

# THE EVOLUTION OF TÓPOI IN THE ITALIAN LITERARY TRADITION

**Authors:**

Giorgio CARBONE no. 811974

Gianluca CAVALLARO no. 826049

Remo MARCONZINI no. 883256



# / The literary topos

## ❑ Literary themes:

- Tòpos: 'commonplace'
- Repertoire of **thematic and formal constants** in **Western and Italian literature**
- Tópoi as a **tool** for passing on the literary tradition
- Theópoi go through history and literary phases, but **change** in form and **interpretation**



# / Objectives and research questions

## ❑ Objectives:

1. **Generating corpora** from a **collection of texts** obtained from **heterogeneous sources**
2. **Learning word embeddings** from corpora generated and **processed**, using **word2vec** and **CADE**
3. **Analysing** some particularly long-lived **types**

## ❑ Research questions:

1. How do literary themes **change** in history?
2. How Do Literary Currents Shape Themes?
3. Given some peculiar tòpos of the great authors, what do **they correspond to** in the works of other authors?



# / The data: the SCRIPTA language corpus

- ❑ The main data source was the **SCRIPTA linguistic corpus** of **Prof. Michele Giordano**
- ❑ **3111 texts** of Italian literature, published between **1224** and **1922**
- ❑ **736 unique authors**
- ❑ **133,000,000 words**
- ❑ **Supplemented** with texts after 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**
- ❑ **Generation of corpora collections**
  - **Reconstruction** of **texts** from the SCRIPTA vocabulary
  - Subdivision into different corpora **collections**
- ❑ Corpora **preprocessing**
- ❑ Model **training**
- ❑ **Analysis**

# / Data integration and data augmentation

## □ Data integration

- **Integration of** SCRIPTA tables on **authors, genres and works**
- **Selection and integration of** 72 novels published after **1922**

## □ Data **augmentation** works

- Definition of the **historical period of publication**
- Definition of **literary current/phase**

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

# / Generation of corpora collections

- ❑ Text regeneration from words
- ❑ Creation of 3 corpora collections
  - Corpora of texts by **historical period**
  - Corpora of texts by **literary phase/current**
  - Corpora of texts produced by some **important authors**

## 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_avanguardia_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**

- ❑ For each corpus generated, the following pre-processing operations were performed:
  - Tokenisation
  - Conversion to lower case
  - Removal of non-alphanumeric characters
  - Removal of punctuation
  - Removing stopwords
    - List extension [6]
    - Problem archaic forms



# / Lemmatisation

---

- ❑ Several Python libraries available
  - NLTK
  - Spacy
  - Simplema
- ❑ Comparison of pre-processing results with and without lemmatisation
  - Using all libraries
  - Corpus sampling
- ❑ Analysis of results
  - **Total** number of words
  - Number of **unique** words
  - Most **frequent** words

# / Lemmatisation

- ❑ Pre-processing with lemmatisation **w/NLTK**
  - ❑ Results **similar** to pre-processing without lemmatisation
- ❑ Pre-processing with lemmatisation **w/Simplemma**
  - ❑ Fewer total words
  - ❑ Drastic reduction in the number of unique words
- ❑ Pre-processing with lemmatisation **w/spacy**
  - ❑ Fewer total words
  - ❑ Drastic reduction in the number of unique words

## Without lemmatisation

```
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)]
```

## Lemmatisation with 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)]
```

## Lemmatisation with 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)]
```

## Lemmatisation with 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)]
```

# / Lemmatisation

---

## ❑ NLTK

- Invariance of the number of total and unique words compared to non-lemmatising
- Same frequent words compared to non-lemmatising

## ❑ Spacy and Simplemma

- It returns verbs to their infinitive form which become the most frequent
- Reducing the relative frequency of useful words for subsequent analysis

## ❑ Adopted libraries

### ▪ **Reliability problem**

- For the Italian language
- Problem Evolution of the Italian Language

## ❑ For these reasons we decided to proceed **without lemmatising**

# / Generation of Bigrams

---

## ❑ Motivation:

- Some topos are **difficult to represent** with a single word
- Word2Vec only accepts uni-grams

## ❑ Gensim Bookshop

- Generation by the **Phrases method** [8]
  - Does not consider language
  - Consider the frequency of juxtaposed words

# / Generation of Bigrams

---

- ❑ For each **different subdivision of** the text collection
  - Union of all texts into one corpus
  - Bi-gram generation over the entire corpus
    - Increased consistency of bi-grams generated
  - Using bi-grams identified in the pre-processing phase
- ❑ Training of **Word2Vec** and **CADE** models:
  - ❑ No improvement in results
  - ❑ Display of results without considering bi-grams

# / Model training

---

- ❑ Corpus processed without **bi-grams**
- ❑ Both unaligned and aligned models are created, using **Word2Vec** and **CADE** algorithms
- ❑ By nature, word embeddings are stochastic: to get more reliable results we decide to train **5 word embeddings for each corpus** by combining the results
- ❑ We use the **Skip-Gram** method, which is best suited to semantic tasks [2].
  - ❑ The **Skip-Gram Negative Sampling** strategy was used, generally preferred for its greater reliability in handling infrequent words
- ❑ Based on [4], the values of the remaining parameters were selected

```
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  
                 )
```



## / Question 1 and 2

How do the most enduring literary themes change in different historical periods? Does the cultural-historical context influence the recurring themes?

How do the canons of the different literary currents in Italian literature shape the representation of these common themes?

# / Question 1 and 2 - considerations

- ❑ Analysis conducted on both unaligned and CADE-aligned corpora. The results obtained on the unaligned models are shown, as they are more significant
- ❑ *The tòpos* were analysed both through different **literary currents** and **historical periods**
- ❑ The analysis was conducted from a set of several words for each *tòpos*

# / The shepherd's *tòpos*

- ❑ Search word: **shepherd**
- ❑ Interesting conclusions from the analysis across **historical periods**
- ❑ The figure of the shepherd is linked as much to the rural as to the religious world
- ❑ In the **late Middle Ages**, the words most similar to shepherd are related to the **religious sphere**

## Late Middle Ages

	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

# / The shepherd's *tòpos*

- Subsequently, the figure of the shepherd began to be associated with the rural world. Interesting adjectives appear such as **hillbilly**, **humble** and **meek**

## Six hundred

	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

## Renaissance

	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

## Eighteenth century

	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

# / The shepherd's *tòpos*

- In more recent periods, there is a return to the religious sphere. More derogatory adjectives such as **swineherd**, **shepherd** and **servant** appear, up to **faggot** and **transvestite**

## World War I

The presence of **eumèò**, the 'faithful swineherd of Ulysses', is interesting

	Words	Abs.frequency	Med.similarity
0	anglicano	5	0.722219
1	eumèò	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

	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

## Twenty years of Fascism

## Liberal Italy

	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

## After the Second World War

	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

# / The *tòpos* of love

- ❑ Search word: **love**
- ❑ Interesting conclusions from the analysis across both **historical periods** and **literary currents**
- ❑ From a historical perspective, one can see how love is described as an **ardent, fervent** and **honest** feeling

## Renaissance

	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

## Six hundred

	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

## 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



# / The *tòpos* of love

- ❑ From the First World War onwards, negative feelings are also associated with **unhappiness, jealousy, betrayal** and **repentance**
- ❑ Interestingly, in the period of the First World War, love is associated with the **holocaust**

## Twenty years of Fascism

	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

## World War I and the early post-war period

	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

## After the Second World War

	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

# / The *tòpos* of love

- ❑ Even in different literary currents, love continues to be described as **ardent**, **fervent** and **honest**
- ❑ Two interesting aspects:
  - ❑ In **decadentism**, the word **morro** appears, a sign of a love so intense that it can lead to death
  - ❑ In the **avant-garde of the early 20th century**, love acquired connotations linked to **race** and **Italy**

## Decadentism

	Words	Abs.frequency	Med.similarity
0	amor	5	0.581041
1	morro	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

## Early 20th century avant-gardes

	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	poiché	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

# / The homeland *tòpos*

- ❑ Search word: **homeland, nation, flag**
- ❑ Interesting conclusions from the analysis across **historical periods**
- ❑ Clear difference between the period before and after Italian unification
- ❑ Before 1861, the concept of homeland was linked to those of **exile, freedom** and **citizenship**

## Renaissance

	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

## Full Middle Ages

	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

## Eighteenth century

	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

# / The homeland *tòpos*

- ❑ After the creation of the Kingdom of Italy, among the most similar terms appear **Italy**, **Europe** and a number of words related to the political sphere

## World War I and the early post-war period

	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

Interesting appearance of the topic of **parental authority**

## Liberal Italy

	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

## Twenty years of Fascism

	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

# / The homeland *tòpos*

- The previous themes are also found by searching for the word **nation**

## World War I and the early post-war period

	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

## Liberal Italy

	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

## Twenty years of Fascism

	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

# / The homeland *tòpos*

- ❑ The **flag** also changed meaning with the creation of the Kingdom of Italy. Before 1861 it was a **banner**, a **standard** to be displayed in battle...

## Six hundred

	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

## Renaissance

	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

## Eighteenth century

	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



# / The homeland *tòpos*

- ❑ ... then it is associated with the **waving tricolour**
- ❑ In the more recent period, the presence of the **red** flag, a communist symbol, is also noticeable.

## World War I and the early post-war period

	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

## Liberal Italy

	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

## After the Second World War

	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

# / The *tòpos* of war

- ❑ Search word: **war**
- ❑ Interesting conclusions from the analysis across **historical periods**
- ❑ Still the unification of Italy as a watershed
- ❑ Until then, stories are told of **victories** and **defeats**, **exploits** and **truces**
- ❑ References to wars characteristic of a certain historical period also appear

## Napoleonic period

	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

## Late Middle Ages

	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

Interesting return to the narrative of the Punic Wars

	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

# / The *tòpos* of war

## Liberal Italy

With the unification of Italy comes **secession**, **guerrilla warfare** and **insurrections**

	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

## After the Second World War

The war in **Abyssinia** and **Fascism** holds sway

	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

## World War I, post-World War I and World War II

The word **world** appears, in addition to the nations that played a leading role in the war scenario of the period

	Words	Abs.frequency	Med.similarity
0	conflagrazione	5	0.516260
1	europa	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	europa	5	0.641724
6	battaglia	5	0.640732
7	austria	2	0.610381
8	storia	2	0.609037
9	francia	3	0.618931

# / The woman's *tòpos*

- ❑ Search word: **woman**
- ❑ Interesting conclusions from the analysis across **literary currents**
- ❑ In **humanism**, **woman** is a **chick**, a **wise** and **shrewd young girl**
- ❑ In the **Baroque** period, **woman** is portrayed as a **virgin** and **honest** figure

## Humanism

	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

## Baroque

	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

# / The woman's *tòpos*

- ❑ **Classicism** recovers the figure of the **princess** and **queen**
- ❑ In the **Enlightenment** period, women are **excellent, attractive** and **virtuous**
- ❑ In **Romanticism**, the figure of the woman is associated with **Prassede**, a character from The Betrothed who is extremely **bigoted** and **demure**

## Enlightenment

	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

## Classicism

	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

## Romanticism

	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

# / Question 1 and 2 - conclusions

- ❑ Interesting conclusions for some literary *tòpos*
- ❑ *The tòpos* shown are the simplest ones, easily connoted from a historical or literary point of view
- ❑ Despite numerous attempts, no interesting conclusions have been reached on more complex *tòpos*
- ❑ The extremely varied content of each corpus made it complicated to identify specific characteristics for complex *tòpos*
- ❑ A more specific choice of books and a more accurate subdivision of the corpus could lead to better results



## **/ Question 3**

**Is it possible, using aligned corpuses of various authors from the Italian literary tradition, to identify correspondences between peculiar *tópoi* or concepts?**

## **/ Question 2 - considerations**

---

- ❑ Analysis conducted considering the corpora of several authors aligned via CADE
- ❑ For each author, the most representative concepts and characters were evaluated

# / Pirandello's mask

---

- ❑ For **Luigi Pirandello**, the **mask** is associated with the **shattering of the ego** and the **adaptation of the individual according to the context** in which he finds himself
- ❑ You can see how the mask is **shapeless** and **insubstantial**

## 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

# / Pirandello's mask

- In other authors, the Pirandellian mask becomes a **figure**, a **stain**, a **shell**. It is **compact**, **impenetrable** and often denotes feelings of **jealousy** and **inferiority**

## Francis Petrarch

	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	cirscritta	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

# / Manzoni's Bravi

---

- ❑ **Bravi** is a familiar name from **The Betrothed**: in the 16th and 17th centuries, **thugs in the** service of lords, often **executors of orders and crimes**, were so called

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

# / Manzoni's Bravi

- ❑ For other authors, the good continue to be **servants, helpers**, in some cases called **mules** or **mastiffs**
- ❑ Very interesting correspondence for **Primo Levi**: the good become the **officers** and **soldiers** of the Auschwitz concentration camp

## 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

## **/ Question 2 - conclusions**

---

- ❑ Despite some matches, the analysis yielded unsatisfactory results
- ❑ It was not possible to establish correspondences between characters
- ❑ Probably, a deeper knowledge of each author's thought would allow better identification of characters and concepts to be analysed for more meaningful results

# / Conclusions and future developments

---

## Conclusions

- ❑ The analyses conducted did not lead to the desired results
- ❑ The very varied corpuses have complicated the identification of the characteristics of the different *tòpos*, especially for the more complex ones
- ❑ A more restrictive selection of texts and a more judicious division of books into different historical and literary periods could lead to more meaningful conclusions

## Future developments

- ❑ Use of different algorithms, methods and parameters for training models
- ❑ Deepening the use of *word phrases* in models
- ❑ Improvement of corpus quality with targeted operations, for example:
  - Improving Old Italian Language Management (expanding the list of stopwords)
  - Use of paraphrases for older texts



# / Bibliographic references

---

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: literary language corpus: <https://parolescritte.it>
4. 50 tòpoi in Italian literature: <http://www.letteratura-italiana.com/pdf/letteratura%20italiana/13%2050%20topoi%20della%20letteratura%20italiana.pdf>
5. Genesini Pietro, Italian Literature 123, Padua, 2022 URL: <http://www.letteratura-italiana.com/pdf/letteratura%20italiana/01%20GENESINI%20Letteratura%20123.pdf>
6. Extension of the stopwords list. URL: <https://raw.githubusercontent.com/stopwords-iso/stopwords-it/master/stopwords-it.txt>
7. History of Italian Literature URL: [https://it.wikipedia.org/wiki/Storia\\_della\\_letteratura\\_italiana](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>