Bauhaus-Universität Weimar
Faculty of Media
Degree Programme Computer Science and Media

Mining Rhetorical Devices by means of Natural Language Processing

Master's Thesis

Viorel Morari

1. Referee: Prof. Dr. Benno Stein
2. Referee: Dr. Andreas Jakoby

1. Advisor: Khalid Al-Khatib

Submission date: January 8, 2018

Declaration

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Weimar, January 8, 2018

. . . .

Viorel Morari

Abstract

The style of texts is one of the most important criteria responsible for effective communication. It contributes to the brevity, correctness, and ornament of the conveyed message, and therefore, influences how the target readers comprehend and accept the message. This thesis aims at mining the "style" of the language based on rhetorical devices – techniques used to convey the meaning or heighten the emotional effect of an utterance. Previous studies have already proven the effectiveness of using rhetorical devices in the task of analyzing the stylistic aspect of the language. However, the ambiguous nature of the language and its volatility, in particular, makes the automatic identification of rhetorical devices in texts a challenging task. Still, with the existing technologies in computational linguistics, it is possible to reach decent results to that end. In this thesis, we build a system for detecting syntax-based rhetorical devices in text documents. Then, based on the empirical distribution of the detected devices in collections of documents, each of which represents a different attribute of text (type, genre, topic, and author), we identify the stylistic patterns which, as we show, characterize these attributes. Overall, we developed a novel framework for detecting rhetorical devices, we built an extensive dataset for evaluating our framework, and presented new patterns and intriguing insights based on the detection results. These contributions should foster the research in the "style" area of computational linguistics and help to stimulate the development of various applications such as writing assistant system.

Contents

Chapter 1

Introduction

Style and Rhetorical Devices

This thesis studies one of the main elements of rhetoric[1]: the "style" , i.e., how people encode the messages of their texts and deliver them in different linguistic variations, considering their goal (e.g., persuasion), audience (e.g., expert), and type of text (e.g., novel), amongst others.

The "style" element is highly correlated with the effectiveness of communication(Burton [2007]). Favoring a particular context-depending style could play a significant role in achieving the goal of communication, the persuasion for instance. This holds true for over two millennia when the "style" emerged as a subject of study in the days of Aristotle.

Manual analysis of the *style* of speeches or written texts is considered a challenging task; Rhetoric in general, and "style" in particular, covers a broader range of choices that people follow. In texts, these choices influence the whole set of text's levels: the lexical, syntactical, semantic, and pragmatic. Anyway, different theoretical models have been proposed for identifying and analyzing the texts' "style". The most salient of these models is the one which is based on, what is called, *rhetorical devices*[2]. Rhetorical devices are techniques used to convey the meaning or heighten the emotional effect of an utterance. These techniques have been shown to be reliable and meaningful to quantify the text "style" according to several previous studies (Java [2015], Gawryjołek et al. [2009]).

The long and fruitful history of rhetoric has witnessed the development

[1] The other elements are: invention, arrangement, memory, and delivery.
[2] Also referred to as *rhetorical figures*

of a huge number of rhetorical devices. The extensive compiled lists of the studied devices include on average more than five hundred of such figures[3] (Lawrence and Reed [2017]). Although those devices can be grouped based on the similarity of their characteristics (e.g., repetition of words or structure), a vast number of devices require a tailored process and substantial effort to be analyzed (e.g., irony and sarcasm).
Overall, identifying the rhetorical devices is a complex task for many reasons; some devices might be spread across phrases, sentences, and even paragraphs. Also, a broad range of contextual information has to be taken into account (e.g., the background of the authors and readers). Moreover, identifying whether a span of text contains a specific device could be subjective to a wide extent.

Automatic Detection of Rhetorical Devices

Forasmuch as the rapid development of automated systems for text processing, the need for proposing a way of analyzing the style *automatically* became necessary. No doubt that the analysis of style is fundamental for different applications such as the assessment of writing effectiveness and quality.

However, the outlined complexity and diversity of rhetorical devices, we discussed above, affect the way that computer scientists deal with the task of "style analysis". Despite the extensive study of rhetorical devices in humanity and communication theory (see section *§Related Work*), as for computer science area, and in particular, Natural Language Processing (NLP) community, only a few research studies have investigated the task of "style analysis" by means of rhetorical devices (Java [2015]). Instead, the task has been tackled based on a shallow linguistic set of style indicators (called features), e.g., the number of words and punctuation per sentence, percentage of words with-/without vowel, and the percentage of function words[4] per sentence (Mann and Thompson [1988]). In fact, some studies have gone further with features that capture different content and structural information such as character n-grams and production rules (Strommer [2011]). In this manner, despite the "shallowness" of their approach, researchers are able to avoid the complexity of analyzing the rhetorical devices and at the same time achieve sufficient effectiveness in many NLP applications; especially those with classification settings, like genre classification, perspectives, hyperpartisan identification, and authorship attribution.

[3] on Silva Rhetoricae

[4] Function word – word that expresses a grammatical or structural relationship with other words in a sentence.

Nevertheless, this success comes with limitations. Namely, a classification system based on shallow linguistic analysis is merely able to explain its decisions, e.g., why and how a text is of high quality; and therefore, it couldn't provide any suggestions for improving the writing. Often, such suggestion play a crucial role in many applications like writing assistance systems.

To overcome these limitations, and to be able to work towards our envisioned systems (see *§Envisioned Applications*), we propose an explanatory approach[5] for style detection and analysis, relying solely on rhetorical devices.

Research Questions

In this work, we view the rhetorical devices in respect to their linguistic levels (e.g., semantic), and concentrate on 26 rhetorical devices which deal with the *syntax level* (arrangement of words). The reason for this choice is the advanced development of the methods that analyze the text structure. The available technologies for semantic and pragmatic levels are still far from maturity(Java [2015]).

In general, this thesis targets three research questions:

1. How to identify syntax-based rhetorical devices in texts?
2. What are the frequent usage patterns of rhetorical devices in high quality texts?
3. How different are these patterns within and across texts' types, genre, topics, and authors?

Approach

We address our research questions by proposing a two-stage approach. In the first stage, we develop and implement new algorithms for detection of rhetorical devices in texts. The algorithms are based on several linguistic rules we develop for targeted devices. These rules are implemented based on several computational linguistics frameworks. More specifically, we rely on Apache UIMA to build the backbone of our approach, which is used for preparing, processing and annotating the rhetorical devices. Also, we use the output of

[5] An explanatory model is a useful description of why and how a thing works or an explanation of why a phenomenon behaves the way it does.

Stanford Parser as the input of our rules. After developing the algorithms, we evaluate their effectiveness. To this end, we create a dataset of 1658 texts, each of which is labeled with one of the 26 rhetorical devices. We report the performance measures for each rhetorical device based on precision, recall, and F1-score metrics. In the second stage, we apply our algorithms for detection of rhetorical devices to unlabelled texts that are grouped into four different sets: types, genre, topic, and author. Based on the distribution of the identified rhetorical devices in each group, we assess the frequency fluctuations across the four sets, identify the *significant* differences, and measure the effect size of the differences. Also, we show frequent patterns of rhetorical devices and infer valuable insights about the results.

Figure 1 illustrates the pipeline of our approach.

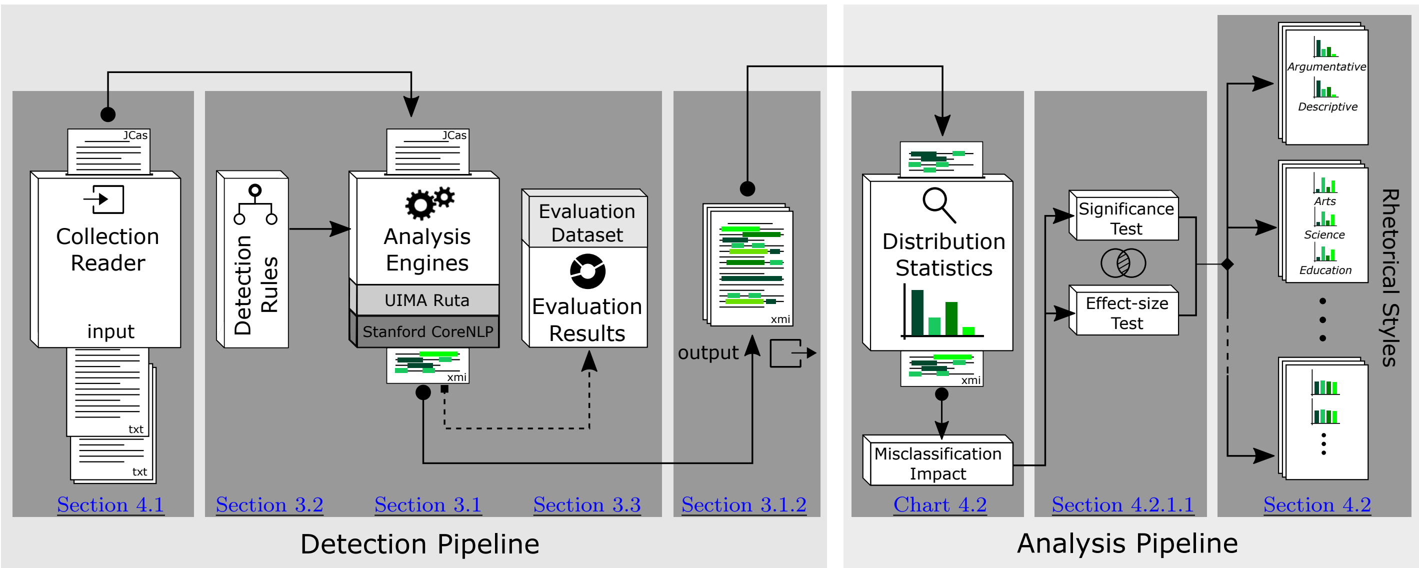


Table 1.1 shows the dimensions of our analysis.

Table 1.1: Data dimensionalities considered in our experiments. Green cells indicate the variety of types of data we target in this research.

Language	Mode	Communic.	Medium	Type	Genre	Topic	Author	Audience
English	Written	Monological	Newspaper	Descriptive	Editorial	Education	Identity	U.S.
German	Spoken	Dialogical	Presidential Debates	Argumentative	Review	Science	Age	Europe
Chinese			Encyclopedia	Narrative	Biography	Art	Gender	Middle East
Spanish			Forum	Expository	Debate	Politics	Type	Russia

Envisioned Applications

Overall, we see the importance of our approach in different scenarios. Automatic detection of rhetorical devices in texts is useful as a tool to analyze and suggest particular writing styles which are suitable for specific scopes. For instance, aspiring writers might study the styles of top writers and masterpieces, and deduce the writing techniques which don't seem evident at first glance. On the other hand, the analyzed patterns of rhetorical devices might be also used in modern Natural Language Generation systems (NLG) for improving the quality of the synthesized texts (e.g., making them more persuasive). Anyway, the detection of rhetorical devices can be engaged in other NLP-oriented applications like authorship attribution, sentiment analysis or identification of partisan discourses. Either as the main model for style analysis or side by side with the shallow model. Figure 1.1 shows an envisioned style suggestion system.

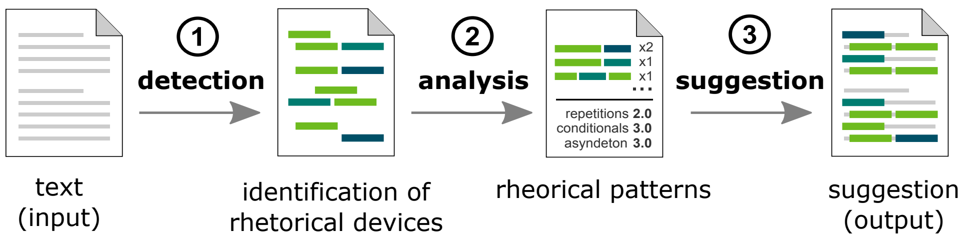


Figure 1.1: Style suggestion system pipeline

Contributions

The contribution of our thesis is threefold:

1. A new and publicly available framework for the detection of syntax-based rhetorical devices in texts. The framework is based on a well-defined pipeline which facilitates further enhancement (e.g., identifying new rhetorical devices) and customization.

2. A new dataset that comprises 26 sets of texts, each of which includes 60 texts for a specific rhetorical device. To our knowledge, this is the largest dataset for rhetorical devices.

3. A deep analysis of the rhetorical devices patterns in presidential debates and newspaper articles within and across different texts' types, genres, topics, and authors.

Chapter Summaries

In the next chapter, we dive deeper into rhetoric as an art form by exploring its origins, evolution, and adaptation to modern society. Also, the second chapter brings to attention the related work and advancements of rhetoric in NLP by drawing the parallels between them and this research. We use the correlations to show that there are unexplored directions in computational rhetoric in which our work fits. In the third chapter, we lay out the architecture of our rhetoric identification framework and discuss the underlying tools and concepts. A detailed descriptive immersion into all the engaged rhetorical devices represents the core of this chapter. Here, we also point out the refinements introduced later in the process of development; list the encountered problems, assumptions, and limitations that should be considered for following prospects. A large part of the third chapter discusses the evaluation results of our system. We present a detailed report of the misidentified instances of rhetorical devices and suggest further improvements. We then proceed to the analysis part in chapter 4, in which we focus on the structure of rhetorical patterns, distribution, and frequency of detected rhetorical devices in texts. We identify the significant differences in the distributions of devices across various types of articles in a newspaper corpus and a Presidential Debates dataset. Lastly, in chapter 5, we lay out our conclusions, discuss the achievements and the future work.

Chapter 2

Background and Related Work

2.1 Background

We start this section by discussing the origin, evolution and historical importance of rhetoric. In the second part, we make a transition from antiquity to modern times. We present the logical link between rhetoric and computer science and discuss the impact they had on each other.

2.1.1 Historical Background

Rhetoric is a form of discourse which aims to persuade the audience by using the language effectively. As innocuous as it may seem, in a specialist's "hands", rhetoric can be easily transformed in a tool of manipulation of its listeners or readers. Therefore, that's no wonder that it has been used since ages by skilled orators throughout the history. However, back in time, when it all started, rhetoric was considered more of a heuristic strategy of understanding the intent of the orator and clarifying the arguments rather than a means to disguise the truth. Although, nowadays, rhetoric in general does not represent something people are aware of most of the times, it consisted one of the central pylon in education and was studied intensively from Ancient Greece up to the 20th century (McKay and McKay [2010]).

Even though, rhetoric, as a solid art form began in Ancient Greece, the first examples in the history of humanity, were found in Mesopotamia(c. 2285–2250 BC) (Hallo [2004]). About 200 years later, in Ancient Egypt, rhetoric started to round up with the aid of well-educated people in the society who considered it to be a skill of a very high importance in that society, having an enormous respect for an eloquent, meaningful and calculated speech. They even had a set of rules intended to formalize rhetoric as a discipline. One of the rules which is highly appreciated today, states that "knowing when not to speak is essential, and very respected, rhetorical knowledge." Therefore, for Egyptians, rhetoric was more of a means to wisely control the speech rather than to enrich it or send a message, a "balance between eloquence and wise silence." (Hutto [2002]).

In Ancient Greece though, in the very beginning, rhetoric was not so warmly greeted by the society. With democracy being the system that offered a chance to any free man to get into politics, a clear, eloquent and persuasive speech was a must-have skill to absorb power and become influential. Rhetoric started to be seen as a tool to grow the electorate and convince them to vote for or against a particular piece of legislation (McKay and McKay [2010]). This created the optimal conditions for rhetoric to evolve as a discipline, and soon, small schools specialized in rhetoric, started to appear. The teachers in such schools were called sophists[1]. They were skilled orators and were promoting their abilities in the society as men who are able to debate and win an argument on any topic using confusing analogies and clever wordplay, regardless of whether they have any prior knowledge about it (McKay and McKay [2010]). Consequently, considering the demand and circumstances, they were paid heavily to teach this mastery those who were aiming for a successful political career. Nevertheless, by the rest of the society, sophists were considered ordinary impostors, men who would juggle with the truth just for financial prosperity (McKay and McKay [2010]). Even so, the impact of sophists on rhetoric was so important that it probably would not evolve such rapidly without their contribution.

Apart from the actual value of rhetoric, it required a formalization as an art form, a predefined set of "rules" intended to help anybody to use it effectively. Such an interpretation was defined by Quintilian, a Roman rhetorician. In his 12-volume textbook on rhetoric, *Institutio Oratoria*, he proposes five "canons" to follow for a correct rhetorical training: *inventio* (invention), *dispositio* (arrangement), *elocutio* (style), *memoria* (memory), and *actio* (delivery). Out of these, the most interesting for the scope of this research is elocutio, which relates to the style of arguments in speech, and namely, various rhetorical techniques and figures of speech used to persuade the audience.

In general, when we imagine a persuasive speech, we think about the meaningful and powerful message behind. Even though, *what* one tries to say

[1] Sophist - any of a class of ancient Greek teachers of rhetoric, philosophy, and the art of successful living prominent about the middle of the fifth-century b.c. for their adroit subtle and allegedly often specious reasoning [Merriam-Webster.com [c]]

constitutes the backbone of persuasiveness, *how* it is presented – elocutio – provides the factor that keeps the audience connected to the message (McKay and McKay [2010]).

In ancient Greece, Theophrastus and Demetrius, two pupils of Aristotle, divided the style of a speech(i.e., elocutio) into four categories: *correctness*, *clarity*, *appropriateness* and *ornament* (Kirchner [2007]); the latter is our subject of interest.

As the name suggests, ornament uses a series of techniques to "decorate" that particular piece of work and make it interesting for the audience to read or listen to. These techniques are generally referred to as figures of speech, figures of thought and tropes (Cicero [1954]). *Figures of speech* relate to "verbal expressions", grammatical patterns and arrangement, while *figures of thought* denote the composition and presentation of ideas (Burton [2007]). Modern rhetoric classifies tropes as a separate category in figures of speech. In this work we will only consider the *figures of speech* (also referred to as *rhetorical devices*).

In his work *Institutio Oratoria*, Quintilian denotes a figure of speech as to be "the term employed when we give our language a conformation other than the obvious and ordinary". A reinterpretation of this definition is later formulated by Corbet (1990), who defines a figure of speech as "any artful deviations from the ordinary mode of speaking or writing". He correctly limits the definition to an "artful deviation", since the type of deviation establishes the category a figure of speech may fall into *tropes* or *schemes*.

Tropes involve a change in the meaning of the words, a seamless "substitution of one word for another" (Quintilianus [1921]). Quintilian argues that a trope should not be considered a figure of speech because "a figure does not necessarily involve any alteration either of the order or the strict sense of the words"; yet, in a general sense, "figure" as a name, is common to both (Quintilianus [1921]). In practice, a trope is built by substituting the original, proper word (*verbum proprium*) with a figurative one (*verbum improprium*). The degree of relatedness between the proper and figurative words, create a large variety of tropes, starting with equivalence relation between the interchangeable words (synonymy), leading up to a total contrast (antonymy) (Müller [2006]). For example, we can replace "Joe is our Usain Bolt" with "Joe is our best runner", and the semantic congruence between "Usain Bolt" and "best runner" is evident. On the other hand, when we refer to a messy place saying "Wow, that's almost clean!", the dissimilarity becomes obvious, yet both are valid instances of tropes (Müller [2006]).

A higher importance for this research though, are the schemes – a rearrangement of the words in a sentence such that the meaning remains unaltered. Unlike tropes, the schemes usually do not result in semantic changes and as the name suggests, it is just a minor deviation from the original "scheme" of the sentence; Plett [1977] calls it "the smallest deviant language unit". Nevertheless, if applied correctly, they are able to strengthen any argument and amplify its emotional signal.

2.1.2 Communication Theory

Communication, as we generally know it, can be defined as a direct (mutual) transmission of thoughts, ideas and meanings from sender to receiver (Craig [2006]). Given as granted, the role of communication in our society is immense; it holds the people together and acts as an engine of evolution, helping us to learn, build and propagate the knowledge to the future generation. In his *Essay concerning Human Understanding (1960)*, John Locke, a famous English philosopher, argues that communication is not about the words, it is about the people and their own interpretation of them (Craig [2006]). Words do not carry any meaning on their own; it is, therefore, our task to match the right idea with the received set of words. Language is highly ambiguous, and if we fail to use a term with connotation everyone else is used to, then we fail to communicate.

Through the prism of communication theory, rhetoric emerged in ancient Greece in the context of public speaking, consisting the main occupation in their society. Communication and rhetoric go hand in hand, with the latter being a supplement and a catalyst for effective communication. In Europe, rhetoric continued to increase its evolutionary pace in educational system and public communication and was seen as a fundamental skill to success in business (Craig [2006]). On the other hand, people were aware of the power and confusion it induces; the effects of rhetoric on communication have been acknowledged as harmful, causing misinterpretation of the intent. Locke considers rhetoric to be detrimental for communication, as he "warned against common abuses" of language when using rhetoric, like substitutions of words and deviations from the formal meaning of the terms (Craig [2006]). He recognizes the benign contribution of clarity and order, "which promote understanding", yet regarding the usage of other rhetorical instruments, in particular, the figurative language, he condemns them saying that it is an "instrument of error and deceit" (Craig [2006]).

In the beginning of the 20th century, education started to get more focused on individual's career. Consequently, rhetoric and communication as an art form got shifted away from the top priorities of a well-educated man. Along with the rise of the media, rhetoric suffered important changes in how it is perceived by the crowd. During the mid of the 20th century, gradually, people started to shift the attention from what rhetoric originally means – analysis of the speech/speaker and its intent, on the potential meaning hidden in the message and the subjective interpretation of the audience, listener or reader (Gaines [2006]).

As our society became more consumer-oriented, digital media was in a search for methods to effectively persuade and seduce the audience to follow their instructions. Slowly but surely, rhetoric and persuasion started to regain the lost popularity; now in a refreshed, readapted form. The need for persuasion was not only present in consumer advertising and political speeches, but also in legal communication (e.g., attorney argumentation), in the mass-media messages regarding environmental or biological dangers, or in health communication as campaigns oriented in preventing diseases by informing the population (O'Keefe [2006]).
On the other hand, the effectiveness of rhetoric was questionable without scientific proofs. Therefore, certain areas of practical activity (i.e., commercials, political communication), which involved rhetoric, started to conduct multiples studies to quantify the usefulness of persuasion. For instance, extensively researched was the question whether one's attitude towards an ad (i.e., evaluation of the emotions an ad can evoke, apart from the product being advertised) affects the overall effectiveness of the said advertisement. Unsurprisingly, the results have shown that persuade-oriented ads are more likable and effective; however, the attractiveness decreases once the advertised object becomes well-known for the public (O'Keefe [2006]).

In general, men tend to underestimate or even decline the power of emotions. In the era of technology, everything seems to be driven by science, strong proofs, pure reason and logical argumentation. However, the human nature imposes that in a battle between emotion and rationality, emotion is a sure winner, most of the times (McKay and McKay [2010]). Advertisers know this fact, and they try to exploit it every time the chance shows up. That's why, nowadays, most qualitative TV commercials address the "why?"(i.e., why the consumer needs the product in the first place) side and oftentimes disregard the "what?" aspect of the advertised object (the banal description of the product).

2.1.3 Computer Science

The advent of computers in the second half of the 20th century opened new horizons in the research of language and established a potential link between human and computer to that end. One of the pioneers in this field was Alan Turing, who, in 1950 enounced the first ideas in his article *Computing Machinery and Intelligence*. Later on, the first practical foundation was laid by the Georgetown-IBM experiment, which was a powerful introduction of *machine translation* and consisting of an automatic translation of more than sixty Russian sentences into English (Gordin [2015], Nye [2016]). Even though it was a rather simple, rule-based approach with a carefully selected dataset, the researchers were very optimistic about the further development of machine translation; claiming that it can be considered a solved problem within three to five years (Reifler [1960], Hutchins [2005]). After just one decade of intensive research, when machine translation failed to generate accurate translations on the spot, the authors realized that this task is going to take much longer than it was originally set; ultimately, machine translation was abandoned.

The constant increase in processing power (Moore's law), the progress made in linguistics (Chomsky's theories) and the introduction of machine learning algorithms, led to a revolution in the language processing field, formally referred to as Natural Language Processing (NLP). As defined by Liddy [2001], Natural Language Processing is a set of techniques intended to analyze and represent naturally occurring texts, aiming to reach "human-like language processing for a range of tasks or applications". As the evolution of NLP was not steady throughout the last decades of the 20th century, multiple fields influenced its development. With *linguistics* being the main contributor, originally, NLP was named Computational Linguistics; and the names remained interchangeable even today (Liddy [2001]). Linguistics is the study and analysis of the language focusing on its form, meaning, and context (Martinet [1960]). "Computational side" of NLP, deals with representation and analysis of data; and cognitive psychology brings the human factor into the mixture, in particular, it addresses the cognitive usage of the language and aims to "model the use of language in a psychologically plausible way" (Liddy [2001]). All of these combined helped to relaunch NLP in the 80's and put it into a different perspective as a successful division of computer science.

2.2 Related Work

One of the earliest research in the computational analysis of rhetorical language is *Rhetorical Structure Tool* (RST), published by Mann & Thompson in 1988, which aims to facilitate the manual analysis of rhetorical structure in text (Mann and Thompson [1988]). It investigates the rhetorical relations between parts of text and uses them to theorize the organization of discourse in the said text. The arrangement of relations constitutes a basis to form the connections between two neighboring chunks of text. They call them *nucleus* and *satellite*, and the text is structured just by analyzing the relationships between these two entities Gawryjołek et al. [2009].

An important concept introduced by Mann & Thompson is the organization of texts in small units – text spans. They are meant to ease the job by establishing clear boundaries of these units and directly engaging them in rhetorical structure identification. In this case, the smallest text unit to work on is a clause, yet because of the complexity clause identification implies, they offer full control to the user to correct or improve the segmentation within the sentence. In contrast, we aim for an entirely automatic annotation of rhetorical devices; therefore we reserve no space for potential misidentification of clause boundaries in a sentence. As basic as it may seem, defining the correct boundaries of a clause or phrase in a sentence, is a rather complex task in NLP. Although, multiple rule- and memory-based methods promise decent results, they do not reach the level of accurateness acceptable for stylometric processing of text (Tjong and Sang [2001], Leffa [1998], Orăsan [2000]). The detection of syntax-based rhetorical devices in text requires a highly qualitative method for clause/phrase identification since any misplaced boundary might easily affect the detected device within the said span of text. Additionally, the list of rhetorical devices we consider relevant in this thesis does not imply any constraints for text spans smaller than a sentence. That's why we define the smallest text span for this research to be a sentence.
It is worth mentioning, also, that Mann & Thompson accounts for the relations existing between non-overlapping units of text. We, on the other hand, consider that a text unit, by nature, might combine more than one stylometric feature and therefore, detected rhetorical devices might overlap.

Argamon et al. [2007] introduces a lexical-based stylistic classification intended to improve the discrimination of texts in the evaluation corpora, along with a set of dimensions. These dimensions are, in fact, a predefined set of stylistic-based classification tasks which target authorship and gender attribution, personality typing, sentiment analysis and scientific rhetoric. That is, relying mainly on lexical features of the texts, they try to classify them across the listed dimensions. Implementing a rather "unexceptional" model to deal with the bold tasks set ahead, they report increased accuracy overall in the classification tasks; and in some cases the improvements are significant. As they initially note, it is not the approach; it is the correct choice of attributes and the "semantic organization of their possible values" that made the difference and helped to achieve the goals Argamon et al. [2007]. It is a valuable thought for us that suggests to carefully study the features and specifics of each rhetorical device before implementing it into the model. More relevantly is their finding based on the results that certain stylistic classification tasks require different kinds of lexical features; furthermore, the blind usage of irrelevant features might not only appear useless but be detrimental to the overall performance of the model.

A much more relevant work in this direction is presented by Gawryjołek et al. [2009]. It seems to be one of the first detailed analysis on authorship attribution based on rhetorical device. Although, the main focus is put on schemes: *anaphora*[2], *isocolon*[3] and *epizeuxis*[4] amongst others; he also considers the influence of tropes on authorship attribution, namely, *oxymorons*[5]. As shown, in order to detect them, certain devices (anaphora, for instance) require low-level text units. Gawryjołek makes use of *BreakIterator*[6] and parsing trees to identify phrases, clauses, and sentences in the text. He then introduces the concept of sliding "window" meant to ease the job of detection by iterating over the sentence units sequentially and checking for the presence of specific devices. Even though his method of segmenting the sentence into smaller units seemed to work well on some instances, he agrees that such a crucial task requires more advanced algorithms. Identification of tropes is even a harder task, in part because natural language is ambiguous by nature and computers do not handle the meaning so effortlessly as humans do. In his thesis, Gawryjolek builds an algorithm able to detect oxymorons by engaging the semantical relations implemented in *WordNet*[7] dictionary, in particular, the antonyms. It has proven to be a rather volatile and raw approach, yet it is a good start in the right direction. Though we do not account for tropes in this research, his

[2]Anaphora - repetition of a word or expression at the beginning of successive phrases, clauses, sentences, or verses especially for rhetorical or poetic effect (Merriam-Webster.com [a]).

[3]Isocolon - a figure of speech or sentence having a parallel structure formed by the use of two or more clauses, or cola, of similar length (Dictionary.com).

[4]Epizeuxis - a form of repetition in which one word or a short phrase is repeated in succession with no other words in between (Literarydevices.com).

[5]Oxymoron - a combination of contradictory or incongruous words (such as cruel kindness) (Merriam-Webster.com [b]).

[6]The BreakIterator class implements methods for finding the location of boundaries in a text (Java Documentation).

[7]WordNet is a large lexical database of English. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept (Fellbaum [2005]).

work should be taken into consideration for future improvements.

A different perspective on the computational analysis of rhetoric in language is conveyed by Strommer [2011]. He presents a method to assess the authorial intent in text documents by exploiting shallow attributes in rhetorical devices. The main distinction between Strommer [2011] and Gawryjołek et al. [2009] is that the former looms a semantical rather than syntactical analysis. In fact, he decides to dedicate the research to just one device - rhetorical anaphora (referred to as *epanaphora*), a figure of repetition - considered one the most important to signal the authorial intent of repetition or emphasis (Burton [1996]). On the one hand, restricting the domain of explored rhetorical devices allows to obtain a fine-tuned model and carry out a much exhaustive research in this direction; on the other hand, the range of possible applications are hereby limited and specialized. Nonetheless, the usage of shallow attributes provides a highly versatile approach, able to detect and distinguish between two types of epanaphora: accidental and intentional. Obviously, the spotlight is on the latter because an accidental repetition (the essence of anaphora) is likely to be meaningless or unintentional. Even though the dissimilarity between them is evident and easily spotted by humans, for computers, it appears to be a tough task. As the author shows, the key lies in the correct selection of attributes of repetition (i.e., shallow attributes) representative for repetition-based rhetorical devices, including epanaphora. N-gram overlap (number of identical and consecutive pairs of tokens within sentences), tuples length (number of sentences contained within a single instance of epanaphora) and the gaps between constituents (sentences) of epanaphora seem to be the most reliable attributes to identify the types of epanaphora. In particular, large tuples (at least 4), medium n-gram length (in particular – 4), a short median gap (smaller than 2) and short sentences (smaller than 15 tokens), appear to be the best model configuration able to detect intentional epanaphoras. This is a valuable finding for the scope of this thesis since we consider a subset of repetition-based rhetorical devices which employ similar attributes, and therefore, their post-refinement is a crucial step which determines the validity of detection. As we infer from Strommer [2011], the accurate localization of rhetorical devices in the text depends on a lot of factors, including: punctuation (i.e., exclamation signs make the appeal to emotion), document type, placement in the document or relative to other devices, and writing style (narrative, poetry).

With authorship attribution leading in the rank of applications based on stylistic analysis of language, it comes naturally that the majority of studies and noteworthy advancements have been specifically focused in this direction. One of the most recent and comprehensive research on authorship attribution is presented by Java [2015]. Basically, he expands the previous work of Gawryjołek et al. [2009] by implementing a fully automated approach, able to detect a predefined set of rhetorical devices. The novelty consists in the use of Machine Learning algorithms, having as the task, to classify the anonymous input texts according to their rightful authors. In general, modern authorship attribution methodologies employ manifold training instances for each class (i.e., documents by specific author), aiming to establish an attribution model wherein each new input text is seen as an entity to contribute to the general model. During the training phase, each document is characterized by a vector of attributes which describes the feature set chosen to represent the model (see figure 2.1). Then, based on the document vectors, the classification algorithm is "taught" to develop an attribution model. It is worthy to mention that in order to obtain a reliable attribution model, a large set of training instances should "feed" the model Java [2015].

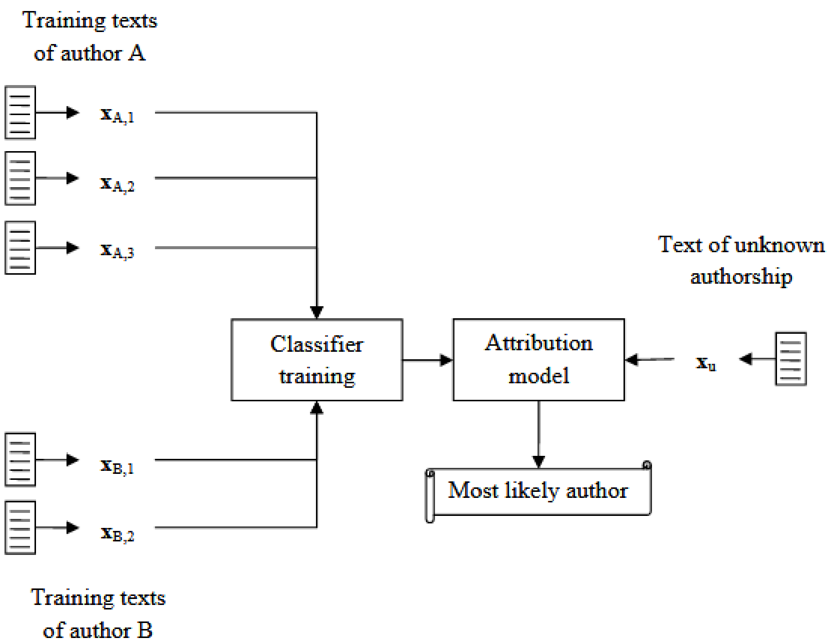


Figure 2.1: Typical architecture of instance-based approaches (from Stamatatos [2009]).

An indispensable tool which connects most of the studies in NLP nowadays is the natural language parser. Starting with its development in the 1990s, it has been one of the greatest contribution to NLP throughout the years. A parser is software package which gets sentences in its natural form as input, and generates their grammatical structure by grouping words together (as "phrases") and tagging each of them as part of speech (POS) (Java [2015], Klein and Manning [2003]). The grammatical structure of sentence is laid out as a hierarchical collection of POS tags wherein each node is divided into its subordinate children. As it can be seen in the figure 2.2, such representation adopts a Penn Treebank (Marcus et al. [1993]) style tree. Java [2015], em-

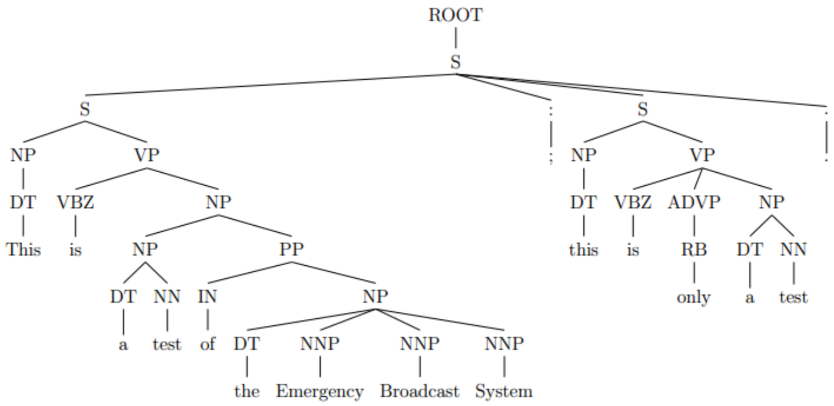


Figure 2.2: Parse tree for the sentence "This is a test of the Emergency Broadcast System; this is only a test." (from Java [2015]).

ploys Stanford PCFG (probabilistic context-free grammar) parser to identify sentence constituents: *phrases* and *clauses*. He relies on a set of heuristic rules inferred from the typical configuration of S-nodes (clauses) and *P-nodes (phrases). Even though this is vindicated by the parser's performance which works well on unsophisticated sentences, our tests on sample sentences from news articles (our analysis corpus) influenced our decision to lower-bound the text unit size to a sentence. However, we rely on the Stanford PCFG parser to POS annotate the sentence words; tags which we engage to construct the detection algorithms around. We will discuss more thoroughly about the contribution of Stanford PCFG parser and other variations of it in the next chapter (see *§Detection of Rhetorical Devices*).

One last idea that we would like to draw attention to, with reference to Java [2015], is the method of evaluation. In NLP field, there is a clear dichotomy between *manual* and *automatic* evaluation. In order to assess the developed system in its entirety, beyond the intended goals or requirements, manual evaluation implies hiring a group of human "judges" (or assessors) which will thoroughly evaluate the system output following the predefined instructions and ultimately, will convey its usefulness by mimicking real-world users and usage scenarios of said application. Despite its magnitude, manual evaluation is a rather expensive task, both in time and resources. Furthermore, it is prone to generate inconsistent results, since human beings are known to be quite unpredictable when it comes to opinions and judgements. Automatic evaluation, instead, is more often preferred as a reliable alternative in modern NLP studies. Scholars develop evaluation algorithms able to simulate, to some extent, the behavior of users (Clark et al. [2010]). Besides the aforementioned types of evaluation, in the context of this research, it is worth to discuss about *intrinsic* and *extrinsic* evaluation. In fact, more often than not, automatic and manual evaluation go hand in hand with intrinsic and extrinsic, respectively. It is due to the fact that intrinsic evaluation pertains strictly to the functionality of the evaluated system without generalizing; in other words, it evaluates a certain feature by considering the predefined criteria (i.e., evaluation corpus, configuration). Conversely, extrinsic evaluation takes a broader view and assesses the system considering the user factor and complex usage scenarios. As an analogy, in the context of a Rhetorical Devices Detection System, intrinsic evaluation might raise questions like: How well the detected devices convey the authorial intent? What is the combination of rhetorical devices used to persuade the reader? Alternatively, extrinsic evaluation touches more abstract aspects like: To what extent the identified devices influence user's emotions or make him more gullible? What are the user judgments about the author considering the rhetorical terms used? (Clark et al. [2010]).

Java [2015] approaches an intrinsic and automatic evaluation to assess the performance of rhetorical devices detection system. He starts off by collecting valid samples of rhetorical devices to be evaluated, from various reliable sources like Bible, literature, political speeches and common sayings amongst others. The resulted dataset is unevenly distributed, in the sense that the number of gathered samples of each rhetorical device ranges from 25 to 66. The performance measures, precision, and recall are determined using the system output on the evaluation dataset. Similarly, for the scope of this thesis, we lean towards an automatic type of evaluation. We consider it to be a good fit for the case when the simulation of user's interaction (especially, respecting the assumed user's incertitude regarding rhetorical devices) and the scrupulosity of usage scenarios are not vital for an accurate evaluation of our system. Therefore, as in Java [2015], we build up lists of evaluation samples of each rhetorical device and assess the performance measures based on that. We will elaborate more on this topic in the corresponding chapter (see *§Evaluation*).

Chapter 3

Detection of Rhetorical Devices

3.1 Pipeline Components

3.1.1 Stanford CoreNLP Suite

The focus in this thesis falls on syntax-based rhetorical devices; therefore, it is critical to rely on a good syntactical parser which we could use as a headstone to our approach. As already discussed, the contribution of natural language parsers in modern NLP is inestimable, especially in the context of low-level sentence analysis. We make use of Stanford CoreNLP suite of human language analysis tools, to perform two crucial tasks: POS (parts-of-speech) tagging and Dependency Parsing.

We employ a probabilistic context-free grammar parser (PCFG) as part of Stanford CoreNLP suite, to generate the parts-of-speech tag for each word in the sentence under analysis. Since a significant portion of all the considered rhetorical devices is repetition-based, we rely on the distribution and frequency of POS tags in the sentence to validate the presence of a particular device. In other words, we analyze the recurrent tags or combination of tags by respecting the guidelines/patterns inferred from the definition of the device under consideration. This straightforward approach guarantees a high accuracy, being provided well-defined, formal representations of rhetorical devices for the detection algorithm. However, formalization of the style of writings is not always a trivial task, even if a comprehensive definition is available. In such cases, supplementary rules and tools are needed to disambiguate the definition and constrain the algorithm. Besides the fundamental role it plays in the detection of repetition-based rhetorical devices, the tagger is directly engaged in the identification of omission-based devices amongst others. In particular, it checks for the presence of certain POS tags in the sentence to validate the rhetorical device.

Despite the richness of information introduced by Stanford PCFG, when dealing with high-level stylometric text features, it comes in handy to get the underlying relations between the identified POS tags in the sentence. We allot this task to Stanford Dependencies parser, a member of Stanford CoreNLP family. It extracts and creates representations of grammatical relations from a sentence in an intelligible form and ready to be used by people without any linguistic expertise (de Marneffe and Manning [2008]). A schematic view of Stanford Dependencies for a sample sentence is shown in the figure 3.1. Overall, the tool is able to recognize about 50 grammatical relations, depending

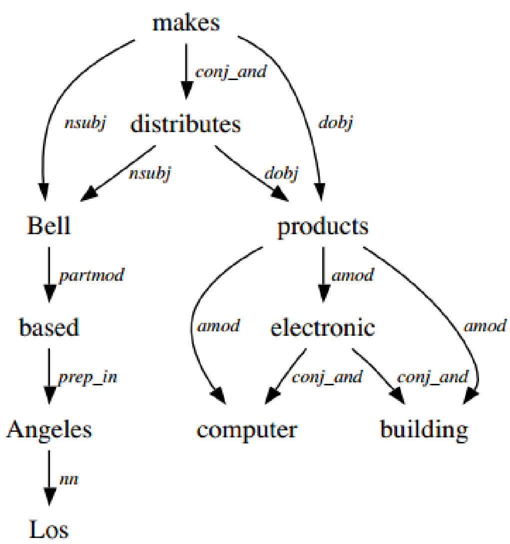


Figure 3.1: Graphical representation of Stanford Dependencies for the sentence: *Bell, based in Los Angeles, makes and distributes electronic, computer and building products.* Graph nodes denote the words in the sentence and the edge labels describe the grammatical relations between them (from de Marneffe and Manning [2008]).

on the level of granularity desired for a specific representation (ranging from a rather superficial to a more semantic-oriented representation) (de Marneffe and Manning [2008]). In the plain form, which we regard, the relations are described as pairs consisting out of two components: *governor* and *dependent*, as follows, *abbreviated_relation_name(governor, dependent)*. For instance, a nominal subject relation identified in the sentence "Trump defeated Clinton" is represented as *nsubj(defeated, Clinton)*.
How do these dependencies contribute to a stylometric analysis of the text, is

the question we consider in the section *§Detection of Schemes*.

3.1.2 Apache UIMA

The backbone of our approach is Apache UIMA™ (Unstructured Information Management Architecture), a versatile platform which can be engaged in a broad set of information management operations, ranging from robust text analysis to integration with search engines (Ferrucci and Lally [2004]). It is capable of analyzing large volumes of (unstructured) information to discern, organize and deliver relevant knowledge about the input to the end-user. For instance, an application which processes tons of articles to discover pieces of evidence which (dis)prove a certain argument. Apache UIMA includes a set of interdependent and easily-customizable components intended to facilitate the text analysis process. Below, we list the main UIMA components along with their tasks (Wachsmuth [2015]):

- Type systems – defines the existing annotation types along with their attributes.
- Collection readers – reads the input files iteratively.
- Primitive analysis engines – smallest text analysis entity; defines the analysis algorithm configuration.
- Aggregate analysis engines – builds up the pipeline consisting of primitive analysis engines.

A typical pipeline for text analysis tasks is illustrated in the figure 3.2.

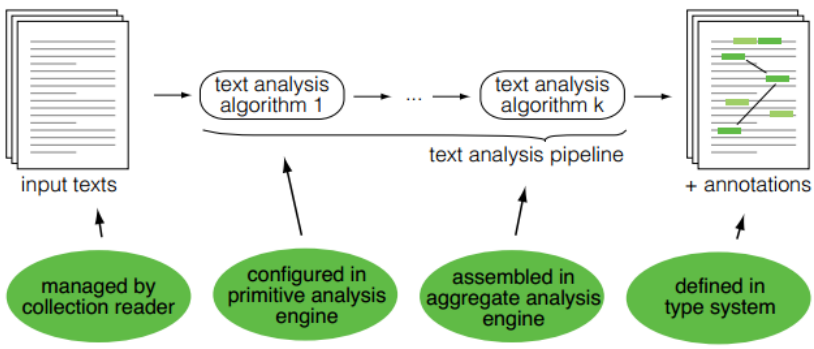


Figure 3.2: General UIMA text processing pipeline (from Wachsmuth [2015]).

As output, UIMA generates metadata XMI files which are rich in annotations of the requested type. Additionally, the generated files might contain intermediate annotation types which were used by the analysis engines to reach the target annotation type. These annotations, in the output files, can be later accessed by their type or position - information which is readily available in the resulting documents.

In order to infer the correct judgments about the rhetorical style of a certain domain of writings, it is primary to analyze large sets of documents. Moreover, the advantage from the centralization of the entire process (starting with the input and ending with ready-to-evaluate annotations) makes Apache UIMA the best fit in the context of this thesis.

3.1.3 Apache UIMA Ruta

A distinctive feature of this work is the inherent implementation of rule-based algorithms used in the detection of particular rhetorical devices. In general, writing algorithms for text processing applications is a tedious and error-prone task; usually, that is due to the erratic nature of the language. Apache Ruta™ (Rule-based Text Annotation) comes to alleviate this job by introducing an "intuitive and flexible domain specific language for defining patterns of annotations" (Kluegl et al. [2016]). A rule-based algorithm is, in fact, a set of Ruta rules able to annotate any span of text which satisfies the implied rules. In other words, Ruta allows the user to define patterns of annotations which are checked against predefined text spans; if the pattern applies, then the corresponding actions of the underlying rule are executed on the matched span of text. To exemplify, consider the simple task of heuristic identification of sentence boundaries in a text. While in conventional programming languages it might be a little cumbersome, Ruta is able to handle this in two lines of code:

```
DECLARE Sentence;
PERIOD #{→ MARK(Sentence)} PERIOD;
```

This rule annotates any text surrounded by `PERIOD` annotations as a sentence. Inherently, Ruta contains a set of basic universal annotation types like `PERIOD`, `COMMA`, `W1` and `SW2` amongst others, which aim to facilitate the development of simple rules. Yet, any other particular annotation types, not recognizable by the framework, must be declared in advance (e.g., `Sentence`).

Although Ruta is relatively new to the NLP community – an early version being published in 2009 under the name TextMarker (Kluegl and Atzmueller [2009]) – the rapid pace of development and the efforts that have been dedicated to it, helped Ruta to reach the maturity phase very quickly. It supports a wide range of actions and conditions that significantly simplify the rule-writing process. The conditions and actions are usually interrelated such that it is possible to simultaneously create multiple annotations of different types. These annotations can either be final or as constituents for an annotation of interest.

Considering the variety of rhetorical devices we plan to detect, it is easy to see why we favor Ruta over the conventional approaches. For instance, to identify an enumeration of annotations of the type `Person`, it is enough to execute the following rule:

```
DECLARE PersonEnum;
(Person COMMA)+{→ MARK(PersonEnum,1,2)} Person;
```

This rule translates as follows: find and annotate the patterns of text which start with `Person` annotation followed by a `COMMA` iteratively, until one of the inner rule elements does not match anymore. Evidently, the complexity of the rules increases in the case of much more intricate rhetorical devices; yet the efforts required to write such rules are tolerable compared to the clumsiness conventional programming languages have to offer. On top of that, UIMA Ruta integrates smoothly with Apache UIMA. Following a similar idea, Ruta consists of two components: an Analysis Engine, able to understand and execute the rule-based scripting language and a Workbench. Besides the Ruta Workbench, oriented to facilitate the development process, it is possible to "attach" Ruta rules to the UIMA Primitive Analysis Engines, part of the text analysis pipeline. Additionally, the predefined Type Systems and Analysis Engines are cross-compatible entities within both frameworks (Kluegl et al. [2016]).

Doubtless, adopting such a technology has its own drawbacks. First of all, detection of rhetorical devices requires fine-tuned algorithms because of the complexity involved; therefore, in many cases, it is practically impossible to cover all the conditions and subtleties which describe a rhetorical device just by using UIMA Ruta. Secondly, the constraints imposed by the Ruta actions and conditions, limit the generalization of the device in a way that more than one rule is necessary to reach the desired result. That's why, to not lose the floppiness granted by Java and the practicability of UIMA Ruta, based on the implementation rigidity, we divide the set of rhetorical devices into two categories: Ruta "friendly" and Java "friendly".

3.2 Detection of Schemes

Schemes represent one of the two categories of rhetorical devices which deal with the arrangement of the words in a sentence (Burton [2007]). Since "arrangement" is a general term to describe the schemes, literature agrees on four types of possible arrangements of words in a sentence; this is, four types of schemes: *balance*, *inversion*, *omission* and *repetition*.
Balance-focused schemes are focused on the rhythm of the thought; they can cause a sense of equivalence between the exposed ideas and is a good tool emphasize multiple parts of a fragment. For instance, it is easy to identify the two contrasting ideas in the famous quote of Neil Armstrong:

Example (1) *That's one small step for man, one giant leap for mankind.*

The schemes of inversion consider the changes in the grammatically correct order of the words in a sentence. Their primary goal is to introduce diversity in the monotonous sentence flow and thus, to draw the attention to some specific parts of the sentence. To illustrate this, consider the following example from John Milton's poem *Paradise Lost*:

Example (2) *Of man's first disobedience . . . / Sing, Heavenly Muse*

Omission-based schemes, also called schemes of subtraction, create an impression of incompleteness by removing intuitive words from an utterance and "asking" the reader to supply them by herself. It is often used to indicate an unfinished thought or a pause (case of Ellipsis[1]); conversely, it can have the effect of spontaneous multiplicity (Asyndeton[2]), like in the example below:

Example (3) *He came, he saw, he conquered.*[3]

The schemes of repetition are the most frequent type of rhetorical devices, in part because of their distinctive and powerful effect. In particular, they aim to produce a strong emotional impact on the reader, an emphasis or amplification (Burton [2007], Corbett [1990]). According to Aristotle, repetition is the key to a persuasive speech. By recapping the important aspects of an argument through properly created phrases and clauses, further embeds the idea into the listener, ultimately resulting in persuasion (Fahnestock [2003]). An example which illustrates the emotional potency of repetitions is the famous line from *King Lear* by Shakespeare (Müller [2006]):

[1] Ellipsis - omission of a word or short phrase easily understood in context (Burton [2007]).

[2] Asyndeton - the omission of conjunctions between clauses, often resulting in a hurried rhythm or vehement effect (Burton [2007]).

[3] Julius Caesar: *Vini, Vidi, Vici.*

Example (4) *Never, never, never, never, never.*

In general, computers are great at identifying recurrent patterns in a set of unstructured data (Strommer [2011]). Relying on this intuition and concluding the similar works mentioned in the previous chapter (*§Related Work*), we expect the detection algorithms of this type of rhetorical devices to deliver the most accurate results in comparison to the rest.

We find this categorization of rhetorical devices of a high importance and the traits of each class will be later used in the analysis phase when we try to infer the nature of the writings by considering the characteristics of each. Even though it's a well-established classification, upon which linguistic experts and literature agree, in this thesis, we regard the stylistic aspect of rhetoric, and the style, fortunately, is a broad term which leaves room for interpretation. Therefore, we consider appropriate to enlarge the domain of schemes by including a set unreferenced rhetorical devices, able to carry an intrinsic value or effect. Indeed, it is perhaps not correct to refer to them as rhetorical devices, yet we will further consider them as such just because they represent elements of style.

The *custom* category, as we name it, consists preponderantly of conditionals which entail the causality aspect of the language. As Aristotle considers, causality implies "explanation", it includes "an answer to a why question"[4]. Respectively, conditionals should pertain to texts of argumentative nature; therefore, we engage them as a category of rhetorical devices aiming to assess the argumentativeness of writings.
Besides conditionals, this category involves passive voice and comparative and superlative adjectives/adverbs. We believe that all of them are rhetorically valuable.

To correctly identify rhetorical devices in text, it is imperative to establish a well-defined and accurate formalization for each one of them. Here, we consider the formalism introduced by Harris and DiMarco [2009], as shown in the table 3.1.

The majority of the definitions we employ in constructing the formalisms and the detection algorithms are taken from Silva Rhetoricae - one of the most comprehensive guide on rhetorical devices (Burton [2007]).

[4]https://en.wikipedia.org/wiki/Causality

Table 3.1: Formalism for representing rhetorical devices (Adapted from Harris and DiMarco [2009])

Element	Meaning
Cl	clause
Phr	phrase
W	word
N	noun
Vb	verb
CC	conjunction
COMMA	comma
...	arbitrary intervening material
[...]	word boundaries
{...}	phrase or clause boundaries
$_{\mathrm{a,\ b,\ \dots}}$	identity $_{\mathrm{a}\,=\,\mathrm{b}}$, nonidentity $_{\mathrm{a}\,\neq\,\mathrm{b}}$

3.2.1 Schemes of Balance

Rhetorically, schemes of balance control the rhythm of thought. It is an essential tool for causing equivalence within the enunciated ideas. In this section, we individually describe each instance in this category.

Enumeration

Definition 3.2.1. Enumeration is a rhetorical device used mainly to list a series of details, words or phrases. [5]

By itemizing a sequence of terms, enumeration carries, in fact, the effect of amplification or division. In particular, it divides the main subject into its subsequent adjuncts, the cause, into its effects or an antecedent, into its consequents (Burton [2007]). Peter Mack, in his book *A History of Renaissance Rhetoric 1380-1620* (2011), argues that enumeration is a sort of "argumentation, in which all the possibilities are set out, and all but one are eliminated." (Mack [2011]). It should, therefore, signal the presence of argumentation side in writings.
Respecting the formalism mentioned above (table 3.1), and considering the definition, we formally define enumeration as follows:

```
< ... W [CC | COMMA] W ... >
```

[5] https://literarydevices.net/enumeration/

The detection algorithm is therefore straightforward, especially within UIMA Ruta environment. We build up the rule which targets spans of texts starting with a comma and ending with a conjunction in close proximity. The window between the first comma and the last conjunction represents, in fact, the enumeration constituents. We set the length of this window to range between 2 up to 5 tokens as we focus on short enumerations, on the word-level.

Example (5) ***Diligence**, **talent** and **passion** will drive anybody to success.*

Isocolon

Definition 3.2.2. Isocolon is a rhetorical device that involves a series of similarly structured elements having the same length. A kind of parallelism.

Formally,

$$< \ \dots \ <\mathrm{Phr}>_a<\mathrm{Phr}>_a \ \dots \ > \ \text{or} \ ,$$
$$< \ \dots \ <\mathrm{Phr}>_a<\mathrm{Phr}>_a<\mathrm{Phr}>_a \ \dots \ > \ \text{or} \ ,$$
$$< \ \dots \ <\mathrm{Phr}>_a<\mathrm{Phr}>_a<\mathrm{Phr}>_a<\mathrm{Phr}>_a \ \dots \ >$$

The term "Isocolon" comes from Greek, and it literally can be translated as "equality of members". At its foundation, isocolon has two core concepts: (1) the symmetrical construction between its constituents and (2) the uniform distribution of those constituents along the entire span of text isocolon covers; that is, parallel constructions. Aristotle refers to this device as an element of style which produces balance and rhythm in speech (Aristotle and Roberts [2004]). Therefore, the device can control the tempo of expressed thoughts by gently reinforcing a parallel nature in the writer's claims. However, linguists agree that an abusive or clumsy use of isocola "can create too glaring a finish and too strong a sense of calculation." (Farnsworth [2011]).

Before proceeding to its implementation, it is essential to analyze the definition further. First of all, it states that the device constituents should have the same length, which becomes a rather contestable claim when consulting multiple sources. For instance, Henry Peacham, a famous rhetorician of the renaissance rhetoric, argues that the members of isocolon should be of about equal number of syllables, "yet the equalitie of those members or parts, are not to be measured upon our fingers". In other words, if the rhythm is preserved even with slight variations between the lengths of constituents, then such changes can be neglected. Furthermore, loosening this parameter would permit the detection of more true instances of isocolon. This claim is supported by the next example:

Example (6) ***Fill the armies**, **rule the air**, and **pour out the munitions***[6].

This example of isocolon consists of three parallel clauses which preserve the same rhythm throughout the sentence. While the first two clauses (i.e., *fill the armies, rule the air*) have the same length, the third one (i.e., *pour out the munitions*) is larger by one token; and thus, disqualifying the sentence as a valid example of certain type of isocolon (considering the strict definition).

Returning to our approach, by default, we consider negligible, a deviation from the reference isocolon member (first element in an isocolon candidate) by three tokens.
Within this context, it is worth to mention that literary, there are three types of isocolon, depending on the number of parallel structures involved: *bicolon* (two grammatically equal structures), *tricolon* (three grammatically equal structures) and *tetracolon* (four grammatically equal structures). Since inherently, they carry the same rhetorical effect, we do not consider separate instances of isocolon within our implementation.

Bearing in mind the intricacies of isocolon, the detection algorithm is rather complex and cumbersome in part. Our approach is exclusively based on POS tags, as they define the grammatical structure of the sentence.

By definition, isocolon is not fixed on any level; it may extend across sentences as well. Therefore, the detection process is divided into two phases: paragraph level and sentence level. On the paragraph level, we target consecutive combinations of sentences up to 4 units (largest possible isocolon – tetracolon) and check for the presence of grammatically equal structures in a decreasing combination size manner. In other words, while no isocolon is detected, decrement the set size by one and repeat the process. Likewise, on the sentence level, we count the sequential matching pairs of POS tags to attest the presence of isocolon. For instance, example 3.3 will be identified as isocolon because of the matching pairs of POS tags:
Another detail that influences the validity of isocolon candidate is its coverage. This is, the share of the matching pairs of tags in the sentence(s) under analysis. It is easy to see the importance of this parameter, especially in the context of long sentences. Since we treat all the POS tags equally, it is quite probable that short recurrent patterns of insignificant tags (denoting conjunctions or

[6] W. Churchill at Manchester, 29 January, 1940

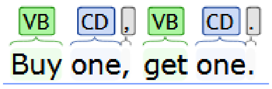


Figure 3.3: POS tags of the sentence *Buy one, get one.*

punctuation marks) in a larger sentence, will blindly convince the algorithm to consider it as a valid instance of isocolon. Consequently, observing the training data, we set a coverage threshold of 55% of the sentence, for the candidate to be considered a valid isocolon. In other words, if the detected matching pairs of tags take less than 55% of the span under consideration, then the sentence candidate is ignored.

Pysma

Definition 3.2.3. Pysma is the act of asking multiple questions successively (which would together require a complex reply).

Formally,

$$< \mathrm{Cl?} > < \mathrm{Cl?} > \ldots$$

As most of the balance schemes, pysma aims to emphasize specific ideas or parts of text which carry a certain value for the reader. The author lists a set of questions in one place to hone his speech by making it "very sharp and vehement" (Peacham [1593]). This technique of piling a sequence of questions requiring an answer represents another rhetorical device called *Erotema*; however, while the latter accepts a single word as an answer to all the questions, pysma, requires separate reactions for each of them (Peacham [1593]).

Example (7) *In what place did he speake with them? with whom did he speake? did he hire them? whom did he hire, and by whom? To what end, or how much did he give them?* [7].

As Peacham argues, this rhetorical device is multipurpose in a sense that it can be used by the author to complain, to provoke, to confute, to insult, to draw attention or to confirm (Peacham [1593]).

[7] Cicero for Roscius (Peacham [1593])

Example (8) *Will the Lord absent himselfe for ever, and will hee be no more intreated? Is his mercy cleane gone for ever? and is his promise come utterly to an end for evermore? hath God forgotten to be gracious? and will hee shut up his loving kindnesse in displeasure?* [8].

Bearing in mind the definition, the implementation of pysma is rather straightforward. We engage the ready-to-use annotations of question marks provided by UIMA Ruta to detect that sequence of questions. By default, we limit the range of consecutive questions from 2 up to 10.

3.2.2 Schemes of Omission

Omission-based schemes are used to produce a sense of incompleteness in the expressed ideas by removing instinctive words from an utterance. This section introduces them more in detail.

Asyndeton

Definition 3.2.4. Asyndeton denotes the omission of conjunctions between clauses, often resulting in a hurried rhythm or vehement effect.

Fomally,

$$\{<\mathrm{Cl}_a>\ \mathtt{COMMA}\ <\mathrm{Cl}_b>\ \mathtt{COMMA}\ <\mathrm{Cl}_c>\ \ldots\},\ \text{or}$$
$$\{<\mathrm{Phr}_a>\ \mathtt{COMMA}\ <\mathrm{Phr}_b>\ \mathtt{COMMA}\ <\mathrm{Phr}_c>\ \ldots\}$$

Being a figure of omission, asyndeton lacks in conjunctions, yet by no means accidental. Adopting such a construction, the sentence, sometimes, becomes impressive by asking the reader's imagination to infer the missing connectors and therefore fixing the attention upon the main subject (Johnson [2016]). It is important to mention though that, as a rule, an omission of conjunctions can easily happen between single words or phrases, and therefore slightly disregarding the definition. That's because irrespective of the length of the sentence component, the effect of asyndeton remains unaltered.

Example (9) *I came, I saw, I conquered.*[9]

[8] Psal., an example from the Sacred Scripture

[9] Caesar: Veni, vidi, vici.

Example (10) *If, as is the case, we feel responsibility,* ***are ashamed, are frightened,*** *at transgressing the voice of conscience, this implies that there is One to whom we are responsible, before whom we are ashamed, whose claims upon us we fear.*[10]

Generally, asyndeton has the effect of unintended multiplicity or a "staccato-like rhythm that results in clear-cut brevity and celerity of speech" (Grün-Oesterreich [2007b]). Peacham interprets the aim of this device as a method to avoid "tedious repeating of a conjunction" intended to introduce clarity and brevity in the speech or writings. However, he advises to avoid the interplay of opposite terms (i.e., war/peace, life/death) (Peacham [1593]).

The definition of asyndeton is rather exact and precise, without many intricacies we should care about in the process of detection. Therefore, a combination of simple rules in UIMA Ruta should do the job. The first rule targets spans of text wherein at least two commas are separated by single words. This pattern would work to identify the asyndeton present in the Example (9); but will fail to detect it across phrases or clauses, as in Example (10). That's why an additional rule is necessary to solve this issue by including annotations larger than a word, provided by the Stanford Parser. In this step, we consider all the phrase and clause annotations detected in the sentence and employ the same rule. This combination of rules covers most of the encountered asyndeton patterns.

Zeugma

Definition 3.2.5. Zeugma is a rhetorical device which shortens the sentence by removing (redundant) syntactic units from a sentence "in favor of a remaining one used to complete the meaning of two or more congruent words or clauses" Plett [2007].

In fact, zeugma cannot be treated as a well-defined, unique rhetorical device, since depending on the position of the removed syntactical units, separate subtypes of zeugma can be identified. Thus, if the *governor*[11] is left in the first part of the sentence, it is called *prozeugma*; if it governs a set of words, phrases or clauses and is placed after them – *hypozeugma*; if the governor occurs either at the very end or very beginning of the sentence, it is termed *epizeugma*; and if it is in the middle – *mesozeugma*. Although, all of them carry some rhetorical effect and should be studied thoroughly, in this work, we focus on

[10] John Henry Newman

[11] word able to govern congruent clauses

hypozeugma and epizeugma only.

Hypozeugma

Definition 3.2.6. Hypozeugma is placing last, in a construction containing several words or phrases of equal value, the word or words on which all of them depend.

Formally,

$$< \ldots \ \mathrm{N}_a \ , \ \mathrm{N}_b \ , \ \mathrm{N}_c \ \ldots \ \mathrm{Vb}> \ \text{or} \ ,$$
$$< \ldots \ <\mathrm{Phr}_a> \ , \ <\mathrm{Phr}_b> \ , \ <\mathrm{Phr}_c> \ \ldots \ \mathrm{Vb}>$$

One thing to observe is that hypozeugma, like any other zeugma, is a form of *ellipsis*[12] so that the sentence becomes more concise and efficient. By placing the governor after all the dependents (i.e., set of words, phrases or clauses which are governor dependent) it creates a more stylish and dramatic effect.

Example (11) *Friends, Romans, countrymen, lend me your ears ...*[13]

Example (12) *Assure yourself that Damon to his Pythias, Pylades to his Orestes, Titus to his Gysippus, Theseus to his Pyrothus, Scipio to his Laelius, was never found more faithful than Euphues will be to his Philautus.*[14]

The algorithm for detecting hypozeugma is much more intricate compared to asyndeton, in part because it requires a different approach to detect its constituents. That is, the first phase is to identify the governor and its dependents in the sentence. We rely on Stanford Dependencies to extract the existing grammatical relations. Based on them, we further annotate potential zeugma constituents. In particular, out of all the detected relations we consider only the nominal[15] (passive[16]) subject and clausal[17] (passive[18]) subject

[12]a rhetorical device wherein a lexeme is deliberately omitted

[13]first line of a speech by Mark Antony in the play Julius Caesar, by William Shakespeare

[14]John Lyly, Euphues

[15]A nominal subject is a noun phrase which is the syntactic subject of a clause (de Marneffe and Manning [2008])

[16]A passive nominal subject is a noun phrase which is the syntactic subject of a passive clause (de Marneffe and Manning [2008]).

[17]A clausal subject is a clausal syntactic subject of a clause, i.e., the subject is itself a clause (de Marneffe and Manning [2008]).

[18]A clausal passive subject is a clausal syntactic subject of a passive clause (de Marneffe and Manning [2008]).

relations. From our observations, these relations convey the most from the zeugma nature.
Afterwards, we consider the span of text covered by the governor-dependents relation and use it as an annotation within UIMA Ruta. Respectively, in the second stage of the detection algorithm, we "convert" the hypozeugma definition into Ruta rules by referring strictly to the zeugma constituents that have already been identified. In turn, the rule becomes simple – we look for consecutive nouns split by conjunctions or commas, as this is the common syntactical pattern for hypozeugma within the zeugma constituents (see examples above).

Epizeugma

Definition 3.2.7. Epizeugma is placing the verb that holds together the entire sentence (made up of multiple parts that depend upon that verb) either at the very beginning or the very ending of that sentence.

Formally,

```
<Vb ...> or ,
<... Vb>
```

Unlike hypozeugma, epizeugma allows the governor to be at the very beginning of the sentence; furthermore, even if it can be placed after all the dependents, its position should be last in the sentence. The rhetorical effect it carries is similar to the rest of zeugmas except that the grammatical structure of the sentence gives a flavor of aristocracy and nobleness, since this style characterizes more the epoch of Shakespeare rather than modern writings.

Example (13) *Fades beauty with disease or age.*[19]

Example (14) *Neither a borrower nor a lender be.*[20]

The constraints about the position of the governor make the detection of epizeugma much simpler. We take as candidates the sentences with a single governor, and then, its position is responsible for the validation of the current sentence as epizeugma. It is important to mention though, that following the definition, the governor can be placed "either at the very beginning or very ending" of the sentence. Conducting a set of test on our training data, we consider appropriate to slightly relax the position constraints and take as valid,

[19] Silva Rhetoricae
[20] Lear 1.3 qtd. in Garrett Epp

candidates whose governor does not exceed one-fifth of the sentence length in words from the beginning or end. In other words, if the governor is placed in the first or the last one-fifth of the sentence and no other governor is detected, then it is a valid instance of epizeugma.

3.2.3 Schemes of Repetition

In this section we will present the repetition schemes regarded in our research. These schemes can provoke a strong emotional impact by repeating essential words/ideas in a sentence.

Epanalepsis

Definition 3.2.8. Epanalepsis is the repetition at the end of a line, phrase, or clause of the word or words that occurred at the beginning of the same line, phrase, or clause.

Formally,

$$< \text{[W]}_a \ \ldots \ \text{[W]}_a >$$

Epanalepsis, also called *resumptio*, is a rhetorical device which encircles the sentence with the same word(s). Although its contribution to the connotation of the sentence is negligible, and can easily be avoided, epanalepsis is generally used to emphasize the importance of a statement and can provoke strong affections like love or hate (Peters [2007]). Peacham argues that authors make use of epanalepsis to position essential words in the most sensible places of the sentence (beginning and ending); such that putting a word at the beginning of the sentence allows it to be "considered", and "in the end, to be remembered" (Peacham [1593]).

Example (15) ***Believe*** *not all you can hear, tell not all you* ***believe.***[21]

Example (16) ***A lie*** *begets* ***a lie.***[22]

Again, as in the case of epizeugma, "beginning" and "ending" of a sentence are variable measures. That's why, to preserve the consistency amongst definitions, we divide the sentence into five equal parts and denote the beginning to be the first one fifth and the ending the last one-fifth of the sentence length.

[21] Native American proverb
[22] English proverb

Next, we simply scan these portions of the sentence for the presence of a duplicate term. Evidently, a preprocessing step is required, which consists in filtering out the most frequent English words which do not carry relevant significance (also called, *stopwords*) to avoid potentially biased results.

Mesarchia

Definition 3.2.9. Mesarchia is the repetition of the same word or words at the beginning and middle of successive sentences.

Formally,

$$<[W]_a \ \dots \ [W]_b \ \dots><[W]_a \ \dots \ [W]_b \ \dots>$$

As a figure of repetition, mesarchia's main purpose is to emphasize certain statements, phrases or words in text. A distinguishable feature of this device is its dimension, in a sense that it targets successive sentences (i.e., on the paragraph level). In fact, mesarchia is a member of a set of rhetorical devices which assess the repetition of words across sentences in certain places. It is somewhat unpopular amongst writers compared to other devices of repetition (i.e., anaphora), yet we consider it should bring some value and clarity regarding the repetition type and effect.

Example (17) ***And they shall dwell*** *in the land that I have given unto Jacob my servant, wherein your fathers have dwelt;* ***and they shall dwell*** *therein, even they, and their children, and their children's children for ever: and my servant David shall be their prince for ever.*[23]

Example (18) ***I was*** *looking for* ***a piece*** *of paper.* ***I was*** *anxious for* ***a piece*** *to write on.* ***I was*** *in need of* ***a piece*** *to start my butterfly census project.*

The detection algorithm is similar to the first device in this category – epanalepsis, except that here, we consider pairs of sentences instead of single units. Likewise, we select the necessary parts of the sentences (beginning and middle) and compare against each other to attest the presence of recurrent words. Since the algorithm checks consecutive sentences, we observed that relaxing the sentence constituents' boundaries, in particular making the beginning of the sentence larger, improves the overall detection results. Therefore, we divide the sentences by a factor of four and designate the beginning to represent

[23] Ezekiel 37:25

the first quarter, and the middle – the second and third quarters.

As in the case of epanalepsis, it is necessary to filter out the stopwords from the analyzed sentences. However, this means that combinations of two or more such words like "we have" or "there is", would be eliminated as well. As we regard various parts within consecutive sentences, such constructions might occur more frequently (compared to epanalepsis) and ultimately, consist a source of erroneous identification. As we show in Example (18), the combination of stopwords "I was" is part of the repetition pair which attests the rhetorical figure mesarchia in this sentence, and its removal would directly invalidate the example.
Therefore, we decide to filter out just the single instances of stopwords repeated in the beginning and middle of consecutive sentences. Similarly, we employ this approach for the rest of rhetorical devices in this group (i.e., epiphoza, mesodiplosis and anadiplosis).

Epiphoza

Definition 3.2.10. Epiphoza denotes the repetition of the same word or words at the end of successive sentences (Bullinger and Delmarva Publications [1970]).

Formally,

$$< \dots \ [W]_a > < \dots \ [W]_a >$$

Epiphoza (also called epiphora or epistrophe) is a rhetorical device that acts as an "ornament of great eloquence" (Peacham [1593]). By repeating the same word at the very end of successive sentences, epiphoza highlights the value of that word as being important, and thus, it may "longer hold in the mind of the reader" (Peacham [1593]). However, as with many other rhetorical terms, epiphoza should not be excessively used, as this might not only appear rhetorically pointless but also it would sound unpleasant.

Example (19) *O* ***apple!*** *wretched* ***apple!*** *Miserable* ***apple!***[24]

Example (20) *Are they Hebrew?* ***so am I****; Are they Israelites?* ***so am I****; Are they the seed of Abraham?* ***so am I****.*[25]

[24] Vinsauf
[25] 2 Cor. 11:22

Programmatically, the algorithm doesn't differ much from the previous one (i.e., mesarchia). After selecting the key spans of the sentences under analysis, we check for repetitive words within. Once a pair a sentences contain the same word or group of words within their last quarter (of their lengths), then we attest them as epiphoza instances.

Mesodiplosis

Definition 3.2.11. Mesodiplosis is the repetition of the same word or words in the middle of successive sentences.

Formally,

$$< \dots \ [W]_a \ \dots > < \dots \ [W]_a \ \dots >$$

Just like the rest of the devices in this category, mesodiplosis can bring an emphasized dramatic effect into play. The repetition of an important word in the middle of the sentences strengthens its meaning and makes it easier to remember by creating a parallel structure.

Example (21) *We are troubled on every side, yet* ***not*** *distressed; we are perplexed,* ***but not*** *in despair; Persecuted,* ***but not*** *forsaken; cast down,* ***but not*** *destroyed.*[26]

Example (22) *There's* ***no time*** *like the future! There's* ***no time*** *like the past! .*[27]

To detect instances of mesodiplosis, we "overload" the same algorithm used in the detection of epiphoza and mesarchia, with small modifications regarding the span of text wherein the duplicate words are sought.

Anadiplosis

Definition 3.2.12. Anadiplosis represents the repetition of the last word (or phrase) from the previous line, clause, or sentence at the beginning of the next.

Formally,

$$< \dots \ [W]_a > < \ [W]_a \ \dots >$$

[26] 2 Corinthians 4:8-9
[27] https://dailytrope.com/category/mesodiplosis/

Anadiplosis, which Putterham names "redouble" (Puttenham [1970]), is generally used as a technique to couple successive textual units in a smooth manner by repeating the last term at the beginning of the next line (Grün-Oesterreich [2007a]). Its purpose is to provoke strong amplification effect and an appeal to pity (Cicero [1954]). According to Peacham, anadiplosis does not only sound agreeable but also powerful, by increasing the "weight" of the repeated term. He would rather call it "the Rhetoricall Eccho" since it "carrieth the resemblance of a rebounded voyce, or iterated sound." (Peacham [1593]).

Example (23) *I will life my eyes unto the hills, from whence cometh* ***my help. My help*** *cometh from the Lord . . .*[28]

Example (24) *I am* ***Sam; Sam*** *I am.*[29]

In order to detect such patterns in text, we consider the ending and beginning of consecutive sentences, analyzing each for the presence of duplicate words within the boundaries of the regarded spans of text. The device is validated as soon as at least one pair of words is repeated amongst the spans under analysis.

Diacope

Definition 3.2.13. Diacope denotes the repetition of a word with one or more in between, usually to express deep feeling.

Formally,

$$< \dots \ [W]_a \ \dots \ [W]_a \ \dots >$$

Diacope comes from the Greek word thiakhop, which literally means "to cut into two". That is, it "cuts" the repeating word into two and place them in the proximity of each other. To illustrate this with an example, let us take the famous line from William Shakespeare's *Hamlet* "to be, or not to be!". We can easily observe how does diacope work and what are its constituents.
The main focus of diacope is to emphasize certain ideas or subjects; it helps to make the statement rhythmic and memorable. Therefore, frequently enough, diacope occurs in advertising, slogans, catch-phrases, and music amongst others.
Literary, there are two categories of diacope, function of its intrinsic purpose: *vocative* and *elaborative*. The former aims to emphasize important ideas or

[28] 2 Corinthians 4:8-9
[29] Dr. Seuss, Green Eggs and Ham

terms. In the Example (25) the author tries to accentuate how horrific the event is:

Example (25) ***The horror!*** *Oh,* ***the horror!***[30]

Elaborative diacope might appear much more interesting since it conceptualizes the first term and rather than repeating its exact form, it adds additional details and descriptions to it, with the end-goal to make the utterance easier to grasp. The elaborative diacope in the Example (26), introduces a much detailed description of the appearance of the subject:

Example (26) *He is standing with a lovely woman. A* ***tall, well-dressed and beautiful*** *woman.*[30]

Although it is a viable rhetorical method to further describe novel terms and persuade the reader, elaborative diacope is, at this step, hard to detect computationally because it would require a complex mechanism to draw the relations between its constituents. Therefore, we aim to assess the rhetorical value of vocative diacope.

The first step in the identification of diacope is to filter out the punctuation marks, non-alphanumeric characters, numbers, and stopwords because a sequence of such tokens would falsely convince the algorithm about the presence of diacope in the respective sentence. Next, the program iterates over all the tokens and checks its following neighbors to find a potential match. An important detail here is to define the number of neighbors it has to check, since as per definition, the repetition occurs with "one or more words in between". By default, we set the size of the window of intervening words to range between 1 up to 5 words; this appears to be the most advantageous compromise between *precision* and *recall* (see *§Evaluation*). If any pair of duplicate words is found within *five* words from the current checked one, then the sentence is annotated as diacope.

Epizeuxis

Definition 3.2.14. Epizeuxis consists in the repetition of words with no others between, for vehemence or emphasis.

Formally,

$$< \ [W]_a \quad [W]_a \ >$$

[30]https://literarydevices.net/diacope

Even though all the schemes of repetition are more or less used with the intention to accentuate specific subjects or ideas in a statement, perhaps, the most intense to that end is epizeuxis. The immediate succession of the key words, induce a strong emotional appeal, able inspire and motivate the audience (Hauser [2002]). Taking advantage of its versatility, epizeuxis is capable of intensifying affections of any kind, be it of joy, hatred, grief, or admiration. Peacham, poetically compares epizeuxis' pleasant affection to quaver[31] in music, its grief to "a double sigh of the heart" and its affections of anger to "a double stabbe with a weapons point" (Peacham [1593]). However, one should be very careful when dealing with syllable-rich words because of the time it would take to repeat them (Peacham [1593]).

Example (27) ***Awake, awake*** *and stand up O Jerusalem.*[32]

Example (28) *Never give in –* ***never, never, never, never,*** *in nothing great or small, large or petty, never give in except to convictions of honour and good sense. Never yield to force; never yield to the apparently overwhelming might of the enemy.*[33]

Considering the simplicity of the definition, the computational detection of epizeuxis is rather simple by adopting the same algorithm responsible for diacope identification. That is, we remove the mandatory window of intervening words and check for the duplicates in immediate succession. Evidently, before proceeding, it is important to remove the noise in the form of punctuation marks, stopwords and the rest aforementioned. The epizeuxis is attested as soon as at least two consecutive duplicates are found.

Polysyndeton

Definition 3.2.15. Polysyndeton is a rhetorical term which employs many conjunctions between clauses, often slowing the tempo or rhythm.

Formally,

$$\{<\mathrm{Cl}_a>\ \mathrm{CC}\ <\mathrm{Cl}_b>\ \mathrm{CC}\ <\mathrm{Cl}_c>\ \ldots\},\ \text{or}$$
$$\{<\mathrm{Phr}_a>\ \mathrm{CC}\ <\mathrm{Phr}_b>\ \mathrm{CC}\ <\mathrm{Phr}_c>\ \ldots\}$$

[31] a musical note having the time value of an eighth of a whole note [https://www.vocabulary.com/dictionary/quaver]

[32] Esay in Esa.46 qtd. in Peacham

[33] Winston Churchill

Polysyndeton, also called "cople-clause" by Putterham, is a rhetorical device which connects clauses, phrases or words by the recurrent use of conjunctions. Structurally, polysyndeton is the opposite of asyndeton, and just like the latter, the majority of times, writers make use of the conjunctions deliberately and on purpose. It can slow down the rhythm by guiding the reader's attention to individual details and introducing a flavor of solemnity, or it can speed up the pace by spawning emotions of ecstasy or a forthcoming surprise. For instance, Jane Austen is able to convey a sense of excitement in the next example from *Pride and Prejudice*:

Example (29) *Mrs. Hurst* ***and*** *her sister allowed it to be so–but still they admired her* ***and*** *liked her,* ***and*** *pronounced her to be a sweet girl,* ***and*** *one whom they would not object to know more of.*

Conversely, polysyndeton can be intentionally used to overwhelm the reader or create an atmosphere of boredom by using conjunctions in places where commas are needed and thus leaving the reader with a tight room to breathe mentally (Corbett [1990]).

Example (30) *I got into my old rags and my sugar-hogshead again,* ***and*** *was free* ***and*** *satisfied. But Tom Sawyer he hunted me up* ***and*** *said he was going to start a band of robbers,* ***and*** *I might join if I would go back to the widow* ***and*** *be respectable. So I went back.*[34]

The detection algorithm is implemented within Ruta environment with the help of Stanford Parser. In the first step, we select the valid POS units such as clauses, phrases, and single words. Afterwards, we build a set of rules that match the syntactical pattern described by definition. This is, the rules check for sequences of clauses, phrases, and words split by conjunctions. If at least two conjunctions follow in immediate succession, we consider it as a valid instance of polysyndeton.

3.2.4 Custom Schemes

This category of schemes includes devices which are not broadly recognized as part of rhetoric, yet we suspect they carry a rhetorical value which cannot be ignored. Below, we prove our point by comprehensively describing each instance along with their potential rhetorical impact.

[34] Mark Twain's *The Adventures of Huckleberry Finn*

Conditionals

Definition 3.2.16. A Conditional is a two-clause structure in which one of the clauses is introduced by "if" or by a term which has a meaning similar to "if" (i.e., "only if" or "except if") (Declerck and Reed [2001]).

The given definition is a very rough representation of conditionals, yet it seems to be a common ground for the majority of linguists. The difficulty in defining conditionals lies in their inherent versatility and criteria that can be used to categorize them. As Wierzbicka (1997) correctly points out, "the meaning of the English word condition is semantically more complex than that of IF" (Wierzbicka [1997]). It is enough to illustrate this with valid semantical relations between the conditional and subordinate clauses:

Example (31) *Do it and/or I'll leave you.*

Example (32) *If that witness is speaking the truth, I'm the next President of the U.S.*[35]

The first example shows a form of so-called *paratactic* conditional, which asks the reader to interpret it as a conditional even without obvious conditional indicators (like *if* or *unless*). In the second example, the reader is invited to infer from the obvious falsity of the conclusion (also referred to as Q-clause), that the premise (P-clause) is also incorrect.

Most of the conditionals are introduced by characteristic connecting devices like *if, even if, unless, providing/provided (that), on condition that, as/so long as, in case.* In the rest of the cases, they are either simply omitted (paratactic conditionals) or the conditional sense is made by inversion (e.g. *Had he seen this, he would have been curious.*) (Declerck and Reed [2001]). In this thesis, we will address the conditionals introduced by the most frequent connecting devices: *if, unless* and *whether*. This decision is endorsed by the frequency count performed on a corpus of 300 editorials (∼12000 sentences):

[35] Declerck and Reed [2001]

Table 3.2: Distribution of types of conditionals in a corpus of 300 editorials.

Conditional type	Count
If	558
Whether	50
Unless	22
Lest	3
In case	1
Only/Even if	2

If Conditional

Doubtless, *if-conditional* is the most frequent and flexible type of conditionals; it can introduce nearly any P-clause in the variety of existing conditionals without any loss of meaning. The only limitation would be that it is not semantically equal to *unless, in case* and *lest* (Declerck and Reed [2001]).

Logically, *if-conditionals* are defined in terms of truth relations:

If P, then Q[36]

That is, three cases are possible: (1) P and Q are both true, (2) P and Q are both false, (3) P is false and Q is true (Wierzbicka [1997]).

Because of their nature, *if-conditionals* often create a suppositional or theoretical world, which is either unrelated to the actual world or totally dissimilar. Thus, this type of conditionals is virtually divided by linguists into *factual-P* and *counterfactual-P* conditionals. Counterfactual-P (or simply counterfactuals) conditionals raise a lot of interest from domains like history, political science and recently computational linguistics (Janocko et al. [2015]), and we will refer to it in the context of rhetoric detection in a separate section (see *§Counterfactuals*). Irrespective of the factuality of *if-conditionals*, linguists define four types of them.

Before proceeding to the computational identification of each type separately, it is important to prepare the terrain by establishing an algorithm for finding *if-conditional* sentences along with their constituents: *P-clause* and *Q-clause* (or *protasis* and *apodosis*; we use the former terminology). If the

[36] P – premise or conditional clause; Q – conclusion or head clause

first task is rather effortless, since the presence of an *if* in a sentence, generally, signals an *if-conditional*, the second one is much more complicated. After selecting all the candidates – sentences that contain *if* – it is critical to correctly identify their constituents (i.e., P and Q clauses). To accomplish this, we again rely on Stanford Dependencies. Following the observations on the grammatical structure of *if-conditional* sentences, we have concluded that, most of the times, a clause is uniquely identified by the nominal or clausal subject relations (*nsubj, nsubjpass, csubj* and *csubjpass*) identified by Stanford Dependencies. A nominal/clausal subject is a noun phrase which is the syntactic/clausal subject of a clause (de Marneffe and Manning [2008]). This results in individual *governor-dependent* pairs, for each clause in a sentence. For example, computing the dependencies on the sentence:

$$\underbrace{\textit{I would have been happy,}}_{\textit{Q-clause}}\ \underbrace{\textit{if she had helped me.}}_{\textit{P-clause}}$$

Figure 3.4: Constituents of a conditional sentence.

generates two different nominal subject relations (in which the first is the governor and the second is the dependent; the attached numbers designate the token's position in a sentence):

- *nsubj(happy-5, I-1)*
- *nsubj(helped-10, she-8)*

The boundaries of the P and Q clauses are then established in the following way:

1. get the span of text between *if* and the next governor *(helped-10)* $\rightarrow$ P clause.
2. locate the next closest governor and extract the span of text between it (as the upper limit) and the token placed at most 4 tokens backwards(lower limit) from the identified governor (i.e., the span of text between *I* and *happy*) $\rightarrow$ Q clause.

It is important to mention that the obtained clause might lack in meaning and cannot form an independent unit because of the heuristics assumed with respect to its boundaries. However, as we will see later, this is by no means an obstacle in establishing its true nature.

In some cases, especially when dealing with long sentences, there might be more than two clauses; therefore we extract an additional remaining clause (R clause) for testing purposes.

If-conditional type 0

Definition 3.2.17. Zero conditionals express general truths, events in which the premise always causes the conclusion to happen.[37]

Formally,

If ***simple present****, then* ***simple present***

Zero conditionals are always involved in situation referring to the real world, common sense and natural events. The temporal space the event happens is now or always and the event is real and possible.[38]

Example (33) *If you don't brush your teeth, you get cavities.*

Example (34) *If you heat ice, it melts.*

As the English grammar states[39], zero conditional is formed when both of the clauses (P and Q) are in the simple present tense. That is, the core of the detection algorithm consists in three simple Ruta rules:

1. Confirm the present tense of the P-clause.
2. Confirm the present tense of the Q-clause.
3. Assure that the Q-clause doesn't contain modal verbs (as this will interfere with other conditionals).

As indicators of present tense and modal verbs we consider the Penn Treebank tag set[40] (Marcus et al. [1993]) computed by Stanford Parser. A detailed explanations of all the tags can be found in the Appendix 2.

An approximate illustration of the detection algorithm is given below:

If ***[VB/VBP/VBZ]****, then* ***[VB/VBP/VBZ]***

[37] https://www.grammarly.com/blog/conditional-sentences/
[38] http://www.ef.com/english-resources/english-grammar/zero-conditional/
[39] http://www.ef.com/english-resources/english-grammar/zero-conditional/
[40] https://gist.github.com/nlothian/9240750

If-conditional type 1

Definition 3.2.18. First conditional is used to refer to situations which are very likely (yet not guaranteed) to happen in the future.[41]

Formally,

*If **simple present**, then **simple future***

First conditional is also called "real" conditional since its outcome if often plausible and real. However, its result will happen in the future only if certain conditions are met by that time.[42] We use such sentences to warn or advise somebody. In the suppositional world it creates, the conditions are quite realistic and achievable, which makes us believe, the event is very likely to happen.

Example (35) *If it rains, you will get wet.*

Example (36) *I will be mad if he is late again.*

The detection algorithm is analogous to zero conditional, except that here, we have to impose simple future in the Q-clause. In English, future simple tense is built by combining a modal verb with the infinitive form of the main verb (*will + get*, in the Example (35)). Therefore, we look for the following patterns:

*If [**VB/VBP/VBZ/VBG**], then [**MD+VB**]*

If-conditional type 2

Definition 3.2.19. Second conditional sentences expresses consequences that are totally unrealistic or will not likely happen in the future.[43]

Formally,

*If **simple present**, then **present conditional** or **present continuous conditional***

[41] https://www.grammarly.com/blog/conditional-sentences/
[42] http://www.myenglishpages.com/site_php_files/grammar-lesson-conditionals.php
[43] https://www.grammarly.com/blog/conditional-sentences/

Contrary to the first conditional, this one is called "unreal" conditional, as it is used for impossible events. The usage of auxiliary modal verbs like *could, would, should*, in the Q-clause automatically shifts the perspective by expressing an unrealistic outcome. It refers to a hypothetical imaginary world which is utterly unrelated to the actual one. Hence, this conditional, along with the next one, is oriented towards counterfactuals.

Example (37) *If it rained, you would get wet.*

Example (38) *If I was the Queen of England, I would give everyone a chicken.*

As we advance with the type of conditional, the requirements a sentence has to fulfill become stricter. The advantage is that more relevant instances are, allegedly, going to be detected (i.e., higher precision); however, this means that more true instances of second conditional will be discarded (i.e., lower recall), because of the subtleties of the language.

As previously, the detection consists in "projecting" the definition onto Ruta rules. That is, the rules consider the verb forms in the required tenses in each of the clauses separately. On top of that, it is imperative to assure the presence of an auxiliary modal verb in the Q-clause. This leads to the following formalization:

*If [**VBD**], then [**MD+VB**]*

If-conditional type 3

Definition 3.2.20. Third conditional sentences are used to explain that present circumstances would be different if something different had happened in the past.[44]

Formally,

*If **past perfect**, then **perfect conditional** or **perfect continuous conditional***

Within third conditional, both the premise and the result refer to the past. In

[44] https://www.grammarly.com/blog/conditional-sentences/

the suppositional world it shapes, it describes an impossible event in the past and its hypothetical outcome, also in the past. None of the constituents makes sense in the present moment since it is already too late for them to happen. Rhetorically, third conditionals might evoke emotions of regret or guilt.[45]

Example (39) *If I had worked harder, I would have passed the exam.*

Example (40) *If he had been careful, he wouldn't have had that terrible accident.*

Even if it sounds slightly more complicated to detect, we show that it is enough to adapt the detection algorithm from previous conditionals by replacing the syntactical search pattern. Unlike second conditionals, we look for a bond of verb forms (specifically, *past tense + past participle*) in the P-clause, as well. Formally, the algorithm consists in:

*If [**VBD** + **VBN**], then [**MD** + **VBN**]*

If Counterfactual

Definition 3.2.21. "Counterfactuals are statements that examine how a hypothetical change in a past experience could have affected the outcome of that experience." (Janocko et al. [2015])

The assumed environment or world in which the premise and conclusion of conditionals are set is an important parameter which defines their nature. In particular, as we observed, it is the type of the temporal space which in large measure controls the verb forms, tenses and modal auxiliaries in conditionals. Literary, there are four large categories which describe the conditional "world": *factual, theoretical, neutral* and *non-neutral* (Declerck and Reed [2001]). The last category – non-neutral, also has four subtypes; and though each of them present interest on their own, in this thesis we will focus on *counterfactuals*. Returning to the definition, a sentence is called counterfactual, or *counterfactual-P conditional* (Declerck and Reed [2001]), if the hypothetical world within the P-clause is supposed to be contrary to the fact; this is, completely different from the actual world (Declerck and Reed [2001]). For instance, in the sentence *If I were you, I wouldn't come*, the P-clause ("If I were you") is assumed to be true in a theoretical counterfactual P-world, yet false in the real world.

[45] http://www.ef.com/english-resources/english-grammar/type-3-conditional/

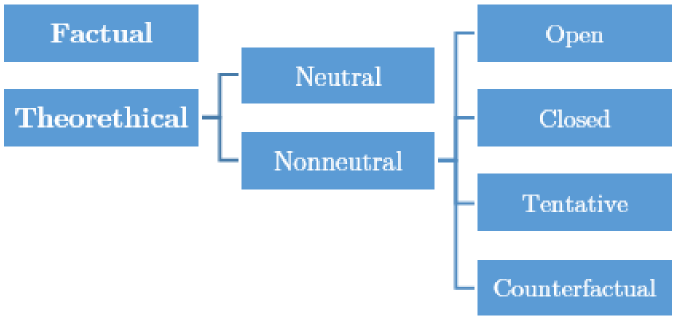


Figure 3.5: Semantical categorization of conditionals (based on P-clause type).

Counterfactual thinking might seem irrelevant in the context of rhetoric, yet just like the latter, it is able to impact the personalities and emotions of the audience. A research conducted in 2010, concluded that practicing counterfactual thinking around certain aspects or events in subjects' life, helps them to assign more meaning and importance to the said events (Kray [2010] [as cited in Janocko et al. 2015]). Furthermore, it has been shown that counterfactual thinking is able to improve team-working skills; and that people suffering from schizophrenia are significantly less likely to employ counterfactuals (Hooker and Park [2000] [as cited in Janocko et al. 2015], Galinsky and Kray [2004] [as cited in Janocko et al. 2015]).

It is important to mention that counterfactuals and if-conditionals are not always grammatically separable. Very often, counterfactuals take the form of traditional second and third conditionals, especially when they imply impossible events (Webb [2012]). However, we cannot consider them identical, since counterfactuals are much more grammatically diverse and relaxed in terms of definition.

Our approach to detect such conditional is based on two assumptions, which, from our observations, hold true in the majority of cases.
First, we assume that the governor verb of the P-clause ought to be in the past tense (i.e., past simple and past participle); that's intuitive since there is highly unlikely for a counterfactual premise to happen in the future, it's about a "hypothetical change in a past experience" (Janocko et al. [2015]). However, as we show in *§Assumptions and Limitations* section, there is a special type of counterfactuals which contradicts the above claim.
The second hypothesis we adopt concerns the structure of the Q-clause, and namely, it must contain one past tense modal verb like *should have, could*

have, and *would have*. Being also called "modals of lost opportunity", these constructions are used to express the present emotions about a past decision, or even describe situations when we imagine that the past was different[46]. Thus, they represent the key ingredient in the detection of counterfactuals. Formally, the algorithm will look for:

If ***VBD + VBN****, then* **past modals**

Example (41) *I would have been happy, if she had come.*

Having fixed all the conditions required to identify counterfactuals, we proceed to implement them as a chain of Ruta rules. Considering the versatility of counterfactuals and the complexity of language in general, we did not expect, any impressive results from our approach; however, the evaluation experiments (table 3.4) show some decent results on that. On the other hand, some studies report an accuracy of little over 90% in identifying conditional counterfactuals (Janocko et al. [2015]).

Unless-conditional

Definition 3.2.22. Unless conditional is a restricted version of if-condtional, in a sense that its intrinsic meaning is narrowed down to "Q in the case other than P"(Declerck and Reed [2001]).

Literary, unless conditional is equivalent to *if. . . not* or except it and it is mostly used when referring to present situations.

Example (42) *You can't go on vacation* ***unless*** *you save some money.*

In rhetoric, unless conditional does not have any special implications compared to other types of conditionals. Nonetheless, since it is a rather frequent conditional (see table 3.2), we aim to assess its rhetorical influence within our data.

To the best of our knowledge, *unless* cannot be used outside the context of conditionals. Therefore, we build the detection algorithm based on this statement and simply extract the sentences which contain at least a single instance

[46] https://www.espressoenglish.net/past-modals-should-have-could-have-would-have/

of *unless*. These sentences are directly annotated as unless conditionals.

Whether. . . or-conditional

Definition 3.2.23. Whether. . . or conditionals are "used to express alternative (disjunctive) conditions."(Declerck and Reed [2001]).

Whether. . . or conditionals are part of a separate category, called "alternative-P" conditionals. It is worth mentioning though, that even if they seem to be semantically identical, *if* and *whether. . . or* cannot always substitute each other; that is, we cannot use *if. . . or* in such "alternative-P" sentences (Declerck and Reed [2001]). Declerck and Reed also interestingly observe that, albeit the conditions are disjunctive (i.e., only one at a time can be satisfied), the meaning of *whether A or B* is not "either if A or B" but "both if A and if B".

Example (43) ***Whether*** *you are overweight* ***or*** *not, it is always better to watch your diet.*[47]

Example (44) ***Whether*** *you did it* ***or*** *Mary (did it), the whole class will be punished.*[47]

The detection algorithm adopts the exact same structure used in the case of unless conditionals. Once the candidate sentences are identified, they are automatically annotated as "whether. . . or" conditionals.

Comparative Adjectives and Adverbs

Definition 3.2.24. Comparative adjectives and adverbs are used to compare differences between the two objects/states they modify.

Grammatically, the difference between adjectives and adverbs consists in the fact that the former (Example (45)) describes the objects by directly addressing the nouns, while the latter (Example (46)) characterizes the actions or states and refers to the verbs (or any other POS except nouns and pronouns).

Example (45) *My house is* ***larger*** *than yours.*

Example (46) *Tim works* ***harder*** *than me.*

[47] Declerck and Reed [2001]

As insignificant as they may seem, comparatives, in fact, categorically influence our day-to-day lives. Humans tend to make decisions based on comparisons between their constituents and some benchmarks they set. We do this involuntarily, and advertisers know about it by making use of this weakness to promote their "greatest", "biggest" and "shiniest" products. In the context of comparatives, they place their products on a higher pedestal by directly engaging a direct comparison with their concurrent. That's why we consider important to evaluate the rhetorical value of apparently "rhetoricless" devices.

On top of its value, comparatives are fairly easy to detect just by making use of Stanford Parser. Inherently, it can produce annotations comparative and superlative adjectives/adverbs. In particular, we look for tags like **JJR** and **RBR**, as they denote the comparative form of adjectives and adverbs respectively. Of course, this "shortcut" comes with sizeable drawbacks in terms of performance (see *§Evaluation*). Nevertheless, as an initial step in assessing their rhetorical impact, the results look decent.

Superlative Adjectives and Adverbs

Definition 3.2.25. Superlative adjectives and adverbs are used to describe an object/action which is at the upper or lower limit of quality.

Example (47) *My brother is the **shortest** person in the team. (superlative adjective)*

Example (48) *Mrs. Smith talks **most quietly**. (superlative adverb)*

We suspect superlative adjectives and adverbs to carry a much more emotionally charged value than comparatives do.
Firstly, there is a concept called superlative rhetoric, in which one engages superlative forms of words or ideas and use them to persuade the target audience. For instance:

Example (49) *This phone is **the first** to outperform the iPhone.*[48]

Example (50) *Sign up here for **the best** newsletter on the web!*[48]

It is important to note that superlative rhetoric is not always about the adjectives or adverbs in their superlative form.

[48] http://www.danieltrichards.com/superlative-rhetoric-the-first-best-only-post-youll-ever-need-to-read-period/

Example (51) ***Only** subscribers to our email list get the discount.*[48]

The idea behind this technique is simple: "creating urgency encourages action" (Richards [2014]). The moment when people feel that some occasion is about slip away, they start to do something about it.

Although experts strongly advise against the frequent use of superlatives, by arguing that "We do not wish to give the impression that we live in a constant state of excitement." (Wadsworth [1950] [as cited in Gingell 2016]), our social lives and media carelessly make use of them, in a hope to persuade people. For instance, Donald Trump seems to be very fond of superlatives, as he constantly used them to put himself "on a higher shelf"; proclaiming himself as "the most militaristic person ever" and declaring that he would become "the greatest jobs president that God ever created" (Gingell [2016]).

Secondly, superlatives introduce a nonstandard concept of double superlatives, a literary device consisting in the usage of the both *most* and the suffix *-est* (the same is valid for comparatives). They are preferred for their distinct rhetorical force and emphasis.

Example (52) *Mirror, mirror, on the wall, who's the **most baddest** angry young man of all?*[49]

Although they are rarely used in contemporary English, it is important to mention them in the context of superlatives and rhetoric, in general.

Again, to detect such constructions, we rely on the annotations of superlative forms of adjectives and adverbs produced by the Stanford Parser. In this case, we search for **JJS** and **RBS** tags and annotate them as superlative adjectives and adverbs respectively.

Passive Voice

Definition 3.2.26. A passive voice is a type of a clause or sentence in which the focus is put on the main action, or object of the said sentence rather than its subject. This is, the subject receives the action of the verb.[50]

Literary, two types of passive voice can be distinguished: *short form* and *long form*. By far, the most commonly used form in English is short passive

[49] Donald Barthelme, "Before the Mirror." Sixty Stories. G.P. Putnam's Sons, 1982

[50] https://literarydevices.net/passive-voice/

(also called, *agentless*), wherein the action is performed by an unknown character (i.e., agent); for instance, "The problem **is solved**". In the long form of passives, the object of the verb in an active sentence becomes the subject; for instance "The problem **is solved by us**" (Nordquist [2017]).

Passive voice can rhetorically be useful when the author wants to move the spotlight form the subject to an action. Even though, style guides do not endorse the usage of passive voice in technical writings, since it slows the reading pace by making the text lengthy, passive voice has a large presence in literature and scientific papers (Nordquist [2017]). That's because, in such writings, the subject might be less important than the action it performs. Furthermore, passive voice is useful in this context, to avoid taking responsibility for the carried actions[50].

Grammatically, a passive voice construction is formed by using an appropriate inflection of the verb *to be* in combination with a *past participle*. This formalization is type-independent, in a sense that it covers both forms of passive: short and long.
Algorithmically, we "translate" the grammatical formalization in Ruta rules able to detect such patterns. In the first stage, we build a short lexicon containing all the forms of the verb *to be*, which later serves us to detect and annotate them in the analyzed sentence. Next, we consider any *past participle* verb situated at most four words to the right, apart from the detected form of the verb *to be*. Considering that we don't preprocess the input sentence, the default window size of four words seems to be appropriate, since punctuation marks and stopwords might occur in between the two constituents of a passive voice instance.

3.2.5 Assumptions and Limitations

Isocolon

As it was already mentioned, isocolon is a type of parallelism which consists of a sequence of equivalent phrases or clauses, both in terms of length and structure. One assumption, and therefore a relaxation, we adopt regarding the detection algorithm, directly affects the interpretation of the POS tags; tags on which we rely to identify the parallel structures in the sentence. In particular, we reduce the range of the POS tag forms (i.e., singular, plural and proper for nouns, the verb tenses as well as the adjective and adverb comparison degrees) to a unique base form. For instance, the POS tag corresponding to the past participle of a verb - VBN would be replaced with another tag denoting the verb's infinitive form, or VB. This step is necessary as we directly compare the constituents of the existing parallel structures against each other, and a minor modification of the tag form would instantly signify the lack of parallelism between the engaged structures.

Example (53) *Fill the armies, rule the air, pour out the munitions, strangle the U-boats, sweep the mines, plow the land, build the ships, guard the streets, succor the wounded, uplift the downcast, and honor the brave.*[51]

To exemplify this, let us consider Example (53); the whole sentence consists of equivalent clauses, but in some of them, the author refers to plural nouns (e.g., armies, munitions) while in the others - to singular (e.g., air, land). As Penn Treebank (see Appendix 2) defines individual tags for each form, the detection of isocolon in the above example would be saddled.

If-Conditionals

A noteworthy limitation of our approach concerning conditionals lies in our method of detection third conditionals with the help of Stanford Parser. Grammatically, the main verb structure of the Q-clause of a third conditional must be in the perfect conditional or perfect continuous conditional, as in "If this thing had happened, that thing *would have happened.*" Respectively, the perfect (continuous) conditional of any verb consists of:

would + have + past participle

We observed that the verbs whose form remain invariant between the past tense and past participle, tend to confuse the Stanford Parser in a way which directly affects the performance of our algorithm. For example, the words "crash" and "pass" are identical in their forms for all of the grammatical tenses (table 3.3).

Table 3.3: Tenses forms of verbs *crash* and *pass*

Infinitive	Present Participle	Past Tense	Past Participle
crash	crashing	crashed	crashed
pass	passing	passed	passed

[51] Winston Churchill, speech given in Manchester, England, on January 29, 1940

However, the output generated by the Stanford Parser is inconsistent between these two:

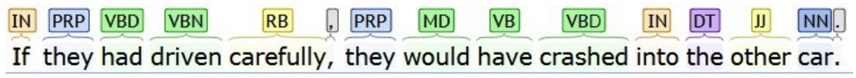


Figure 3.6: POS tags[52]of the sentence *If they had driven carefully, they would have crashed into the other car.*

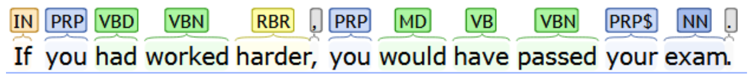


Figure 3.7: POS tags[52] of the sentence *If you had worked harder, you would have passed your exam.*

As we can see, the tense of "passed" is determined correctly in the context of third conditional, yet "crashed" is considered to be in the past tense.

Another limitation, which was anticipated, is given by the fact that some conditionals do not respect the common pattern defined by the majority. In other words, they don't conform to the conditional usage of "if" (Declerck and Reed [2001]). For example, we cannot put the equal sign between the examples Example (54) and Example (55).

Example (54) *John asked **if** he could come now.*[53]

Example (55) ***If** John asked he could come now.*[53]

Likewise, counterfactual conditionals might behave similarly. Although there are inherently considered a kind of ad *absurdum* conditionals, linguists agree that they can be interpreted as counterfactuals because of the way Q-clause sounds (Declerck and Reed [2001]).

Example (56) *If you're the General Manager here, I'm Dracula!*[53]

We are aware of these type of anomalies as they cannot be easily detected. Fortunately, they constitute just a tiny amount of conditionals compared to the common patterns we consider in our approach.

[52]generated by http://nlp.stanford.edu:8080/parser/index.jsp
[53]from Declerck and Reed [2001]

3.3 Evaluation

The evaluation phase consists in assessing the performance measures of our approach in detecting the listed rhetorical devices. As mentioned in chapter 2.2 (*§Related Work*), we adopt an automatic type of evaluation; that is, we simulate, to some extent, the behavior or trained annotators, which would, in the case of manual evaluation, assess the system's performance (Clark et al. [2010]).

This first step towards this type of evaluation is to collect the examples of rhetorical devices, against which our system will be evaluated. It is imperative for the built dataset, to contain only verified and actual instances of rhetorical devices of each kind. Therefore, it is our primary focus to consider only trustworthy sources with some expertise in the domain of rhetoric. Each file in our dataset would consist of the instances of one particular type of device. Most of the examples are collected from literature, Bible, political speeches, commercials and websites like Silva Rhetoricae (Burton [2007]) and RhetFig (Bowden and Harris [2011]) on less known websites have been individually examined to follow the rules imposed by definition.
To build a balanced dataset of rhetorical figures, we set a benchmark of 60 instances per device. This is significantly more than the sample size considered Java [2015], so we predict much more evenhanded results. As expected, some of the rhetorical devices (specifically, *mesarchia*) in our list turned out to be so uncommon that we barely could find 20 valid examples; still, when possible we exceeded the average number per category (superlative and comparative adjectives/adverbs).
The second, and final step in determining the evaluation measures, consists in the actual execution of the program, having as input the collection of files containing instances of each rhetorical device separately. We compute the precision and recall for each device individually, which further allows us to calculate the F1-score. Before jumping to the analysis of the performance results, we will briefly define them in the context of rhetoric detection.

- *Precision*: The number of correctly identified instances of the tested rhetorical device, divided by the total number of identified instances of the said device. Formally,

$$Precision = \frac{true\ positives}{true\ positives\ +\ false\ positives}$$

  where *true positives* denotes the number of correctly identified examples, and *false positives* – the number of misclassified instances of the same device. A high precision means that the majority of the devices were correctly identified than misidentified (Java [2015]).

Table 3.4: Evaluation measures results of our system

Device	Total No.	Precision	Recall	F1-score
Anadiplosis	60	0.76	0.73	0.74
Asyndeton	60	0.25	0.95	0.4
Comparative Adjective	67	0.51	0.61	0.56
Comparative Adverb	71	0.6	0.62	0.61
Diacope	60	0.75	0.73	0.74
Enumeration	60	0.76	0.93	0.84
Epanalepsis	60	0.63	0.83	0.72
Epiphoza	60	0.61	0.93	0.74
Epizeugma	60	0.68	0.7	0.69
Epizeuxis	60	0.79	0.77	0.78
Hypozeugma	60	0.61	0.8	0.69
If Conditional One	60	0.78	0.78	0.78
If Conditional Three	60	0.86	0.65	0.74
If Conditional Two	60	0.82	0.75	0.78
If Conditional Zero	60	0.71	0.76	0.73
If Counterfactual	60	0.84	0.87	0.85
Isocolon	180*	0.57	0.83	0.68
Mesarchia	20	0.45	0.85	0.59
Mesodiplosis	40	0.28	0.68	0.4
Passive Voice	60	0.79	0.98	0.87
Polysyndeton	60	0.77	0.7	0.73
Pysma	60	1	1	1
Superlative Adjective	70	0.62	0.73	0.67
Superlative Adverb	70	0.63	0.5	0.56
Unless Conditional	60	1	1	1
Whether Conditional	60	1	0.83	0.91

* includes samples of bicolon(60), tricolon(60) and tetracolon(60)

- *Recall*: The number of correctly identified instances of the tested rhetorical device, divided by the total number of instances that should have been identified. Formally,

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

where *true positives* denotes the number of correctly identified examples, and *false negatives* – the number of instances which have not been annotated as the device under evaluation. A high recall suggests that most of the devices were correctly identified (Java [2015]).

- *F1-score*: The harmonic average of *precision* and *recall.* It is generally used to assess the system's accuracy. Formally,

$$F1\text{-}score = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

In the next sections, we will analyze and interpret the performance measures, as summarized in the table 3.4.

3.3.1 Schemes of Balance

Enumeration

The built set of Ruta rules having as task the detection of enumeration performs rather well as we can infer from the evaluation measures. However, it is interesting to consider the valid samples of enumeration which the algorithm failed to detect. After analyzing the missed examples, we concluded that the issue arises in the case of enumerations whose constituents (or enumerated items) are larger than five tokens. Let us consider the Example (57):

Example (57) *You will find that they will buy your surplus land, make blossom the waste places in your fields, and run your factories.*[54]

The enumeration is not detected in this sentence, since the length of "make blossom the waste places in your fields" is larger than five tokens. On the other hand, increasing this threshold accounts for more false positives. Furthermore, as conveyed by the recall value, the majority of enumerations are characterized by short item lengths (1 up to 5 tokens).

[54] Booker T. Washington Delivers the 1895 Atlanta Compromise Speech

The second remark regarding enumerations is that occurrences of this device might also overlap with other rhetorical devices such as polysyndeton or isocolon, because of similar characteristics. For instance, the English grammar tolerates the usage of commas before conjunctions like:

Example (58) *And Joshua, and all of Israel with him, took Achan the son of Zerah, and the silver, and the garment, and the wedge of gold, and his sons, and his daughters, and his oxen, and his asses, and his sheep, and his tent, and all that he had.*[55]

As a result, this would drastically bias the precision measure; so we consider natural and appropriate to allow such intersections between rhetorical figures.

Isocolon

The subtleties of isocolon caused it to be the most difficult rhetorical device to detect. As the definition implies – similarly structured elements having the same length – it was imperative to consider not only the word count but also the structure of its constituents; this is, analysis of the POS tags. To reiterate, our detection algorithm relies on sequential matching pairs of POS tags as on the sentence level, so on the paragraph level. To make any distinction, or in other words, to isolate individual parallel elements, we depend on punctuation marks and on POS tags, which interrupt the current parallel element.

Example (59) *Fill the armies, rule the air, and pour out the munitions.*[56]

In Example (59), the isolation of isocolon's constituents is based on commas; thus, the extracted clauses are ready to be analyzed. Although our approach seems intuitive, it is very volatile; this is reflected in the precision value. Since we depend solely on the parser's POS tags, slight deviations or misinterpretations of the function of certain words automatically discards the isocolon candidate.

Example (60) *It takes a licking, but it keeps on ticking!*[57]

The structural similarity in Example (60) is obvious; however, it is not so evident for the detection algorithm. The problem lies in the POS tags

[55] The Bible

[56] W. Churchill at Manchester, 29 January, 1940

[57] advertising slogan of Timex watches

attributed to the words *licking* and *ticking*. While the former has been considered a noun (NN), the latter is treated as a verb (VBG). On top of that, in the first clause, the article "a" is a determiner (DT), whereas "on", in the second clause, is a preposition (IN). These intricacies are hard to detect relying exclusively on the parsing tree, so currently, we just ignore them.

3.3.2 Schemes of Omission

Asyndeton

The evaluation results of this rhetorical device somewhat puzzling. On the one hand, the algorithm manages to obtain a good recall measure (95%); this is, to detect almost all the relevant instances from the dataset of valid examples of asyndeton; on the other hand, it highly prone to misidentification when determining the precision (25%). As we later observed, this is explained by the fact that the abundance of commas, which implies the lack of conjunctions, is not enough to uniquely identify asyndeton from the whole list of samples of rhetorical devices. The majority of misclassified asyndetons, in fact, are instances of enumerations, which are also characterized by similar criteria; yet they are not equivalent in rhetorical value, so we cannot tolerate the intersections of examples.

Example (61) *Old McDonald had* ***a pig, a dog, a cow and a horse***[58]

The Example (61), is annotated as an asyndeton even though it is a valid example of enumeration. Rhetorically, this example has the effect of amplification (by virtually increasing the size of Old McDonald's farm), which innately describes the enumerations.

Hypozeugma

As described in section *§Schemes of Balance*, the detection process of hypozeugma is fragmented into two stages; this means that each stage is potentially a source of errors. The performance results look decent; still, we would like to discuss some of the issues spotted.

[58]Old MacDonald Had a Farm: Traditional Children's Music
[59]by http://brat.nlplab.org/index.html

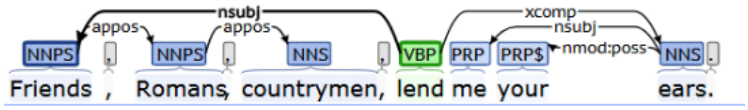


Figure 3.8: Graphical visualization[59] of the typed dependencies in the sentence *Friends, Romans, countrymen, lend me your ears.*

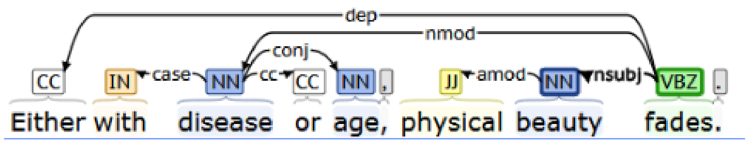


Figure 3.9: Graphical visualization[59] of the typed dependencies in the sentence *Either with disease or age, physical beauty fades.*

Both of the examples in the figures above (3.8 and 3.9) are valid instances of hypozeugma; yet while the first (3.8) is correctly identified, the second instance (3.9) is missed. The reason lies in the unexpected detection of the nominal subject relation (*nsubj*) generated by Stanford Dependency. That is, as illustrated in 3.8, the parser covers the relevant portion of the sentence (i.e., *Friends, Romans, countrymen, lend*) which meet the Ruta rules (see *§Schemes of Balance*) to be successfully annotated as hypozeugma. Conversely, the nominal subject relation in the missed instance, is built between adjacent words (i.e., *beauty fades*), and therefore Ruta is unable to recognize the pattern which characterizes hypozeugma, in the extracted *nsubj* relation. In general, this appears to be the most common type of mistake which leads to misclassification; however, we do not see any better alternatives to solve this problem.

Epizeugma

As for epizeugma, our system performs comparably well to its closest by structure device – hypozeugma. We would like to remind the reader that even though these two rhetorical devices are closely related, we come with a different approach in the detection of epizeugma. In particular, we no longer rely on the computed dependency relations in the sentence; instead, we consider the constraints about the position of the governor as the definition states ("either very beginning of very ending of the sentence"). This allows for a more robust and reliable implementation.

In an attempt to fulfill the definition and consider just a single governor verb which "holds the entire sentence" (see *§Schemes of Balance*), we filter out the candidates with more than one verb. This, however, turns out to be the common reason for undetected instances of epizeugma.

Example (62) ***Take****, if you* ***must****, this little bag of dreams.*[60]

The sentence in Example (62) is a valid instance of epizeugma; since the verb which governs the entire sentence is at the very beginning of it. However, it is a missed instance by our system because of the intervening construction "if you must", which contains a modal verb. We considered a relaxation of the algorithm which would allow modals as the second verb in the sentence; still, it doesn't completely solve the problem because many other verb types are possible as well.

3.3.3 Schemes of Repetition

Epanalepsis

The detection algorithm identifies the majority of epanalepsis instances correctly (recall - 83%); however, regarding the precision measure, a lot of epizeuxis instances are misclassified as epanalepsis. That is due to the correspondence between their intrinsic structure, and namely, epizeuxis happens when a series of words are repeated with no other in between whereas epanalepsis, denotes the repetition of the same word at the beginning and end of the sentence. To illustrate this, let us consider the Example (63):

Example (63) ***Alone****, alone, all, all* ***alone.*** (epizeuxis)

It is obvious that this confusion cannot be avoided and since they carry different rhetorical values, we disregard the annotation of this example as epanalepsis.

A different source of misidentifications comes from our assumptions regarding the boundaries of the beginning and end of the sentence. As stated in section *§Schemes of Repetition*, we split the sentence into five parts and mark the start to be the first one fifth and the ending – the last one-fifth of the sentence. As expected, this assumption accounts for some false positives like in Example (64).

[60] William Butler Yeats

Example (64) *When the **teacher** is in the classroom, the students study more quietly than when the **teacher** is not there.*

In this sentence, the repetition of "teacher" is rather unintentional. Still, the algorithm considers it a valid instance of epanalepsis, which ironically, conforms the definition (to some extent).

Mesarchia

As mentioned at the very beginning of this chapter (3.4), mesarchia is the rhetorical device with the least number of instances (20) in the evaluation dataset. Therefore we admit that the performance results might be slightly biased in comparison to the rest.

As per our observations, the only source of erroneous identification lies in the heuristics regarding the beginning and middle of the sentence. Therefore, since the definition lacks in exactness concerning these concepts, a minor deviation (from the predefined beginning and middle) in the position of the repeating words across sentences will instantly disqualify valid instances of mesarchia.

Epiphoza

Most of the misidentified epiphozas either comply with the definition, but overlap with other devices of the same nature, or, they are completely accidental by misinterpreting the ending of consecutive sentences.

Example (65) *If I am right then the whole **world** will applaud. If I am wrong then the whole **world** will despair.*

For instance, the above pair of sentences is originally a bicolon; however, given that the word "world" is contained within the endings of these sentences, the couple is annotated as epiphoza. A contraction of the boundaries is a potential solution for this problem, although detrimental consequences are expected. We would like to restate that following our observations on rhetorical devices which depend on the correct fragmentation of the sentence, show that the adopted size of the sentence portions represents a compromise between precision and recall.

Mesodiplosis

As in the case of previous rhetorical devices, some instances of mesodiplosis were missed by the algorithm because they did not obey the boundary constraints. It is important to mention that the evaluation measures are determined on just 40 samples of mesodiplosis. Another problem, as we observed, consists in the recurrent usage of conjunctions, prepositions and other stopwords which we disregard in the preprocessing phase, as indicators of validity of mesodiplosis.

Example (66) *One, **but not** two. Three, **but not** four.*

This is a missed instance of mesodiplosis since "but" and "not" are part of the stopwords list on which we rely to filter them out. In an attempt to remove this restriction, allegedly, we would accept much more false positives (since the repetition of stopwords in most of the cases is unintentional (Strommer [2011])). It is possible though, to reduce the negative impact of stopwords by considering a whitelist of the most frequent stopwords; however, this step would require significant efforts and ultimately, it does not guarantee, by any means, the reliability.

Anadiplosis

Anadiplosis involves the repetition of the same word(s) at the beginning and end of successive sentences. Our system can correctly identify the majority of the instances of anadiplosis (recall – 77%; precision – 80%). As for the most of the devices of repetition previously mentioned, the algorithm fails to identify the examples wherein the parts of the sentence inferred virtually from the context don't correspond with our norms defined in the algorithm. Analyzing the missed instances of anadiplosis, we conclude that pairs of short sentences are affected in the first place.

Example (67) *Insecurity breeds **suspicion and fear. Suspicion and fear** breed violence.*

Let us take Example (67); since the instances are relatively short, the boundary selection algorithm marks just the last and first word in the sentences as their ending and beginning, respectively; thus, "fear and Suspicion" forms a pair of distinct words, resulting in a missed instance of anadiplosis.

Diacope

As we mentioned in *§Schemes of Repetition*, diacope consists in the repetition of a word with some intervening words in between. Concerning its detection performance, the length of the intervening words is often the cause of misidentification

Example (68) *I knew it; born in a hotel room - and goddamn it - died in a hotel room.*[61]

As we allow not more than five tokens between the repeating terms, Example (68) is missed by the algorithm. Evidently, by increasing the size of the intervening window we obtain a much higher recall value; however, as Strommer [2011] show in his experiments on a similar topic, as the window length between the repeating words increases, the rhetorical value of the device decreases (i.e., more likely to be unintentional).

Epizeuxis

Unlike diacope, by definition, this rhetorical device doesn't tolerate any intervening words between the repeating ones (although exceptions exist). As it was predictable, the algorithm is performing rather well (precision 79%; recall 80%). Still, we would like to highlight some "questioning" instances of epizeuxis.

Example (69) *Scotchy, scotch, scotch.*

Although Example (69) was correctly identified, our algorithm doesn't take the full length of the sentence as epizeuxis. Since we account for identical words in a sequence, the diminutive form "scotchy" is not annotated as a part of rhetorical device; however, from a pragmatic point of view, the whole sentence should be taken as epizeuxis.

An important remark to make here is that we consider only single consecutive words in our approach. Since there is no large consensus regarding the number of words which makes up the constituent of repetition, we take as valid the smallest unit – a single word. This interpretation, however, disregards consecutive phrases, which can also be an instance of epizeuxis (Example (70)).

Example (70) *Over and over and over again.*

[61] Eugene O'Neill

Polysyndeton

The abundance of conjunctions in a sentence signals the presence of polysyndeton. This is the premise we rely our algorithm upon. We would like to remind the reader, that our approach considers only the conjunctions between clauses, phrases, and words. It means that any pair of conjunctions within the boundaries of a clause or phrase, with more than one word in between, is neglected by the algorithm.

Example (71) *If there be cords,* ***or*** *knives, Poison,* ***or*** *fire,* ***or*** *suffocating streams, I will not endure it.*

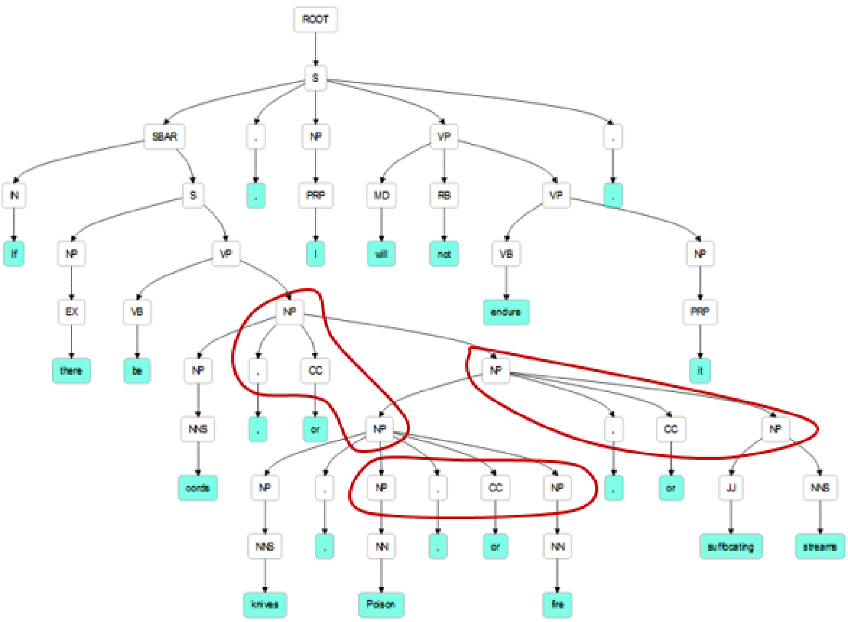


Figure 3.10: Parse tree[62]of the sentence *If there be cords, or knives, Poison, or fire, or suffocating streams, I will not endure it.* We can observe the intervention of commas between the phrases and thus, the algorithm fails to detect it as a valid instance of polysyndeton.

For instance, Example (71) is missed because it contains no consecutive phrases. Let us take a look at figure 3.10; we can spot the intervention of commas between the phrases and thus, the algorithm fails to detect it. As we already mentioned, English grammar does not encourage or discourage the use

[62]by Ben Podgursky, http://nlpviz.bpodgursky.com/

of commas before the conjunctions, so they easily can be eliminated; however, we wanted to keep the originality of the examples, to be aware of the limitations of our approach. It is worth mentioning that instead of commas, we could as well find different types of tokens; so a simple solution like tolerating non-alphabetic characters would not solve the problem.

3.3.4 Custom Schemes

If-conditional type 0

As in case of all the conditionals, zero conditional is described by an exact set of rules which makes the targets of the algorithm less fuzzy in comparison to the rest of rhetorical devices. However, since the detection performance depends upon the correct extraction of the two conditional constituents, premise (P-clause) and conclusion (Q-clause), this constitutes a major source of errors.

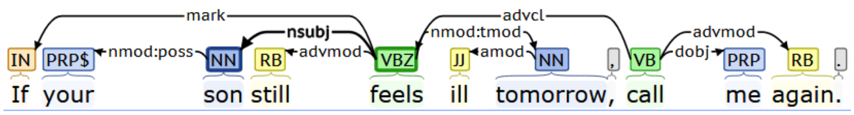


Figure 3.11: Graphical visualization of the typed dependencies in the sentence *If your son still feels ill tomorrow, call me again.*

The conditional in figure 3.11, for instance, is missed because the Q-clause cannot be identified here. The first nominal subject relation (*nsubj(feels, son)*) permits the extraction of the P-clause; still, the lack of a second *governor-dependent* relation makes this example invalid.

Another ignored instance is caused by the erroneous interpretation of the POS tags.

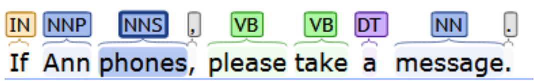


Figure 3.12: POS tags visualization of the sentence *If Ann phones, please take a message.*

Because of the ambiguous nature of the word "phones", the POS tagger treats it as a noun, which in turn, invalidates the P-clause along with the whole example.

If-conditional type 1

The evaluation measures of this device are slightly better (precision 78%; recall 78%) in comparison to zero conditional. Nevertheless, the detection algorithm still suffers from an inaccurate selection of the conditional constituents.

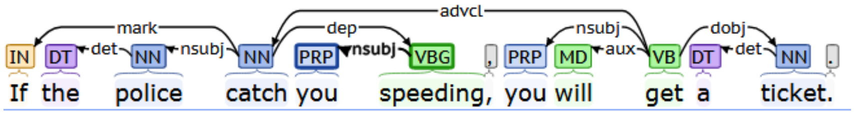


Figure 3.13: Graphical visualization of the typed dependencies in the sentence *If the police catch you speeding, you will get a ticket.*

In figure 3.13, we can observe three nominal subject relations: *nsubj(catch, police)*, *nsubj(speeding, you)*, and *nsubj(get, you)*. Even though our algorithm is able to extract all of them, the problem lies in the labels assigned to each relation. As we mention in *§Custom Schemes*, in case of more than two relevant dependency relations in a sentence, we denote the P-clause to cover the relation which follows immediately after the "if" particle, in this case - *nsubj(catch, police)*; next, the Q-clause is selected heuristically, based on the distance from the P-clause (the closest is favored). This assumption fails in this case since the Q-clause would cover the relation *nsubj(speeding, you)*, which is not a relevant dependency relation, unlike the last one - *nsubj(get, you)*.

If-conditional type 2

As per definition, the second conditional happens when the P-clause is in the simple past tense, and the Q-clause contains an auxiliary modal verb (e.g., would, could) along with the infinite form of the main verb[63]. Unfortunately, exceptions exist, and they constitute an important source of errors in identification. Let us take an example of such exception and discuss the implications.

Bearing in mind the aforementioned definition, the detection algorithm would expect a modal verb in the Q-clause (i.e., *he had have*); however, a verb in the past tense "had", doesn't fulfill the strict rule of the implementation and the sample is neglected. Such subtleties are difficult to catch since pragmatically,

[63] https://www.ego4u.com/en/cram-up/grammar/conditional-sentences

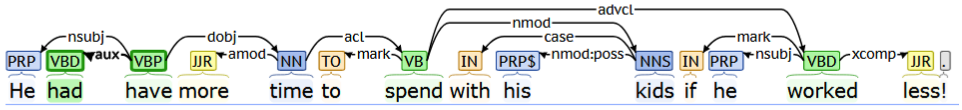


Figure 3.14: Graphical visualization of the typed dependencies in the sentence *He had have more time to spend with his kids if he worked less!*

replacing "had" with "would" doesn't change the outcome of the conditional sentence.

If-conditional type 3

Third conditional is defined by much more strict rules which ultimately, helps us to cut off more potential false positives. As we can infer from the table of evaluation results (table 3.4), the precision measure is inversely proportional to the degree of conditionals. On the other hand, by tightening the rules, we allow fewer exceptions to pass; that is, a lower recall value.

Most of the missed instances of the third conditional are due to incorrect selection of the Q-clause. Like in case of first conditionals, a third (or more) relevant dependency relation (*nsubj(pass), csubj(pass)*) introduces uncertainties concerning the selection of the Q-clause.

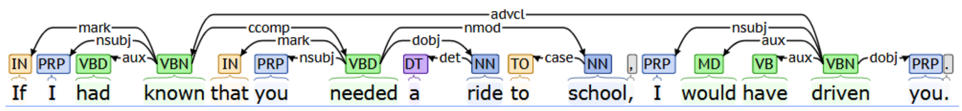


Figure 3.15: Graphical visualization of the typed dependencies in the sentence *If I had known that you needed a ride to school, I would have driven you.*

The figure 3.15 depicts a scenario in which the relation *nsubj(needed, you)* is favored (by our assumptions) over the correct relation to describe the Q-clause - *nsubj(driven, I)*. This limitation should be taken in consideration for future development as it discards a significant portion of examples from all the conditionals. A potential improvement would be to try pairing the P-clause with all the Q-clause candidates until the newly created combination satisfies the conditional definition.

If Counterfactual

In section *§Custom Schemes*, which introduces the counterfactuals, we make an important observation that counterfactuals can easily take the form of second or third conditional, from the grammatical point of view. Therefore, during the evaluation phase, we allow potential intersections between instances of these devices. That is, if a third conditional has been annotated as counterfactual and we can manually attest its presence in the annotation, then such an instance is considered a true positive.

To illustrate the limitations of the adopted algorithm, let us consider the following example:

Example (72) *If I went to Mars, I would meet a Martian.*

At first glance, the temporal space in which the events are set refers to a hypothetical world in the past. Furthermore, the "ingredients" required to build an if-counterfactual are all present (Figure 3.16).

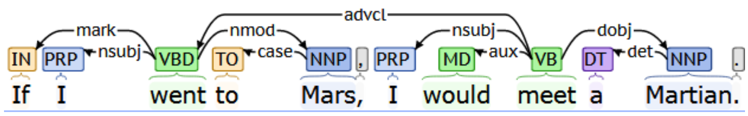


Figure 3.16: Graphical visualization of the typed dependencies of the Example (72)

Nevertheless, this is a misidentified instance of second conditional. In our view, even if hypothetically the premise (i.e., *if I went to Mars*) is set in the past, we cannot take it as an impossible scenario. Indeed, it doesn't meet the factuality in the current moment (since the subject is not on Mars), yet it could in the future. Moreover, "would" expresses the intention of the subject, not a sentiment of regret or grief, as it usually does in the case of counterfactuals. Such exceptions are context related and therefore, much harder to detect. Fortunately, as inferred from the performance results, they are not so frequent.

Unless-conditional

The detection algorithm of unless-conditional is straightforward (see section *§Custom Schemes*). It can correctly identify all of the instances of unless-conditional in the evaluation dataset. We would like to reiterate that we depend exclusively on the presence of *unless* in the sentence under analysis.

Even if this assumption seems to be successful in the task of detection of such conditionals, it would still miss cases in which *unless* is inferred pragmatically. For instance, within imaginary conditionals (which are interpreted as having no truth relation to the actual world at all (Declerck and Reed [2001])), *unless* is interchangeable with *except if* without any loss in meaning.

Example (73) *I wouldn't be here* ***unless/except if*** *I was interested, would I?*

Counterfactual conditionals, on the other hand, might inflict a different interpretation of *unless*, and namely, "if it hadn't been the case that", thus, interchangeable with *if... not*, but not with *except if* [Declerck and Reed, 2001].

Example (74) *I couldn't have finished this in time unless you'd helped me. (=... if it hadn't been for the fact that you helped me)*

Whether... or-conditional

As in the case of unless conditionals, we consider the same assumption in the detection algorithm; this is, we validate the sentence if it contains *whether... or* constructions. Our approach can deliver a perfect precision value and a high recall value. The reason behind some of the missed instances is that we stick to the literal definition of this device, which explicitly refers to *whether... or* pattern. This, however, results in ignoring cases of conditional in which *whether* alone is enough to express the conditional.

Example (75) *Brad wonders whether he'll get the job.*

Despite the fact that Example (75) sounds incomplete, it is a perfectly valid instance of *whether... or* conditional, in which the *or* part, is supplied by the reader by inference.

Comparative and Superlatives adjectives/adverbs

In the detection of comparative (and superlative) adjectives and adverbs, we engage an off-the-shelf approach provided by the Stanford POS tagger. As described in section *§Custom Schemes*, we directly annotate the generated tags associated with comparative and superlative adjectives/adverbs as a rhetorical device. Obviously, we did not expect impressive results considering the

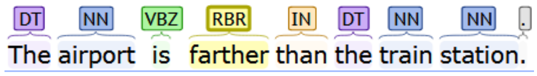


Figure 3.17: POS tags visualization of the sentence *The airport is farther than the train station.*

simplicity of our approach, and indeed the performance is only satisfactory. Nonetheless, bearing in mind the rhetorical impact they bring, comparatives and superlatives are not to be neglected.

Let us take the example from the figure 3.17, where "father" is evidently supposed to be an adjective in the comparative form, since it describes and compares two nouns "airport" and "train station". However, it is treated as a comparative adverb RBR.

It is important to point out that the detection algorithm of these rhetorical devices employs the latest version of the Stanford POS tagger (3.8.0). The conducted performance tests between this version and the POS tagger (3.5.2) which is inherently part of our pipeline, show the advantages of the former, in the detection of these rhetorical devices (see Appendix 1). Further updates of the Stanford CoreNLP suite will, allegedly, solve this problem and improve the detection results; until then, we should, probably, seek other methods to identify these rhetorical devices.

Passive Voice

Our system reports promising results for both, precision and recall. We will discuss both of the cases, to give a comprehension image of the limitations of our algorithm.
All of the instances but one were correctly identified in the file containing samples of this rhetorical device (recall 99%). The single missed instance had been incorrectly tagged and therefore, neglected. In particular, as we can see in figure 3.18, the parser erroneously treats "drunk" as an adjective while the detection algorithm expects a past participle verb.

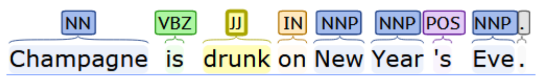


Figure 3.18: POS tags visualization of the sentence *Champagne is drunk on New Year's Eve.*

What concerns the misidentified instances (i.e., precision), a major source of errors represents the grammatical exceptions which describe passive voice.

For example, intransitive verbs – which have a meaning on their own without being accompanied by a direct object[64] – do not have passive forms (since they cannot take a direct object). As a result, constructions like "was gone" or "is vanished" cannot be part of the passive voice, even though they satisfy the grammatical rules. Let us take a look at of the misidentified example in figure 3.19.

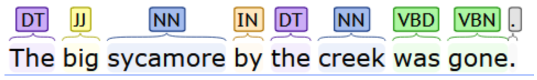


Figure 3.19: POS tags visualization of the sentence *The big sycamore by the creek was gone.*

We can observe that the construction "was gone" meets the algorithm's requirements and therefore, the sentence is labeled as passive voice. However, it is evident that the verb "gone" doesn't require a direct object for the action to be completed, which means passive voice cannot be detected here.

3.4 Summary

In this chapter we first, introduce the tools we rely upon in the process of detection of rhetorical devices; we individually approach each of the involved devices by discussing its rhetorical effect and presenting an outline of how we identify them both algorithmically and formally. Additionally, we examine the assumptions and limitations of our approach as well as some potential improvements. In the last part of this chapter, we present the performance results of our system by analyzing separately, the flaws and misidentifications of each rhetorical device. We bring concrete examples which are missed or erroneously identified by our system, to support the presented evaluation measures.

[64]intransitive verb. (n.d.). Dictionary.com Unabridged. Retrieved October 31, 2017 from Dictionary.com website http://www.dictionary.com/browse/intransitive-verb)

Chapter 4

Analysis of Rhetorical Devices

In this chapter we aim to assess the practicability of our system in a task to analyze and distinguish the style traits of writings across multiple dimensions. In the first stage, we study the rhetorical aspect of writings categorized by genre, topic and author. For this scope, we consider the publicly available corpus of articles published by New York Times (NYT). In the second stage of our analysis we focus on the rhetorical style in the presidential debates by Donald Trump and Hillary Clinton.

The first part of this chapter addresses the preprocessing steps adopted to collect the data and build the relevant subsamples ready to serve as input for our system. In the second part, we analyze the results by presenting our insights and findings regarding the rhetorical style within the compared types of articles and debates.

4.1 Subsampling the NYT corpus

The NYT Corpus is drawn from the historical archive of the New York Times (NYT) and contains every[1] article published in the New York Times between 1st of January 1987 and 19th of June 2007 (Sandhaus [2008]). Overall, the corpus consists of 1.8 million articles. Each document includes various kinds of metadata provided by the NYT staff, which structures the whole corpus on a wide range of dimensions. This diversity of material, created by professional writers and editors, happens to be a good choice for testing our system's capabilities.

Based on the meta-information, we start off by forming the clusters of

[1] excluding wire services articles that appeared during the covered period

articles depending on their topic, genre, and author (originally indicated by metadata tags as taxonomic classifiers, types of material and byline, respectively). The extracted sub-corpus, with further preprocessing, is later used to perform our analysis experiments. Mainly, the analysis on the NYT corpus is divided into three phases, depending on the subsampling method of our sub-corpus:

- Random subsampling with 1000 articles per topic/genre/author.
- Article-length based subsampling with 600 articles per topic/genre/author.
- Controlled subsampling via matching with dozens of articles per topic/-genre/author.

Even though the NYT corpus seems to be rather comprehensive (1.8 million articles), extracting the relevant subsamples of data turned out to be difficult in some of the cases above. In general, the incompleteness of author metadata (about 40% of all articles are left unsigned, as they were initially published (Sandhaus [2008])) and the imbalanced distribution of articles across categories, greatly influence the uniform subsampling within each considered dimension (i.e., topic/genre/author).

To preserve the consistency between the approached subsampling methods, we restrict each dimension to three instances; that is, our sub-corpus consist of three different topics, genres, and authors.
As expected, the random data subsampling was the easiest to perform. As we mentioned, we limit the number of articles per category to 1000. Referring to previous work in this direction, which analyzes the performance of an authorship attribution system with less than 500 articles per author (Java [2015]), we consider that 1000 articles should be more than sufficient to infer the rhetorical style from a particular type of writings. This results in 3000 randomly selected articles in each of the tested dimension (i.e., topic, genre, and author). For the second method of subsampling, we extract only the articles that have comparable length, expressed in word count. The challenge here is to find the smallest possible range of permitted article word-length, across all the dimensions. Since, a review is expected to be much shorter than an editorial, after computing the average word-lengths per dimension, we decide on articles with a length between 400 and 800 words. This appears to be the smallest range which allows us to collect subsamples of 600 articles per type.

Perhaps the most challenging and direct way of assessing our system's performance, is analyzing the dataset generated by a controlled method of subsampling. In general, a controlled experiment is the one in which everything is kept constant (i.e., variables that might influence the outcome) except the single independent variable. The biggest advantage of a controlled experiment is that it excludes much of the uncertainty with regards to the outcome; and therefore the results are much more reliable (Helmenstine [2017]). Conversely, in an uncontrolled experiment the variables are unpredictable, which might result in a confusing results.
Correlation errors between variables and subsequent biases can be introduced by so-called *confounding variables* (or *confounder*). These variables are not part of the experiment itself, yet they are able to influence both the dependent and independent variables. Let us take our setup as an example: if we want to find out whether the rhetorical style of writings is solely dependent on their authors; various confounding variables like genre, topic or audience might have a direct impact on the outcome of this analysis.

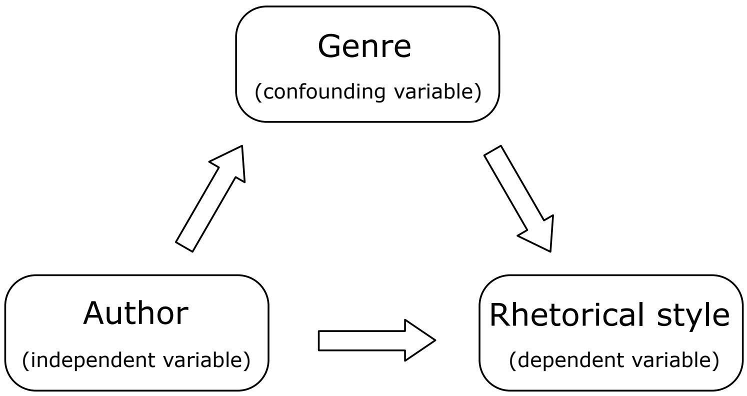


Figure 4.1: Impact of a confounding variable in an uncontrolled study. "Genres", as a confounder, influences the outcome of an experiment which tries to determine whether the rhetorical style of writings is dependent on their authors.

The negative effect of cofounders can be diminished via randomized sampling of data (our first method of subsampling), stratification[2] and matching, amongst others. Because of the erratic nature of the NYT corpus, we subsample the data by matching. To keep the confounders under control, by matching, one has to "ensure an equal distribution between exposed and unexposed variables" considered to be confounding (de Graaf M [2011]).

[2]Stratification – "allows to control for confounding by creating two or more categories or subgroups in which the confounding variable either does not vary or does not vary very much." (Tripepi G [2010])

Returning to our corpus, the first step in conducting a controlled experiment is to identify the confounding variables. We focus our analysis on three dimensions: topic, genre, and author. Therefore, depending on the article dimension we analyze in a certain moment, the others two, we regard as confounders. In other words, if we intend to explain the rhetorical style of articles on two different topics, then, besides the equal distribution of articles, they should be of the same genre and written by the same author; that is, we vary the considered topics and keep the confounding variables constant. In this manner, we can obtain much more accurate and reliable results on the style of the analyzed dimension.

Of course, such a constrained subsampling comes with its downsides. It is difficult to get a large number of samples to match all the imposed restrictions. Either the corpus, from which the samples are drawn, should be large enough, or the data to be distributed uniformly within the considered dimensions. Out of all the articles in the NYT corpus, through matching, we were able to obtain significantly smaller datasets compared to the other two methods of subsampling. Even so, in the next section of this chapter, we will see that the obtained datasets are enough to infer relevant judgments about the rhetorical style of each dimension. Since the number of articles in random and article-length are fixed (1000 and 600 respectively), table 4.1 shows the distribution of articles in controlled subsampled dataset.

Table 4.1: Distribution of articles in controlled subsampled dataset

	Dimension	Confounders	Articles
Authors	Hevesi Dennis	biography-business	9
		biography-newyork	10
		biography-obituaries	10
		obit-newyork	10
		obit-obituaries	10
	Lewis Paul	biography-business	10
		biography-newyork	10
		biography-obituaries	10
		obit-newyork	10
		obit-obituaries	10
	Martin Douglas	biography-business	10
		biography-newyork	10
		biography-obituaries	10
		obit-newyork	10
		obit-obituaries	10
Genres	Biography	freedman-news	6
		norris-markets	10
		wade-health	10
	Editorial	freedman-news	10
		norris-markets	10
		wade-health	10
	Review	freedman-news	7
		norris-markets	10
		wade-health	7
Topics	Arts	goodman-review	10
		martin-biography	10
		saxon-biography	10
		saxon-obituaries	10
	Education	goodman-review	10
		martin-biography	9
		saxon-biography	10
		saxon-obituaries	10
	Science	goodman-review	5
		martin-biography	10
		saxon-biography	10
		saxon-obituaries	10

4.2 Interpretation of Results

In this chapter we aim to assess the practicability of our system in a task to analyze and distinguish the style traits of writings across multiple dimensions. In the first stage, we study the rhetorical aspect of writings categorized by genre, topic and author. For this scope, we consider the publicly available corpus of articles published by New York Times (NYT). In the second stage of our analysis we focus on the rhetorical style in the presidential debates by Donald Trump and Hillary Clinton.

The first part of this chapter addresses the preprocessing steps adopted to collect the data and build the relevant subsamples ready to serve as input for our system. In the second part, we analyze the results by presenting our insights and findings regarding the rhetorical style within the compared types of articles and debates.

4.2.1 NYT Corpus

In this subchapter, we perform a comparative study of distributions of rhetorical devices between the considered methods of subsampling. We start by presenting our interpretation of results in random and article-length based subsampled dataset. Further on, the controlled experiment allows us to perform a very meticulous analysis of the rhetorical style of writings. Even slight variations in the frequency of a particular device might influence the style adopted in the regarded type of articles. Although we cannot consider all the linguistic aspects in defining a rhetorical style, as a professional rhetorician would do; our knowledge base accumulated during this study supported by empirical tests should suffice to conduct a qualitative analysis.
A general observation, which might seem discouraging, is that the same pattern of the distribution of rhetorical devices is maintained across all the considered sub-corpora. However, as we will see, the variations of individual devices between corpora help us to distinguish and infer the judgments regarding the rhetorical style of the analyzed types of writings.

4.2.1.1 Random and Article-length based subsampling methods

The distributions of rhetorical devices in corpora based on random and article-length subsampling don't reflect any striking findings. Mainly, that is due to their structure, including large amounts of articles by different authors on various topics. As a result, these confounding variables (i.e., authors and topics) introduce undesirable noise in the distribution. Thus, concluding over a specific style would be unreliable. Charts 4.2, 4.3, 4.4, 4.5, and 4.6, present our partial results on these subsamples. The detected rhetorical patterns and combinations of devices are shown in Appendix 3.

Figure 4.2: Distribution of rhetorical devices amongst authors, in random and article-length subsampled dataset.

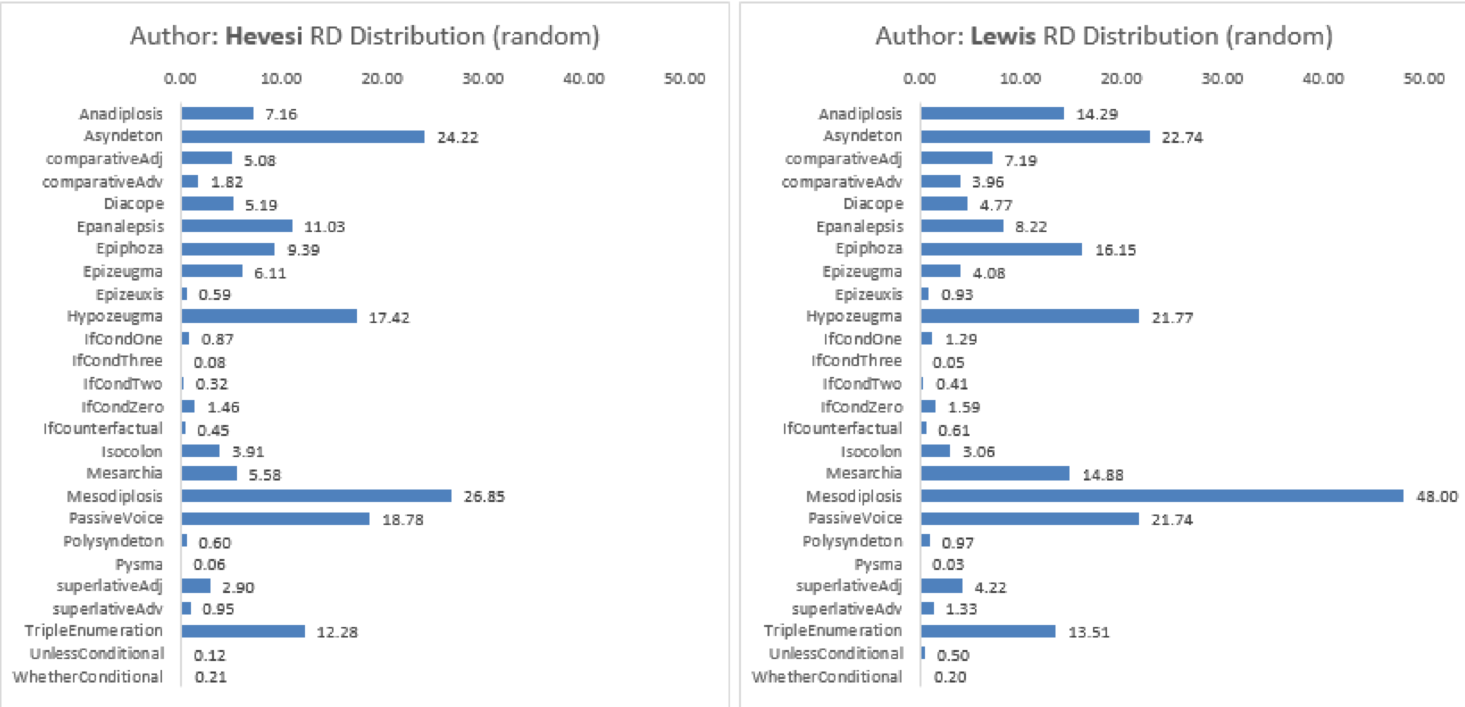


Figure 4.3: Distribution of rhetorical devices amongst authors, in random and article-length subsampled dataset.

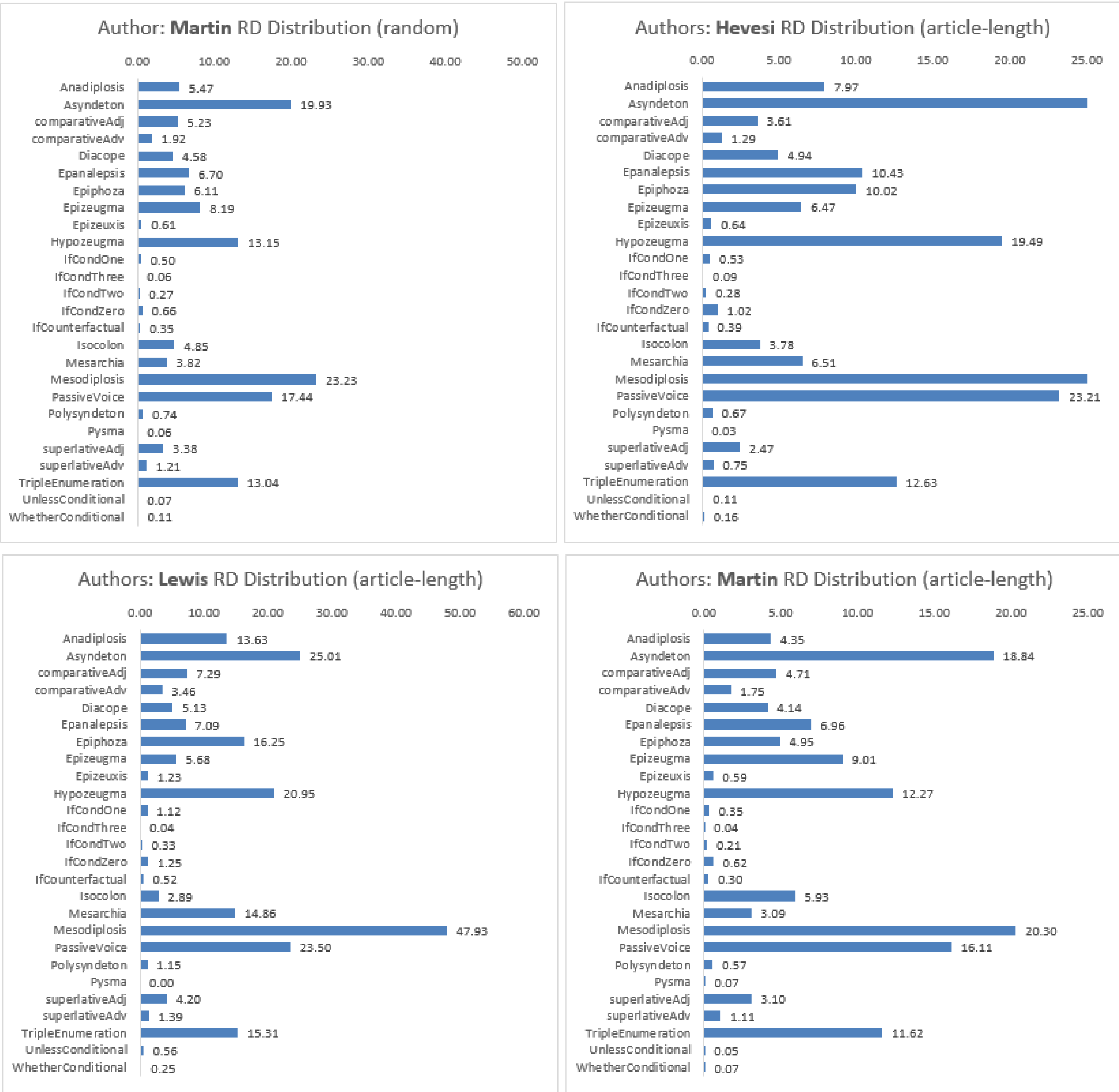


Figure 4.4: Distribution of rhetorical devices amongst genres, in random and article-length subsampled dataset.

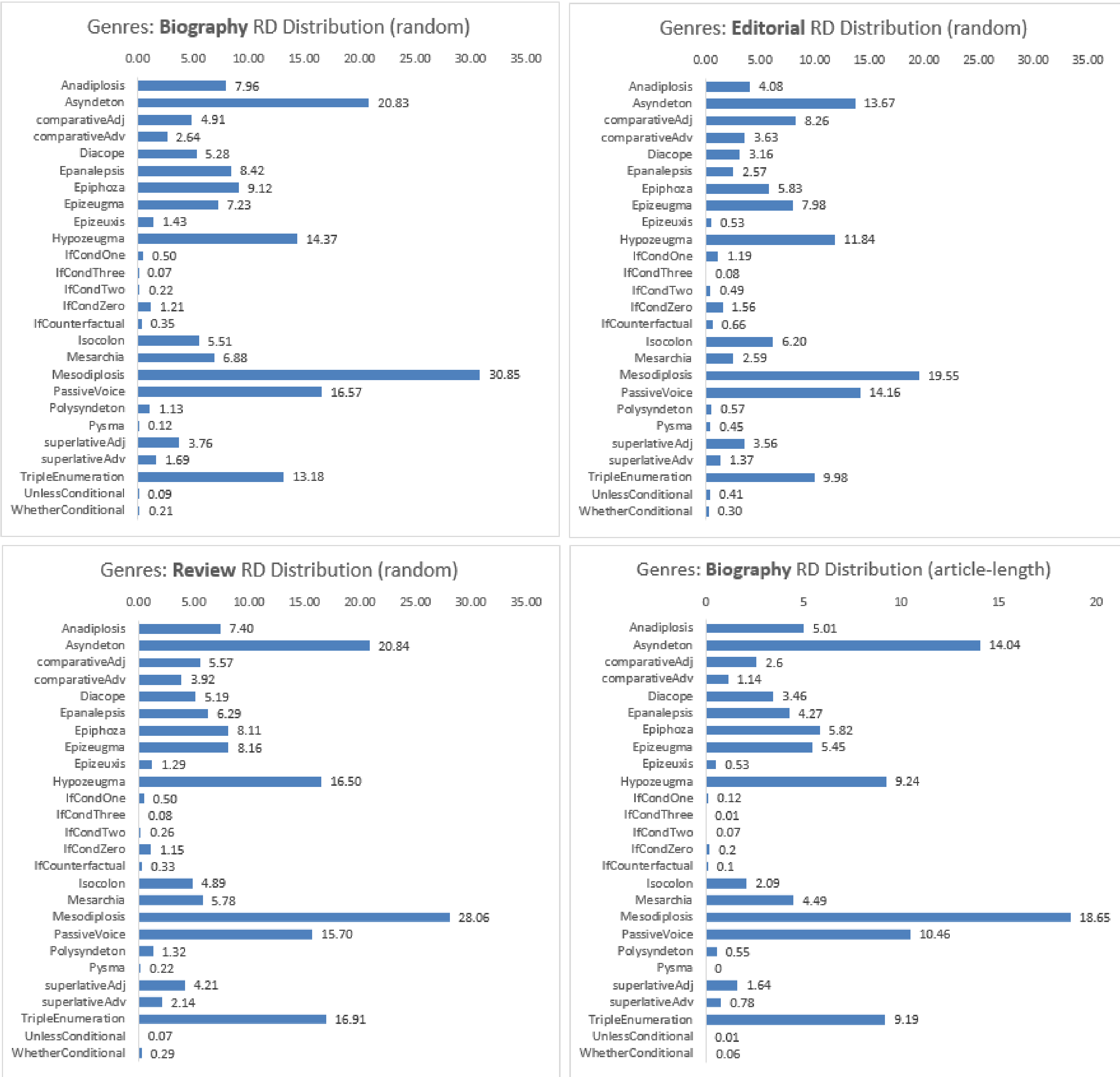


Figure 4.5: Distribution of rhetorical devices amongst genres and topics, in random and article-length subsampled dataset.

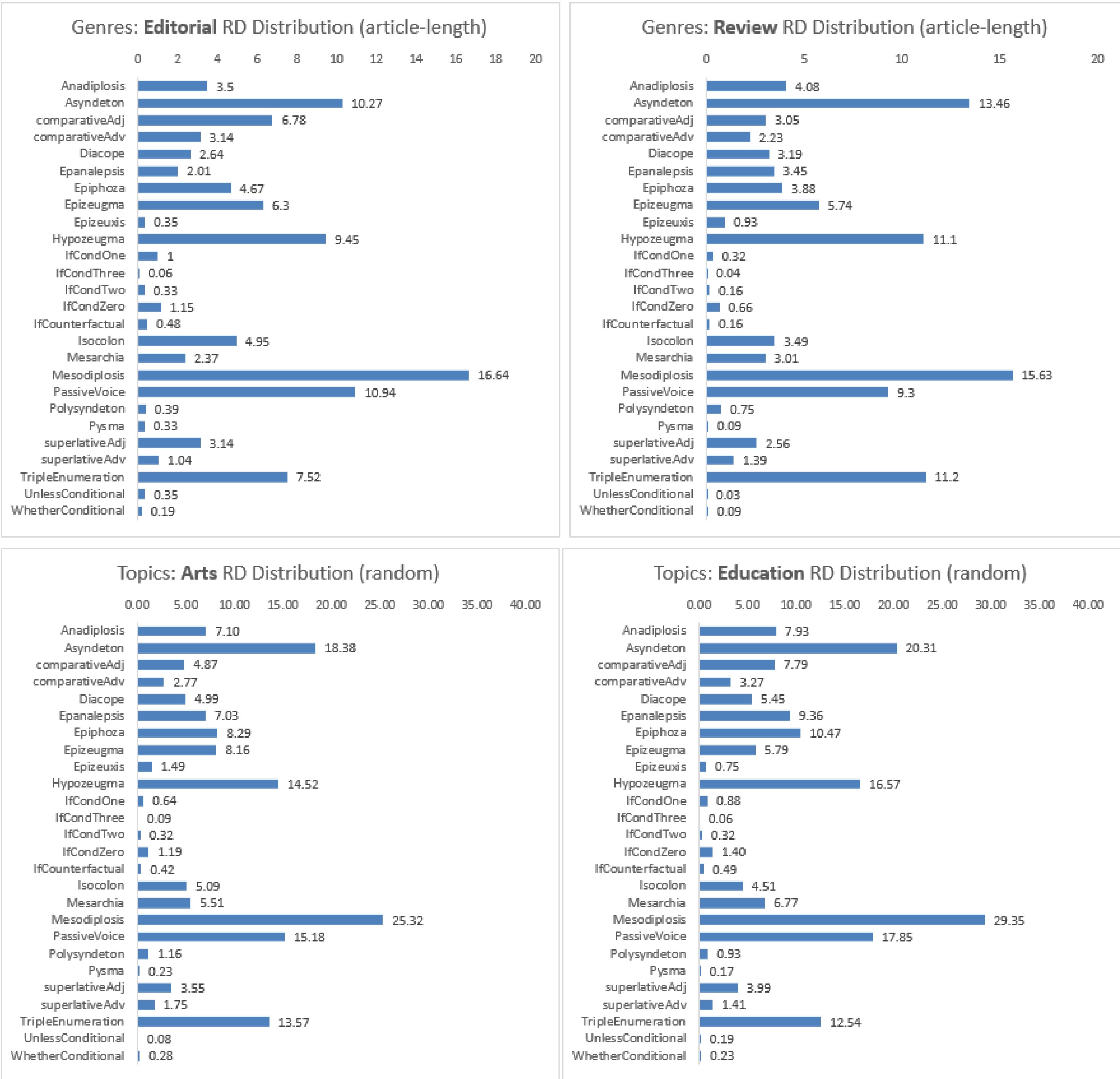
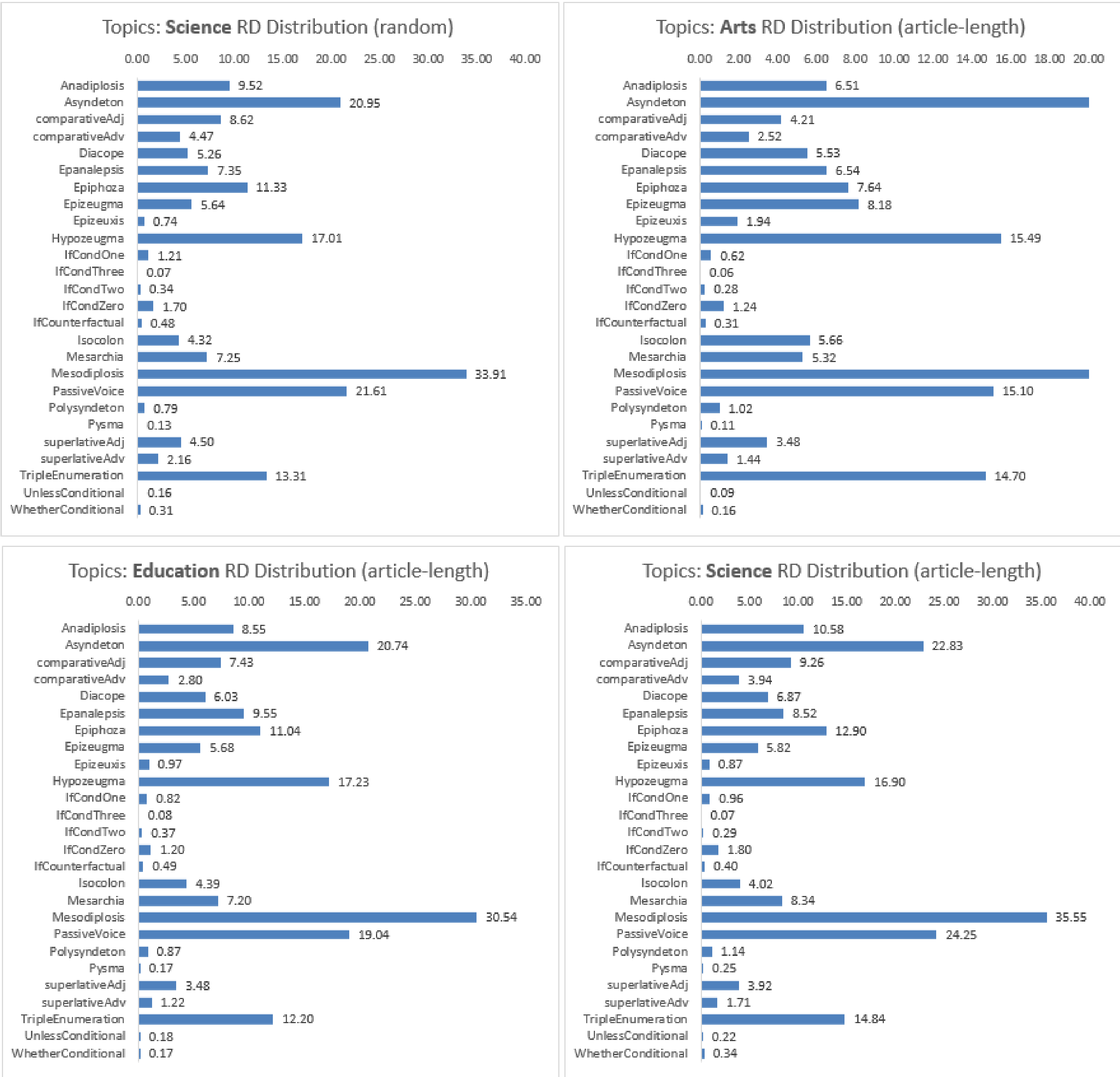


Figure 4.6: Distribution of rhetorical devices amongst topics, in random and article-length subsampled dataset.



At this point, it is interesting to observe the fluctuations of scheme categories within the analyzed type of data. In particular, tables 4.2 and 4.3 show how the style varies from author to author. We can identify a preference for omission-based rhetorical devices in some cases (*Hevesi Dennis*) and a frequent usage of repetition schemes in others (*Lewis Paul*). We expect such patterns to be much more salient in the controlled subsampling corpus.

In addition to individual distributions, we were also interested in analyzing the most frequent combination of rhetorical devices which describe a particular style. Appendix 3 presents the top 4 recurrent combinations by two, three and four rhetorical devices within each category for both random and article-length subsampled datasets. Evidently, the top combinations are bonds between the most frequent devices in a particular dimension. As the generic distribution holds steady across all dimensions, we get recurrent combinations of rhetorical devices from which it is hard to infer any relevant conclusions.

To examine the rhetorical style on a larger scale, we analyze the frequency of devices per category of schemes. We have four types of schemes: balance, omission, repetition, and custom. In the custom category, we can identify passive voice, comparatives (which include all the comparative and superlative adjectives/adverbs) and conditionals. Tables 4.2 and 4.3 present the distribution per category of schemes.

Besides the analysis of the distributions of rhetorical devices between two adjacent types of data, it is critical to determine how (di)similar those datasets are in terms of rhetoric. Symmetrical distributions can only mean that our system cannot distinguish between the styles of the analyzed types of articles. To catch the relevant differences between the set of genres, topics, and authors, we compare the rhetoric distribution of all the possible combinations iteratively within each dimension (e.g., Review-Editorial, Editorial-Biography, Science-Arts, and so on). More precisely, on each such pair, we perform two tests: *significance* and *effect-size*.

Significance test defined for two distributions – expected and observed – is the probability p on the null hypothesis that the observed frequency follows the expected frequency. The significance level α is the probability of rejecting the null hypothesis considering it to be true (Schlotzhauer [2007]). We use the *chi-square* statistical method with a significance level $\alpha = 0.01$, which means there is at most 1% chance of incorrectly rejecting the null hypothesis (i.e., observed distribution follows the expected). Respectively, a value of probabil-

Table 4.2: Distribution of types of schemes in random subsampled dataset

Category	**Normalized distribution (%)**		
Author:	*Hevesi Dennis*	*Lewis Paul*	*Martin Douglas*
conditionals	3.51	4.65	2.02
comparatives	10.74	16.70	11.74
passive voice	18.78	21.74	17.44
balance	16.25	16.60	17.96
repetition	66.40	108.20	51.26
omission	47.75	48.58	41.26
Genre:	*Biography*	*Editorial*	*Review*
conditionals	2.65	4.69	2.69
comparatives	12.99	16.82	15.85
passive voice	16.57	14.16	15.70
balance	18.81	16.64	22.02
repetition	71.08	38.87	63.45
omission	42.43	33.50	45.51
Topic:	*Arts*	*Education*	*Science*
conditionals	3.02	3.57	4.26
comparatives	12.95	16.46	19.75
passive voice	15.18	17.85	21.61
balance	18.90	17.21	17.77
repetition	60.89	71.01	76.14
omission	41.06	42.67	43.60

Table 4.3: Distribution of types of schemes in article-length based subsampled dataset

Category	**Normalized distribution (%)**		
Author:	*Hevesi Dennis*	*Lewis Paul*	*Martin Douglas*
conditionals	2.58	4.08	1.63
comparatives	8.12	16.34	10.67
passive voice	23.21	23.50	16.11
balance	16.44	18.19	17.62
repetition	70.75	107.27	44.96
omission	52.73	51.65	40.13
Genre:	*Biography*	*Editorial*	*Review*
conditionals	1.09	4.49	2.27
comparatives	11.72	17.78	14.32
passive voice	19.88	13.80	14.42
balance	21.46	16.14	22.92
repetition	81.35	41.07	54.20
omission	54.64	32.80	47.01
Topic:	*Arts*	*Education*	*Science*
conditionals	2.76	3.32	4.09
comparatives	11.65	14.93	18.82
passive voice	15.10	19.04	24.25
balance	20.47	16.77	19.12
repetition	58.50	74.75	84.77
omission	43.88	43.65	45.55

ity p lower than the set threshold α sheds doubts on the validity of the null hypothesis.

Since we don't possess any information about the expected distribution of rhetorical devices within each type of articles, we compare the observed distributions obtained by our system. It is important to mention that instead of the raw frequency of devices within each type, we should take into account the evaluation measures (i.e., precision and recall) associated with each device. Thus, we get a normalized and objective overview of the significance measure for the compared datasets. An approach in this direction is proposed by Al-Khatib et al. [2017]. The paper describes a method to assess the impact of classification errors of argumentation strategies by means of significance test. Namely, they suggest to compute a confidence interval for each rhetorical device *rd* in the analyzed type of data *t*. This interval derives from the precision and recall values of each device. The *lower bound* is calculated with $n \cdot$ *precision(rd)*, and the *upper bound* with *n/recall(rd)* (Al-Khatib et al. [2017]). Ultimately, the mean of the confidence interval serves as input for the chi-square test.
Following this approach yields the significant correlation between the analyzed types of articles function of the significance level α. As we show above, a significance level set to 0.01 should be sufficient enough to ensure the confidence in our results. Yet, we vary this coefficient (between 0.1 and 0.001) to consider the dynamics of significance across different types of articles. The results of chi-square test between frequencies of rhetorical devices within the random and article-length subsampled corpora are shown in table 4.4.

As we can observe, all of the analyzed pairs of types of articles are significantly different, according to the chi-square test. Such results were anticipated because statistically, "significant differences are more likely to occur with large sample sizes"[3]. Additionally, we should note that the sub-corpora under analysis consists of articles by different authors on various topics and genres. Therefore, this might be a source for dissimilarity in the distribution of rhetorical devices. Conversely, we expect much more exact results for controlled subsampled sub-corpus.

As we show, significance test is more often than not dependent on the sample size we regard. That's why we consider calculating the correlation between the compared distributions regardless of the sample size. For this scope, we use the effect size Cramer's V test. The primary advantage of this test is that it is

[3]https://www.fort.usgs.gov/sites/landsat-imagery-unique-resource/statistical-interpretation

Table 4.4: Significance test results on random and article-length based subsampled corpora

Dataset	P-value	Independence
Authors		
Hevesi vs. Lewis	∼0	TRUE
Lewis vs. Martin	∼0	TRUE
Martin vs. Hevesi	∼0	TRUE
Genres		
Biography vs. Editorial	∼0	TRUE
Editorial vs. Review	∼0	TRUE
Review vs. Biography	∼0	TRUE
Topics		
Science vs. Education	∼0	TRUE
Education vs. Arts	∼0	TRUE
Arts vs. Science	∼0	TRUE

able to reveal the practical difference rather than plain statistical differences[4]. Effect-size test measures the amount of impact an independent variable has on a dependent variable (Murphy and Myors [1998]). The results are usually represented in fuzzy terms like small (0.1), medium (0.3) or large (0.5)[4]. Between two types of articles, we aim for significantly different distributions ($p < 0.1$) and a small effect size. These conditions should be enough to eliminate any uncertainties regarding the difference in rhetorical style between the compared datasets.

Table 4.5, which present the results of the effect size test between random and article-length based subsampling distributions, support our hypothesis that the rhetorical style varies from type to type.

[4]https://www.fort.usgs.gov/sites/landsat-imagery-unique-resource/statistical-interpretation

Table 4.5: Effect-size test results on random and article-length based subsampled datasets

	RANDOM		ARTICLE-LENGTH BASED	
Dataset	**Cramer's V value**	**Effect***	**Cramer's V value**	**Effect***
Authors				
Hevesi vs. Lewis	0.17	SMALL	0.16	SMALL
Lewis vs. Martin	0.22	SMALL	0.22	SMALL
Martin vs. Hevesi	0.10	SMALL	0.13	SMALL
Genres				
Biography vs. Editorial	0.11	SMALL	0.21	SMALL
Editorial vs. Review	0.18	SMALL	0.16	SMALL
Review vs. Biography	0.16	SMALL	0.12	SMALL
Topics				
Science vs. Education	0.06	SMALL	0.05	SMALL
Education vs. Arts	0.09	SMALL	0.10	SMALL
Arts vs. Science	0.11	SMALL	0.12	SMALL

* Effect's color is directly proportional with its value, i.e., a lighter nuance denotes a larger effect

4.2.1.2 Controlled subsampling method

As we already mentioned at the beginning of this subchapter, the controlled experiment requires an in-depth analysis. We try to adopt a bottom-up analysis approach, meaning that we start by focusing on individual rhetorical devices and end with more generalized observations. It is worth to consider that we analyze the normalized distributions of devices; that is, we divide individual counts over the total number of sentences per type of documents under analysis. Our observations are based on individual and scheme category distributions. Also, we evaluate the results of the significance and effect-size tests. We disregard the analysis of the frequent combinations of rhetorical devices as it is redundant in most of the cases.

4.2.1.2.1 *Authors* dataset

The distribution of rhetorical devices between authors seems to be somewhat balanced. Still, we would like to highlight the disproportionate usage of epiphoza. To remind the reader, this scheme consists of a repetition of the same word at the end of successive sentences. Being a figure of repetition, it is generally used to emphasize the key ideas. We can observe that *Martin* benefits from epiphoza much less (6.49%) than Lewis (12.99%) and *Hevesi* (10.74%). Interestingly enough, this trend is characteristic for repetition schemes in general, for *Martin*. *Lewis* and *Hevesi* adopt a more persuasive method of writing, with 81.93% and 70.99% repetition-based schemes usage, compared to 55.49% in *Martin*'s articles. Thus, we can conclude that a sparse usage of epiphoza is not accidental by any means, and explains the rhetorical strategy of the respective author.

Table 4.6: Normalized distribution of epiphoza, repetition-based schemes and conditionals amongst authors

	EPIPHOZA	REPETITION SCHEMES	CONDITIONALS
Author	**Distribution (%)**	**Distribution (%)**	**Distribution (%)**
Hevesi Dennis	10.74	70.99	0.84
Lewis Paul	12.99	81.93	0.39
Martin Douglas	6.49	55.49	1.14

On the other hand, *Lewis* barely considers conditionals in his writings (table 4.6). This style, however, cannot be identified when we look at the distribution of conditionals in the article-length based subsampled data (table 4.3). Apparently, the type of confounders in the author category determines Lewis to employ fewer conditionals.

Most of the articles in this category are on biographies and obituaries. These genres are pretty similar in content as they discuss the most important achievements in the subject's life[5] [6]. The only difference is that biographies are more detailed in this sense. As a result, we do not see radical changes in the rhetorical style within the mentioned authors. Relying on the results of the significance test, in table 4.7, we can infer that the distributions of rhetorical devices are different amongst authors. However, it is worth to say that the confidence of this hypothesis is not the highest for two pairs of authors: *Hevesi-Lewis* and *Martin-Hevesi*. Still, given effect-size results, we can claim that the rhetorical style differs amidst *authors*.

[5]https://en.wikipedia.org/wiki/Biography
[6]https://en.wikipedia.org/wiki/Obituary

Table 4.7: Significance and effect-size test results of *Authors* dataset

	SIGNIFICANCE		EFFECT-SIZE	
Datasets	**P-value**	**Independence**	**Cramer's V value**	**Effect**
Hevesi vs. Lewis	0.015	TRUE*	0.1	SMALL
Lewis vs. Martin	$\sim$0	TRUE	0.15	SMALL
Martin vs. Hevesi	0.017	TRUE*	0.1	SMALL

* for $\alpha > 0.001$

4.2.1.2.2 *Genres* dataset

Unlike the *authors* dataset, the *genres* are much more diverse in the types of articles (i.e., *author-topic* confounder variance). This is reflected, for instance, in the usage of comparative rhetorical devices. Regardless of the genre, articles written by *Norris* on the topic *Markets* contain considerably more comparatives in contrast with the distribution of the same group of devices within other topics. Table 4.8 illustrates a minimal difference of more than 4% for the *Reviews* between *Health* articles by *Wade* and *Markets* by *Norris*; and a maximal difference of about 17% for *Editorials* between the same groups of articles. Considering the topics involved, it seems natural that an article on *Markets*, which, allegedly, talks about selling strategies, money, and business in general, would contain more comparatives that a *Health* article. As we mentioned in the section *§Custom Schemes*, when we present the role of comparatives in rhetoric, this is the favorite device of advertisers. It is part of the strategy to increase the value of their products and boost the sales. *Markets* are as about sales as *Health* are about illnesses. Therefore, this style is completely justified.

Table 4.8: Normalized distribution of comparatives within confounders across genres

Confounders	**Distribution (%)**		
Genre:	*Biography*	*Editorial*	*Review*
freedman-news	11.65	25.57	11.75
norris-markets	22.59	30.06	20.99
wade-health	12.04	12.97	16.40

Another point we would like to make concerning comparatives is that *Reviews* and *Editorials* lead in the usage frequency of this particular device. We believe this is due to the opinionated nature of both *Reviews* and *Editorials*. While *Biographies*, usually, describes facts and events in someone's life, in *Reviews* and *Editorials* articles, authors present their opinion either with regard to a hot topic[7], or a relevant product like a book or a movie. Therefore, the content of the latter is more dynamic and prone to contrasting ideas, which explains a higher frequency of comparatives.

Table 4.9: Normalized distribution of comparatives, epanalepsis and conditionals amongst genres

	COMPARATIVES	EPANALEPSIS	CONDITIONALS
Genre	**Distribution (%)**	**Distribution (%)**	**Distribution (%)**
Biography	14.07	7.09	3.45
Editorial	23.16	2.71	5.95
Review	16.29	6.05	3.41

Reviews and *Biographies* share the same feature: they both comprehensively describe and evaluate an entity. We assume that in this process of evaluation there is more place for repeating specific ideas to embed them into the reader's mind. We can easily imagine a review or a bio wherein particular characteristics of the entity are replicated to make the point. Starting with this assumption, we can observe a meaningful difference in the usage of epanalepsis between *Review* or *Biography* and *Editorials* (table 4.9). Epanalepsis consists in the repetition at the end of the sentence, the word that occurred at the beginning of it. Through this technique, author strengthens his utterance concerning the described entity. The argumentative nature of *Editorials* (Al-Khatib et al. [2017]), on the other hand, involves less epanalepsis.

However, as *Editorials* are more argumentative, we expect them to be richer in conditionals. Indeed, if we take a look at table 4.9, we can note that authors prefer to use substantially more conditionals in *Editorials* compare to other genres. The most sporadic usage of conditionals is observed in *Biographies*. We assume, that's because these type of articles consist of general truths about entities, which cannot be subject of dispute.

Based on the observations above, we can identify different rhetorical styles depending on the type of the data: argumentative and descriptive. Since *Biographies* and *Reviews* are more explanatory in their content, epanalepsis is

[7] https://en.wikipedia.org/wiki/Editorial

much frequent there. Conversely, as we saw, argumentative articles like *Editorials* are more abundant in conditionals and comparatives. More specifically, argumentative articles employ more counterfactuals that descriptive ones (see Appendix 4). Again, that might be due to a more stable nature of the descriptive articles.

As an empirical proof to our observations, both the significance and effect-size tests suggest that the rhetorical style varies considerably between at least two genres. With a very high confidence, *Biography* and *Editorial* adopt different rhetorical styles. The same applies to *Editorial* and *Review*. What was unexpected though, this does not hold true for *Review* and *Biography*. As we previously mentioned, these two genres are similar in at least one dimension: they are both descriptive by nature. Hence, we assume that explains the similarities in style within *Review* and *Biography*.

Table 4.10: Significance and effect-size test results of *Genres* dataset

	SIGNIFICANCE		EFFECT-SIZE	
Datasets	**P-value**	**Independence**	**Cramer's V value**	**Effect**
Biography vs. Editorial	∼0	TRUE	0.16	SMALL
Editorial vs. Review	∼0	TRUE	0.14	SMALL
Review vs. Biography	0.68	FALSE	0.07	SMALL

4.2.1.2.3 *Topics* dataset

Starting with an analysis of the distribution of rhetorical devices between confounders (i.e., *author-genre* pairs), it is interesting to observe how the rhetorical style in *Biography* articles varies amongst authors (table 4.11). In comparison with *Saxon*, *Martin* employs much more conditionals and comparatives in his bio writings. Moreover, this style is characteristic for *Martin* across all the topics in the genre *Biography*. This observation suggests that either there are no clear rhetorical guidelines for authors to pertain to – thus they develop their proprietary style – or *Biographies* don't necessarily require a persuasive rhetorical style.

The schemes of balance are able to control the rhythm of expressed ideas by creating a sense of equivalence between them. The usage frequency of such schemes differs within *Reviews* by *Goodman*, across all topics. In particular, table 4.12 shows a little scarcer usage of balance schemes in *Education* compared with *Arts* or *Science*. Conceptually, arts and science are part of a vast

Table 4.11: Normalized distribution of comparatives and conditionals within *Biography* confounder, across topics

	COMPARATIVES			CONDITIONALS		
Confounders	**Distribution (%)**			**Distribution (%)**		
Topics:	*Arts*	*Education*	*Science*	*Arts*	*Education*	*Science*
martin-biography	11.95	10.94	12.24	3.41	1.51	1.19
saxon-biography	6.15	6.14	12.50	0.00	0.44	0.00

ecosystem called education[8], and it is perhaps not fair to directly compare their rhetorical style. However, since both arts and science are more specialized fields, we assume they involve much more creativity and contrasting thoughts than education. In this context, balance-focused schemes represent an excellent tool to efficiently analyze and emphasize equal concepts in an idea.

Table 4.12: Normalized distribution of balance schemes within *Reviews* by *Goodman*

Confounder	**Distribution (%)**		
Topics:	*Arts*	*Education*	*Science*
goodman-reviews	21.03	20.51	21.26

Likewise, the contrasts between *Arts* and *Science* are rhetorically fueled by conditionals. In table 4.13, we can spot a notable difference between the distributions of conditionals in these two topics compared to *Education*. Again, the reasonable explanation to support this behavior is that the nature of *Arts* and *Science* articles allows for more debatable matters. Both areas place important questions for all of us: "What is true? Why does it matter? How can we move society forward?"(Maeda [2013]). Both study the problems comprehensively and wander endlessly in a search for answers. Therefore, a conditional-rich article in arts or science should be absolutely encouraged.

Concluding with this dataset, we should mention that unlike *genres* and *authors*, the distribution of rhetorical devices amongst *topics* is rather similar. This is shown in the results of the significance test (table 4.14). Even though we obtain a favorable verdict from the effect-size test, it is not enough to raise doubts about the significance of datasets. The rhetorical style amidst the

[8]https://en.wikipedia.org/wiki/Education

Table 4.13: Normalized distribution of conditionals across topics

Topic	Distribution (%)
Arts	1.83
Education	1.26
Science	1.56

considered topics appears to be significantly alike. In particular, a very tight relation is between the styles of *Science* and *Education* articles. Although a more diverse dataset would be required to conclude anything with certainty, we guess that generally, rhetorical style is merely influenced by topics. As we show, it is possible to identify specific rhetoric divergences in some cases; still, rhetorical style seems to be primarily genre- and author-dependent.

Table 4.14: Significance and effect-size test results of *Topics* dataset

	SIGNIFICANCE		EFFECT-SIZE	
Datasets	**P-value**	**Independence**	**Cramer's V value**	**Effect**
Science vs. Education	0.70	FALSE	0.09	SMALL
Education vs. Arts	0.26	FALSE	0.10	SMALL
Arts vs. Science	0.19	FALSE	0.10	SMALL

4.2.2 Presidential Debates corpus

The Presidential Debates dataset is drawn from the American Presidency Project (APP) – "a leading source of presidential documents on the internet" [Woolley and Peters, 1999]. We only refer to debates of Donald Trump and Hillary Clinton with an assumption that these candidates are as opposite as possible to create a strong opinion dissonance; which ultimately leads to a dynamic rhetorical style.

The dataset contains both versus and individual debates against other candidates. Thus, we segment the dataset into four chunks of debates: Hillary $\rightarrow$ [9] Trump, Hillary $\rightarrow$ others, Trump $\rightarrow$ Hillary and Trump $\rightarrow$ others. Table 4.15 presents a general overview of the built corpus. Below, we present our

[9] the arrow indicates the source and destination of debate arguments.

analysis of the rhetorical style in all these subsets.

Table 4.15: General statistics on presidential debates dataset

Debate Type	Turns	Sentences	Avg. sent./turn	Avg. words/sent.
Clinton → Rest	1216	7187	5.91	17.47
Clinton → Trump	226	1227	5.43	15.69
Trump → Clinton	342	2023	5.92	10.95
Trump → Rest	778	3884	4.99	10.48

It is a common assumption that Trump is very fond of comparatives (especially, superlatives) as he uses them very frequently (Gingell [2016]). So, the first question we were interested in, is whether we can observe such behavior in his rhetorical style. Indeed, if we consider the simple count of comparatives (without normalization), he employs them more often (table 4.16); however, relating to the total number of devices, Clinton appears a much more avid user of this rhetorical style.

Table 4.16: Total occurrences and normalized distribution of comparatives in versus debates between Hillary Clinton and Donald Trump

Debate Type	Occurences*	Distribution (%)
Clinton → Trump	135	11.00
Trump → Clinton	142	7.02

* for reference, consider table 4.15

A recent study by Robin Raskin from *The Huffington Post* and *VisibleThread* – a company that "analyzes documents much like tools that analyze data", ran an algorithmic evaluation of the acceptance speeches of both candidates (Raskin [2016]). The results concluded that Clinton's language is 13% "clearer and more direct" than Trump's. The comprehension of Clinton's speech has been determined to be at a 5th-grade education level (in the US), compared to 8th-grade comprehension level of Trump's speech. Also, the analysis concluded that Clinton uses less passive voice (3.39%) than Trump (8.8%). Table 4.17 presents the results of this study.

Before discussing our results in relation to this study, it is important to mention that their dataset is utterly different in type and size. Therefore, it is

Table 4.17: Clinton's and Trump's acceptance speeches analysis results by VisibleThread

Candidate	Sent.	Long Sent. (%)	Passive voice (%)	Grade Level (US)
Hillary Clinton	413	7.26	3.39	5
Donald Trump	341	16.42	8.8	8

perhaps not fair to directly compare their findings with ours.

An interesting discovery which explains the comprehension level of their speeches, suggests that Clinton uses asyndeton more often than Trump (table 4.18). We would like to remind the reader that this rhetorical device is responsible for brevity and rhythm. Hence, it might be a crucial factor in making the speech easier to grasp.

Table 4.18: Normalized distribution of asyndeton, passive voice and balance schemes in versus debates between Clinton and Trump

	ASYNDETON	VOICE	BALANCE SCH.
Debate Type	**Distribution (%)**	**Distribution (%)**	**Distribution (%)**
Clinton → Trump	15.24	8.07	17.69
Trump → Clinton	10.83	5.29	19.92

As for passive voice, our analysis contradicts the study above. Table 4.18 shows a much higher rate of passive voice adoption by Hillary Clinton. Again, this can be easily influenced by the type of data (debates vs. acceptance speech). Based on our partial knowledge in this area, we suspect that in a direct debate with Trump, Clinton ought to use more passive voice because of her ties with Democratic Party, a member of which is Barack Obama – president of the US from 2009 to 2017. Allegedly, in a debate, Trump targets the whole Democratic Party with questions wherein Clinton is not the direct object. In our opinion, this explains the difference in passive voice usage between candidates.

Furthermore, table 4.18 emphasizes another contrast between Trump and Clinton, and namely, the sentence length. As reported in the conducted study, Hillary's sentences are short, concise and simple (Raskin [2016]). We assume that this fact explains the rare usage of balance schemes (in particular, isocola) in Hillary's speech compared to Trump (table 4.18). Balance schemes aim to emphasize multiple parts of a longer sentence (or fragment) to preserve the equivalence between the exposed ideas. Therefore, we suspect it is more characteristic to Trump's rhetorical style.

In general, considering the significance and effect-size tests results (tables 4.19 and 4.20), we conclude that unlike Hillary Clinton, Donald Trump doesn't change his rhetorical strategy depending on the opponent. Even though the confidence level is not high, it is evident that Clinton prefers to vary the rhetorical style function of the opponent.

Table 4.19: Significance test results of presidential debates between Hillary Clinton and Donald Trump

Debate Type	*Clinton → Rest*	*Clinton → Trump*	*Trump → Clinton*	*Trump → Rest*
Clinton → Rest		TRUE*	TRUE	TRUE
Clinton → Trump	TRUE*		TRUE	TRUE
Trump → Clinton	TRUE	TRUE		FALSE†
Trump → Rest	TRUE	TRUE	FALSE†	

* for $\alpha > 0.01$
† for $\alpha > 0.1$

Table 4.20: Effect-size test results of presidential debates between Hillary Clinton and Donald Trump

Debate Type	*Clinton → Rest*	*Clinton → Trump*	*Trump → Clinton*	*Trump → Rest*
Clinton → Rest		SMALL	SMALL	SMALL*
Clinton → Trump	SMALL		SMALL	SMALL
Trump → Clinton	SMALL	SMALL		SMALL
Trump → Rest	SMALL*	SMALL	SMALL	

* close to medium effect (Cramer's V value = 0.19)

4.3 Efficiency

Although we do not focus much on the efficiency aspect of our system, it is essential to discuss this, in particular for further improvements. An efficient program is considered the one whose resource consumption is lower or equal to some acceptable level[10]. Of course, "acceptable" is an abstract term which

[10] https://en.wikipedia.org/wiki/Algorithmic_efficiency

depends on different factors like type of problem, available computational resources, and complexity amongst others. Roughly, "acceptable" means that the program will run a reasonable amount of time on the given machine[10].

Before measuring the execution time of our software, let us list the computer specs on which the efficiency test is performed:

- CPU: 2 ∗ Intel® Xeon® E5-2620 v2 @ 2.10GHz
- Memory: 128Gb
- Display: NVIDIA® GeForce® GTX480

For simplicity, we measure the efficiency of our rhetoric detection system on a single document which contains one sentence (20 words long). The time it takes to process our sample sentence is *5.8 seconds*. Out of these, the program takes about 1.7 seconds to initialize the UIMA components: collection reader and analysis engine. Adding another sentence increases the execution time to 0.4 seconds on average. Bearing in mind that the reliability of results is directly proportional to the input corpus size, there is, obviously, room for improvements in this direction.

One of the main factors which slow down our system's efficiency in terms of runtime is the Stanford CoreNLP suite. In particular, the parser needs to load its models, parse each token iteratively in the sentence and build its grammatical structure. Likewise, dependency relations graph is generated by using different models. These operations are time-expensive, yet we heavily rely on their performance to identify the rhetorical devices. We suspect that further updates should incrementally improve the overall efficiency.

Besides the third-party software we rely upon, the complexity of some algorithms responsible for rhetorical devices identification might also affect the efficiency of our system. For example, isocolon's algorithm involves a series of nested loops which tend to increase its runtime complexity. On top of that, the number of rhetorical devices we regard in this thesis is much higher than in the existing works of the same nature (Java [2015], Gawryjołek et al. [2009]). Each sentence in the document is checked by every algorithm for rhetorical devices. There are, of course, more elegant ways to deal with such issues, yet for now, we consider them for further improvements.

Chapter 5

Conclusion

In this thesis, we have investigated the style of writing across multiple dimensions based on automatic detection of rhetorical devices; we present a new approach to process large corpora of texts, find and annotate the existing rhetorical devices there, and compute insightful statistics on the rhetorical patterns found in the texts.

The generated statistics which consider the significance and effect-size tests, help us to deduce and interpret the difference in rhetorical style employed by particular authors across genres, topics, and presidential debates.

We started with the premise that, historically, rhetoric is a powerful tool to decorate the language and to make it more efficient to persuade the audience. Rhetoric is a broad ecosystem which treats multiple aspects of the language. This thesis, on the whole, focuses on the *style* side of rhetoric. The *§Background* section discusses the origin of rhetoric and its evolution up to modern times in a detailed fashion. After stating the role of rhetoric in computational linguistics and outlining the noteworthy works in this direction in section (*§Related Work*), we proceed to discuss our approach.

The *§Detection of Rhetorical Devices* chapter presents our approach to detect rhetorical devices. We start with introducing the third-party tools used in our framework, then, we present the selected rhetorical devices, and describe the detection rules. After that, we report on the performance of our approach in terms of precision, recall, and F1-score measures. The *§Analysis of Rhetorical Devices* chapter discusses a series of experiments for style exploring in two different corpora: newspaper articles (NYT) and presidential debates. The experiments have led to several findings that we discussed and interpreted in details.

In our analysis, we considered a threefold subsampling of the NYT corpus: random, article-length, and controlled subsampling via matching. Using the random and article-length subsampled datasets, we didn't find many compelling patterns of rhetorical devices across texts aspects (e.g., genre). Presumably, that's due to large amounts of data characterized by multiple confounders. This problem is solved with matching in the controlled subsampled dataset. Consequently, by carrying out the experiments on the said dataset, we were able to identify more patterns of style across genre, type, author, and debaters.

Based on the results of our analysis on the controlled subsampled dataset, we concluded that rhetorical style in articles is largely influenced by their genres and authors. Apparently, the topic of texts has the least influence on their rhetorical style. As a concrete example, we have observed that *Arts* articles, unlike *Science* and *Education* ones, do not share a common rhetorical style; while some authors seem to agree to use more conditionals and comparatives, others try to avoid such rhetorical techniques. This again, confirms our hypothesis that the style is more author- than topic-dependent.

Another example of findings concerns the *genres* this time, particularly, the distribution of comparatives. We discover a higher frequency of comparatives in *Review* and *Editorial* articles compared to *Biographies*. This was anticipated as the former genres are more opinionated by nature, and this allows for more freedom to interpret and describe the subject more creatively via comparatives. In contrast, *Biographies* approach the entity directly, without much room for improvisation.

As for the *authors* dataset, we concluded that each author has its unique rhetorical style. This is shown in the results of the significance and effect-size tests. Unfortunately, after matching (see *§Subsampling the NYT corpus*), we end up with similar topics as confounders: *biography* and *obituary*. Therefore, the system couldn't catch significant oscillations in style across articles of the same author. Nonetheless, we observed that some authors favor the usage of repetitions towards a more persuasive style of writing, while others avoid conditionals in particular genres.

The analysis of the presidential debates between Hillary Clinton and Donald Trump was much more intriguing and controversial in patches. Specifically, contradicting the common belief, we concluded that Trump employs comparatives less often than Clinton. We believe that's due to the nature of debate texts; since other studies found Trump to use visibly more comparatives that

Clinton (Raskin [2016]). Overall, we inferred that while Clinton varies her style considering the opponent, Trump doesn't. This finding might suggest that because of a larger experience in politics, Hillary learned to adjust her speech style based on her intentions and target audience.

Future Work

In general, the scarcity of data in the controlled experiment – as consequence of *matching* (see table 4.1) – can be considered a limitation of our analysis. We suspect that a larger and more diverse dataset should be able to truly assess our system's capacity and yield much more valuable insights. Therefore, part of the future work should focus on corpus diversification, in particular for the controlled experiment. Also, it would be interesting to see whether the rhetorical style depends on the corpus origin, target audience or quality.

As we mention in the *§Introduction* chapter, from the rhetorical perspective, the missing part of this work is semantical rhetoric. There is no doubt that rhetorical devices in this category should contribute to a more accurate identification of the rhetorical style. Therefore, we believe that semantic-based rhetorical devices must be part of future enhancements in this direction.

Lastly, we would like to point out minor improvements to be considered in future work. First, we can adopt a dynamic window-size approach presented in Gawryjołek et al. [2009] and Java [2015]. A window length dependent on the type of rhetorical device should improve the accuracy of detection. Also, besides the sentence-level detection, we might obtain insightful results on the paragraph level as well. Secondly, even if the distributions of devices represent one of the most important indicators to infer the style, there are other stats like placement, rhetorical evolution (Gawryjołek et al. [2009]) or flows. These statistics could provide an additional perception of the rhetorical style of writings.

Appendices

Appendix 1: Performance comparison between versions of Stanford Parser

Table 1: Performance measures comparison between two versions of Stanford Parser (3.5.2 and 3.8.0), in the detection of comparatives and superlatives adjectives/adverbs. A slight performance improvement is brought by the 3.8.0 version.

Device	Precision	Recall	F1-score
Stanford Parser 3.8.0			
Comp. Adjective	0.51	0.61	0.56
Comp. Adverb	0.6	0.62	0.61
Super. Adjective	0.62	0.73	0.67
Super. Adverb	0.63	0.5	0.56
Stanford Parser 3.5.2			
Comp. Adjective	0.48	0.44	0.46
Comp. Adverb	0.61	0.79	0.7
Super. Adjective	0.48	0.63	0.56
Super. Adverb	0.65	0.49	0.57

Appendix 2: Penn Treebank II tagset

Table 2: Penn Treebank II - word level tags

Tag	Definition
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun (prolog version PRP-S)
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non 3rd person singular present
VBZ	Verb, 3rd-person singular present
WDT	Wh - determiner
WP	Wh - pronoun
WP$	Possessive wh - pronoun (prolog version WP-S)
WRB	Wh - adverb

Appendix 3: Combinations of RD in random and article-length based subsampled dataset

Table 3: Distributions of combinations of RD amongst authors in random subsampled dataset

Author	Group	Combination	Count
Hevesi Dennis	by 2	Asyndeton - Hypozeugma	4963
		Asyndeton - PassiveVoice	2518
		Asyndeton - Enumeration	2433
		Hypozeugma - PassiveVoice	1227
	by 3	Asyndeton - Hypozeugma - PassiveVoice	708
		Asyndeton - Hypozeugma - Enumeration	626
		Asyndeton - PassiveVoice - Hypozeugma	569
		Asyndeton - Enumeration - Hypozeugma	470
	by 4	Asyndeton - Hypozeugma - Enumeration - PassiveVoice	66
		Asyndeton - PassiveVoice - Enumeration - Hypozeugma	58
		Asyndeton - Enumeration - PassiveVoice - Hypozeugma	48
		Asyndeton - Hypozeugma - Enumeration - comparativeAdj	48
Lewis Paul	by 2	Asyndeton - Hypozeugma	3417
		Asyndeton - Enumeration	1575
		Asyndeton - PassiveVoice	1375
		Hypozeugma - Enumeration	1112
	by 3	Asyndeton - Hypozeugma - Enumeration	547
		Asyndeton - Hypozeugma - PassiveVoice	465
		Asyndeton - Enumeration - Hypozeugma	374
		Asyndeton - PassiveVoice - Hypozeugma	370
	by 4	Asyndeton - PassiveVoice - Hypozeugma - Enumeration	57
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	45
		Asyndeton - Hypozeugma - Enumeration - comparativeAdj	38
		Asyndeton - Hypozeugma - PassiveVoice - comparativeAdj	36
Martin Douglas	by 2	Asyndeton - Hypozeugma	4025
		Asyndeton - Enumeration	2613
		Asyndeton - PassiveVoice	2251
		Hypozeugma - Enumeration	1059
	by 3	Asyndeton - Hypozeugma - Enumeration	546
		Asyndeton - Hypozeugma - PassiveVoice	474
		Asyndeton - Enumeration - Hypozeugma	436
		Asyndeton - Enumeration - PassiveVoice	432
	by 4	Asyndeton - Hypozeugma - Enumeration - comparativeAdj	45
		Epizeugma - Asyndeton - Enumeration - PassiveVoice	42
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	40
		Asyndeton - Enumeration - PassiveVoice - Hypozeugma	39

Table 4: Distributions of combinations of RD amongst authors in article-length subsampled dataset

Author	Group	Combination	Count
Hevesi Dennis	by 2	Asyndeton - Hypozeugma	2322
		Asyndeton - PassiveVoice	1277
		Asyndeton - Enumeration	1006
		Hypozeugma - PassiveVoice	632
	by 3	Asyndeton - Hypozeugma - PassiveVoice	415
		Asyndeton - PassiveVoice - Hypozeugma	283
		Asyndeton - Hypozeugma - Enumeration	283
		Asyndeton - Enumeration - Hypozeugma	184
	by 4	Asyndeton - Hypozeugma - Enumeration - PassiveVoice	32
		Asyndeton - PassiveVoice - Enumeration - Hypozeugma	31
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	28
		Asyndeton - Hypozeugma - PassiveVoice - Enumeration	24
Lewis Paul	by 2	Asyndeton - Hypozeugma	1502
		Asyndeton - Enumeration	856
		Asyndeton - PassiveVoice	775
		Hypozeugma - Enumeration	501
	by 3	Asyndeton - Hypozeugma - Enumeration	248
		Asyndeton - Hypozeugma - PassiveVoice	211
		Asyndeton - PassiveVoice - Hypozeugma	172
		Asyndeton - Enumeration - Hypozeugma	153
	by 4	Asyndeton - PassiveVoice - Hypozeugma - Enumeration	33
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	27
		Epizeugma - Asyndeton - Enumeration - PassiveVoice	23
		Asyndeton - Hypozeugma - PassiveVoice - comparativeAdj	20
Martin Douglas	by 2	Asyndeton - Hypozeugma	1785
		Asyndeton - Enumeration	1103
		Asyndeton - PassiveVoice	962
		Enumeration - PassiveVoice	432
	by 3	Asyndeton - Hypozeugma - Enumeration	214
		Asyndeton - Hypozeugma - PassiveVoice	207
		Asyndeton - Enumeration - PassiveVoice	194
		Asyndeton - Enumeration - Hypozeugma	168
	by 4	Asyndeton - PassiveVoice - Enumeration - Hypozeugma	16
		Asyndeton - Enumeration - Hypozeugma - comparativeAdj	15
		Asyndeton - Hypozeugma - Enumeration - comparativeAdj	15
		Epizeugma - PassiveVoice - Asyndeton - Enumeration	15

Table 5: Distributions of combinations of RD amongst genres in random subsampled dataset

Genre	Group	Combination	Count
Biography	by 2	Asyndeton - Hypozeugma	3863
		Asyndeton - Enumeration	2502
		Asyndeton - PassiveVoice	1900
		Hypozeugma - Enumeration	1133
	by 3	Asyndeton - Hypozeugma - Enumeration	615
		Asyndeton - Hypozeugma - PassiveVoice	466
		Asyndeton - Enumeration - Hypozeugma	451
		Asyndeton - PassiveVoice - Hypozeugma	400
	by 4	Asyndeton - Hypozeugma - Enumeration - PassiveVoice	56
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	53
		Asyndeton - Enumeration - PassiveVoice - Hypozeugma	48
		Asyndeton - Hypozeugma - PassiveVoice - Enumeration	46
Editorial	by 2	Asyndeton - Hypozeugma	1228
		Asyndeton - Enumeration	734
		Hypozeugma - Enumeration	449
		Asyndeton - PassiveVoice	428
	by 3	Asyndeton - Hypozeugma - Enumeration	176
		Asyndeton - Enumeration - Hypozeugma	122
		Asyndeton - Hypozeugma - PassiveVoice	122
		Asyndeton - PassiveVoice - Hypozeugma	93
	by 4	Asyndeton - Hypozeugma - Enumeration - comparativeAdj	16
		Asyndeton - Hypozeugma - PassiveVoice - Enumeration	12
		Asyndeton - Hypozeugma - Enumeration - superlativeAdj	11
		Asyndeton - Enumeration - Hypozeugma - comparativeAdj	11
Review	by 2	Asyndeton - Hypozeugma	2800
		Asyndeton - Enumeration	2441
		Asyndeton - PassiveVoice	1327
		Hypozeugma - Enumeration	1282
	by 3	Asyndeton - Hypozeugma - Enumeration	610
		Asyndeton - Enumeration - Hypozeugma	429
		Asyndeton - Hypozeugma - PassiveVoice	350
		Asyndeton - PassiveVoice - Hypozeugma	246
	by 4	Asyndeton - Hypozeugma - Enumeration - comparativeAdj	47
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	45
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	44
		Asyndeton - Enumeration - Hypozeugma - comparativeAdj	41

Table 6: Distributions of combinations of RD amongst genres in article-length subsampled dataset

Genre	Group	Combination	Count
Biography	by 2	Asyndeton - Hypozeugma	1459
		Asyndeton - Enumeration	1069
		Asyndeton - PassiveVoice	777
		Hypozeugma - Enumeration	389
	by 3	Asyndeton - Hypozeugma - Enumeration	244
		Asyndeton - Hypozeugma - PassiveVoice	204
		Asyndeton - Enumeration - PassiveVoice	199
		Asyndeton - Enumeration - Hypozeugma	177
	by 4	Epizeugma - Asyndeton - Enumeration - PassiveVoice	49
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	33
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	20
		Diacope - Asyndeton - Hypozeugma - Enumeration	19
Editorial	by 2	Asyndeton - Hypozeugma	828
		Asyndeton - Enumeration	443
		Hypozeugma - Enumeration	310
		Asyndeton - PassiveVoice	272
	by 3	Asyndeton - Hypozeugma - Enumeration	117
		Asyndeton - Enumeration - Hypozeugma	88
		Asyndeton - Hypozeugma - PassiveVoice	73
		Asyndeton - Hypozeugma - comparativeAdj	57
	by 4	Asyndeton - Hypozeugma - Enumeration - superlativeAdj	9
		Asyndeton - Enumeration - Hypozeugma - comparativeAdj	9
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	8
		Asyndeton - Enumeration - Hypozeugma - superlativeAdj	7
Review	by 2	Asyndeton - Hypozeugma	1207
		Asyndeton - Enumeration	1002
		Hypozeugma - Enumeration	573
		Asyndeton - PassiveVoice	482
	by 3	Asyndeton - Hypozeugma - Enumeration	265
		Asyndeton - Enumeration - Hypozeugma	190
		Asyndeton - Hypozeugma - PassiveVoice	115
		Asyndeton - PassiveVoice - Hypozeugma	99
	by 4	Asyndeton - Hypozeugma - Enumeration - superlativeAdj	16
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	15
		Asyndeton - Enumeration - Hypozeugma - comparativeAdj	14
		Asyndeton - Hypozeugma - Enumeration - comparativeAdj	13

Table 7: Distributions of combinations of RD amongst topics in random subsampled dataset

Topic	Group	Combination	Count
Arts	by 2	Asyndeton - Hypozeugma	3137
		Asyndeton - Enumeration	2496
		Asyndeton - PassiveVoice	1562
		Hypozeugma - Enumeration	1305
	by 3	Asyndeton - Hypozeugma - Enumeration	626
		Asyndeton - Enumeration - Hypozeugma	431
		Asyndeton - Hypozeugma - PassiveVoice	412
		Asyndeton - PassiveVoice - Hypozeugma	342
	by 4	Asyndeton - Hypozeugma - PassiveVoice - Enumeration	57
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	55
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	52
		Asyndeton - Hypozeugma - Enumeration - comparativeAdj	45
Education	by 2	Asyndeton - Hypozeugma	3676
		Asyndeton - Enumeration	1956
		Asyndeton - PassiveVoice	1693
		Hypozeugma - Enumeration	1056
	by 3	Asyndeton - Hypozeugma - Enumeration	498
		Asyndeton - Hypozeugma - PassiveVoice	493
		Asyndeton - Enumeration - Hypozeugma	402
		Asyndeton - PassiveVoice - Hypozeugma	397
	by 4	Asyndeton - Hypozeugma - Enumeration - comparativeAdj	47
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	46
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	45
		Asyndeton - Hypozeugma - PassiveVoice - comparativeAdj	42
Science	by 2	Asyndeton - Hypozeugma	3540
		Asyndeton - Enumeration	2143
		Asyndeton - PassiveVoice	1801
		Hypozeugma - Enumeration	1164
	by 3	Asyndeton - Hypozeugma - Enumeration	583
		Asyndeton - Hypozeugma - PassiveVoice	521
		Asyndeton - Enumeration - Hypozeugma	420
		Asyndeton - PassiveVoice - Hypozeugma	402
	by 4	Asyndeton - Hypozeugma - PassiveVoice - Enumeration	59
		Asyndeton - Hypozeugma - Enumeration - comparativeAdj	56
		Asyndeton - Enumeration - PassiveVoice - Hypozeugma	48
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	48

Table 8: Distributions of combinations of RD amongst topics in article-length subsampled dataset

Topic	Group	Combination	Count
Arts	by 2	Asyndeton - Hypozeugma	1235
		Asyndeton - Enumeration	918
		Asyndeton - PassiveVoice	541
		Hypozeugma - Enumeration	502
	by 3	Asyndeton - Hypozeugma - Enumeration	222
		Asyndeton - Enumeration - Hypozeugma	164
		Asyndeton - Hypozeugma - PassiveVoice	155
		Asyndeton - PassiveVoice - Hypozeugma	118
	by 4	Asyndeton - PassiveVoice - Enumeration - Hypozeugma	21
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	17
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	17
		Asyndeton - Enumeration - Hypozeugma - PassiveVoice	14
Education	by 2	Asyndeton - Hypozeugma	1932
		Asyndeton - Enumeration	917
		Asyndeton - PassiveVoice	907
		Hypozeugma - Enumeration	526
	by 3	Asyndeton - Hypozeugma - PassiveVoice	289
		Asyndeton - Hypozeugma - Enumeration	263
		Asyndeton - PassiveVoice - Hypozeugma	215
		Asyndeton - Enumeration - Hypozeugma	186
	by 4	Asyndeton - Hypozeugma - Enumeration - PassiveVoice	27
		Asyndeton - PassiveVoice - Hypozeugma - Enumeration	23
		Asyndeton - Hypozeugma - PassiveVoice - comparativeAdj	21
		Asyndeton - PassiveVoice - Hypozeugma - comparativeAdj	20
Science	by 2	Asyndeton - Hypozeugma	1357
		Asyndeton - Enumeration	893
		Asyndeton - PassiveVoice	816
		Hypozeugma - Enumeration	439
	by 3	Asyndeton - Hypozeugma - PassiveVoice	237
		Asyndeton - Hypozeugma - Enumeration	221
		Asyndeton - Enumeration - Hypozeugma	171
		Asyndeton - PassiveVoice - Hypozeugma	158
	by 4	Diacope - Epizeugma - Asyndeton - PassiveVoice	41
		Diacope - PassiveVoice - Epizeugma - Asyndeton	33
		Asyndeton - Hypozeugma - Enumeration - PassiveVoice	27
		Asyndeton - Hypozeugma - PassiveVoice - Enumeration	25

Appendix 4: Controlled experiment: distribution of rhetorical devices

Figure 1: Distribution of rhetorical devices amongst authors, in controlled subsampled dataset.

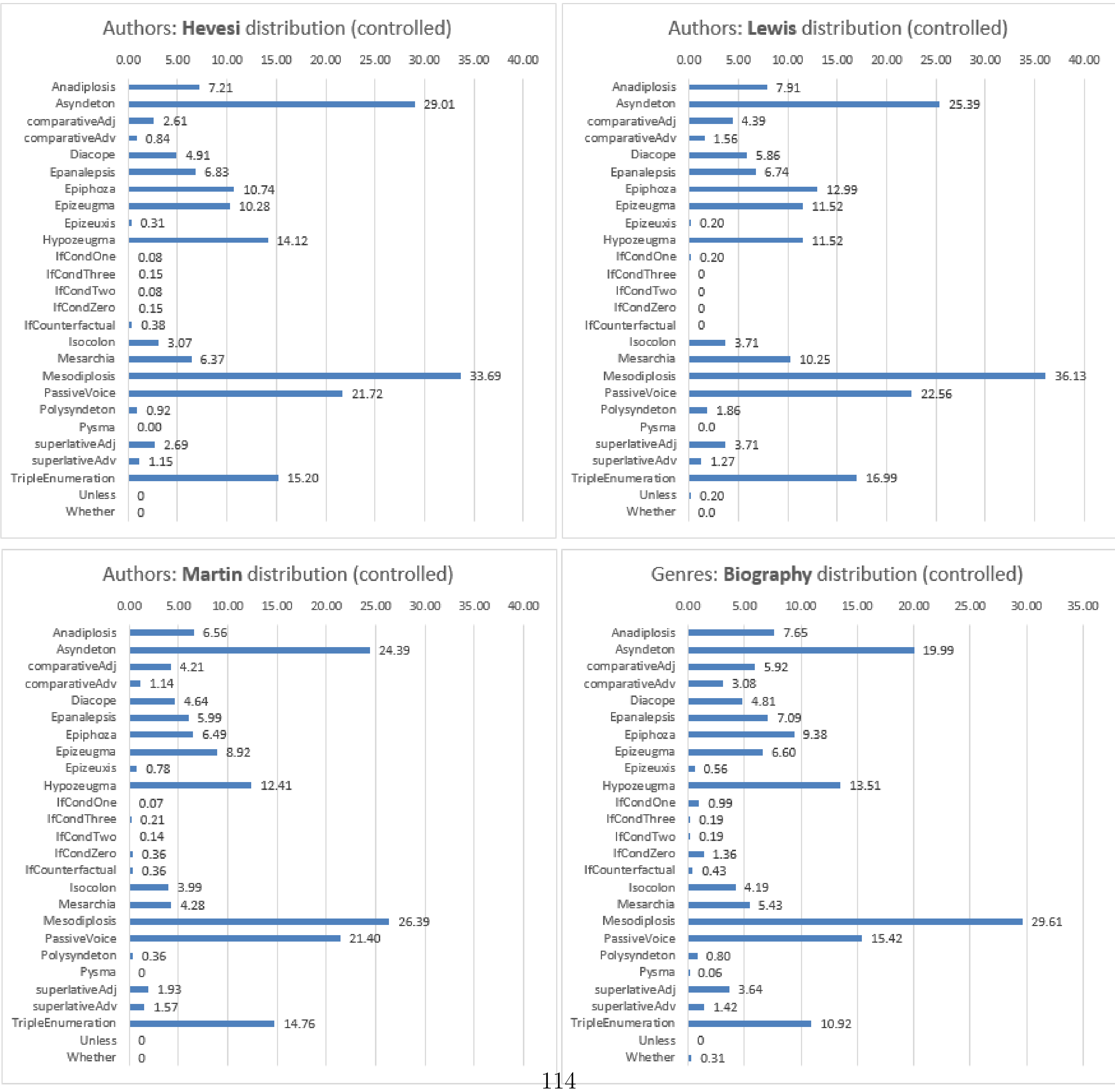


Figure 2: Distribution of rhetorical devices amongst authors, in controlled subsampled dataset.

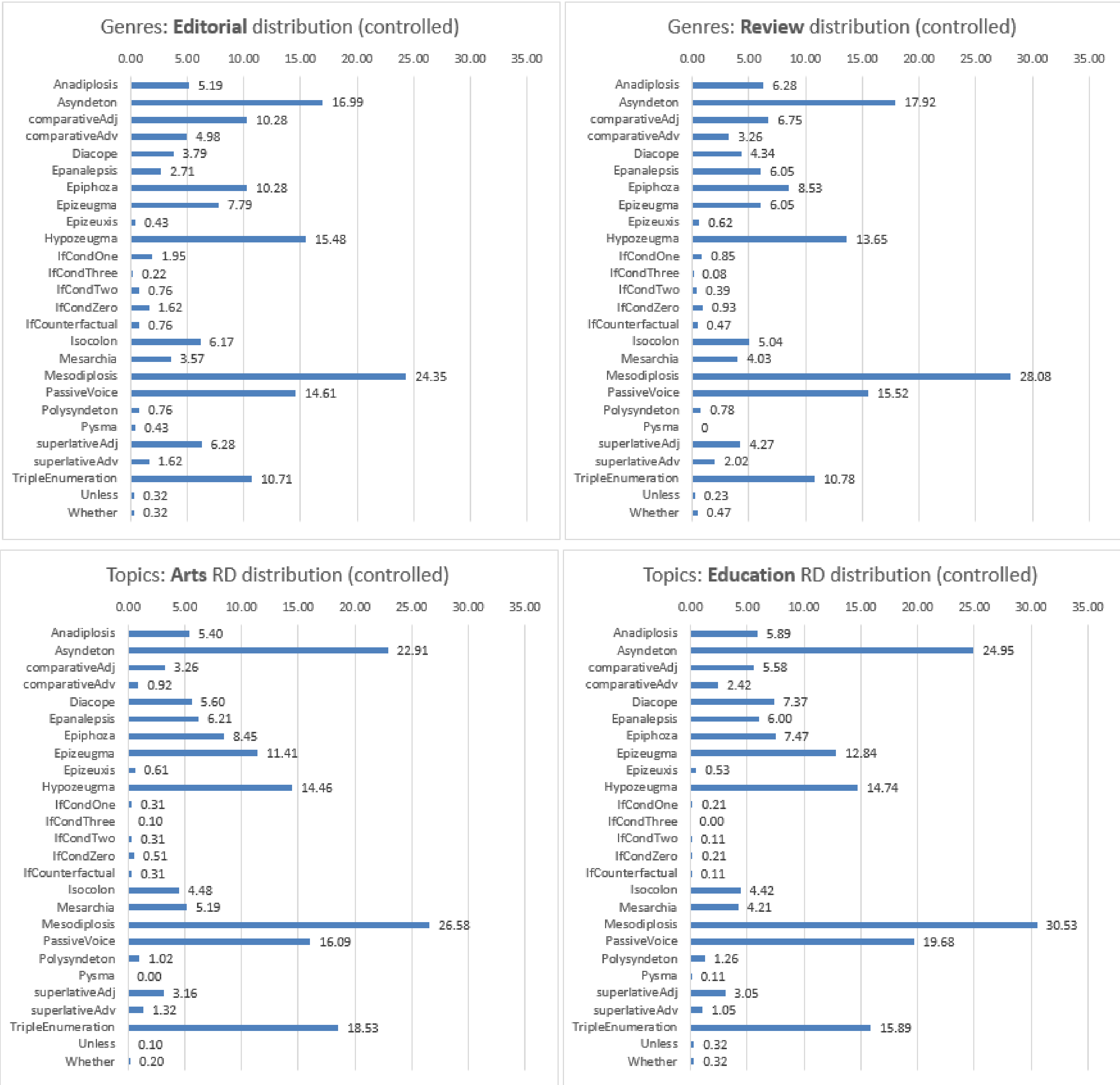
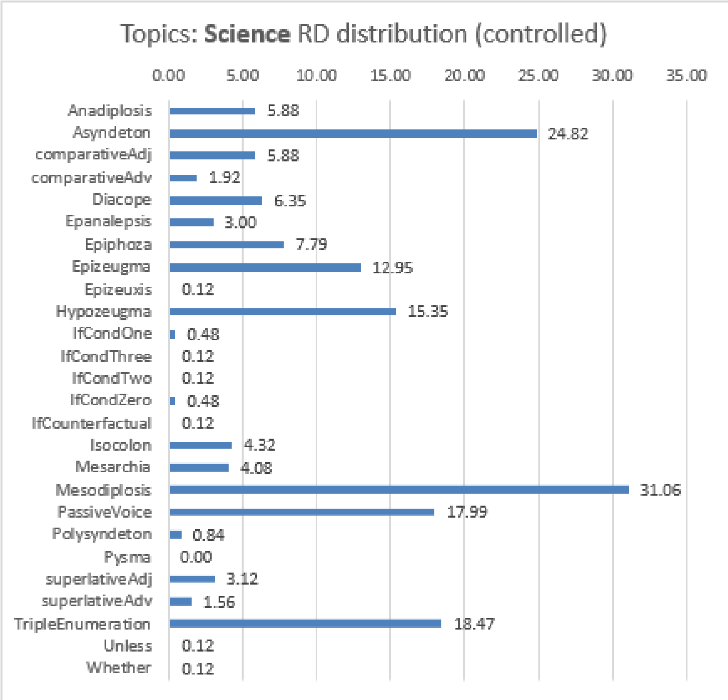


Figure 3: Distribution of rhetorical devices amongst authors, in controlled subsampled dataset.



Bibliography

Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein. Patterns of Argumentation Strategies across Topics. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 17)*, pages 1362–1368. Association for Computational Linguistics, September 2017. URL http://aclweb.org/anthology/D17-1142. 4.2.1.1, 4.2.1.2.2

Shlomo Argamon, Casey Whitelaw, Paul Chase, Sobhan Raj Hota, Navendu Garg, and Shlomo Levitan. Stylistic text classification using functional lexical features: Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 58 (6):802–822, April 2007. ISSN 1532-2882. doi: 10.1002/asi.v58:6. URL http://dx.doi.org/10.1002/asi.v58:6. 2.2

Aristotle and W.R. Roberts. *Rhetoric*. Dover thrift editions. Dover Publications, 2004. ISBN 9780486437934. URL https://books.google.de/books?id=V_IMCohto0YC. 3.2.1

Pilon M.-A. Bowden, B. and A. Harris. The rhetorical figure ontology project. https://artsresearch.uwaterloo.ca/rhetfig/, 2011. visited on 25.10.17. 3.3

E.W. Bullinger and I. Delmarva Publications. *Figures of Speech Used in the Bible*. Baker Book House, 1970. URL https://books.google.de/books?id=CNrNBgAAQBAJ. 3.2.10

G. Burton. Figures of repetition on silva rhetoricae, 1996. Accessed on 18.09.2017. 2.2

G. Burton. The forest of rhetoric (silva rhetoricae), 2007. Accessed on 16.08.2017. 1, 2.1.1, 3.2, 3.2, 1, 2, 3.2, 3.2.1, 3.3

Cicero. Ad c. herennium: de ratione dicendi (rhetorica ad herennium). Harvard University press, 1954. translated by Caplan, Harry. 2.1.1, 3.2.3

A. Clark, C. Fox, and S. Lappin. *The Handbook of Computational Linguistics and Natural Language Processing*. Wiley-Blackwell, 2010. ISBN 9781444324044. 2.2, 3.3

E. P. J. Corbett. *Classical rhetoric for the modern student*. USA: Oxford University Press, 3 edition, 1990. 3.2, 3.2.3

Robert T. Craig. Communication. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2006. URL `http://www.oxford-rhetoric.com/entry?entry=t223.e45`. 2.1.2

Jager K J Zoccali C Dekker F W de Graaf M, A. Matching, an appealing method to avoid confounding? 2011. 4.1

Marie-Catherine de Marneffe and Christopher D. Manning. Stanford typed dependencies manual, September 2008. 3.1.1, 3.1, 3.1.1, 15, 16, 17, 18, 3.2.4

R. Declerck and S. Reed. *Conditionals: A Comprehensive Empirical Analysis*. Beitrage Zur Alexander-Von-Humboldt-Forschung. Mouton de Gruyter, 2001. ISBN 9783110171440. URL `https://books.google.de/books?id=RtyewWWe1LAC`. 3.2.16, 3.2.4, 35, 3.2.4, 3.2.4, 3.2.4, 3.2.22, 3.2.23, 3.2.4, 47, 3.2.5, 3.2.5, 53, 3.3.4

Dictionary.com. Isocolon. URL `http://www.dictionary.com/browse/isocolon`. Accessed on 09.09.2017. 3

Jeanne Fahnestock. Verbal and visual parallelism. *Written Communication*, 20(2):123–152, 2003. doi: 10.1177/0741088303020002001. URL `https://doi.org/10.1177/0741088303020002001`. 3.2

W. Farnsworth. *Farnsworth's Classical English Rhetoric*. David R. Godine, Publisher, 2011. ISBN 9781567923858. URL `https://books.google.de/books?id=1tmyiqjvpnYC`. 3.2.1

Christiane Fellbaum. Wordnet and wordnets. In Keith et al. Brown, editor, *Encyclopedia of Language and Linguistics*, pages 665–670. Elsevier, 2005. 7

David Ferrucci and Adam Lally. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. *Natural Language Engineering*, 10(3-4):327–348, sep 2004. doi: 10.1017/S1351324904003523. 3.1.2

Robert A. Gaines. Speech. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2006. URL `http://www.oxford-rhetoric.com/entry?entry=t223.e232`. 2.1.2

A. D. Galinsky and L. J. Kray. From thinking about what might have been to sharing what we know: The effects of counterfactual mindsets on information sharing in groups. *Journal of Experimental Social Psychology*, pages 606–618, 2004. 3.2.4

Jakub J. Gawryjołek, Randy A. Harris, and Chrysanne DiMarco. An annotation tool for automatically detecting rhetorical figures. In *Proceedings, CMNAIX (Computational Models of Natural Argument)*, Pasadena, CA, July 2009. 1, 2.2, 4.3, 5

James Gingell. Why superlatives are the absolute worst (unless you're donald trump). https://www.theguardian.com/media/mind-your-language/2016/apr/15/why-superlatives-are-the-absolute-worst-unless-youre-donald-trump, 2016. visited on 24.10.17. 3.2.4, 4.2.2

Michael D. Gordin. *Scientific Babel: How Science Was Done Before and After Global English*. University of Chicago Press, Chicago, Illinois, 2015. ISBN 9780226000299. 2.1.3

Andrea Grün-Oesterreich. Anadiplosis. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2007a. URL http://www.oxford-rhetoric.com/entry?entry=t223.e9. 3.2.3

Andrea Grün-Oesterreich. Asyndeton. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2007b. URL http://www.oxford-rhetoric.com/entry?entry=t223.e28. 3.2.2

William W. Hallo. The birth of rhetoric. In Carol S. Lipson and Roberta A. Binkley, editors, *Rhetoric before and beyond the Greeks*, pages 25–46. State University of New York Press, 2004. ISBN 0-7914-6099-1. 2.1.1

R. Harris and C. DiMarco. Constructing a rhetorical figuration ontology. In *In Symposium on Persuasive Technology and Digital Behaviour Intervention*, 2009. 3.2, 3.1

G.A. Hauser. *Introduction to Rhetorical Theory: Second Edition*. Waveland Press, 2002. ISBN 9781478608943. URL https://books.google.de/books?id=IWAWAAAAQBAJ. 3.2.3

Anne Marie Helmenstine. What is a controlled experiment? - thoughtco. thoughtco.com/controlled-experiment-609091, 2017. visited on 14.11.17. 4.1

Roese N. J. Hooker, C. and S. Park. Impoverished counterfactual thinking is associated with schizophrenia. *Psychiatry*, 2000. 3.2.4

J. Hutchins. The history of machine translation in a nutshell. 2005. 2.1.3

David Hutto. Ancient egyptian rhetoric in the old and middle kingdoms. *Rhetorica: A Journal of the History of Rhetoric*, 20(3):213–233, 2002. ISSN 07348584, 15338541. doi: 10.1525/rh.2002.20.3.213. URL `http://www.jstor.org/stable/10.1525/rh.2002.20.3.213`. 2.1.1

Anthony Janocko, Allegra Larche, Joseph Raso, and Kevin Zembroski. Counterfactuals in the language of social media: A natural language processing. *Project in Conjunction with the World Well Being Project.*, August 2015. 3.2.4, 3.2.21, 3.2.4, 3.2.4

James Java. *Characterization of Prose by Rhetorical Structure for Machine Learning Classification.* PhD thesis, Nova Southeastern University, August 2015. URL `http://nsuworks.nova.edu/gscis_etd/347`. 1, 1, 1, 2.2, 2.2, 2.2, 3.3, 3.3, 4.1, 4.3, 5

Java Documentation. Breakiterator. URL `https://docs.oracle.com/javase/7/docs/api/java/text/BreakIterator.html`. Accessed on 09.09.2017. 6

R. Johnson. *The Alphabet of Rhetoric*. BiblioLife, 2016. ISBN 9781358725661. URL `https://books.google.de/books?id=Z8FfvgAACAAJ`. 3.2.2

R. Kirchner. Elocutio: Latin prose style. In W. Dominik and J. Hall (Eds.), editors, *A companion to Roman rhetoric*, pages 181–194. Blackwell Publishing, Malden, MA, USA, 2007. 2.1.1

Dan Klein and Christopher D. Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 423–430, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. doi: 10.3115/1075096.1075150. URL `https://doi.org/10.3115/1075096.1075150`. 2.2

Peter Kluegl and Martin Atzmueller. Textmarker: A tool for rule-based information extraction, 2009. 3.1.3

Peter Kluegl, Martin Toepfer, Philip-Daniel Beck, Georg Fette, and Frank Puppe. Uima ruta: Rapid development of rule-based information extraction applications. *Natural Language Engineering*, 22:1–40, 1 2016. ISSN 1469-8110. doi: 10.1017/S1351324914000114. URL `https://journals.cambridge.org/article_S1351324914000114`. 3.1.3, 3.1.3

George Linda G. Liljenquist Katie A. Galinsky Adam D. Tetlock Philip E. Roese Neal J. Kray, Laura J. From what might have been to what must have been: Counterfactual thinking creates meaning. *Journal of Personality and Social Psychology*, pages 106–118, 2010. 3.2.4

Visser Jackya Lawrence, Johna and Chris Reed. Harnessing rhetorical figures for argument mining. *Argument & Computation*, 8:289–310, 2017. 1

V. J. Leffa. Clause processing in complex sentences. In *In Proceedings of the First International Conference on Language Resource & Evaluation*, volume 1, pages 937–943, May 1998. 2.2

E.D Liddy. Natural language processing. In NY. Marcel Decker, editor, *Encyclopedia of Library and Information Science*. 2 edition, 2001. 2.1.3

Literarydevices.com. Epizeuxis. URL `http://www.literarydevices.com/epizeuxis/`. Accessed on 09.09.2017. 4

P. Mack. *A History of Renaissance Rhetoric, 1380-1620*. Oxford-Warburg studies. Oxford University Press, 2011. ISBN 9780191804588. URL `https://books.google.de/books?id=TBpwrgEACAAJ`. 3.2.1

John Maeda. Artists and scientists: More alike than different. - scientific american. `https://blogs.scientificamerican.com/guest-blog/artists-and-scientists-more-alike-than-different/`, 2013. visited on 18.11.17. 4.2.1.2.3

W. Mann and S. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Journal of Advertising Research*, 1988. ISSN 1613-4117. doi: 10.1515/text.1.1988.8.3.243. 1, 2.2

Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of english: The penn treebank. *Comput. Linguist.*, 19(2):313–330, June 1993. ISSN 0891-2017. URL `http://dl.acm.org/citation.cfm?id=972470.972475`. 2.2, 3.2.4

André Martinet. *Elements of General Linguistics*. Faber and Faber Ltd., London, 1960. ISBN 978-0571058167. 2.1.3

Brett McKay and Kate McKay. Classical rhetoric 101, 2010. URL `http://www.artofmanliness.com/2010/11/14/classical-rhetoric-101-an-introduction/`. Accessed on 14.08.2017. 2.1.1, 2.1.2

Merriam-Webster.com. Anaphora, a. URL https://www.merriam-webster.com/dictionary/anaphora. Accessed on 09.09.2017. 2

Merriam-Webster.com. Oxymoron, b. URL https://www.merriam-webster.com/dictionary/oxymoron. Accessed on 09.09.2017. 5

Merriam-Webster.com. Sophist, c. URL https://www.merriam-webster.com/dictionary/sophist. Accessed on 18.09.2017. 1

Wolfgang G. Müller. Style. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2006. URL http://www.oxford-rhetoric.com/entry?entry=t223.e237. 2.1.1, 3.2

K.R. Murphy and B. Myors. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests, Fourth Edition*. Lawrence Erlbaum Associates, Inc., 1998. 4.2.1.1

Richard Nordquist. Passive Voice Definition and Examples. https://www.thoughtco.com/passive-voice-grammar-1691597, 2017. visited on 30.10.17. 3.2.4

Mary Jo Nye. Speaking in tongues: Science's centuries-long hunt for a common language. *Distillations*, pages 40–43, 2016. 2.1.3

Daniel J. O'Keefe. Persuasion. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2006. URL http://www.oxford-rhetoric.com/entry?entry=t223.e183. 2.1.2

Constantin Orăsan. A hybrid method for clause splitting. In *In Proceedings of ACIDCA 2000 Corpora Processing*, pages 129–134, Monastir, Tunisia, March 2000. 2.2

Henry Peacham. *The Garden of Eloquence*. Scholars' Facsimiles & Reprints, Inc. 1977, Delmar, New York 12054, 1593. 3.2.1, 3.2.1, 7, 3.2.2, 3.2.3, 3.2.3, 3.2.3, 3.2.3

Heiner Peters. Epanalepsis. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2007. URL http://www.oxford-rhetoric.com/entry?entry=t223.e81. 3.2.3

Heinrich F Plett. Die Rhetorik der Figuren. Zur Systematik, Pragmatik und ästhetik der locutio. pages 125–165, 1977. 2.1.1

Heinrich F Plett. Zeugma. In Thomas O. Sloane, editor, *Encyclopedia of Rhetoric*. Oxford University Press, February 2007. URL http://www.oxford-rhetoric.com/entry?entry=t223.e258. 3.2.5

Puttenham G. Puttenham, R. The arte of english poesie. Cambridge England: University Press, 1970. 3.2.3

Marcus Fabius Quintilianus. Institutio oratoria. volume 3. Harvard University press, 1921. translated by Butler, H.E. 2.1.1

Robin Raskin. Hillary clinton's acceptance speech as seen by the algorithms. the huffington post. `https://www.huffingtonpost.com/robin-raskin/hillary-clintons-acceptan_1_b_11265336.html`, 2016. visited on 18.11.17. 4.2.2, 4.2.2, 5

Erwin Reifler. The solution of mt linguistic problems through lexicography. In *Proceedings of the National Symposium on Machine Translation*, February 1960. 2.1.3

T. Daniel Richards. Superlative rhetoric: The first, best, only post you'll ever need to read. period. `http://www.danieltrichards.com/superlative-rhetoric-the-first-best-only-post-youll-ever-need-to-read-period/`, 2014. visited on 24.10.17. 3.2.4

Evan Sandhaus. The new york times annotated corpus ldc2008t19. dvd. *Philadelphia: Linguistic Data Consortium*, 2008. 4.1

Sandra Schlotzhauer. *Elementary Statistics Using JMP*. SAS Publishing, 2007. ISBN 1599943751, 9781599943756. 4.2.1.1

Efstathios Stamatatos. A survey of modern authorship attribution methods. *J. Am. Soc. Inf. Sci. Technol.*, 60(3):538–556, March 2009. ISSN 1532-2882. doi: 10.1002/asi.v60:3. URL `https://doi.org/10.1002/asi.v60:3`. 2.1

Claus W. Strommer. *Using rhetorical figures and shallow attributes as a metric of intent in text*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 2011. 1, 2.2, 3.2, 3.3.3, 3.3.3

Erik F. Tjong and Kim Sang. Memory-based clause identification. In *In Proceedings of CoNLL-2001*, Toulouse, France, 2001. 2.2

J Dekker F W Zoccali C Tripepi G, Jager K. Stratification for confounding – part 1: The mantel-haenszel formula. 2010. 2

Henning Wachsmuth. Text analysis and feature extraction in aitools 4 based on apache uima. *Webis Group*, 2015. 3.1.2, 3.2

A. P. Wadsworth. Machester guardian, 1950. 3.2.4

Jed Webb. Conditional sentences in english 5. `https://www.usingenglish.com/articles/conditional-sentences-in-english-5.html`, 2012. visited on 28.10.17. 3.2.4

Anna Wierzbicka. Conditionals and counterfactuals: conceptual primitives and linguistic universals. In Angeliki Athanasiadou and Renè Dirven, editors, *On Conditionals Again*, pages 16–59. Amsterdam/Philadelphia: John Benjamins, 1997. 3.2.4, 3.2.4