Rassegna Italiana di Linguistica Applicata

BULZONI EDITORE

Anno LIII

Settembre-Dicembre 2021/3
ISSN 0033-9725

Rassegna Italiana di Linguistica Applicata

Quadrimestrale di ricerca linguistica e glottodidattica
Anno LIII

3/2021

BULZONI EDITORE

ISSN 0033-9725

00185 Roma, via dei Liburni, 14
http://www.bulzoni.it
e-mail: bulzoni@bulzoni.it

Sommario

EDITORIALE

SEZIONE MONOGRAFICA
Nuovi repertori dei linguaggi giovanili: voci e scritture

a cura di
Johanna Monti, Francesca Chiusaroli, Maria Pia di Buono, Maria Laura Pierucci

SEZIONE MISCELLANEA

RECENSIONI

Dominique Brunato
Istituto di Linguistica Computazionale "A. Zampolli"

Felice dell'Orletta
Istituto di Linguistica Computazionale "A. Zampolli"

Andrea Mattei
Università di Pisa

ANALISI DELLA SCRITTURA GIOVANILE DA UNA PROSPETTIVA LINGUISTICO-COMPUTAZIONALE: IL CASO DI STUDIO DELLA FANFICTION

Abstract

This paper presents a study aimed at characterizing the linguistic style of an emerging literary genre of the web, particularly appreciated by teens and young adults: fanfiction. By relying on Natural Language Processing approaches, and in particular on the methodology of linguistic profiling applied to a novel corpus of Italian fanfiction stories inspired by the fantasy saga "Harry Potter", we investigate the relationship between linguistic style and 'success', measured in terms of number of reviews obtained by the readers. We show that it is possible to detect a set of features, among a wide set of linguistic ones modeling lexical, morpho-syntactic and syntactic phenomena, which help more in discriminating between 'successful' and 'unsuccessful' fanfics.

1. Un approccio stilometrico basato sulle tecnologie del linguaggio per caratterizzare il genere della fanfiction

Negli ultimi anni, la crescente disponibilità di strumenti e tecnologie di Trattamento Automatico del Linguaggio (TAL) in grado di ricostruire il profilo linguistico di un testo in maniera articolata e affidabile, anche in relazione a corpora di ampie dimensioni, ha contribuito a richiamare l'attenzione della linguistica tradizionale verso i metodi statistico-quantitativi, stimolando un rinnovato interesse per i temi classici della disciplina affrontata dalle sue molteplici prospettive. La sociolinguistica è indubbiamente tra i settori che hanno accolto con più favore l'uso di approcci linguistico-computazionali, come dimostra la recente introduzione

del termine *Computational Sociolinguistics*, che identifica "an emerging research field that integrates aspects of sociolinguistics and computer science in studying the relation between language and society from a computational perspective" (Nguyen *et al*. 2016). Metodologie di analisi di corpora testuali basate sulla ricostruzione del profilo linguistico del testo tramite strumenti di TAL fanno oggi da sfondo a studi che si propongono di descrivere le varietà d'uso della lingua, da quelle più tradizionali alle diverse forme di Comunicazione mediata dal computer, così come lo stile di un autore o di gruppi di autori accomunati da fattori biologici, psicologici e sociologici quali l'età, il genere, la provenienza geografica (van Halteren 2004; Montemagni 2013; Daelemans 2013).

Il presente contributo si inserisce in questo più ampio contesto di ricerca e propone un caso di studio volto a caratterizzare lo stile di scrittura di un genere narrativo emergente, particolarmente amato e sperimentato dai giovani, che appare tra i più rappresentativi dei generi legati alla diffusione di Internet: la *fanfiction*. Come suggerisce il nome, tale genere identifica "l'insieme delle produzioni narrative scritte dai fan di un'opera appartenente al mondo letterario, cinematografico, televisivo o di qualsiasi altra natura, prendendo spunto dalle storie o dai personaggi di un lavoro originale" (Calabrese, Conti 2019: 7). Sebbene precedenti all'avvento dell'era digitale, tali produzioni oggi trovano diffusione principalmente in rete e sono fruibili all'interno di una determinata comunità *fandom* (Sindoni 2015). In particolare, in questo lavoro si illustreranno i risultati di un'indagine di monitoraggio linguistico del testo che ha preso in considerazione un ampio corpus rappresentativo di questo genere, composto da più di 16mila racconti pubblicati online e ispirati alla saga fantasy "Harry Potter", con l'obiettivo di studiare quanto, e in che modo, lo stile linguistico di una storia influisca sul grado di apprezzamento da parte del lettore.

Come verrà approfondito nel paragrafo 4, l'indagine è stata condotta seguendo un approccio metodologico innovativo – ispirato al modello dell'analisi multidimensionale sviluppatasi originariamente nel contesto della linguistica dei corpora (Biber 1993; 1998) – che pone al centro dell'analisi il testo linguisticamente annotato tramite una catena di moduli di TAL allo stato dell'arte per la lingua italiana, allo scopo di consentirne l'estrazione di un vasto spettro di caratteristiche linguistiche identificative delle tendenze stilistiche sottostanti. La scelta delle caratteristiche oggetto di monitoraggio e di valutazione in relazione al loro coinvolgimento nel 'successo' di una storia, è stata motivata dalla portata informativa che tali caratteristiche hanno già dimostrato di avere in un'ampia varietà di scenari teorici e applicativi, accomunati dall'interesse a modellare fenomeni relativi alla forma linguistica del testo, piuttosto che al suo contenuto. Tra questi

possiamo citare: la valutazione della complessità linguistica nel sistema (inteso, in senso ampio, come lingua, genere o registro testuale) con l'obiettivo, ad esempio, di automatizzare misure di leggibilità del testo (Collins-Thompson 2014), e della complessità rispetto alla percezione del lettore (Brunato *et al.* 2018); la caratterizzazione dell'interlingua a partire dall'identificazione di tratti linguistici tipici della lingua materna all'interno dei testi prodotti da parlanti/scriventi in una seconda lingua (Malmasi *et al.* 2017); la discriminazione tra testi scritti da donne e da uomini in base a caratteristiche stilometriche (Herring, Paolillo 2006); o, ancora, lo studio del processo di apprendimento e di evoluzione delle abilità di scrittura in apprendenti la lingua a diversi livelli di scolarità (Miaschi *et al*. 2021; Weiss, Meurers 2019).

In quanto segue viene innanzitutto presentato il corpus di storie che fa da sfondo a questo studio, descrivendone le diverse fasi di raccolta e i criteri di selezione dei racconti in base agli scopi della ricerca. Viene poi introdotta la metodologia di ricostruzione del profilo linguistico del testo adottata per le analisi dei dati e infine discusse le caratteristiche linguistiche risultate maggiormente correlate con il successo ottenuto da una *fanfic*.

2. Il corpus di partenza

I testi del corpus su cui è stata condotta la nostra ricerca sono stati raccolti da efpfanfic.net, un portale attivo dal 2001 che permette agli utenti di pubblicare e commentare racconti amatoriali in lingua italiana. Il sito contiene due macro-sezioni: una per racconti originali, divisi per genere letterario, l'altra dedicata alle *fanfiction*. Queste ultime sono distinte in base al medium dell'opera di riferimento e secondo ogni specifica proprietà intellettuale. I racconti sono pubblicati in capitoli, ciascuno caricato singolarmente su una pagina web contrassegnata da un ID unico per quel capitolo. All'inizio di ogni pagina sono disponibili vari metadati sul capitolo in questione, tra cui: titolo, nickname dell'autore, data di pubblicazione, numero di recensioni. Inoltre, è presente un quadrato colorato in verde, giallo, arancio o rosso che rappresenta il *rating* della storia, ovvero una stima data dall'autore sulla potenziale conflittualità delle tematiche trattate e la crudezza delle scene rappresentate. Se la storia è composta da più di un capitolo, è possibile navigare tra i capitoli tramite un menù a tendina. Il testo dei racconti è racchiuso in una sezione all'interno della quale gli scrittori possono modificare il codice html, per assicurarsi maggior libertà nell'impaginazione e nello stile di formattazione. Gli utenti della piattaforma possono recensire i vari capitoli, lasciando un commento e una

valutazione, che può essere positiva, negativa o neutra. Le recensioni possono essere visualizzate divise per capitolo, o complessivamente per ogni storia.

Per l'obiettivo della nostra ricerca, che intende monitorare l'effetto dello stile linguistico nel successo di una *fanfic*, si è deciso di raccogliere e analizzare racconti di utenti ispirati a una stessa opera letteraria, così da limitare quanto possibile la variabilità di contenuti. La scelta è ricaduta sulla saga *fantasy* Harry Potter della scrittrice britannica J. K. Rowling. Trattandosi infatti di un romanzo molto popolare tra i giovani, si è potuto raggiungere un campione molto numeroso, sia in termini di testi analizzabili (pari a circa 55mila storie), sia in termini di utenti potenzialmente interessati a leggere e votare questi testi. Per l'estrazione dei testi è stata effettuata un'operazione di web scraping, scrivendo in Python due *crawler* – programmi informatici (detti anche *spider*) che esplorano il web alla ricerca di informazioni utili per la costruzione di grandi archivi o per la realizzazione o l'aggiornamento di un motore di ricerca – che si appoggiano sul framework *open source* Scrapy2[1]. Il primo *spider* scorre l'elenco delle storie e scarica il loro primo capitolo, insieme a una serie di metadati su di esse, fra cui la lista degli ID dei capitoli successivi. L'ID del primo capitolo è stato utilizzato come riferimento per raggruppare racconti appartenenti alla stessa storia. Le informazioni vengono salvate in un file .json, che viene letto dal secondo *spider* per trovare i link di tutti i capitoli 'figli'. I due *spider* sono fondamentalmente identici in termini di dati estratti, in modo da mantenere una struttura dati coerente. Nel dataset così creato, il record corrispondente a un capitolo comprende: ID ed ID di riferimento, ossia dei codici numerici assegnati dal sito per identificare la pagina web di ogni testo pubblicato; Titolo; Rating; Data di pubblicazione; Nickname dell'autore; Numero di capitoli della storia; Testo del racconto; Numero di recensioni ricevute in totale dalla storia, divise in positive, critiche e neutre; Numero di recensioni ricevute dal singolo capitolo in questione; Testo delle recensioni più recenti.

I *crawler* hanno scaricato 54.717 storie, per un totale di 197.310 singoli capitoli, con una media di circa 3,6 capitoli per storia calcolata sul totale del sito. Il corpus scaricato ha una dimensione pari a circa 426 milioni di token, con una media di 2.162 token per capitolo. I dati sono stati salvati in due file .json: il primo (di dimensione pari a 533 MB) contenente le

[1] Dal momento che il corpus non può essere reso liberamente disponibile per le politiche di privacy del sito, è possibile scaricare i due crawler al seguente indirizzo: https://github.com/AndreMatte97/Fanfiction

informazioni relative al primo capitolo di ogni storia; il secondo di 1,86 GB, con all'interno le informazioni su tutti gli altri capitoli. Una volta ottenuto il corpus completo, le storie sono state divise per numero di capitoli, eliminando i racconti iniziati negli anni 2018 e 2019, per cercare di limitare l'analisi a testi presumibilmente conclusi. Per ogni classe di libri è stato calcolato il numero medio di recensioni ricevute, il numero di storie con un numero di recensioni sopra o sotto la media e la distribuzione dei libri per numero di recensioni. Nella tabella 1 è presente un estratto delle statistiche ottenute, che riporta queste informazioni per le classi con un numero di storie superiore a 300.

# Capitoli	# Libri	# Medio Rec.	# Rec > Media	# Rec <= Media
1	37783	4,96	14667	23116
2	2721	6,78	925	1796
3	2349	9,88	854	1495
4	1703	13,0	559	1144
5	1510	15,86	550	960
6	1068	20,57	373	695
7	913	24,96	335	578
8	683	27,03	235	448
9	595	33,88	210	385
10	653	39,44	220	433
11	454	39,68	175	279
12	431	47,78	161	270
13	328	53,28	106	222
14	308	62,20	106	202

Tabella 1. Alcune statistiche tratte dal corpus raccolto[2]

Si può notare come, all'aumentare del numero dei capitoli, la quantità di libri diminuisca in modo irregolare, mentre il numero medio di recensioni aumenti quasi linearmente, come prevedibile: ogni capitolo è pubblicato singolarmente e quindi porta con sé recensioni aggiuntive. Il rapporto tra numero di libri recensiti un numero di volte superiore e inferiore alla media rimane invece consistentemente intorno allo 0,5.

[2] Per ogni classe di libri, contraddistinta dal numero di capitoli di cui si compone il libro, è riportato il numero totale di libri di quella classe, il numero medio di recensioni ottenute dai libri per classe, il numero di libri con un numero di recensioni inferiore o superiore alla media.

3. La selezione delle storie oggetto di analisi

Ai fini della nostra indagine si è deciso di restringere l'attenzione alle storie presenti nel corpus composte da un singolo capitolo e scritte prima del 2018, per evitare l'inclusione di storie non ancora concluse. Una volta identificata la sezione dei testi di interesse, che si compone di 16.587 storie, si è posto il problema di definire un criterio che rendesse possibile quantificare un aspetto altamente soggettivo del testo, quale il grado di apprezzamento da parte del lettore. Il successo di un'opera letteraria può dipendere infatti da molteplici fattori, sia intrinseci – legati alla trama, all'intreccio, alla caratterizzazione dei personaggi – sia estrinseci, legati alla popolarità dell'autore, alle strategie editoriali, all'aderenza dei temi trattati con questioni culturali e socio-politiche di attualità. Il celebre sociologo della letteratura Robert Escarpit (1972) riflette inoltre su quella sorta di predisposizione, che è propria delle grandi opere letterarie, a subire il 'tradimento' da parte da parte del pubblico che, a decenni o secoli di distanza, può riconoscervi delle intenzioni che l'autore non ha mai voluto inserire e proiettare su di essa le proprie attese.

Come già anticipato nell'introduzione, in questo studio abbiamo voluto sondare se esistano aspetti dello stile linguistico di una *fanfiction* che contribuiscono maggiormente al suo apprezzamento da parte del pubblico. Seppur ristretto l'ambito di osservazione allo stile di scrittura del testo, rimane comunque complessa la questione di come giudicare se un'opera sia di successo, dal momento che, sempre adottando la prospettiva escarpitiana, si può valutare il successo dell'oggetto librario sia in termini di successo commerciale sia come valore culturale dell'opera letteraria. A nostra conoscenza, se numerose sono le riflessioni sulla qualità dell'opera letteraria in ambito di critica della letteratura, più limitati sono i contributi che hanno indagato questo tema dal punto di vista quantitativo, proponendo definizioni di successo in qualche modo 'misurabili'. In particolare, nell'ambito degli studi più affini al nostro, condotti con approcci di stilometria computazionale, si possono citare due recenti lavori che hanno affrontato questi aspetti. Entrambi hanno preso in considerazione un ampio corpus di libri di narrativa inglese appartenenti a vari generi letterari tradizionali e scaricabili liberamente grazie al progetto Gutenberg[3], con l'obiettivo di sviluppare un un modello 'predittivo' della scrittura di successo sulla base della distribuzione di caratteristiche formali di varia

[3] http://www.gutenberg.org/

natura estratte dal testo. Nello studio di Ganjigunte e colleghi del 2013, i testi del corpus sono stati distinti in due classi (successo e non) in base al numero di *downloads* registrati dal sito Gutenberg per ciascun libro. In uno studio successivo di Maharjan *et al*. del 2017, gli autori hanno invece proposto una metrica più soggettiva, che valuta il successo in base al numero di recensioni presenti per ciascun romanzo all'interno di un sito web dedicato proprio alla pubblicazione di recensioni di opere letterarie da parte dei lettori.

Per le nostre analisi ci siamo ispirati proprio a questa seconda metodologia, decidendo tuttavia di utilizzare come indicatore del successo il numero totale di recensioni associate a ciascuna *fanfiction* nel portale efpfanfic.net e non soltanto quelle positive. Due motivazioni principali hanno indotto questa scelta: in primo luogo è stato notato che la maggioranza delle recensioni rilevate sono state scritte a scopo di apprezzamento, con soltanto uno 0,73% di recensioni negative sulle 900mila circa totali. Da un punto di vista statistico è quindi possibile eliminare la distinzione tra le varie tipologie di *feedback* e prendere in considerazione semplicemente la quantità totale di giudizi ricevuti. Inoltre, si può ritenere che anche una recensione negativa provi che la storia in questione sia stata letta e abbia suscitato un certo interesse nel lettore. In base a queste considerazioni, abbiamo quindi distinto i testi in due classi, testi di 'successo' e 'insuccesso', dove la prima classe raccoglie tutte le storie che hanno ricevuto un numero di recensioni più alto della media di tutte le storie con la stessa quantità di capitoli (per un totale di 14.486 storie), mentre la seconda è composta da tutte le storie che non hanno ricevuto nessun feedback, venendo quindi ignorate dai lettori, ed è composta da 2.101 testi. Infine, tutti i testi delle due classi sono stati trattati tramite l'applicazione di alcune espressioni regolari, allo scopo di rimuovere errori e inconsistenze nell'uso della punteggiatura, delle lettere maiuscole e dei caratteri speciali. Lo scopo è stato quello di aumentare l'affidabilità del processo di annotazione linguistica automatica, e di conseguenza l'estrazione delle caratteristiche linguistiche oggetto di monitoraggio linguistico, secondo la metodologia descritta nel paragrafo successivo.

4. La metodologia di monitoraggio linguistico del testo

Il monitoraggio linguistico del testo basato sull'uso di tecnologie di Trattamento Automatico del Linguaggio può essere concepito come un *framework* di analisi che, a partire da un processo incrementale di annotazione linguistica automatica del testo, permette di renderne esplicita

l'informazione associata ai diversi livelli di descrizione linguistica (quali lessico, morfosintassi, sintassi) così da consentire l'estrazione di un ampio repertorio di parametri – che derivano proprio dai livelli di annotazione presenti – e che permettono di caratterizzare l'uso della lingua nel testo di riferimento (Montemagni 2013). Tale metodologia, soprattutto se applicata comparativamente per confrontare le distribuzioni statistiche dei parametri linguistici estratti da corpora rappresentativi di varietà di lingua diverse, è propedeutica allo studio della variazione stilistica tra generi, registri e, più in generale, tipi testuali differenti definiti in ottica funzionale. Il monitoraggio linguistico si ispira infatti ai fondamenti dell'approccio multidimensionale/multifattoriale inaugurato dai lavori pionieristici di Douglas Biber alla fine degli anni '90, e originariamente concepito per indagare la distinzione tra registro scritto e parlato sulla base dell'evidenza empirica ricavata da ampi corpora di lingua inglese. Gli assunti di tale approccio sono ben evidenti in questa citazione, tratta da (Biber 1993: 219), secondo cui "linguistic features from all levels function together as underlying dimensions of variation, with each dimension defining a different set of linguistic relations among registers".

All'interno di questa tradizione di studi, le tecnologie linguistico-computazionali offrono oggi un importante valore aggiunto, che è quello di poter estendere la copertura dei fenomeni linguistici oggetti di monitoraggio in maniera affidabile e spesso multilingue, anche in relazione a corpora di dimensioni molto più estese di quelle del passato. Infatti, come osserva Argamon (2019), gli studi tradizionali nell'ambito della *Computational Register Analysis* erano necessariamente ristretti alla valutazione di pochi indicatori, quali la distribuzione di categorie funzionali o di sequenze di *n-grams* di caratteri, il cui calcolo non richiedeva strumenti di trattamento del testo particolarmente sofisticati e specifici per l'analisi delle peculiarità di una determinata lingua. Al contrario, gli approcci attuali allo studio della variazione linguistica possono fare affidamento su tecnologie del linguaggio molto più robuste, in grado di adattarsi con minime variazioni a domini e tipologie testuali diverse e di trattare in maniera più affidabile fenomeni linguistici anche devianti dall'uso standard della lingua. Tali progressi hanno reso possibile l'automatizzazione del processo di estrazione di indicatori relativi alla forma linguistica del testo che spaziano tra i livelli di descrizione della lingua e modellano un'ampia varietà di fenomeni. In questo filone di ricerche, si inserisce lo strumento utilizzato per la nostra analisi, Profiling-UD (Brunato *et al*. 2020), che traduce operativamente i presupposti degli attuali approcci al monitoraggio linguistico del testo. Lo strumento, infatti, permette di annotare un testo, o una collezione di testi, fino al livello dell'analisi sintattica a dipendenze e di estrarne un ricco

repertorio di caratteristiche linguistiche, più di un centinaio, con un focus particolare su caratteristiche che modellano la struttura morfosintattica e sintattica della varietà di lingua rappresentata dal corpus analizzato. La peculiarità di Profiling-UD è di essere uno strumento multilingue, in quanto l'annotazione multi-livello sottostante al testo, che costituisce l'input per il componente di monitoraggio linguistico del testo, è condotta da un parser statistico allo stato dell'arte descritto in Straka *et al.* (2016), che restituisce come risultato un testo linguisticamente annotato rispetto al *framework* di annotazione definito nel progetto delle Universal Dependencies (UD)[4] (Nivre 2015). Tra le motivazioni di questo progetto, tutt'ora in corso, vi è stata la definizione di un formalismo di analisi morfosintattica e sintattica condiviso tra lingue, in grado di ricondurre ad una rappresentazione uniforme costruzioni che esprimono lo stesso fenomeno (es. la determinazione, la subordinazione) indipendentemente dalla sua realizzazione interlinguistica[5]. Sebbene non utilizzata ai fini della nostra indagine, riteniamo che la prospettiva multilingue al monitoraggio linguistico resa possibile da questo strumento possa aprire la strada a studi di più ampio raggio sulle varietà di scrittura giovanili, anche in ottica di multilinguismo.

4.1 Le caratteristiche linguistiche oggetto di monitoraggio

In questo paragrafo diamo una breve descrizione delle caratteristiche linguistiche estratte da Profiling-UD e utilizzate per analizzare i testi del nostro corpus. Come riportato dagli autori, tali caratteristiche possono essere ricondotte a sette gruppi che identificano altrettante tipologie di fenomeni linguistici, ognuno dei quali derivante da un livello diverso di annotazione linguistica, ovvero:

1. caratteristiche di base: si tratta di indicatori superficiali del testo, quali la lunghezza media dei documenti, delle frasi e delle parole, la cui estrazione richiede la sola segmentazione del testo in frase e in parole ortografiche (*tokens*), fasi che vengono tipicamente operate dai componenti di base di una catena di analisi linguistica

[4] https://universaldependencies.org/

[5] Resta comunque ammessa la possibilità di estensioni alle categorie morfosintattiche e sintattiche previste dagli schemi 'universali' UD per rendere conto delle specificità di ciascuna lingua.

automatica quali il *sentence splitter* e il *tokenizzatore*;

2. varietà lessicale: questo aspetto è misurato da un indice standard di ricchezza lessicale del testo quale il rapporto tra parole tipo e parole unità (*Type/Token Ratio*), computato per porzioni di testo di lunghezza fissa, ovvero i primi 100 e 200 tokens. Inoltre, dal punto di vista lessicale, viene calcolata la percentuale di parole presenti nel Vocabolario di Base di De Mauro (2000) e all'interno dei tre repertori d'uso (lessico fondamentale, alto uso e alta disponibilità);
3. informazione morfosintattica: a questa macro-classe di fenomeni appartengono caratteristiche linguistiche diverse, quali la distribuzione delle 17 categorie morfosintattiche definite dal tagset universale di UD[6], la densità lessicale calcolata come rapporto tra parole piene (aggettivi, avverbi, sostantivi e avverbi) sul totale di *tokens* del testo e la distribuzione di verbi lessicali e ausiliari in base ai tratti morfosintattici di tempo, modo e persona;
4. informazione legata alla struttura del predicato: rientrano in questo gruppo le caratteristiche relative al numero medio di teste verbali[7] in una frase, parametro che fornisce una prima indicazione sul numero di proposizioni (principali e subordinate) nel testo, la distribuzione di radici verbali sul totale delle radici sintattiche e l'arità verbale, calcolata come il numero medio di dipendenti (senza distinzione tra argomenti e aggiunti) per testa verbale;
5. informazione legata alla struttura sintattica locale e globale: si tratta di caratteristiche che computano la profondità media dell'albero sintattico della frase (calcolata come numero di relazioni di dipendenza sintattica che intercorrono tra il *token* radice e un elemento "foglia", ovvero un *token* della frase senza dipendenti); la lunghezza media delle clausole (equivalente al rapporto tra il numero totale di *tokens* della frase e il numero di teste verbali e copulari); la lunghezza delle relazioni di dipendenza sintattica (ovvero il numero medio di *tokens* che separano la testa e il dipendente in una frase); la profondità media delle catene nominali, dove una catena nominale è presente quando una testa

[6] https://universaldependencies.org/u/pos/

[7] Nell'analisi sintattica a dipendenze, le relazioni sintattiche sono definite in termini di relazioni di dipendenza binaria tra due elementi della frase di cui uno svolge il ruolo di elemento reggente o testa sintattica (es. il verbo) e uno di elemento dipendente (es. il soggetto).

sintattica nominale viene modificata ricorsivamente da complementi nominali o aggettivali e, infine, caratteristiche relative all'ordine dei costituenti primari, quali la percentuale di soggetti e oggetti in posizione pre e postverbale;

6. distribuzione delle relazioni sintattiche: per ciascuna relazione sintattica definita dallo schema di annotazione UD (es. soggetto nominale e frasale, oggetto diretto e indiretto, modificatore aggettivale, avverbiale, ecc.) ne viene calcolata la frequenza media nel corpus analizzato;
7. uso della subordinazione: questo fenomeno viene modellato in termini di: rapporto tra frasi principali e subordinate, posizione della subordinata rispetto alla principale e profondità media delle catene subordinanti che, in analogia a quelle nominali, sono calcolate come numero medio di archi sintattici ricorsivamente incassati e dipendenti dalla testa sintattica che introduce la frase subordinata.

L'immagine sottostante mostra una visualizzazione grafica dell'albero a dipendenze di una frase tratta dal corpus e linguisticamente annotata secondo il formato UD. Data questa annotazione, con Profiling-UD, è possibile ricavare, ad esempio, che la frase è lunga 27 tokens e contiene parole mediamente lunghe 4,7 caratteri. Rispetto alla distribuzione delle categorie morfosintattiche, si conta ad esempio, il 14,8% di aggettivi, il 3,7% di avverbi, il 7,4 % di pronomi, il 18,5 % di sostantivi e l'11,1% di verbi. Il link più lungo (esclusa la punteggiatura) è pari a 7 token ed quello che collega il dipendente con il ruolo di soggetto ('donna') alla sua testa, in questo caso la radice (*root*) della frase, ossia il verbo della principale ('continuava'). Data la presenza di due subordinate (l'infinitiva retta dalla principale, marcata come *xcomp*, e la relativa soggetto, marcata come *acl:relc*), il rapporto principali/subordinate è pari allo 0,33%.

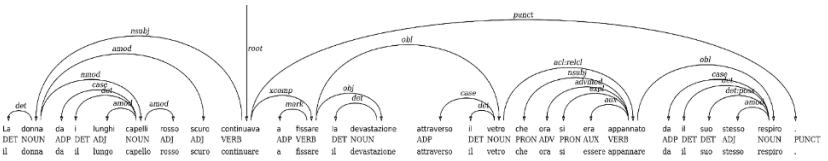

Figura 1. Esempio di visualizzazione grafica dell'albero sintattico a dipendenze di una frase del corpus secondo lo schema di annotazione UD.

5. Cosa rende una storia di successo? Un'analisi dei risultati del monitoraggio

Per ciascuna delle caratteristiche linguistiche descritte nella sezione precedente, sono state calcolate la media e la deviazione standard nelle due classi in cui è stato suddiviso il corpus di partenza, ovvero testi di successo e non. È stata poi valutata la significatività della variazione della media di ogni caratteristica tra le due classi, utilizzando un test statistico non parametrico, il test dei ranghi con segno di Wilcoxon, allo scopo di identificare quali parametri linguistici siano maggiormente identificativi delle storie più apprezzate dal pubblico. Il test di Wilcoxon, infatti, verifica la validità di un'ipotesi (detta 'ipotesi nulla') secondo cui è equamente probabile che un valore scelto casualmente da una popolazione di dati sia minore oppure maggiore di un altro valore casuale scelto da una seconda popolazione di dati: in altre parole, che le due popolazioni campione abbiano la stessa distribuzione statistica. Il test restituisce come risultato un numero compreso tra 0 e 1, detto *p-value*: se sufficientemente basso è possibile confutare l'ipotesi nulla e quindi accettare l'ipotesi di ricerca, ovvero che i due campioni hanno effettivamente distribuzioni diverse. In quanto segue vengono presentati i risultati di questo confronto.

Innanzitutto, si è osservato che ben 126 caratteristiche (pari all 57% delle 219 analizzate) sono distribuite differentemente in modo statisticamente rilevante tra storie di successo e insuccesso. La Tabella 2 riporta una selezione delle più significative, distinte in base al livello di annotazione sintattica da cui derivano.

	- Insuccesso -	+ Successo +
Caratteristica linguistica	Media (±Dev Std)	Media (±Dev Std)
Caratteristiche di Base		
1. Lunghezza media testi (in token)	1401 (±1940)	2120 (±2718)
2. Lunghezza media frasi (in token)	78,4 (±116,7)	125,1(±153,6)
3. Lunghezza media frasi per testo	20,18 (±12,39)	17,38 (±6,43)
4. Lunghezza media parole per testo	4,50 (±0,250)	4,52 (±,193)
Caratteristiche Lessicali		
5. % Token nel VdB	85,7 (±5,1)	84,8 (±4,7)
6. % Lemmi nel VdB	73,4 (±7)	70,1 (±7)
7. % VdB_Lessico Fond	61 (±7,5)	57,1 (±7,7)
8. % VdB_AltoUso	3,1 (±1)	3,1 (±1)

9. % VdB_AltaDisp	8,5 (±2,4)	9,1 (±2,5)
10. Densità Lessicale	0,498 (±0,033)	0,503 (±0,031)
Caratteristiche Morfo-Sintattiche		
11. % Aux 3PerPl	13,2 (±9,00)	11,8 (±7,56)
12. % Aux 3PerSin	54,4 (±17,1)	53,2 (±15,6)
13. % Aux 2PerSin	6,30 (±8,12)	7,86 (±8,53)
14. % Auc Imperfetto	38,5 (±24,8)	31,3 (±22,2)
15. % Aux Presente	52,4 (±26,0)	60,0 (±23,7)
16. % Verbi al Gerundio	5,65 (±3,81)	6,27 (±4,00)
17. % Sostantivi	13,8 (±2,32)	13,5 (±2,11)
18. % Verbi	12,5 (±1,79)	12,3 (±1,66)
19. % Aggettivi	4,74 (±1,41)	4,59 (±1,20)
20. % Preposizioni	10,8 (±1,86)	10,4 (±1,78)
21. % Puntegg. Bilanciata	1,75 (±1,98)	2,29 (±2,36)
22. % Virgole	6,52 (±2,71)	7,11 (±2,83)
23. % Punti fine frase	5,52 (±2,08)	6,13 (±2,10)
Caratteristiche Sintattiche		
24. Lungh. media link sintattici	2,78 (±0,438)	2,72 (±0,385)
25. * Link sintattico più lungo	1,19 (±2,38)	0,687 (±1,33)
26. Profondità max albero	3,96 (±1,45)	3,58 (±0,857)
27. Num. Teste verbali/frase	2,63 (±1,72)	2,26(±0,9)
28. % Prop. Principali	48,8 (±9,9)	49,9 (±9)
29. % Prop. Subordinate	51,2 (±9,9)	50,1 (±9)
30. Lungh. media catene subord.	1,31 (±0,15)	1,30 (±0,13)

Tabella 2. Una selezione delle caratteristiche linguistiche monitorate, distinte per livelli di annotazione linguistica, che risultano distribuite diversamente nei testi di successo e insuccesso. Per ogni caratteristica è riportata la media e, in parentesi, la deviazione standard. Tutte le differenze tra le due classi sono altamente significative (p-value < 0,001), ad eccezione di quella contrassegnata da asterisco che è significativa con p-value < 0,05.

È possibile notare come le storie di successo siano mediamente più lunghe, sia in termini sia di numero di *token* che di frasi (caratteristiche 1, 2 nella tabella); tali frasi, tuttavia, sono generalmente più brevi (3), un dato che suggerisce una preferenza per uno stile di scrittura più asciutto e scorrevole da parte dei lettori. Uno sguardo agli indici relativi al profilo

lessicale rivela tuttavia una disposizione positiva verso testi che impiegano un vocabolario mediamente meno frequente, come indicato dalla distribuzione leggermente più bassa di parole e lemmi appartenenti al Vocabolario di Base della lingua italiana (5, 6) e in particolare del Lessico Fondamentale (7).

Le due classi mostrano variazioni anche al livello della morfologia flessiva: *fanfiction* di successo utilizzano più spesso verbi alla seconda persona (13), una caratteristica tipica della scrittura narrativa e presumibilmente collegata all'uso del discorso diretto. Al contrario, osserviamo una distribuzione maggiore di verbi alla terza persona, in particolare ausiliari, sia plurali (11) che singolari (12) nei testi di minor successo, che può suggerire un uso più frequente del discorso indiretto. Per quanto riguarda la distribuzione delle categorie morfosintattiche, è presente una differenza significativa nell'utilizzo dei più comuni segni di punteggiatura, virgole (22) e punti fermi (23), i quali sono assai più frequenti nelle *fanfiction* più recensite. Queste caratteristiche si collegano allo scarto in termini di lunghezza dei periodi precedentemente osservato, poiché testi con frasi più corte necessitano di più segni di punteggiatura per separarle. Notiamo inoltre che i segni bilanciati (21), ossia parentesi e virgolette, sono più presenti nei testi di successo, un dato che rafforza l'ipotesi circa un impiego più ampio del discorso diretto all'interno di questa classe di testi.

A livello della struttura sintattica, osserviamo relazioni sintattiche mediamente più corte nei testi di successo, sia considerando la lunghezza media di tutte le dipendenze sintattiche (24), sia il valore del *link* sintattico più lungo (25). Negli studi sulla valutazione della leggibilità dei testi si rilevano solitamente dipendenze sintattiche più lunghe e alberi sintattici più profondi in testi complessi (Collins-Thompson 2014). Entrambe queste caratteristiche (24, 26) hanno invece valori inferiori nei racconti più recensiti dai lettori, suggerendo che uno stile di scrittura di successo tende a essere caratterizzato da una struttura sintattica più semplice. È interessante notare come questi risultati, seppur preliminari, vadano in direzione opposta rispetto a quelli riportati nel lavoro già citato di Ganjigunte e collaboratori su un corpus di romanzi classici della letteratura inglese (cfr. §3.1) , in cui quelli giudicati di successo sono risultati essere meno correlati con il punteggio di leggibilità del testo. Infine, rispetto all'uso della subordinazione, le proposizioni subordinate (29) sono più frequenti delle principali (28) nei testi dallo scarso successo, mentre si verifica una ripartizione pressoché bilanciata tra ipotassi e paratassi in quelli di successo. Anche la lunghezza media delle catene di subordinazione (30) è ripartita in maniera simile all'interno delle due classi ma è comunque minore e più stabile nei testi di successo.

5.1 Ordinamento delle caratteristiche linguistiche in base all'indice di variabilità nei testi di successo e insuccesso

Per approfondire le differenze nel profilo linguistico discusse al punto precedente, abbiamo infine calcolato i coefficienti di variazione σ* per ciascuna delle caratteristiche linguistiche che sono risultate avere una diversa distribuzione nelle due classi. Il coefficiente di variazione è una misura della variabilità relativa interna a un campione e viene calcolato come il rapporto tra la deviazione standard e la media all'interno di una classe. Pertanto, tramite questo indice è possibile quantificare la dispersione dei valori intorno alla media in modo standardizzato e comparare la stabilità di caratteristiche riferite a dati misurati su scale diverse. L'assunto di base è che una caratteristica che risulti molto dispersa in una classe di testi e molto stabile nell'altra abbia una probabilità più alta di rappresentare significativamente quest'ultima.

Operativamente, per ciascuna classe di testi, abbiamo calcolato il coefficiente di variazione delle caratteristiche linguistiche la cui distribuzione è risultata significativamente differente tramite il test di Wilcoxon e le abbiamo ordinate per valori di coefficiente crescente[8]. Da questo confronto è emerso che lo stile dei racconti di successo è, in generale, più stabile e standardizzato: nella maggioranza dei casi le caratteristiche hanno un indice di variazione minore per le storie di successo. Ciò può essere dovuto in parte allo sbilanciamento dei due campioni: in un dataset ampio come quello dei testi di successo, l'impatto degli *outlier* (valori anomali) è presumibilmente minore. Tuttavia, essendo la maggior parte delle caratteristiche riportate come distribuzioni percentuali sul totale dei *token*, questo effetto è attenuato per buona parte delle rilevazioni e non dovrebbe incidere in maniera rilevante sulla differenza degli indici di stabilità. Si osserva inoltre che le caratteristiche con indice di variabilità minore (quindi posizionate in cima a entrambi i *ranking*) occupano posizioni simili, con sfasamenti di due livelli al massimo per le prime quindici posizioni. La prima caratteristica a mostrare maggior differenza è la lunghezza media della clausola, con un indice di variazione doppio all'interno dei testi di minor successo.

Proprio per rendere più chiare le differenze di variabilità tra testi di successo e di insuccesso, si è infine deciso di ordinare le caratteristiche linguistiche per differenza di posizione (DeltaRank) decrescente all'interno

[8] Per ragioni di spazio omettiamo di riportare le tabelle delle 126 caratteristiche linguistiche ordinate per coefficiente di variazione per entrambe le classi.

dei relativi ordinamenti. Per ciascuna caratteristica, la differenza è stata calcolata sottraendo l'indice posizionale che essa occupa nel *ranking* dei testi di successo da quello che occupa nei testi di insuccesso. Ne consegue che le caratteristiche con DeltaRank più alto sono quelle con maggiore stabilità nei testi che hanno avuto successo, mentre quelle con DeltaRank più basso (negativo) sono caratterizzate da una maggiore stabilità nei testi che non hanno ricevuto recensioni. Le caratteristiche linguistiche con DeltaRank che si aggira intorno allo 0 sono invece quelle che hanno variabilità simile all'interno di entrambe le classi: pertanto, si può ritenere che siano tratti peculiari del genere *fanfiction*.

La Tabella 3 riporta un estratto delle prime dieci caratteristiche emerse da questo confronto, ordinate in maniera decrescente sulla base della differenza di posizione all'interno dei due *ranking*. Nella colonna Delta% sono riportate anche le differenze fra gli indici di variazione di tali caratteristiche. Come si può notare, fra quelle più stabili all'interno dei testi di successo troviamo sia caratteristiche di base, quali il numero di frasi per documento (8) e la lunghezza media delle frasi in numero di *token* (5), sia caratteristiche sintattiche come la lunghezza media della clausola (1), la media delle profondità massime degli alberi sintattici e il numero di teste verbali per frase. Compaiono inoltre nelle prime posizioni della tabella, la distribuzione di numeri ordinali (3) e la presenza di nomi composti e espressioni polirematiche (6, marcate dalla dipendenza 'dep_dist_ flat:name').

Caratteristica Linguistica	RankINS	RankSUC	DeltaRank	Delta%
1. Lunghezza media clausola	52	16	36	17,66%
2. Link sintattico più lungo	58	36	22	12,68%
3. % Numerali Ordinali	124	105	19	78,74%
4. Teste verbali/frase	84	66	18	25,66%
5. Lunghezza media frasi	80	63	17	24,37%
6. % dep_dist_flat:name	126	114	12	110,31%
7. % Aux tempo presente	72	65	12	10,18%
8. Numero frasi/doc	120	115	5	26,10%
9. % Modificatori nominali	55	51	4	4,81%
10. % Aux 2perSing	112	109	3	20,27%

Tabella 3. Estratto delle prime dieci caratteristiche, ordinate per differenza di posizione nei rispettivi *ranking* calcolati sulla base del coefficiente di variazione crescente. Per ciascuna caratteristica, viene riportata la posizione nel *ranking* nelle classi dei testi di successo e dei testi di insuccesso, il valore assoluto tra della differenza (DeltaRank) e la differenza tra gli indici di variazione (Delta%).

6. Conclusioni

In questo contributo, abbiamo illustrato i fondamenti di una innovativa metodologia di analisi e monitoraggio linguistico del testo che, a partire dall'*output* di strumenti di annotazione linguistica automatica, permette di ricostruire un profilo linguistico multi-livello di testi rappresentativi di una specifica varietà d'uso della lingua. Tale metodologia è stata applicata allo studio del genere della *fanfiction*, un genere letterario emergente tra i più rappresentativi della scrittura del web 2.0. In particolare, attraverso la raccolta di un nuovo corpus per la lingua italiana composto da oltre 16,000 racconti amatoriali scritti da giovani internauti e ispirati al romanzo Harry Potter, abbiamo voluto sondare le potenzialità di questa metodologia nel caratterizzare una dimensione molto soggettiva di un testo, quale il grado di apprezzamento che otterrà dal lettore.

I risultati raggiunti, per quanto preliminari, hanno mostrato che le *fanfiction* di maggior successo tra il pubblico sono mediamente più lunghe ma al tempo stesso fanno uso di frasi più brevi e di una struttura sintattica più semplice, contengono un lessico più variegato, con una maggiore percentuale di parole non presenti all'interno del Vocabolario di Base e presentano una distribuzione maggiore di tratti tipici del discorso diretto. Inoltre, tali storie sono caratterizzate da indici di variazione inferiori per la maggior parte delle caratteristiche linguistiche monitorate, mostrando quindi una maggiore stabilità stilistica rispetto alle storie meno recensite.
A nostro avviso, le prospettive che questo studio propone sono diverse e spaziano in molteplici direzioni. Innanzitutto, sarebbe certamente interessante estendere l'analisi ad altre opere letterarie prodotte dai membri di una stessa *fan community*, prendendo in considerazione tanto storie ispirate ad altri romanzi popolari dei più diversi generi testuali, quanto storie originali scritte dagli utenti. Inoltre, se la presenza di tratti tipici del linguaggio giovanile all'interno del genere della *fanfiction* italiana è stata indagata recentemente a livello lessicale, evidenziando ad esempio l'ampia ricorrenza di usi gergali e giochi di parole (Comandini 2020), l'approccio del monitoraggio linguistico applicato in maniera comparativa a corpora di linguaggio giovanile di domini diversi potrebbe rivelare contaminazioni anche sul piano delle costruzioni sintattiche.

Infine, la disponibilità di uno strumento di monitoraggio linguistico multi-lingue come quello utilizzato in questo studio potrebbe permettere di capire se esiste un 'modello' della scrittura di successo i cui tratti stilistici sono trasversali non solo ai generi testuali, ma anche alle lingue.

Riferimenti bibliografici

ARGAMON S., 2019, "Computational register analysis and synthesis", in *Register Studies*, vol. 1, no.1.

BIBER D., 1988, *Variation across Speech and Writing*, New York, NY, Cambridge University Press.

BIBER D., 1993, "Using register-diversified corpora for general language studies", in *Computational Linguistics*, 19.

BRUNATO D. *et al.*, 2020, "Profiling-UD: a Tool for Linguistic Profiling of Texts", in *Proceedings of The 12th Language Resources and Evaluation Conference*, European Language Resources Association, pp. 7145-7151.

BRUNATO D. *et al*, 2018, "Is this Sentence Difficult? Do you Agree?", in *Proceedings of Conference on EMNLP*.

CALABRESE S., CONTI V., 2019, *Che cos'è una fanfiction*, Roma, Carocci.

COLLINS-THOMPSON K., 2014, "Computational assessment of text readability: a survey of current and future research", in *ITL - International Journal of Applied Linguistics*, vol.165, no.1.

COMANDINI G., 2020, "L'ironia criptica dei linguaggi giovanili sul web. Il caso dei giochi di parole nei fandom*",* in ALLOCCA C., CARBONE F., COPPOLA R., OCCHINI B. (a cura di), *Sottosopra. Indagine su processi di sovversione*, *Quaderni della Ricerca,* 6, Napoli, UniorPress.

DAELEMANS W., 2013, "Explanation in Computational Stylometry", in GELBUKH A. (ed.), *Computational Linguistics and Intelligent Text Processing* (CICLing 2013), Lecture Notes in Computer Science, vol 7817. Springer, Berlin, Heidelberg.

DE MAURO T., 2000, *Grande dizionario italiano dell'uso (GRADIT)*, Torino, UTET.

ESCARPIT R., 1972, *Letteratura e società*, Il Mulino, Bologna.

MAHARJAN S. *et al.,* 2017, "A multi-task approach to predict the likability of books", in *Proceedings of the 15th Conference of the European Chapter of ACL*, Valencia, Spain.

MALMASI S. *et al*., 2017. "A Report on the 2017 Native Language Identification Shared Task", *in Proceedings of the 12th Workshop on Building Educational Applications Using NLP*.

MIASCHI A. *et al*., 2021, "A NLP-based stylometric approach for tracking the evolution of L1 written language competence", in *Journal of Writing Research*.

MONTEMAGNI S., 2013, "Tecnologie linguistico-computazionali e monitoraggio della lingua italiana", in *Studi Italiani di Linguistica Teorica e Applicata*, (SILTA), pp. 145-172.

NGUYEN D. *et al.,* 2016, "Computational Sociolinguistics: A Survey", in *Computational Linguistics*, vol. 42, no. 3, pp. 537–593.

HERRING S. C., PAOLILLO J. C., 2006, "Gender and genre variation in weblogs", in *Journal of Sociolinguistics*, vol. 10, no. 4.

NIVRE J., 2015, "Towards a universal grammar for natural language

processing", in Gelbukh A. (ed.), *International Conference on Intelligent Text Processing and Computational Linguistics*, Berling, Springer pp. 3-16.

Sindoni M.G., 2011, "'I really have no idea what non-fandom people do with their lives.' A multimodal and corpus-based analysis of fan-fiction", in *Lingue e Linguaggi*, vol. 13.

Straka M. *et al,* 2016, "UD-Pipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing", in *Proceedings of the Tenth International Conference on LREC*.

Van Halteren H, 2004, "Linguistic profiling for author recognition and verification", in *Proceedings of ACL*.

Weiss Z., Meurers D., 2019, "Analyzing linguistic complexity and accuracy in academic language development of german across elementary and secondary school", in *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA) at ACL*.

NORME DI STILE *RILA*

I saggi vengono presi in considerazione se seguono le seguenti indicazioni:

a. vanno composti in word, Times New Roman 12;
b. tabelle, rientri, elenchi puntati vanno realizzati usando gli appositi comandi di word e non la barra spaziatrice;
c. immagini in toni di grigio devono essere di qualità adeguata ad una facile lettura, visto che la rivista non è a colori;
d. tener conto del formato e della dimensione della pagina di RILA.

Per il layout:

a. sopra il titolo: indirizzo postale completo e la posta elettronica, per i contatti;
b. sotto il titoloriportare il nome completo dell'Autore in MAIUSCOLETTO (ctrl+maiusc+K), seguito dalla istituzione di appartenenza in *corsivo*;
c. se l'articolo è firmato da due o più autori, in una nota a pie di pagina indicare chi ha scritto cosa;
d. sotto l'autore un Abstract di non oltre 120 parole, carattere 12, in corsivo, rientro di 1,25 cm nella prima riga;
e. la scritta ***Abstract*** va in corsivo, grassetto;
f. non devono essere indicate Keywords;
g. paragrafatura:rientro di 1,25 cm della prima riga di ogni capoverso, sia nel testo sia nelle note;
h. paragrafi di primo livello hanno numero+punto e sono in corsivo nero: ***2. La complessità...***;
i. quelli di secondo livello non hanno punto dopo l'ultimo numero e sono in corsivo chiaro: *2.1 La complessità...*;
j. quelli di terzo livello sono in tondo chiaro: 2.1.2 La complessità...;
k. titoli dei paragrafi: due spazi prima del titolo e uno spazio dopo il titolo;
l. tra i titoli dei paragrafi di livelli diversi (esempio: primo – secondo livello, secondo – terzo livello) inserire del testo;
m. l'eventuale appendice va indicata prima della bibliografia.

Quanto allo stile tipografico:

a. parole straniere nel corpo del testo in corsivo;
b. virgolette: nelle citazioni "..."; per le parole da evidenziare'...'; se ci sono virgolette, non c'è corsivo. Evitare « »; l'evidenziazione col neretto è esclusa;
c. interlinea 1; verificare in Layout di pagina che come interlinea sia selezionato 0 in entrambe le caselle;
d. elenchi puntati: spazio sopra e sotto;

e. citazioni: oltre le tre righe di testo, comporre un paragrafo a sé, col margine sinistro rientrato di almeno un centimetro (no rientro nella prima riga) e senza virgolette iniziali e finali, corpo 11, allineate a destra;
f. maiuscolo: sigle come SLIO CLIL, editori come ESI o UTET vanno scritti in maiuscoletto (ctrl+maiusc+K); il MAIUSCOLO PIENO, così come il **neretto,** è escluso dal saggio;
g. le note si usano solo quando non si può usare un riferimento interno tra parentesi; le note vanno messe con l'apposito comando di word che le numera e le colloca a pie di pagina, Times New Roman, rientro prima riga, corpo 10;
h. le didascalie delle figure e delle tabelle vanno sempre indicate sotto queste ultime, Times New Roman 10.

Citazioni

Autore-data tra parentesi, senza virgola fra il nome dell'autore e la data:

- se ci sono le pagine, due punti dopo l'anno e poi uno spazio **(Rossi 1998: 134-135)**;
- se si citano più opere, dello stesso autore o di autori diversi, separare ogni riferimento con un punto e virgola: es. **(Rossi 1993; 1994;Bianchi 1999; 2009)**;
- due opere dello stesso autore pubblicate nello stesso anno vanno distinte con una lettera dell'alfabeto: **(Rossi 1998a) e (Rossi 1998b)**;
- citazione di opere con più di due autori: il primo autore seguito da *et al.* **(Rossi *et al.,* 1999)** purché non si creino ambiguità. La *entry* completa sarà in bibliografia.

Riferimenti bibliografici

Si sconsigliano bibliografie generali, facilmente reperibili on line; sirichiedono invece i riferimenti completi alle opere citate nel saggio, collocatein ordine alfabetico, senza sezioni separate.

È necessario indicare esclusivamente le opere citate nel saggio.

Quanto allo stile:

Titolo ***Riferimenti bibliografici***: corsivo, grassetto, 12, spazio successivo

Nei riferimenti bibliografici: corpo 10, con rientri (sporgente 1 cm)

Volume:

Rossi M., 2014, *Titolo in corsivo,* Città, Editore.

a. niente virgola tra cognome e nome; tutte virgole tra i vari elementi, punto finale;
b. se ci sono più di 3 autori, segnare il primo seguito da *et al.*;
c. curatela senza virgola antecedente: **Bianchi G. (a cura di), 2015, *Titolo del libro,* Città, Editore.**
d. l'anno *dell'edizione originale* non va tra parentesi ma tra virgole, dopo

eventuale (a cura di); l'anno deve essere quello; eventuale traduzione va indicata di seguito, (tra parentesi),

e. il titolo deve essere quello originale;
f. città in italiano;
g. nome editore, senza 'editore', 'edizioni', ecc.

Saggio:

Rossi M., 2014, "Titolo in tondo", in Bianchi G. (a cura di), *Titolo dellibro in corsivo*, Città, Editore, pp. 8-20.

Rossi M., Bianchi G., 2014, "Titolo in tondo", in *nome della rivista in corsivo*, vol. 4, n. 2 (oppure: nn. 1-2), pp. 8-20.

Non serve indicare il volume della rivista, ma se si mette in una entry, va in tutte.

È necessario indicare le pagine dei saggi.

On line

Volumi o saggi reperibili anche on line (in siti stabili, quali ad esempiole versioni on line delle riviste) possono indicare, dopo il punto che seguel'editore o il numero della rivista, la URL dove si può trovare il saggio e la data dell'ultimo accesso.

I saggi vanno spediti a: matteo.santipolo@unipd.it, alberta.novello@unipd.it

STYLE SHEET *RILA*

By sending an essay, the author states that it has not been published in or submitted to other journals.
Linguistic correctness is responsibility of the author, no linguistic editing is provided by RILA.
Essays are submitted in .dox or .doc formats, Times New Roman 12; a pdf copy is welcome, to check layout in case of doubts. Unless specific problems arise, no proofreading by the author is required.
Tables, diagrams, pictures: please consider the width of the page: 11 centimetres; RILA does not use colours.
Essays should not exceed 35.000 characters, spaces included. If tables, pictures, diagrams are included, the number of characters must be proportionally reduced.
A 120-word long abstract in English is required.
Each essay is blind-refereed by 1 member of the scientific committee and, if accepted, by 2 more blind reviewers.
Essays are to be sent to matteo.santipolo@unipd.it, alberta.novello@unipd.it

Layout

a. E-mail and address of the author;
b. Title;
c. Author in SMALL CAPITAL (ctrl+maiusc+K) and affiliation in *italics*;
d. In the case of two or more authors, specify the authorship of each paragraph in a footnote;
e. Abstract: no more than 120 words, Times New Roman 12, in italics, indented of 1,25 cm;
f. ***Abstract*** (just the title) in *italics* and **bold**;
g. No keywords;
h. Every paragraph should be indented of 1,25 cm, both in the text and in the footnotes;
i. Titles at the first level: ***2. The complexity***...;
j. Titles at the second level: *2.1 The complexity*...;
k. Titles at the third level: 2.1.1 The complexity ...;
l. Titles: double space before and a single space after titles;
m. Between the titles write at least one paragraph;
n. Optional appendix: before the references.

Style

a. Foreign words in *italics*;
b. commas: quotations"……."; emphasising words: '…'; no « »;
c. no **bold** in the text, only in the titles of the paragraphs;

d. line-spacing: 1; Layout: 0 and 0;
e. lists: a single space before and a single space after lists;
f. quotations: over 2-3 lines, separate it from the body of the text with double spacing and reduced font (11);
g. capital letters: acronym such as CLILand publishers such as UTET should be written in small capital (ctrl+maiusc+K); CAPITAL LETTERS as well as **bold** should be avoided;
h. footnotes should not be used for references, but only for additional information, Times New Roman, 10;
i. Captions: after figures and tables, Times New Roman 10.

References

- (Rossi 1998), Rossi 1993; 1994), (Rossi 2006a; 2006b);
- in the case of quotations: (Rossi 1998: 134-135);
- 2 authors: (Rossi, Bianchi 1998); more than 2 authors: (Rossi *et al.* 1998).

Bibliography

Indicate only the works mentioned in the essay.

Volumes:

Rossi M., 2014, *Title in italics*, Town, Publisher. [Only commas; full stop at the end]

a. 3 or more authors: Rossi M. *et al.*, 2014, ……..
b. edited books: Rossi M. (ed. / eds.), 2014, ……..
c. use original title and date of the original publication, no titles and date of translations

Essays:

Rossi M., 2014, "Title. No italics",in Bianchi G. (ed.), *Title of the book,* Town, Publisher, pp. 8-20.

Rossi M., 2014, "Title. No italics",in *Journal in italics*, vol. 4, n. 2, pp. 8-20.

Pages required, No number of volume required

On line:

Rossi M.,2014, "Title. No italics", in Bianchi G. (ed.), *Title of the book,* Town, Publisher. URL: www.paolobalboni.info retrieved on March 25th, 2017.

Rassegna Italiana di Linguistica Applicata

Quadrimestrale di ricerca linguistica e glottodidattica

Redazione	Matteo Santipolo
Amministrazione	Bulzoni editore srl
	Via dei Liburni, 14 - 00185 ROMA
	Tel. 06 4455207 - fax 06 4450355

Abbonamenti

Italia	€ 50,00
Estero	€ 85,00
Un fascicolo	€ 20,00
Fascicolo doppio	€ 35,00

Versamento su c.c.p. n. 31054000 intestato a:
Bulzoni editore srl - via dei Liburni, 14 - 00185 Roma
I fascicoli non pervenuti all'abbonato devono essere reclamati esclusivamente entro 30 giorni dal ricevimento del fascicolo successivo.

Stampa: DOMOGRAF sas
Circonvallazione Tuscolana, 38 - 00174 Roma

Autorizzazione del Tribunale di Roma n. 290-93 dell'8 luglio 1993

Referees

Ogni contributo destinato ai fascicoli della rivista Rassegna Italiana di Linguistica Applicata, è sottoposto dal Comitato scientifico alla valutazione di due referees, rispettando il criterio dell'anonimato

Finito di stampare nel mese di ottobre 2022
dalla tipografia DOMOGRAF - Roma

BULZONI EDITORE PERIODICI

ARIEL

Semestrale di drammaturgia
dell'Istituto di studi Pirandelliani

fondata nel 1986
diretta da Paolo Petroni

BIBLIOTECA TEATRALE

Trimestrale di Studi e Ricerche sullo Spettacolo
fondata nel 1971

diretta da
Ferruccio Marotti e Cesare Molinari

IMAGO

studi di cinema e media

Semestrale promosso e curato dal
Dipartimento di Storia dell'Arte e Spettacolo
Sapienza - Università di Roma
e dal Dipartimento Comunicazione e Spettacolo
dell'Università Roma Tre

fondata nel 2010
diretta da Enrico Menduni

LETTERATURE D'AMERICA

Trimestrale dell'Università di Roma
"Sapienza" Facoltà di Scienze Umanistiche

fondata nel 1979 da Dario Puccini
diretta da Ettore Finazzi Agrò

RASSEGNA ITALIANA DI LINGUISTICA APPLICATA

Quadrimestrale di linguistica

fondata nel 1969 da Renzo Titone
direttori scientifici
Paolo E. Balboni e Matteo Santipolo

STUDI (E TESTI) ITALIANI

Semestrale del dipartimento di Studi
Greco-Latini, Italiani, Scenico-Musicali
Sapienza - Università di Roma

fondata nel 2000
diretta da Beatrice Alfonzetti

TEATRO E STORIA

(Nuova serie)

Annuale

diretta da Mirella Schino

€ 20,00