

L'EVOLUZIONE DEI TÓPOI NELLA TRADIZIONE LETTERARIA ITALIANA

Autori:

Giorgio CARBONE matricola n. 811974

Gianluca CAVALLARO matricola n. 826049

Remo MARCONZINI matricola n. 883256



/ Il topos letterario

❑ I tópoi letterari:

- Tòpos: “Luogo comune”
- Repertorio di **costanti tematiche e formali** della **letteratura occidentale e italiana**
- Tópoi come **strumento** per tramandare la tradizione letteraria
- I tópoi attraversano la storia e le fasi letterarie, ma **mutano** in forma e **interpretazione**



/ Obiettivi e domande di ricerca

❑ Obiettivi:

1. **Generare** dei **corpora** da una **collezione di testi** ottenuti da **fonti eterogenee**
2. **Apprendere** i **word embeddings** dai corpora generati e **processati**, mediante **word2vec** e **CADE**
3. **Analizzare** alcuni **tópoi** particolarmente longevi

❑ Domande di ricerca:

1. Come i **tópoi letterari** **mutano** nella **storia** ?
2. Come le **correnti letterarie** plasmano i **tópoi** ?
3. Dati alcuni **tòpos** peculiari dei **grandi autori**, a cosa **corrispondono** nelle opere di altri autori ?

/ I dati: il corpus linguistico SCRIPTA

- ❑ La fonte dati principali è stata il **corpus linguistico SCRIPTA** del **Prof. Michele Giordano**
- ❑ **3111 testi** di letteratura italiana, pubblicati fra il **1224** e il **1922**
- ❑ **736 autori unici**
- ❑ **133'000'000 vocaboli**
- ❑ **Integrato** con testi successivi al 1922

	ID_VOCABOLO	ID_OPERA	VOCABOLO	SEPARATORE
0	243294066	2	Canto	S1P1P
1	243294067	2	Le	S
2	243294068	2	donne	,S
3	243294069	2	i	S
4	243294070	2	cavallier	,S

ID_OPERA	TITOLO	AUTORE	ANNO_PUBBLICAZIONE	GENERE
2	Orlando furioso	Ludovico Ariosto	1532.0	poema cavalleresco
3	Novelle	Matteo Bandello	1554.0	novella o racconto
4	Decameron	Giovanni Boccaccio	NaN	novella o racconto
5	Orlando innamorato	Matteo Maria Boiardo	1494.0	poema cavalleresco
6	Margherita Pusterla	Cesare Cantù	1838.0	romanzo

/ Workflow

- ❑ Data **integration**
- ❑ Data **augmentation**
- ❑ **Generazione** delle **collezioni** di corpora
 - **Ricostruzione** dei **testi** dal vocabolario SCRIPTA
 - Suddivisione in diverse **collezioni** di corpora
- ❑ Corpora **preprocessing**
- ❑ **Addestramento** dei modelli
- ❑ **Analisi**

/ Data integration e Data augmentation

□ Data integration

- **Integrazione** delle tabelle di SCRIPTA relative ad **autori, generi e opere**
- **Selezione e integrazione** di 72 romanzi pubblicati dopo il **1922**

□ Data **augmentation** opere

- Definizione del **periodo storico** di pubblicazione
- Definizione della **corrente/fase letteraria** di appartenenza

ID_OPERA	TITOLO	AUTORE	PERIODO_STORICO	FASE_LETTERARIA
2	Orlando furioso	Ludovico Ariosto	1454_1559_Rinascimento	4_boccaccio_e_umanesimo
3	Novelle	Matteo Bandello	1454_1559_Rinascimento	NaN
4	Decameron	Giovanni Boccaccio	1250_1454_Tardo_Medioevo	4_boccaccio_e_umanesimo
5	Orlando innamorato	Matteo Maria Boiardo	1454_1559_Rinascimento	3_petrarchismo
6	Margherita Pusterla	Cesare Cantù	1814_1860_Risorgimento	8_romanticismo

/ Generazione delle collezioni di corpora

- ❑ Rigenerazione dei **testi** dai **vocaboli**
- ❑ Creazione di **3 collezioni di corpora**
 - Corpora di testi suddivisi per **periodo storico**
 - Corpora di testi suddivisi per **fase/corrente letteraria**
 - Corpora di testi prodotti da alcuni **importati autori**

Autori

Alessandro Manzoni
Dante Alighieri
Dino Buzzati
Francesco Petrarca
Gabriele D'Annunzio
Giacomo Leopardi
Giovanni Boccaccio
Giuseppe Parini
Italo Calvino
Italo Svevo
Ludovico Ariosto
Luigi Pirandello
Pier Paolo Pasolini
Primo Levi
Torquato Tasso
Ugo Foscolo
Umberto Eco
Vittorio Alfieri

Fasi letterarie

1_duecento_poetica_e_prosa	Medievo
2_dolce_stil_novo	
3_petrarchismo	
4_boccaccio_e_umanesimo	
5_manierismo_e_barocco	
6_classicismo_arcadia	
7_illuminismo_neoclassicismo	
8_romanticismo	
9_verismo	
10_decadentismo	
11_avanguardie_primo_novecento	
12_neorealismo	

Periodo storico

1250_1454_Tardo_Medioevo
1454_1559_Rinascimento
1559_1707_Seicento
1707_1796_Settecento
1796_1814_Periodo_napoleonico
1814_1860_Risorgimento
1861_1914_L'Italia_liberale
1915_1922_Italia_nella_Prima_Guerra_Mondiale_e_dopoguerra
1922_1945_Ventennio_Fascista_Seconda_Guerra_Mondiale
1945_presente_La_Repubblica_Italiana

/ Pre-processing

❑ Per ogni corpus generato, sono state effettuate le seguenti operazioni di pre-processing:

- Tokenizzazione
- Conversione in lower case
- Rimozione dei caratteri non alfanumerici
- Rimozione della punteggiatura
- Rimozione delle stopwords
 - Estensione della lista [6]
 - Problema forme arcaiche

/ Lemmatizzazione

- ❑ Diverse librerie Python disponibili
 - NLTK
 - Spacy
 - Simplema
- ❑ Confronto risultati della funzione di pre-processing con e senza lemmatizzazione
 - Utilizzando tutte le librerie
 - Campionamento dei corpus
- ❑ Analisi dei risultati
 - Numero di parole **totali**
 - Numero di parole **uniche**
 - Parole più **frequenti**

/ Lemmatizzazione

- ❑ Pre-processing con lemmatizzazione **w/NLTK**
 - ❑ Risultati **analoghi** al pre-processing senza lemmatizzazione
- ❑ Pre-processing con lemmatizzazione **w/Simplemma**
 - ❑ Minor numero di parole totali
 - ❑ Riduzione drastica del numero di parole uniche
- ❑ Pre-processing con lemmatizzazione **w/Spacy**
 - ❑ Minor numero di parole totali
 - ❑ Riduzione drastica del numero di parole uniche

Senza lemmatizzazione

Parole: 344934
Parole uniche: 50367
Parole più comuni:
[('due', 1297), ('senza', 1216), ('fra', 1150), ('quando', 965), ('ogni', 951), ('così', 950), ('me', 914), ('sempre', 871), ('cosa', 838), ('ancora', 801)]

Lemmatizzazione con NLTK

Parole: 344902
Parole uniche: 50262
Parole più comuni:
[('due', 1297), ('senza', 1216), ('fra', 1150), ('quando', 965), ('ogni', 951), ('così', 950), ('me', 919), ('sempre', 871), ('cosa', 838), ('os', 802)]

Lemmatizzazione con Simplemma

Parole: 360875
Parole uniche: 29090
Parole più comuni:
[('essere', 12083), ('avere', 6158), ('fare', 2793), ('potere', 1755), ('sapere', 1413), ('volere', 1341), ('vedere', 1319), ('due', 1297), ('dire', 1297), ('cosa', 1267)]

Lemmatizzazione con SpaCy

Parole: 344934
Parole uniche: 38502
Parole più comuni:
[('potere', 1650), ('fare', 1618), ('tutto', 1334), ('due', 1297), ('dire', 1274), ('cosa', 1266), ('vedere', 1259), ('volere', 1246), ('sapere', 1224), ('senza', 1216)]

/ Lemmatizzazione

❑ NLTK

- Invarianza del numero di parole totali e uniche rispetto a non lemmatizzare
- Stesse parole frequenti rispetto a non lemmatizzare

❑ Spacy e Simplemma

- Riporta i verbi alla loro forma infinita che diventano le più frequenti
- Riduzione frequenza relativa parole utili per le successive analisi

❑ Librerie adottate

- **Problema di affidabilità**
 - Per la lingua italiana
 - Problema evoluzione della lingua italiana

❑ Per questi motivi abbiamo deciso di procedere **senza lemmatizzare**

/ Generazione di Bigrammi

❑ Motivazioni:

- Alcuni topos sono **difficilmente rappresentabili** con una singola parola
- Word2Vec accetta solamente uni-grammi

❑ Libreria **Gensim**

- Generazione tramite **metodo Phrases** [8]
 - Non considera la lingua
 - Considera la frequenza di parole accostate

/ Generazione di Bigrammi

- ❑ Per ogni **diversa suddivisione** della collezione di testi
 - Unione di tutti i testi in un unico corpus
 - Generazione dei bi-grammi sull'intero corpus
 - Aumento della consistenza dei bi-grammi generati
 - Utilizzo dei bi-grammi individuati nella fase di pre-processing
- ❑ Addestramento dei modelli **Word2Vec** e **CADE**:
 - ❑ Nessun miglioramento dei risultati
 - ❑ Esposizione dei risultati senza considerare i bi-grammi

/ Addestramento dei modelli

- ❑ Corpus processati senza **bi-grammi**
- ❑ Vengono creati modelli sia non allineati che allineati, utilizzando gli algoritmi di **Word2Vec** e **CADE**
- ❑ Per natura, i word embeddings sono stocastici: per avere dei risultati più attendibili decidiamo di addestrare **5 word embeddings per ciascun corpus** combinando i risultati
- ❑ Utilizziamo il metodo **Skip-Gram**, il più adatto a task di tipo semantico [2]
 - ❑ È stata usata la strategia **Skip-Gram Negative Sampling**, preferito generalmente per la maggior affidabilità nella gestione di parole poco frequenti
- ❑ Basandoci su [4], sono stati selezionati i valori dei restanti parametri

```
model = Word2Vec(sentences = sentences,  
                 #window = 5, default value  
                 min_count=10, #not consider word with absolute frequency <10  
                 vector_size=300, #vector size  
                 sg = 1, #skipgram algorithm  
                 hs = 0,  
                 negative = 5, #negative sampling with 5 noise words  
                 workers = 5, #faster process  
                 epochs = 6 #6 iterations  
                 )
```


/ Domanda 1 e 2

Come i tópoi letterari più longevi mutano nei diversi periodi storici?
Il contesto storico-culturale influenza le tematiche ricorrenti ?

Come i canoni propri delle diverse correnti letterarie della
letteratura italiana plasmano la rappresentazione di questi temi
comuni ?

/ Domanda 1 e 2 - considerazioni

- ❑ Analisi condotta sia sui corpus non allineati che allineati tramite CADE. Si mostrano i risultati ottenuti sui modelli non allineati, in quanto più significativi
- ❑ I *tòpos* sono stati analizzati sia attraverso diverse **correnti letterarie** che diversi **periodi storici**
- ❑ L'analisi è stata condotta partendo da un insieme di più parole per ciascun *tòpos*

/ Il tòpos del pastore

- ❑ Parola ricercata: **pastore**
- ❑ Conclusioni interessanti dall'analisi attraverso i **periodi storici**
- ❑ La figura del pastore è legata tanto all'ambito rurale quanto al mondo religioso
- ❑ Nel **tardo medioevo**, le parole più simili a pastore sono legate alla **sfera religiosa**

Tardo medioevo

	Words	Abs.frequency	Med.similarity
0	gattivo	5	0.656023
1	battezzato	5	0.662622
2	ubbidiente	5	0.656196
3	riverito	1	0.640724
4	apostolica	5	0.651634
5	fedelissimo	1	0.626735
6	persecutore	5	0.646884
7	cattolico	3	0.635160
8	religioso	1	0.623095
9	ispirito	1	0.622285

/ Il tòpos del pastore

- ❑ Successivamente, la figura del pastore inizia ad essere associata al mondo rurale. Appaiono aggettivi interessanti come **bifolco**, **umile** e **mansueto**

Seicento

	Words	Abs.frequency	Med.similarity
0	pastor	5	0.643060
1	gregge	5	0.580032
2	ovile	5	0.566644
3	pecorella	5	0.567463
4	greggia	4	0.540363
5	pastori	4	0.550490
6	scorticar	4	0.548306
7	cultor	2	0.541288
8	mandra	4	0.536670
9	agnelle	1	0.535007

Rinascimento

	Words	Abs.frequency	Med.similarity
0	pastor	5	0.625396
1	gregge	5	0.590203
2	ovile	5	0.574538
3	bifolco	5	0.575675
4	lupo	5	0.554852
5	greggie	3	0.558754
6	mandra	1	0.537980
7	armento	2	0.542123
8	androgeo	3	0.537378
9	pecorella	4	0.543468

Settecento

	Words	Abs.frequency	Med.similarity
0	greggia	5	0.560435
1	gregge	3	0.550876
2	tutelare	5	0.571766
3	mansueto	4	0.540882
4	pastor	4	0.548373
5	venerando	2	0.538153
6	giuda	3	0.545793
7	umile	2	0.527816
8	liberator	1	0.533250
9	rispettava	5	0.536125

/ Il tòpos del pastore

- ❑ Nei periodi più recenti, si ritorna alla sfera religiosa. Appaiono aggettivi più dispregiativi quali **porcaro**, **pecoraio** e **servo** fino ad arrivare a **frocio** e **travestito**

Prima Guerra Mondiale

Interessante la presenza di **eumèo**, il «fedele porcaro di Ulisse»

	Words	Abs.frequency	Med.similarity
0	arciprete	2	0.868132
1	annibale	4	0.863771
2	stola	4	0.861070
3	ciro	3	0.862614
4	cantoniere	2	0.861428
5	it	1	0.866589
6	accolsero	1	0.865788
7	romelli	3	0.864293
8	pulpito	1	0.862662
9	addormentava	1	0.862625

	Words	Abs.frequency	Med.similarity
0	anglicano	5	0.722219
1	eumèo	5	0.702061
2	mendico	5	0.738189
3	priamo	4	0.685533
4	avo	5	0.693021
5	eolo	1	0.672593
6	fratel	5	0.688140
7	prisco	1	0.669029
8	nipotino	1	0.668221
9	telemaco	1	0.667017

Ventennio fascista

Italia liberale

	Words	Abs.frequency	Med.similarity
0	mandriano	5	0.585911
1	pastori	5	0.525099
2	porcaro	5	0.531320
3	ameto	5	0.502250
4	contadino	4	0.512448
5	boaro	1	0.502836
6	pecoraio	2	0.501695
7	fraticello	2	0.502071
8	pecore	3	0.506837
9	nuorese	1	0.489044

Secondo dopoguerra

	Words	Abs.frequency	Med.similarity
0	frocio	4	0.704758
1	travestito	4	0.724651
2	paladino	4	0.702842
3	santissimo	1	0.708295
4	probo	4	0.708905
5	investitura	1	0.706455
6	servitore	4	0.709846
7	agnello	3	0.716405
8	adottivo	1	0.702527
9	servo	4	0.703339

/ Il tòpos dell'amore

- ❑ Parola ricercata: **amore**
- ❑ Conclusioni interessanti sia dall'analisi attraverso i **periodi storici** che attraverso le **correnti letterarie**
- ❑ Dal punto di vista storico, si può notare come l'amore sia descritto come un sentimento **ardente, fervente e onesto**

Seicento

	Words	Abs.frequency	Med.similarity
0	amor	5	0.729348
1	pudico	3	0.526763
2	caritade	5	0.535816
3	onestade	2	0.516753
4	gradisce	3	0.518281
5	fraterno	4	0.522209
6	ispira	2	0.525393
7	impudico	4	0.521346
8	contaminar	2	0.515476
9	onestate	2	0.526299

Rinascimento

	Words	Abs.frequency	Med.similarity
0	amor	5	0.627125
1	fervente	5	0.513907
2	affezione	5	0.472854
3	reciproco	5	0.483924
4	ardore	5	0.489205
5	ferventissimo	2	0.475717
6	amar	4	0.472309
7	affectione	2	0.464967
8	amando	3	0.476206
9	fervidamente	2	0.461463

Risorgimento

	Words	Abs.frequency	Med.similarity
0	amor	5	0.600786
1	affetto	5	0.602020
2	conjugale	5	0.512594
3	ardentissimo	5	0.520028
4	sviscerato	5	0.518879
5	sincerissimo	5	0.505034
6	figliale	5	0.513747
7	amaste	2	0.498638
8	fervente	4	0.505010
9	perdonami	2	0.487524

/ Il tòpos dell'amore

- ❑ Dalla Prima Guerra Mondiale in poi vengono associati anche sentimenti negativi, legati all'**infelicità**, alla **gelosia**, al **tradimento** e al **pentimento**
- ❑ Interessante notare come nel periodo della Prima Guerra Mondiale l'amore sia associato all'**olocausto**

Ventennio fascista

	Words	Abs.frequency	Med.similarity
0	affetto	5	0.798223
1	tradita	3	0.752715
2	infelice	2	0.757698
3	amato	4	0.753022
4	rammarico	4	0.746443
5	dimostrava	1	0.744242
6	divenire	1	0.739707
7	odiato	2	0.747546
8	geloso	2	0.742658
9	sacrificio	3	0.742943

Prima Guerra Mondiale e primo dopoguerra

	Words	Abs.frequency	Med.similarity
0	lavami	5	0.568963
1	passione	4	0.526812
2	dedizione	5	0.525074
3	olocausto	5	0.525433
4	sacrificio	4	0.523065
5	peccatrice	5	0.509652
6	irraggiungibile	2	0.517062
7	fraterno	2	0.511204
8	amai	2	0.507468
9	fervore	1	0.510540

Secondo dopoguerra

	Words	Abs.frequency	Med.similarity
0	amare	5	0.570407
1	coniugale	5	0.562748
2	reciproco	4	0.525914
3	disamore	5	0.531162
4	pentimento	1	0.507092
5	carnale	1	0.506550
6	fraterno	2	0.514101
7	disinteressato	3	0.510428
8	avvilimento	4	0.516641
9	amavo	1	0.503905

/ Il tòpos dell'amore

- ❑ Anche nelle diverse correnti letterarie l'amore continua ad essere descritto come **ardente, fervente e onesto**
- ❑ Due aspetti interessanti:
 - ❑ Nel **decadentismo** compare la parola **morirò**, segno di un amore così intenso da poter portare alla morte
 - ❑ Nelle **avanguardie del primo novecento** l'amore acquisisce connotazioni legate alla **razza** e all'**Italia**

Decadentismo

	Words	Abs.frequency	Med.similarity
0	amor	5	0.581041
1	morirò	5	0.555180
2	affetto	2	0.504631
3	fraterno	2	0.517302
4	amando	4	0.508272
5	triforme	4	0.514105
6	studium	5	0.506558
7	nullo	5	0.515546
8	soverchio	1	0.492016
9	amoroso	4	0.503901

Avanguardie primo novecento

	Words	Abs.frequency	Med.similarity
0	italiana	5	0.921791
1	italia	5	0.911734
2	rivoluzione	5	0.910685
3	realtà	3	0.900816
4	poichè	1	0.906275
5	pericolo	5	0.903349
6	razza	2	0.898662
7	eroismo	1	0.899714
8	senso	3	0.900273
9	cioè	1	0.898895

/ Il tòpos della patria

- ❑ Parola ricercata: **patria, nazione, bandiera**
- ❑ Conclusioni interessanti dall'analisi attraverso i **periodi storici**
- ❑ Netta differenza fra il periodo precedente all'unità d'Italia e quello successivo
- ❑ Prima del 1861, il concetto di patria è legato a quelli di **esilio, libertà e cittadinanza**

Rinascimento

	Words	Abs.frequency	Med.similarity
0	libertà	5	0.482865
1	amatore	2	0.460885
2	casato	3	0.454079
3	mantenerla	5	0.459501
4	tirannide	4	0.455882
5	esaltazione	3	0.458977
6	patrie	5	0.464859
7	mutua	3	0.452173
8	tiranni	2	0.454734
9	rebelle	1	0.444805

Pieno medioevo

	Words	Abs.frequency	Med.similarity
0	essilio	5	0.606277
1	ateniesi	3	0.584054
2	natio	4	0.584883
3	origine	4	0.598769
4	fabrizio	3	0.591597
5	familiarità	3	0.592301
6	publica	5	0.591438
7	odii	3	0.580110
8	benivolenza	1	0.575380
9	democrito	3	0.586928

Settecento

	Words	Abs.frequency	Med.similarity
0	concittadini	5	0.463294
1	patrie	4	0.425037
2	cittadinanza	4	0.437209
3	natio	3	0.419183
4	esule	5	0.439768
5	oriondo	3	0.429205
6	pratovecchio	5	0.426667
7	foscarari	1	0.421881
8	gandino	4	0.424970
9	alidosi	1	0.418074

/ Il tòpos della patria

- ❑ Dopo la creazione del regno d'Italia, tra i termini più simili appaiono **Italia**, **Europa** e una serie di parole legate alla sfera politica

Prima Guerra Mondiale e primo dopoguerra

	Words	Abs.frequency	Med.similarity
0	potestà	5	0.591057
1	indipendenza	5	0.581466
2	nativa	5	0.580083
3	umanità	5	0.555170
4	germani	4	0.567030
5	risorta	3	0.553034
6	destini	3	0.565195
7	prosperità	1	0.549767
8	combattendo	1	0.548468
9	averi	3	0.548322

Interessante la comparsa del tema della **patria potestà**

Italia liberale

	Words	Abs.frequency	Med.similarity
0	rendici	5	0.538617
1	nazione	4	0.511126
2	patrio	3	0.507752
3	italia	4	0.504642
4	rivendicarsi	5	0.502145
5	libèrta	1	0.482562
6	glorificando	1	0.482107
7	gloria	2	0.486280
8	patrie	5	0.505363
9	esulato	1	0.476940

Ventennio fascista

	Words	Abs.frequency	Med.similarity
0	industriale	2	0.821594
1	europa	5	0.835042
2	feudale	2	0.817969
3	sorti	3	0.816690
4	ricchezza	2	0.805053
5	stampo	3	0.811512
6	costituzione	1	0.805137
7	religione	1	0.804787
8	popolazione	1	0.803570
9	cultura	2	0.803949

/ Il tòpos della patria

- ❑ Le tematiche precedenti sono ritrovate anche ricercando la parola **nazione**

Prima Guerra Mondiale e primo dopoguerra

	Words	Abs.frequency	Med.similarity
0	proletariato	5	0.711746
1	assolutismo	5	0.700160
2	solidarietà	5	0.699730
3	dinastia	5	0.683707
4	produttori	3	0.679977
5	italianità	2	0.677463
6	internazionale	4	0.679095
7	indipendenza	1	0.668382
8	insurrezione	1	0.666635
9	capitalismo	2	0.667522

Italia liberale

	Words	Abs.frequency	Med.similarity
0	nazionalità	5	0.602440
1	plutocrazia	5	0.575933
2	nazioni	1	0.570988
3	oligarchica	3	0.574760
4	plutocratica	4	0.572381
5	elettorato	5	0.568655
6	smembrata	5	0.572323
7	rivendicarsi	3	0.564098
8	costituitasi	4	0.568730
9	clericume	4	0.569969

Ventennio fascista

	Words	Abs.frequency	Med.similarity
0	monarchia	5	0.937684
1	istituzioni	5	0.940522
2	risorgimento	5	0.940497
3	indipendenza	5	0.932305
4	unità	2	0.927950
5	moderna	3	0.937915
6	tradizionale	3	0.929110
7	istituti	5	0.936902
8	conservazione	2	0.928836
9	economia	2	0.933472

/ Il tòpos della patria

- Anche **bandiera** cambia significato con la creazione del regno d'Italia. Prima del 1861 è un **vessillo**, uno **stendardo** da esporre in battaglia...

Seicento

	Words	Abs.frequency	Med.similarity
0	stendardo	5	0.721515
1	armatura	5	0.666671
2	gonfalone	5	0.649081
3	sbarra	5	0.647597
4	trieme	3	0.629505
5	sopraveste	5	0.643175
6	quadriga	2	0.619638
7	stendardi	2	0.620871
8	guidava	3	0.619319
9	dudone	1	0.617214

Rinascimento

	Words	Abs.frequency	Med.similarity
0	stendardo	5	0.662125
1	sopravesta	5	0.638968
2	centuria	4	0.632649
3	arcata	3	0.624930
4	puliano	5	0.634287
5	capitana	4	0.629944
6	azuro	1	0.615904
7	pagana	1	0.610675
8	impugnò	2	0.614007
9	cimero	3	0.621763

Settecento

	Words	Abs.frequency	Med.similarity
0	vessillo	5	0.595248
1	targone	5	0.585567
2	sguainata	2	0.583686
3	visiera	3	0.582426
4	breccia	4	0.559421
5	tagliangli	1	0.566448
6	guernigione	1	0.563009
7	guarnigion	3	0.558097
8	inespugnabil	2	0.558698
9	sciabla	4	0.551551

/ Il tòpos della patria

- ❑ ...poi viene associata al **tricolore** che **sventola**
- ❑ Nel periodo più recente, si nota la presenza anche della bandiera **rossa**, simbolo comunista

Prima Guerra Mondiale e primo dopoguerra

	Words	Abs.frequency	Med.similarity
0	sventola	5	0.754245
1	tricolore	5	0.742548
2	campidoglio	5	0.672696
3	sventolava	5	0.667345
4	valmy	4	0.643893
5	vessillo	5	0.648845
6	effigie	1	0.642892
7	issato	4	0.645523
8	meteora	2	0.645848
9	bastiglia	3	0.636743

Italia liberale

	Words	Abs.frequency	Med.similarity
0	vessillo	5	0.687883
1	stendardo	5	0.652017
2	tricolore	5	0.666985
3	sventola	5	0.623350
4	inalberata	5	0.619926
5	sventolante	5	0.596344
6	sventolava	5	0.609899
7	sventolerà	5	0.606734
8	bandiere	1	0.576424
9	inalberando	1	0.573380

Secondo dopoguerra

	Words	Abs.frequency	Med.similarity
0	pennone	5	0.658225
1	tricolore	5	0.641758
2	bara	5	0.605591
3	rossa	5	0.599896
4	salve	1	0.580357
5	aquila	2	0.574105
6	divisa	1	0.575744
7	parafulmine	3	0.579777
8	bicipite	1	0.568625
9	equestre	4	0.583981

/ Il tòpos della guerra

- ❑ Parola ricercata: **guerra**
- ❑ Conclusioni interessanti dall'analisi attraverso i **periodi storici**
- ❑ Ancora l'unità d'Italia come spartiacque
- ❑ Fino ad allora si narra di **vittorie e sconfitte, imprese e tregue**
- ❑ Compaiono anche riferimenti alle guerre caratteristiche di un certo periodo storico

Periodo napoleonico

	Words	Abs.frequency	Med.similarity
0	tregua	5	0.550863
1	vittoria	5	0.539529
2	armi	5	0.537516
3	contesa	5	0.522226
4	discordia	5	0.518704
5	alleanza	5	0.535604
6	lega	5	0.510569
7	danni	1	0.512611
8	pace	5	0.523061
9	ostilità	4	0.498956

Tardo medioevo

	Words	Abs.frequency	Med.similarity
0	briga	5	0.575874
1	discordia	5	0.546348
2	triegua	5	0.531149
3	vettoria	3	0.526992
4	tregua	3	0.504783
5	annibale	4	0.505494
6	guera	1	0.510054
7	sconfitta	1	0.508322
8	dissensione	5	0.513684
9	fiaminghi	5	0.512231

Risorgimento

Interessante il ritorno alla narrazione delle guerre puniche

	Words	Abs.frequency	Med.similarity
0	punica	5	0.472369
1	difensiva	5	0.445268
2	crociata	1	0.452217
3	accanita	5	0.441144
4	guerre	5	0.458673
5	perdurante	4	0.435383
6	cartagine	4	0.437072
7	macedonia	1	0.434338
8	lutta	5	0.434404
9	turchia	4	0.431034

/ Il tòpos della guerra

Italia liberale

Con l'unità d'Italia
si parla di
secessione,
guerriglia e
insurrezioni

	Words	Abs.frequency	Med.similarity
0	guerreggiata	5	0.595876
1	secessione	5	0.579482
2	irani	5	0.525427
3	battaglia	5	0.518778
4	guerre	5	0.541156
5	paflagonia	4	0.509272
6	guerriglia	3	0.495785
7	commio	2	0.507281
8	insurrezione	3	0.497181
9	radagasio	1	0.496256

Secondo dopoguerra

Tiene banco la
guerra in
Abissinia e il
fascismo

	Words	Abs.frequency	Med.similarity
0	mondiale	5	0.576602
1	abissinia	5	0.560878
2	africa	5	0.537812
3	fascismo	5	0.512780
4	tifo	3	0.498945
5	guerre	5	0.513271
6	spagna	3	0.499556
7	troia	4	0.496712
8	rivoluzione	5	0.514088
9	prigionia	1	0.486818

Prima Guerra Mondiale, primo dopoguerra e Seconda Guerra Mondiale

Compare la
parola
mondiale, oltre
alle nazioni
protagoniste
dello scenario
bellico del
periodo

	Words	Abs.frequency	Med.similarity
0	conflagrazione	5	0.516260
1	europea	5	0.521375
2	mondiale	5	0.514490
3	germania	3	0.482804
4	turchia	4	0.464106
5	libica	3	0.467070
6	sconfitta	4	0.471986
7	lotta	4	0.471593
8	battaglia	2	0.464164
9	neutralità	2	0.459875

	Words	Abs.frequency	Med.similarity
0	italia	5	0.680465
1	russia	5	0.695360
2	europa	5	0.663876
3	germania	5	0.674757
4	africa	5	0.642764
5	europea	5	0.641724
6	battaglia	5	0.640732
7	austria	2	0.610381
8	storia	2	0.609037
9	francia	3	0.618931

/ Il tòpos della donna

- ❑ Parola ricercata: **donna**
- ❑ Conclusioni interessanti dall'analisi attraverso le **correnti letterarie**
- ❑ Nell'**umanesimo**, la **donna** è una **pulcella**, una **giovinetta saggia** e **accorta**
- ❑ Nel periodo **barocco**, la **donna** è dipinta come una figura **verGINE** e **onesta**

Umanesimo

	Words	Abs.frequency	Med.similarity
0	pulcella	5	0.591015
1	costei	5	0.581541
2	damigella	5	0.574900
3	monna	5	0.571966
4	isposa	1	0.557013
5	innamorata	2	0.554839
6	saggia	1	0.549525
7	giovinetta	1	0.549306
8	accorta	4	0.557097
9	fanciulla	4	0.551015

Barocco

	Words	Abs.frequency	Med.similarity
0	costei	5	0.671297
1	verGINE	3	0.663612
2	innamorata	3	0.659375
3	amata	4	0.662101
4	donzella	3	0.655137
5	dama	5	0.651849
6	giovinetta	5	0.661443
7	fanciulla	4	0.659525
8	innamora	1	0.657258
9	onesta	1	0.646197

/ Il tòpos della donna

- ❑ Nel **classicismo** si recupera la figura della **principessa** e della **regina**
- ❑ Nel periodo **illuminista** la donna è **ottima**, **avvenente** e **virtuosa**
- ❑ Nel **romanticismo**, la figura della donna è associata a **Prassede**, personaggio dei Promessi Sposi estremamente **bigotto** e **pudico**

Illuminismo

	Words	Abs.frequency	Med.similarity
0	ottima	5	0.756136
1	avvenente	4	0.751641
2	egregia	5	0.754042
3	adorata	4	0.748626
4	sventura	2	0.747937
5	sposo	5	0.745428
6	virile	1	0.750061
7	virtuosa	2	0.741387
8	unica	5	0.760584
9	sposa	4	0.756029

Classicismo

	Words	Abs.frequency	Med.similarity
0	principessa	5	0.832233
1	augusta	5	0.819673
2	madre	5	0.813131
3	adorata	5	0.798400
4	bradamante	4	0.778203
5	figliuola	1	0.797922
6	moglie	5	0.786023
7	regina	4	0.784104
8	pulcheria	4	0.768101
9	ii	1	0.770598

Romanticismo

	Words	Abs.frequency	Med.similarity
0	prassede	5	0.570562
1	costei	5	0.577325
2	femmina	4	0.543571
3	fanciulla	5	0.539246
4	dama	5	0.552259
5	pudica	5	0.526367
6	livia	1	0.534009
7	leonilda	2	0.527435
8	ragazza	5	0.534135
9	contadina	3	0.518088

/ Domanda 1 e 2 - conclusioni

- ❑ Conclusioni interessanti per alcuni *tòpos* letterari
- ❑ I *tòpos* mostrati sono quelli più semplici, facilmente connotabili dal punto di vista storico o letterario
- ❑ Nonostante numerosi tentativi, su *tòpos* più complessi non si è arrivati a conclusioni interessanti
- ❑ Il contenuto estremamente variegato di ciascun corpus ha reso complicata l'individuazione di caratteristiche specifiche per *tòpos* complessi
- ❑ Una scelta più specifica dei libri e una suddivisione più accurata dei corpus potrebbe portare a risultati migliori

/ Domanda 3

È possibile, utilizzando corpus allineati di vari autori della tradizione letteraria italiana, individuare corrispondenze tra tópoi o concetti peculiari ?

/ Domanda 2 - considerazioni

- ❑ Analisi condotta considerando sui corpus di diversi autori allineati tramite CADE
- ❑ Per ogni autore sono stati valutati i concetti e personaggi più rappresentativi

/ La maschera di Pirandello

- ❑ Per Luigi Pirandello la **maschera** è associata alla **frantumazione dell'io** e all'**adattamento dell'individuo in base al contesto** in cui si trova
- ❑ Si nota come la maschera sia **informe** e **inconsistente**

Luigi Pirandello

	Words	Abs.frequency	Med.similarity
0	informe	5	0.515715
1	inconsistente	1	0.554545
2	bruta	1	0.547342
3	mostrerà	1	0.539642
4	riveli	4	0.452970
5	nudità	1	0.522253
6	conferiva	3	0.440555
7	distaccare	5	0.435230
8	esisterebbe	1	0.514438
9	fissarla	3	0.465024

/ La maschera di Pirandello

- ❑ Negli altri autori la maschera pirandelliana diventa una **figura**, una **macchia**, un **involucro**. È **compatta**, **impenetrabile** e spesso denota sentimenti di **gelosia** e **inferiorità**

Francesco Petrarca

	Words	Med.similarity
32	volto	0.464169
8	gelosia	0.457815
14	mostri	0.449791
3	bruna	0.447986
17	penso	0.444806

Italo Svevo

	Words	Med.similarity
14	inferiorità	0.533388
28	rigidezza	0.524971
30	rosea	0.508709
20	muovendosi	0.505664
22	oscurava	0.505518

Gabriele D'Annunzio

	Words	Med.similarity
11	indurito	0.543095
4	compatta	0.526775
9	impenetrabile	0.519390
13	magrezza	0.507277
14	maschera	0.503232

Pier Paolo Pasolini

	Words	Med.similarity
10	indecente	0.541669
8	goffa	0.534675
4	cripta	0.533417
15	pettinatura	0.524027
17	rendendolo	0.523083

Dante Alighieri

	Words	Med.similarity
13	figura	0.484128
14	immobile	0.475716
22	signoreggia	0.467974
2	ascosa	0.453767
9	derisa	0.452487

Giacomo Leopardi

	Words	Med.similarity
9	fissa	0.503174
4	circoscritta	0.497070
20	palpabile	0.494521
13	malvagia	0.493976
0	abramo	0.491326

Torquato Tasso

	Words	Med.similarity
13	imagin	0.465204
16	larva	0.458832
10	dogliosa	0.457999
2	angeletta	0.456444
19	menzogna	0.452474

Dino Buzzati

	Words	Med.similarity
10	macchia	0.544638
22	splendevano	0.518783
24	tettoia	0.513378
25	vastità	0.509161
14	orientale	0.504865

Italo Calvino

	Words	Med.similarity
11	involucro	0.531477
13	lucida	0.518793
9	indeterminata	0.511703
21	rasi	0.506885
10	inumano	0.504574

Vittorio Alfieri

	Words	Med.similarity
6	avvilita	0.490364
3	ambiziosa	0.481943
15	ingiusta	0.481730
22	passion	0.476675
4	appaia	0.462176

/ I bravi di Manzoni

- ❑ **Bravi** è un nome noto grazie ai **Promessi Sposi**: nel 16° e 17° secolo erano così chiamati gli **sgherri** al servizio dei signori, spesso **esecutori di ordini e delitti**

Alessandro
Manzoni

	Words	Abs.frequency	Med.similarity
0	vagabondi	5	0.487691
1	accompagnamento	3	0.421896
2	forche	1	0.457767
3	scendono	5	0.421966
4	scompigliata	1	0.440509
5	servitori	1	0.436894
6	trasmutato	1	0.436462
7	travestiti	5	0.422109
8	ronda	2	0.415100
9	camminar	1	0.431270

/ I bravi di Manzoni

- ❑ Per gli altri autori i bravi continuano ad essere dei **servitori**, degli **aiutanti**, in alcuni casi chiamati **muli** o **mastini**
- ❑ Molto interessante la corrispondenza per **Primo Levi**: i bravi diventano gli **ufficiali** e i **soldati** del campo di concentramento di Auschwitz

Luigi Pirandello

	Words	Med.similarity
14	mastini	0.413438
3	bravi	0.394405
0	abituati	0.391045
7	carrettiere	0.388375
16	monelli	0.386815

Primo Levi

	Words	Med.similarity
32	ufficiali	0.394529
28	soldati	0.374173
35	vivi	0.373020
9	dozzina	0.365239
10	fila	0.364154

Dino Buzzati

	Words	Med.similarity
8	bifolchi	0.380241
19	fermi	0.372929
34	valloncello	0.363631
11	cisterna	0.362947
20	formidabili	0.361768

Giacomo Leopardi

	Words	Med.similarity
16	guardassero	0.398753
22	polizia	0.394731
28	servitori	0.393991
7	colloquio	0.390986
10	declamare	0.389572

Ugo Foscolo

	Words	Med.similarity
30	stretto	0.374571
5	chiesto	0.373720
24	servidori	0.373658
18	patibolo	0.373608
4	chiamando	0.370432

Vittorio Alfieri

	Words	Med.similarity
30	ridotti	0.405271
0	aiutanti	0.398692
21	padroni	0.394841
13	eroi	0.382368
10	conversi	0.378412

/ Domanda 2 - conclusioni

- ❑ Nonostante qualche corrispondenza, l'analisi ha portato a risultati poco soddisfacenti
- ❑ Non è stato possibile stabilire corrispondenze fra personaggi
- ❑ Probabilmente, una conoscenza più approfondita del pensiero di ciascun autore permetterebbe di identificare meglio personaggi e concetti da analizzare per ottenere risultati più significativi

/ Conclusioni e sviluppi futuri

Conclusioni

- ❑ Le analisi condotte non hanno portato ai risultati sperati
- ❑ I corpus molto variegati hanno complicato l'identificazione delle caratteristiche dei diversi *tòpos*, specialmente per quelli più complessi
- ❑ Una selezione dei testi più restrittiva e una suddivisione più accorta dei libri nei diversi periodi storici e letterari potrebbe portare a conclusioni più significative

Sviluppi futuri

- ❑ Utilizzo di algoritmi, metodi e parametri diversi per l'addestramento dei modelli
- ❑ Approfondire l'utilizzo delle *word phrases* nei modelli
- ❑ Miglioramento della qualità dei corpus con operazioni mirate, per esempio:
 - Migliorare la gestione della lingua italiana antica (ampliamento lista di stopwords)
 - Utilizzo di parafrasi per testi più antichi

/ Riferimenti bibliografici

1. Compass-Aligned Distributional Embeddings For Studying Semantic Differences Across Corpora – Bianchi F., Di Carlo V., Nicoli P. and Palmonari M.
2. Survey of Computational Approaches to Lexical Semantic Change: <https://arxiv.org/abs/1811.06278>
3. SCRIPTA: corpus linguistico letterario: <https://parolescritte.it>
4. 50 tòpoi della letteratura italiana: <http://www.letteratura-italiana.com/pdf/letteratura%20italiana/13%2050%20topoi%20della%20letteratura%20italiana.pdf>
5. Genesini Pietro, Letteratura Italiana 123, Padova, 2022 URL: <http://www.letteratura-italiana.com/pdf/letteratura%20italiana/01%20GENESINI%20Letteratura%20123.pdf>
6. Estensione delle lista di stopword. URL: <https://raw.githubusercontent.com/stopwords-iso/stopwords-it/master/stopwords-it.txt>
7. Storia della letteratura italiana URL: https://it.wikipedia.org/wiki/Storia_della_letteratura_italiana
8. Learning Embeddings For More Than One Word: <https://towardsdatascience.com/word2vec-for-phrases-learning-embeddings-for-more-than-one-word-727b6cf723cf>