

Contents

Acknowledgements

Acknowledgements	15
------------------------	----

1. Introduction

1. Introduction	21
-----------------------	----

2. Previous Research and Theoretical Framework

2.1 Silver Age: Genre, Novel, and Subgenre of the Novel	33
2.1.1 Introduction	33
2.1.2 Edad de Plata: Silver Age	33
2.1.3 Literary Genre and Novel in this Period	35
2.1.4 The Novel and its Borders	37
2.1.5 Realist and Naturalist Novel	39
2.1.6 Historical and Adventure Novels	41
2.1.7 Comedy Novels	44
2.1.8 Erotic Novels	44
2.1.9 Social Novels	46
2.1.10 One-Author-Labels: <i>Nivola</i> , <i>Greguería</i> , and <i>Episodio Nacional</i>	47
2.2 Genre in Digital Humanities: Methods, Features, and Data Representation	51
2.2.1 Introduction	51
2.2.2 Genre in Computer Science	52
2.2.3 Genre in Digital Humanities	53

2.2.4 The Start of Genre Classification	54
2.2.5 Literary Genre in the Past Years	58
2.2.6 Conclusions and General Patterns	61
2.3 Theory of Genre	63
2.3.1 Basic Distinctions about Concepts and Definitions	65
2.3.2 Models of Genres	68
2.3.3 Macro-Models of Genres	78
2.3.4 Perspectives on Genres	83
2.3.5 Conclusions	90

3. Data: Texts and Metadata

3.1 Corpus of Novels of the Spanish <i>Silver Age</i>: CoNNSA and CoNNSA-canon	95
3.1.1 Introduction	95
3.1.2 Literary Corpora	96
3.1.3 Statistical Population of Authors	97
3.1.4 Statistical Population of Prose	101
3.1.5 Statistical Population of Novels	102
3.1.6 Definition of the Corpora: CoNNSA and CoNNSA-canon	107
3.1.7 Criteria for Selection	111
3.1.8 Digitization Steps	115
3.1.9 Description of the Corpus	116
3.1.10 Gender Distribution	125
3.1.11 Publication of Data	127
3.2 Metadata	131
3.2.1 Introduction	131
3.2.2 Metadata, Distant Reading, and Hypotheses Testing	132
3.2.3 Typology of Metadata	135
3.2.4 Editorial and Administrative Metadata	136
3.2.5 Descriptive Metadata	137
3.2.6 Author's Metadata	139
3.2.7 Text and Genre Metadata	139
3.2.8 Place and Period of the Action	141
3.2.9 Protagonist Metadata	143
3.2.10 Other Literary Metadata	144

3.2.11 Description of the Metadata of CoNSSA	145
3.3 Filtering the Corpus through Classification: Are all Texts in CoNSSA Novels?	155
3.3.1 Introduction	155
3.3.2 Corpus CORDE: <i>Bag of Words</i> and Metadata	155
3.3.3 CORDE 1860–1960 + CoNSSA	159
3.3.4 Binary Class Evaluation of Parameters	160
3.3.5 Multi-Class Evaluation of Parameters	166
3.3.6 Binary Prediction of Genre of Disputed Texts	168
3.3.7 Multi-Class Prediction of Genre of Disputed Texts	171
3.3.8 Conclusions and Modification of the Corpus	173

4. Feature Engineering: Linguistic Annotation and Transformation

4.1 Grammatical, Lexical, Semantic, and Textual Annotation	179
4.1.1 Introduction	179
4.1.2 Format of Annotation	180
4.1.3 Vocabulary and Punctuation	183
4.1.4 Grammatical Annotation	184
4.1.5 Entities	185
4.1.6 Semantic Annotation	186
4.1.7 Pragmatic and Textual Annotation	188
4.2 Transformations of Lexical Data and Linguistic Annotation	191
4.2.1 Introduction	191
4.2.2 Classic Transformation of Frequency of Features	193
4.2.3 Variants	208

5. Analysis of Subgenre Labels

5.1 Evaluation of Consistent Subgenres	223
5.1.1 Introduction	223
5.1.2 Data and Sources	223
5.1.3 Exploration of Sources	226
5.1.4 Exploration of Labels	229

5.1.5 Semantic Mapping of Labels, Description and Standardization trough Thema . .	231
5.1.6 Analysis of Inter-Source Agreement	234
5.1.7 Consistent Labels across Sources	240
5.1.8 Discussion and Proposal of Subgenre Palette	243
5.2 Analysis of Labels of First Editions	247
5.2.1 Description and Comparison of the Labels of the Covers	248
5.2.2 Classification Results Comparing First Publication Labels with other Sources .	249
5.2.3 Association between Author and First Edition Label.	251
5.2.4 Conclusions	255
5.3 The Case of <i>Literary Fiction</i>	257
5.3.1 Introduction	257
5.3.2 Description of <i>Literary Fiction</i>	259
5.3.3 Classification with Linguistic Features	262
5.3.4 Classification with External Metadata	265
5.3.5 Conclusions	267

6. Feature and Labels Selection

6.1 Feature and Parameter Analysis	271
6.1.1 Introduction	271
6.1.2 Feature Evaluation.	273
6.1.3 Knowledge Extraction about Features	278
6.1.4 Parameters Evaluation	285
6.1.5 Conclusions	293
6.2 Identification of Hidden Subgenres	295
6.2.1 Introduction	295
6.2.2 Clustering in Stylometry	297
6.2.3 The Authorial Cue Problem when Clustering Literary Works	300
6.2.4 Evaluation of Parameters of Clustering with Linguistic Features	303
6.2.5 Comparison of Clusters and Evaluation with Metadata	309
6.2.6 Exploration of Meta-Cluster 0: Cluster 50	312
6.2.7 Exploration Meta-Cluster 1: Cluster 217	314
6.2.8 Discussion and Conclusions.	315

7. Analysis of Subgenres

7.1 Classification of the Subgenres	323
7.1.1 Introduction	323
7.1.2 Subgenres as Classes: Multi-Class Classification from Modal Subgenre	324
7.1.3 Subgenres as Categories : Multi-Label Classification	327
7.1.4 Subgenres of Chapters	331
7.1.5 Subgenres as Linguistic and Literary Categories.....	334
7.1.6 Subgenres as Gradual Categories: Probabilities of Classification	341
7.1.7 Conclusions	346
7.2 Easy Genre, Difficult Genre: Why is there Variance in the Results of the Classification?	351
7.2.1 Introduction	351
7.2.2 Year of Publication	352
7.2.3 Number of Texts and Authors	354
7.2.4 Length of Novels	357
7.2.5 Canonicity of the Texts	358
7.2.6 Disagreement on Human Sources	359
7.2.7 Specificity of the Features	361
7.2.8 Correlation and Regression of Variables	363
7.2.9 Conclusions	366

8. Discussion of Tripartite Graph for Genre

8. Discussion of Tripartite Graph for Genre	369
8.1 Introduction	369
8.2 Characteristics of the Genres	370
8.3 Assembling the Graph Model for Genre	372
8.4 The Complete Tripartite Graph Model of Genre: Principles	382
8.5 Evaluation and Interpretation	385
8.6 Further Properties of the Tripartite Graph Model	393
8.7 Using the Tripartite Graph Model of Genre for Descriptions.....	396
8.8 Limitations of the Tripartite Graph Model for Genre	400
8.9 Conclusions	402

9. Conclusion

9. Conclusion	407
---------------------	-----

10. References

10. References	425
----------------------	-----

11. Appendix

11. Empirical Description through the Tripartite Graph Model	447
11.1 Introduction	447
11.2 Realist novel	448
11.3 Social novel	449
11.4 Literary Fiction	449
11.5 Educational Novel	450
11.6 Philosophical Novel	451
11.7 Naturalist Novel	451
11.8 Historical Novel	453
11.9 Memoir	453
11.10 Bucolic Novel	454
11.11 Adventure Novel	455
11.12 Biographic Novel	456
11.13 Modernist Novel	456
11.14 Autobiographical Novel	457
11.15 Comedy Novel	458
11.16 Psychological Novel	459
11.17 Spiritual Novel	459
11.18 Fantasy Novel	460
11.19 War Novel	461
11.20 Dialogue Novel	462
11.21 Mono-Dialogue Novel	463
11.22 <i>Costumbrist</i> Novel	464
11.23 <i>Nivola</i>	464
11.24 Erotic Novel	465
11.25 Episodio Nacional	466

11.26 Poetic Novel	467
11.27 Greguería	468