa la pipì fanno la multa”
16. “come faranno con i tempi che ci troviamo e lo stato che non aiuta”
17. “Quelli dei governi precedenti hanno fatto solo disastri e danni che ci vorranno decenni affinché si sistemino le cose, grazie a quelli del PD and company!”
18. “Povero bambino quanto ha sofferto in quel poco tempo che è rimasto sulla terra”
19. “Ci hanno marciato in tre ed uniamo la falsa della D'urso che fanno quattro”
20. “stiamo parlando di un bambino che hanno ammazzato che gli hanno fatto del male”
21. “Schifosi genitori devono soffrire come quel piccolo cucciolo di Angelo che hanno distrutto la vita”
22. “lo trovo stupido ed è uno di quei casi che il normale buon senso viene a mancare a chi dovrebbe averlo oltre che per natura per legge”
23. “il mio pensiero d'amore va a quell’angioletto che nella sua breve vita ha solo sofferto e che la gente che gli stava intorno ha solo taciuto”

24. “Quando si toccano i bimbi e gli indifesi mi sale una rabbia, che li ammazzerei con le mie mani”
25. “Proprio in questi giorni che lo stato ricorda il sacrifico dei giudici Falcone e Borsellino, complimenti”
26. “Bruciategli qualcosa che lui tiene tanto”
27. “Già dal primo momento che mi scorrevano i post, ogni post pubblicato ne seguono 3 sull’argomento, avevo posto una domanda: ma queste 2 sono fidanzate?”
28. “adesso non facciamo quelli che nessuno sapeva.”
29. “Veramente stiamo toccando il fondo con questa vicenda ridicola, che non se ne può più”
30. “mi dispiace che sono in Sicilia che non si vota”

Per effettuare il test è stato creato uno *script*[34] in *Python 2.7*, utilizzando la funzione *ranksums*[35] importata dal pacchetto *stats* (funzioni statistiche), contenuto nella libreria *spicy*.[36]

La funzione accetta come input due liste (una struttura dati modificabile, che contiene al suo interno dati uguali o diversi tra loro, ordinati in base ad un indice)[37] e restituisce come risultato una tupla (una struttura dati non modificabile, a differenza della lista)[38] composta da due *float*,[39] in cui il primo elemento è il valore *statistic*, mentre il secondo è il *p-value*.

Il *p-value* permette di capire se la differenza tra il risultato osservato e quello ipotizzato è dovuta alla casualità introdotta dal campionamento, oppure se la differenza è statisticamente significativa, ossia difficilmente spiegabile mediante la casualità dovuta al campionamento.[40] Per capire se i due campioni avessero avuto variazioni statisticamente significative è stato analizzato il *p-value* calcolato dalle caratteristiche linguistiche estratte con Profiling-UD.

Per ogni caratteristica linguistica sono state create due liste, una relativa al campione di *che standard* e una relativa al campione di *che polivalente*, entrambe contenenti i valori estratti con Profiling-UD. Successivamente, tramite l’invocazione della funzione *ranksums*, sono stati calcolati i

[34] Script 2 in appendice.
[35] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ranksums.html
[36] Libreria open-source di algoritmi e strumenti matematici, https://www.scipy.org/
[37] https://www.w3schools.com/python/python_lists.asp
[38] https://www.w3schools.com/python/python_tuples.asp
[39] Identificatore del tipo di dati in virgola mobile.
[40] Valore p, in *Wikipedia*, consultato il 28 giugno 2021, https://it.wikipedia.org/wiki/Valore_p

p-value relativi alle caratteristiche linguistiche. Al fine di estrarre solamente la parte della tupla composta dal valore del *p-value* è stata creata la funzione *EstrazionePvalue*. Essa riceve come parametro la tupla e permette di restituire come risultato solamente la parte relativa al valore del *p-value*, accedendo al secondo indice (tupla[1]). Infine, per poter visualizzare facilmente il *p-value* relativo a ciascuna caratteristica linguistica, è stato creato un dizionario[41] contenente come *chiave* la caratteristica linguistica estratta e come valore il relativo *p-value*. Per poter stampare ogni elemento del dizionario è stata utilizzata la funzione *items()* (predefinita in Python), mediante un ciclo *for*.

Il risultato del test di Wilcoxon è stato riportato nella tabella seguente:

CARATTERISTICHE LINGUISTICHE	P-VALUE
N_tokens	0.13928936568684838
Upon_dist_Adj	0.18576685629821565
Upon_dist_Noun	0.3598852439578021
Upon_dist_verb	0.26256051980781836
Lexical_density	0.2339889162810581
Avg_max_depth	0.9528424295801989
Avg_links_len	0.34028846509891686
Avg_prepositional_chain_len	0.34028846509891686
Obj_post	0.19073033347156776
Subj_post	0.5996894845123704
Avg_subordinate_chain_len	0.9351919702393391
Principal_proposition_dist	0.8708108045360416
Subordinate_proposition_dist	0.7787839689422886

Per le caratteristiche linguistiche studiate non sembrerebbe esserci una variazione statisticamente significativa per i campioni analizzati, non avendo registrato un *p-value* inferiore a 0.05.

[41] https://www.w3schools.com/python/python_dictionaries.asp

5. Distribuzione del *che polivalente* all'interno del campione di analisi

Per ottenere la distribuzione del *che polivalente* sono stati analizzati manualmente i 408 commenti estratti da Fanpage e Repubblica, seguendo la classificazione proposta in 2.2.

La distribuzione è stata la seguente:

USO	OCCORRENZE	PERCENTUALE
Relativo	99	24,26%
Relativo + clitico di ripresa	54	13,23%
Relativo-locativo	52	12,74%
Relativo-temporale	120	29,41%
Causale	39	9,55%
Consecutivo-presentativo	24	5,88%
Concessivo	11	2,69%
Consecutivo-esplicativo	3	0,73%

Di seguito sono riportati alcuni esempi:

- Uso relativo: "Adesso siete tutti liberi e avete tutto lo spazio che avete bisogno";
- Uso relativo + clitico di ripresa: "Ai poveri cani, che li date del valium per tenerli bravi";
- Uso relativo-locativo:" Andatevene a casa vostra, che troverete da mangiare, che qua è finita la pacchia";
- Uso relativo-temporale:" Questa e la conseguenza di come si sono comportati in tutti gli anni che sono stati in Italia;
- Uso causale: "mandatela a casa dalla mamma o da Barbara d'uso, che lei la ama tanto";
- Uso consecutivo-presentativo: "brutte schifose ho una rabbia dentro che le farei fuori, e la legge cosa fa?";

- Uso consecutivo-esplicativo:” Ah già, basta attaccare Salvini che si può dire qualsiasi cosa!”
- Uso concessivo: “Spero con vero cuore che ci sia una via di uscita, che poco ci credo”.

5.1 Confronto con l’analisi di UD-pipe

Nonostante UD-Pipe non riconosca il *che polivalente*, per valutare l’efficacia del software si è tenuto conto dei casi in cui è riuscito ad analizzare:

- Il *che* con uso relativo, relativo + ripresa, relativo-locativo e relativo-temporale come pronome relativo;
- Il *che* con uso causale, consecutivo presentativo, consecuitvo esplicativo e concessivo come congiunzione subordinante.

I risultati ottenuti sono stati riportati all’interno della tabella seguente:

USO	ANALISI CORRETTA	VALORE ASSOLUTO
Relativo	88,88%	88 su 99
Relativo + ripresa	87,03%	47 su 54
Relativo-locativo	61,53%	32 su 52
Relativo-temporale	33,33%	40 su 120
Causale	53,84%	21 su 39
Consecutivo-presentativo	33,33%	8 su 24
Concessivo	36,36%	4 su 11
Consecutivo-esplicativo	33,33%	1 su 3

6. Conclusioni

L'analisi di questo elaborato di laurea è stata svolta seguendo un approccio di tipo linguistico-computazionali. La somma dei risultati dell'analisi della distribuzione del *che polivalente*, all'interno dei commenti di Fanpage e Repubblica, mostra un utilizzo maggiore per quanto riguarda l'uso relativo-temporale e l'uso relativo, in commenti di facile lettura dal punto di vista lessicale ma più complicati dal punto di vista sintattico, come si deduce da Read-IT.

Per commentare ulteriormente i risultati è possibile far riferimento a Mattia Filosa, *Analisi del che polivalente all'interno dei commenti di Facebook*, elaborato di laurea triennale, a.a. 2019-2020:

USO	FILOSA	MANNINI	VARIAZIONE
Relativo	32%	24,26%	-7,74%
Relativo + clitico di ripresa	0%	13,23%	+13,23%
Relativo-locativo	2%	12,74%	+10,74%
Relativo-temporale	0%	29,41%	+29,41%
Causale	32%	9,55%	-22,45%
Consecutivo-presentativo	14%	5,88%	-8,12%
Consecutivo-esplicativo	0%	0,73%	+0,73%
Concessivo	0%	2,69%	+2,69%
Finale	6%	0%	-6%
Enfatizzante-esclamativo	14%	0%	-14%

Come si nota dalla tabella, gli usi più utilizzati del *che polivalente* nell'analisi effettuata da Filosa sono quello relativo e causale, con altre variazioni per quanto riguarda gli altri usi. Questo può essere una conseguenza delle dimensioni del campione di analisi, in Filosa è composto da 50 commenti, mentre in questo elaborato di laurea è composto da 408 commenti.

Per quanto riguarda lo studio a livello statistico, è importante analizzare i risultati ottenuti dal test di Wilcoxon. Per le caratteristiche linguistiche studiate, non è stata registrata una variazione statisticamente significativa rispetto a commenti contenenti l'uso del *che standard*, ciò significa che l'uso del *che polivalente* non ha comportato cambiamenti significativi dal punto di vista linguistico. Ciò può accadere principalmente per due motivi:

1. Le 30 frasi contenenti il *che polivalente* e le 30 frasi contenenti il *che standard* non sono sufficienti per ottenere considerazioni statisticamente significative;
2. Rispetto ai parametri studiati non esiste una variazione tra costruzioni contenenti il *che polivalente* e il *che standard*.

Concludendo, sarebbe interessante sviluppare, dal punto di vista linguistico, una forma di riconoscimento da parte di UD-Pipe per quanto riguarda il *che polivalente*. È un fenomeno linguistico in crescita nelle comunicazioni sul web, per cui sarebbe interessante anche svolgere ulteriori analisi dal punto di vista statistico.

7. Bibliografia

Lenci, Montemagni, Pirelli (2016), *Testo e computer. Elementi di linguistica computazionale*, Roma, Carrocci.

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

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

Durant Alan and Lambrou Marina (2009), *Language and Media*, Routledge.

Crystal David, (2011), *Internet Linguistics,* Routledge.

Bagaglini Veronica, PDF, "La relativa non-standard su Facebook".

Filosa Matteo (2020), *Analisi del che polivalente all'interno dei commenti di Facebook*, elaborato di laurea triennale, corso di laurea in Informatica umanistica, Università di Pisa, anno accademico 2019-2020.

8. Sitografia

Manuale di Python: https://www.python.org/doc/

Python: https://www.python.org/

Treccani, Enciclopedia Italiana: https://www.treccani.it/enciclopedia/

We are social: https://wearesocial.com/it/

UD-Pipe: http://lindat.mff.cuni.cz/services/udpipe/run.php

Profiling-UD: http://www.italianlp.it/demo/profiling-UD/

Read-IT: http://www.italianlp.it/demo/read-it/

W3schools : https://www.w3schools.com/

9. Appendice

Script 1

```
import sys
import re
import codecs

def CalcolaRe(testo):

listaMatch=re.findall(r'\d.*(?:RA\so.*ripresa|RR\sloc.*|RR\sLOC.*|RA\soi.*|RA\sOI.*|\bRA\sgen.*|\bRA\sGEN.*|RA\sind.*|\bra\sind\b|RA\sloc.*|RA\sLOC.*|RA\stemp\.|RA\sTEMP.*|RR\so.*ripresa|RR\soi.*|RR\sOI.*|\brr\soi|RR\sgen.*|RR\sGEN.*|RR\sind.*|RR\stemp\.|RR\sTemp.*|RR\sTEMP.*)', testo)
    if not(listaMatch==[]):
        for match in listaMatch:
            print match

def main(file1):
    #leggo file
    fileInput1=codecs.open(file1, "r")
    raw=fileInput1.read()
    CalcolaRe(raw)

main(sys.argv[1])
```

Script 2

```
Test di Wilcoxon
#coding=utf-8
from scipy.stats import ranksums

#standard
```

ntoken_std =
[20,15,20,28,30,24,30,40,23,10,10,21,12,30,24,22,18,15,21,39,20,20,11,14,17,20,43,6,25,8]

adj_std =
[0,0,0,10.714285,33.33333,4.166666,6.666666,2.5,43478260,0,0,0,0,10,4.166666,9.090909,5.5555
55,26.666666,9.523809,12.820512,5,5,0,0,11.764705,10,6.976744,0,8,0]

noun_std =
[20,6.666666,10,17.857142,6.666666,25,20,7.5,2.173913,20,20,14.285714,25,23.333333,12.5,22.7
27272,5.555555,6.666666,14.285714,20.512820,4.545454,1.111111,15,10,18.181818,7.142857,0,1
5,1.627906,0,16,0]

verb_std =
[20,20,25,7.142857,13.333333,8.333333,16.666666,7.5,17.391304,20,20,19.047619,8.333333,6.66
6666,16.666666,13.333333,9.523809,10.256410,15,15,18.181818,21.428571,5.882352,15,11.62790
6,33.333333,8,25]

lex_std =
[0.611111,0.357142,0.421052,0.444444,0.5,0.476190,0.555555,0.416666,0.695652,0.444444,0.444
444,0.368421,0.363636,0.555555,0.5,0.473684,0.352941,0.538461,0.5,0.567567,0.526315,0.42105
2,0.5,0.428571,0.5,0.6,0.380952,0.6,0.434782,0.5]

avg_max_depth_std = [8,5,8,5,4,5,6,8,7,3,3,4,2,5,7,5,4,4,5,6,4,3,4,3,5,5,6,3,5,2]

avg_links_len_std =
[16.470588,23.846153,19.444444,26.153846,27.142857,2.45,34.615384,5.085714,24.545454,3.25,
3.25,23.333333,2.5,28.076923,2.238095,32.777777,23.125,23.333333,23.157894,30.555555,27.777
777,31.666666,17.777777,26.923076,2.6,2.263157,3.048780,2,26.818181,17.142857]

avg_prep_chainlen_std = [0,0,0,1,0,1,0,0,0,0,0,1,0,1,1,1,1,0,1,1,1,0,1,0,0,0,2,0,1,0]

obj_post_std =
[100,0,0,0,100,100,100,100,100,0,0,50,50,0,6.666666,0,100,100,0,100,100,50,0,100,0,6.666666,100,0,50,50]

subj_post_std = [6.666666,0,50,0,0,0,0,0,0,0,0,0,0,50,0,0,0,0,50,50,0,0,0,50,0,0,0,0,0,0]

avg_subordinate_chainlen_std =
[1,1,2,1,1,1,1,1,1.5,1,1,1,0,1,1,2,1,2,1,1,2,1,1,1,1,2,4,2,16.666666,1]

principal_dist_std =
[25,50,0,50,33.333333,33.333333,33.333333,33.333333,25,50,50,33.333333,100,50,33.333333,0,50,33.333333,50,33.333333,33.333333,33.333333,50,50,50,33.333333,20,0,0,50]

subordinate_dist_std =
[75,50,100,50,6.666666,6.666666,6.666666,6.666666,75,50,50,6.666666,0,50,6.666666,100,50,6.666666,50,6.666666,6.666666,6.666666,50,50,50,6.666666,80,100,100,50]

#polivalente
ntoken_pol =
[24,26,15,25,13,14,24,13,24,16,17,8,14,13,11,14,30,15,16,15,15,28,29,20,18,7,32,8,16,10]

adj_pol =
[0,38.461538,0,0,15.384615,0,8.333333,0,0,6.25,0,0,0,0,0,0,33.333333,6.666666,6.25,0,6.666666,10.714285,34.482758,0,0,0,6.25,0,6.25,0]

noun_pol =
[8.333333,23.076923,20,20,23.076923,14.285714,12.5,76.923076,8.333333,18.75,17.647058,12.5,14.285714,15.384615,27.272727,14.285714,23.333333,20,6.25,13.333333,20,17.857142,1.724137,20,2.777777,0,15.625,0,12.5,0]

verb_pol =
[16.666666,38.461538,6.666666,16,15.384615,21.428571,20.833333,23.076923,16.666666,12.5,11

.764705,25,21.428571,23.076923,18.181818,21.428571,10,13.333333,18.75,20,13.333333,10.7142 85,10.344827,15,5.555555,28.571428,12.5,25,6.25,20]

lex_pol = [0.5,0.416666,0.384615,0.4,0.75,0.538461,0.590909,0.545454,0.5,0.466666,0.466666,0.375,0.3846 15,0.416666,0.545454,0.5,0.464285,0.4,0.375,0.333333,0.466666,0.428571,0.413793,0.368421,0.5 29411,0.428571,0.428571,0.571428,0.466666,0.4]

avg_max_depth_pol = [4,4,3,7,4,5,7,5,4,4,3,3,5,6,4,5,7,5,5,5,5,8,6,5,7,3,5,3,3,3]

avg_links_len_pol = [26.666666,25.652173,2.75,26.666666,2.272727,1.75,23.333333,2,26.666666,22.142857,29.28571 4,18.571428,1.75,3.272727,2.7,19.230769,24.074074,20.714285,17.333333,20.714285,18.571428,2 .074074,29.285714,22.777777,2.625,2.5,25,16.666666,27.142857,23.333333]

avg_prep_chainlen_pol = [0,1,0,1,0,0,0,0,0,0,1,0,0,0,0,0,1,0,1,0,1,1,1,0,1,0,0,0,0,0]

obj_post_pol = [0,50,0,25,100,0,6.666666,6.666666,0,50,50,0,100,50,6.666666,0,50,0,6.666666,0,100,50,0,33.333 333,0,50,50,50,50,0]

subj_post_pol = [0,0,0,0,0,0,0,0,0,0,0,0,50,0,0,0,33.333333,0,0,0,0,0,25,33.333333,0,0,0,0,0,0]

avg_subordinate_chainlen_pol = [2,1,1,1.5,1,2,2.5,1,2,1,1,2,2,1,1,1,1,1,1,2,1,1.5,1,1,1,1,2,1,0,1]

principal_dist_pol = [33.333333,0,50,25,50,0,16.66666,33.333333,33.333333,50,50,0,33.333333,33.333333,50,33.3333 33,33.333333,0,50,33.333333,50,25,33.333333,50,0,50,33.333333,50,100,50]

subordinate_dist_pol = [6.666666,100,50,75,50,100,8.333333,33.333333,33.333333,50,50,100,6.666666,6.666666,50,6.66 6666,6.666666,100,50,6.666666,50,75,6.666666,50,100,50,6.666666,50,0,50]

```
#Funzione che estrae il secondo elemento della tupla passata come parametro
def EstrazionePValue(tupla):
    if tupla:
        return tupla[1]
def main ():
    #Calcolo tutte le tuple
    ranks_tok = ranksums(ntoken_pol,ntoken_std)
    #PoS
    ranks_adj = ranksums(adj_pol,adj_std)
    ranks_noun = ranksums(noun_pol,noun_std)
    ranks_verb = ranksums(verb_pol,verb_std)
    #avg_max_depth_std
    ranks_maxDepth = ranksums(avg_max_depth_pol,avg_max_depth_std)
    #avg_prep_chainlen_std
    ranks_prepChainLen = ranksums(avg_prep_chainlen_pol,avg_prep_chainlen_std)
    #avg_subordinate_chainlen_std
    ranks_subChainLen = ranksums(avg_subordinate_chainlen_pol,avg_subordinate_chainlen_std)

    ranks_lexD = ranksums(lex_pol, lex_std)
    ranks_avgLinkL = ranksums(avg_links_len_pol,avg_links_len_std)
    ranks_objPost = ranksums(obj_post_pol,obj_post_std)
    ranks_subjPost = ranksums(subj_post_pol,subj_post_std)
    ranks_principal = ranksums(principal_dist_pol,principal_dist_std)
    ranks_subordinate = ranksums(subordinate_dist_pol,subordinate_dist_std)

    #Estraggo tutti i pValue

    pValue_nToken = float(EstrazionePValue(ranks_tok))
    pValue_adj = float(EstrazionePValue(ranks_adj))
    pValue_noun = float(EstrazionePValue(ranks_noun))
    pValue_verb = float(EstrazionePValue(ranks_verb))
```

```
    pValue_mDepth = float(EstrazionePValue(ranks_maxDepth))
    pValue_prepChainLen = float(EstrazionePValue(ranks_prepChainLen))
    pValue_subChainLen = float(EstrazionePValue(ranks_subChainLen))
    pValue_lexD = float(EstrazionePValue(ranks_lexD))
    pValue_avgLinkL = float(EstrazionePValue(ranks_avgLinkL))
    pValue_objPost = float(EstrazionePValue(ranks_objPost))
    pValue_subjPost = float(EstrazionePValue(ranks_subjPost))
    pValue_principal = float(EstrazionePValue(ranks_principal))
    pValue_subordinate = float(EstrazionePValue(ranks_subordinate))
    #Creo il dizionario con i pValue
    dizionario = dict(Num_Token=pValue_nToken, Distribuzione_Aggettivi=pValue_adj,
Distribuzione_Nomi=pValue_noun,
Distribuzione_Verbi=pValue_verb,Dens_Lessicale=pValue_lexD,
Avg_MaxDepth=pValue_mDepth, Avg_PrepChainLen=pValue_prepChainLen,
Avg_Links=pValue_avgLinkL, Obj_Post=pValue_objPost, Subj_Post=pValue_subjPost,
Avg_SubChainLen=pValue_subChainLen, Prop_Principali=pValue_principal,
Prop_Subordinate=pValue_subordinate)
    print("Risultati Test Wilcoxon:")
    #Stampo tutti i valori
    for x, y in dizionario.items():
        print(x, y)
main()
```