

TNA ACTIVITY REPORT

Calculation of similarities or distances in Peter Ackroyd's historiographic metafiction and lexical diversity in Dan Brown's straightforward storytelling

Author: Khanim Garayeva

Current position: PhD Candidate

Affiliation: University of Szeged, Hungary

Host institution: Laboratorio de Innovación en Humanidades Digitales (LINHD), UNED

Mentor(s): Dr Salvador Ros Munoz and Alvaro Perez Pozo

Period of stay: April 28 – May 28, 2023.

Project description

The project I undertook at LINHD, UNED within CLS INFRA Fellowship constitutes the basis of one of the chapters of my doctoral dissertation. The primary research of my Ph.D Dissertation is the academic study of literary representations of esoteric currents in the selected three novels by the contemporary British writer Peter Ackroyd and American writer Dan Brown. Though they belong to modern English literature and get inspired by the same pedigree, Ackroyd and Brown differ in the literary traditions they represent and styles of fictional utilisation of esoteric conventions. While Peter Ackroyd rewrites the stories that once existed in English history and literature with his specific interpretation, Dan Brown introduces a new look at the existent (hi)stories by creating their alternatives. Apart from conducting the literary examination of the selected novels, my dissertation includes social and cultural analyses of the interests in esotericism and the reasons for urges or drives that make authors take resources in occultism.

At the same time, I am studying the transition that genres of fiction undergo due to the implementation of esoteric conventions.



I argue that esotericism is causing metamorphoses on the genre level of the novels as the historiography in Peter Ackroyd's books and facts in Dan Brown's fiction are being affected by the applied arcane conventions transforming them into historiographic metafiction and alternative histories with straightforward storytelling, respectively. Selected six novels by these authors constitute the primary literature of my research.

Research questions

With the CLS INFRA Fellowship, I tried to introduce computational text analysis methods into my study of literary representations of esotericism in contemporary English literature. The selected six novels served as the primary corpus of my study. Nevertheless, the research questions I want to address require an additional reference corpus, which I plan to create during my Fellowship.

As Peter Ackroyd and Dan Brown have differing approaches to esotericism, I will posit two different research questions per author. Along with many peculiarities, these authors mainly differ in their writing styles. While Peter Ackroyd's writings inhibit historiographical metafictional narration, Dan Brown conveys his ideas in straightforward storytelling. Hence, first, I would like to study the lexical diversity in Dan Brown's novels and observe the changes, if they happened, with time – from his first novel to the last one. Before becoming a best-selling author, Dan Brown has already been publishing novels. Thus, I would like to explore if criticism of his writing style or his later books' worldwide acknowledgment has caused any transformations in his texts. I am aware of a ready Python library for the study of lexical diversity analysis, nevertheless, I am not sure if it is enough or if more sophisticated digital methods are available to undermine this question.

Regarding Peter Ackroyd's historiographic metafictional novels, the Canadian literary theoretician Susana Onega resembles them to North American "fabulation"¹ and Spanish-American "magic realism."²³ However, I have doubts as Peter Ackroyd's novels cannot be termed any closer to 'realism.' Therefore, I would like to create a reference corpus consisting of the texts belonging to the former two sub-genres and calculate the similarities or distances between them and my primary corpus. As these genres appeared towards the

¹ Fabulation is suggested by Robert Scholes to denote the novels that use history as a pretext for the fictional imagination. A more detailed analysis is in Scholes, Robert. *The fabulators*. New York: Oxford University Press, 1967.

² Magic realism is a sub-genre of fiction representing realism combined with mythical elements.

³ Onega, Susana. 1995. "British Historiographic Metafiction." In *Metafiction*, edited by Mark Currie, 92–103. London/New York: Routledge.

middle of the twentieth century and are not a mainstream style, not many authors have written and published them since then. For this reason, it would not be a time-consuming process to create a corpus consisting of such novels.

Craig's Zeta and Oppose

To assess Dan Brown's novels, we have split his bibliography into two parts. The first part consisted of the books – *Angels and Demons*, *The Digital Fortress* and *The Deception Point*, published until his first best-seller. And the second group had his best-sellers, *The Da Vinci Code*, *The Lost Symbol*, *Inferno*, and *Origin*. These two subcorpora were contrasted using the stylo package *Craig's Zeta* available in RStudio. Craig's Zeta is a method working on equal-sized slices of the textual dataset. This function aims to compare two textual datasets by splitting them into equal sections and checking the occurrence of specific words within these sections. Such comparison between the number of sections in which a particular word appears in subcorpora is performed using Craig's formula derived from Burrows' original Zeta formula (Craig and Kinney, 2009; Burrows, 2007). Following this method, we have included the first group of Dan Brown's novels in the primary set and the second group in the secondary group. To have more accurate results, we also modified a list of stopwords into which we added the special names used in the novels and lemmatized the words.

This function has given us a list consisting of two elements. The first element illustrated words or n-grams that Dan Brown statistically preferred in the primary subcorpus. In contrast, the second element included the terms he avoided in the secondary subcorpus. As this measure is symmetrical, it shows the words preferred in the first set by Brown and which he automatically avoided in the second set.

To visualize the result, we have used the function *oppose* in RStudio. This function conducts a comparative analysis between two provided sets of texts. While the default configuration includes 3000 words into one slice of text, we have reduced it to 1000 words, as my corpus is relatively small. The overall configuration that we used is as below:

```
text.slice.length = 1000
text.slice.overlap = 0
rare.occurrences.threshold = 2
zeta.filter.threshold = 0.1
oppose.method = "craig.zeta"
display.on.screen = TRUE
write.pdf.file = FALSE
```

```

write.png.file = TRUE
use.color.graphs = TRUE
titles.on.graph = TRUE
identify.points = FALSE
visualization = "words"
classification = FALSE
plot.token = ""

```

This plot has produced two lists: one containing words significantly favoured by Brown and another consisting of words considerably less preferred when compared to the other set of texts. This analysis generated two opposing vectors as below:

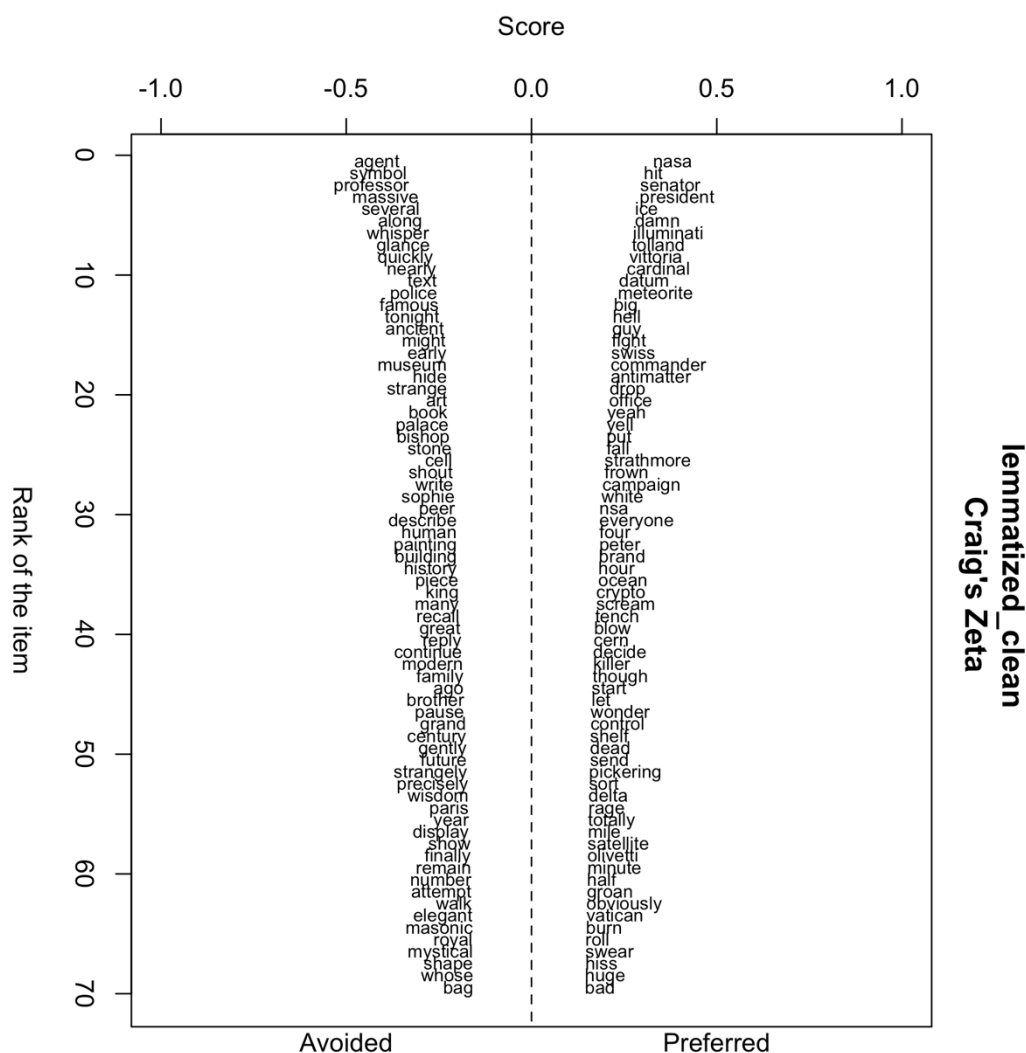


Figure 1: Craig's Zeta. Dan Brown's novels

A similar method called Eder's Zeta was also applied, and the achieved result is:

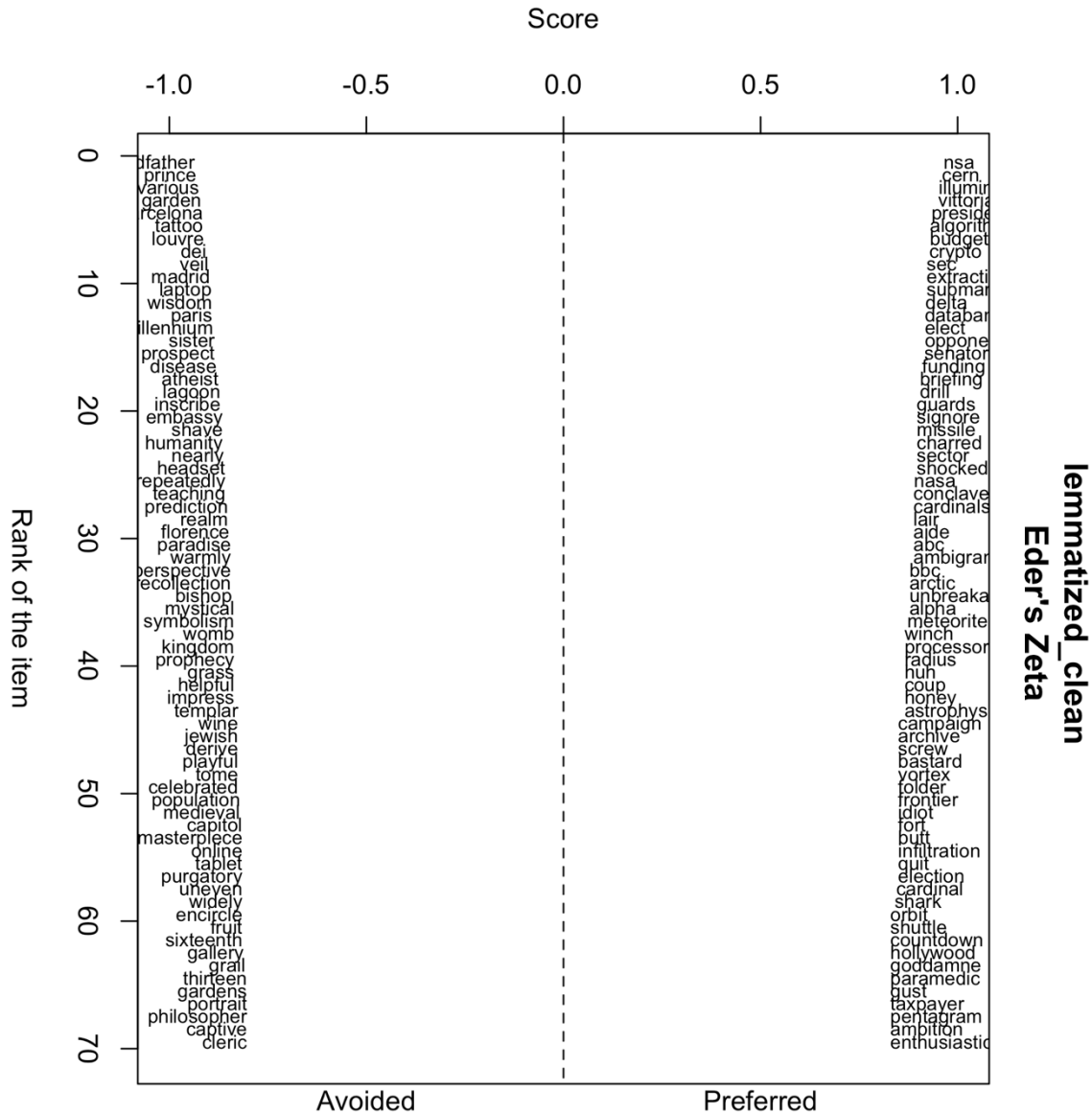


Figure 2: Craig's Zeta. Dan Brown's novels

Topic Modelling: LDA

To address Susana Onega's claim on the similarity of Peter Ackroyd's historiographic metafictional novels and Spanish-American fabulation, I have created a reference corpus consisting of the Spanish stories written before Ackroyd that might serve as inspiration for him. Then, we used these two corpora for topic modelling using Latent Dirichlet Allocation (LDA) in Python. With its statistical modelling technique, topic modelling is constructive for

uncovering the latent or hidden topics within a collection of documents. As it aims to identify the underlying themes or subjects that appear across the corpus of texts, we used the Latent Dirichlet Allocation to mine both datasets. LDA is a widely used topic modelling algorithm based on the assumption that the documents are made up of a mixture of topics, and a distribution of words characterises each topic.

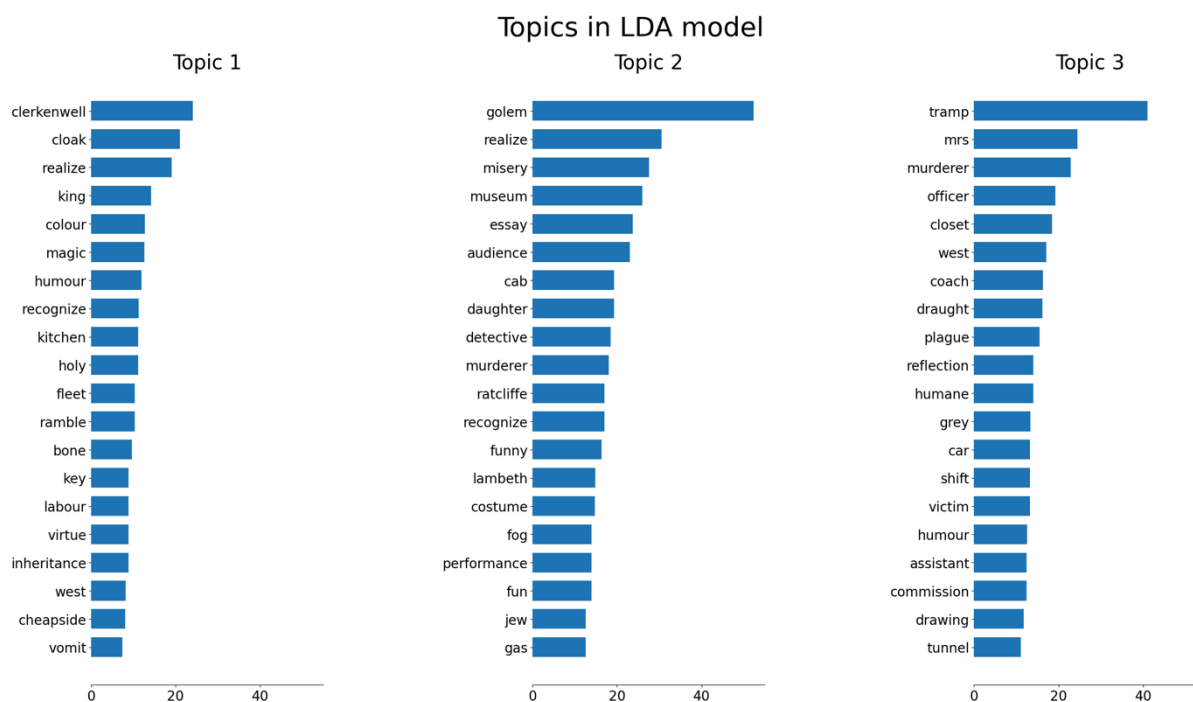


Figure 3: The LDA model of topics in Peter Ackroyd's novels

The LDA algorithm achieves the topics in the above visualisation by iteratively assigning word in documents to topics and adjusting topic-word and document-topic distributions to maximise the likelihood of generating the observed data. At first, we tried to find more topics; however, due to the small volume of the set, we reduced it to only three topics. The same method has been applied to the mixture of Spanish-American texts. However, the outcome was only the topics in English:

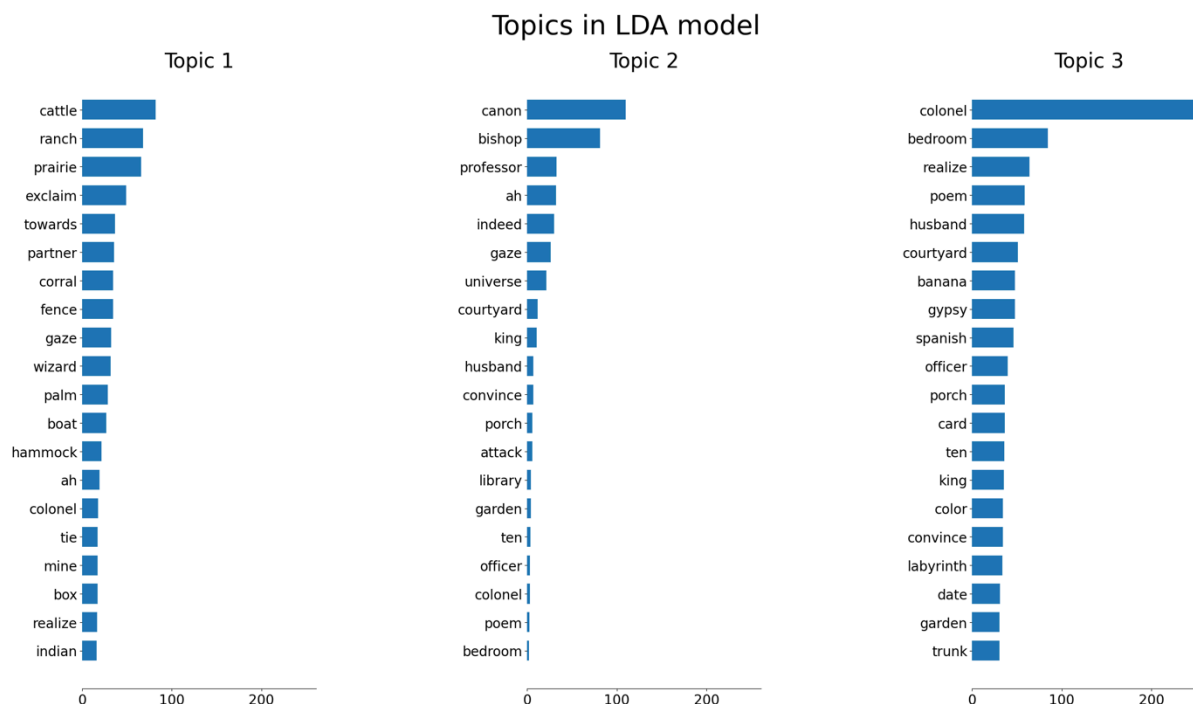


Figure 4: The LDA model of Topics in Spanish-American novels.

As this is a work in progress, we are now trying to make the texts in the Spanish language comparable and relevant for further study. We have tried other settings of LDA like Non-negative Matrix Factorisation's Frobenius norm and generalised Kullback-Leibler divergence or MiniBatchNMF model, and all gave the same results.

While these outcomes that I have showcased in this report by now need further development, the initial stage of this work and the results that we got are much more promising. Other steps include an in-depth lexical and parts of speech analysis of Dan Brown's novels, improvement of the Spanish dataset as a reference corpus for Peter Ackroyd's novels, and their comparative analysis.

In conclusion, I would like to note that I would not have been able to conduct these studies if not for Dr. Salvador Ros's and Alvaro Perez Pozo's expertise and continuing guidance. We have had many discussions at the LINHD Laboratory of UNED and the UNED Madrid Escuelas Pías library and online when necessary. I also had a chance to communicate with Susana Onega herself through the institutional account that LINHD provided me. During my stay, the international CLS INFRA Training School, "Digging for Gold," was taking place at UNED under Dr. Ros' organisation, and I was lucky to attend it as well. The school was fascinating and provided exceptional insights about different digital methods and tools. Meeting various DH scholars and listening to their projects and studies was exceptionally thrilling.

I am enormously grateful for the opportunity that this valuable Fellowship provided me. These are the experiences that add precious milestones to our profiles as scholars.