Collocation in Rhetorical Figures:

A Case Study in Parison, Epanaphora and Homoioptoton

Major Research Project for

Katherine Tu

Department of English

University of Waterloo

June 19, 2019

Acknowledgements

I would like to thank my major research project readers, Randy Allen Harris and Michael Macdonald, for their time.

I would also like to thank James Java for answering my questions about his dissertation; the RhetFig group for being a source of inspiration; Professor Chrysanne Di Marco, who introduced me to the group; Sam Pasupalak for funding the DiMarco Graduate Scholarship in Computational Rhetoric, which in turn helped fund my research; and my good friends, Murphy Berzish, Sylvie Davies and Emily Smith for their encouragement and support.

I cannot overstate my gratitude to Professor Randy Allen Harris, without whose tireless efforts and endless support this project—or any of my projects—would not exist.

Contents

1 Introduction

1.1 Rhetorical figures

The study of rhetorical figures dates back to Aristotle. Rhetoric was—and still is—a rich field with many subfields, with different subfields taking the spotlight at different points in history. Today, figures hardly play a role in rhetorical studies. In the 1500s, however, the study of rhetoric was reduced to the study of elocution, "the study of ornate forms of language," and, from that, the study of figures (Perelman, 1982). "In this way," says Perelman, "classical rhetoric came into being—this rhetoric of figures which led progressively from the degeneration to the death of rhetoric" (Perelman, 1982). In *The Realm of Rhetoric*, he argues that scholars of logic undervalue rhetoric despite its role in argumentation and logic in Antiquity. Perelman proposes a New Rhetoric, which attempts to grant rhetoric as much—or even more—importance than it had in ancient traditions, making it once again relevant to modern scholarship. He does so by first insisting upon the importance of rhetoric in reasoning, in which figures play an important role, and develops a framework for understanding audiences, the latter of which is an important part of rhetorical studies today. In fact, Smith, the President of RhetCanada, defines rhetoric as "the art and study of acts of communication: their forms, strategies, ethics, and effects on various audiences" (n.d.). Clearly, Perelman's New Rhetoric accomplished its goal in reviving rhetoric from the grave of rhetorical figures.

Nowadays, rhetorical figures are largely left to the realm of literary analysis, and when studied at all, the focus is mostly on tropes (semantic figures) like metaphor. However, it is important to analyze figures beyond their aesthetic value in literature, because rhetorical figures are argumentative, persuasive and cognitive, and can "epitomize fundamental mental operations" (Fahnestock, 1999). There is no zero-degree language, or language without figures, and they "do not belong in a separate domain of language use but are rather on a continuum of constructions that are successful in varying degrees as epitomes of their functions" (Fahnestock, 1999). If figures are inseparable from language, then figures travel wherever language travels. Fahnestock points out that "[figures] are typically used in the process of conveying the core of an argument in research articles and in versions of these articles constructed for wider audiences" (2004 qtd. in Harris and Di Marco, 2017).

In fact, she shows this in her book, *Rhetorical Figures in Science*, which includes a study on how the figures incrementum, gradatio and climax are used in evolutionary arguments from Darwin, and how antimetabole is used in the expression of various physical laws, such as Newton's third law of motion and Faraday's laws on electromagnetism. McQuarrie and Mick (1996) study advertisements in terms of rhetorical figures. Rhetorical figures have been utilized in computerized natural language tasks, such as document summarization (see Alliheedi and Di Marco, 2014), analysis of intent (see Strommer, 2011) and authorship attribution (see Java, 2015). Rhetorical figures are used in every domain of research but have seldom been studied outside of literature. Because they are inseparable from natural language, rhetorical figures deserve to be studied—not to the extent that every other subfield of rhetoric is abandoned, but to the extent that the theory of figures becomes a study in its own right.

1.2 Collocation

1a. Ask not what your country can do for you; ask what you can do for your country. (Kennedy and Sorensen, 1961)

1b. Do not ask what your country can do for you, but ask how you can help your country.

What is it about Kennedy's quotation in 1a. that is inarguably more elegant, concise, and memorable than 1b., even though the sentences evoke the same content? Both are examples of antimetabole (reverse repetition of words, with alternating word groups *your country* and *you*), but something about the second quotation is certainly lacking. As it turns out, the second quotation is lacking a number of things, namely: mesodiplosis (medial repetition, as in *what* and *can do for*), epanaphora (initial repetition, as in *ask*), and closer syntactic and prosodic structure between the two phrases. We attribute the characteristics of Kennedy's quotation with collocation, the phenomenon in which multiple figures appear in the same passage.

In her 2005 chapter situating the study of rhetoric in the field of cognitive science, although she does not mention figures explicitly, Fahnestock demonstrates the rhetorical effects of the collocation of figures of parallelism. She gives the following examples, in which 2a. contains the same syllable length (a prerequisite to the figure *isocolon*), 2b. has both syllable length and syntactic pattern (*parison*), and 2c. syllable length, syntactic pattern and semantic domain:

2a. The blue jays chased the finches in the trees.
When the wind turned, the temperature fell.

2b. The blue jays chased the finches in the trees.
The dentist pulled the tooth in her office.

2c. The blue jays chased the finches in the trees.
The sparrows woke the robins in the bush.

The more characteristics the passages have in common, Fahnestock argues, the more "tendency [we have] to group them and to have them perform the same discourse function" (2005). She argues that the similarities also have an effect on the level of cognition: the more similar the phrases, the more redundancy we have in neural processing when we read them.

1.3 Figures of repetition

All of the figures in our study leverage the cognitive affinity of repetition. Rhetorical figures "get their salience, their memorability, and their aesthetic effects because of [these] human neurocognitive affinities" (Harris and Di Marco, 2017). Cognitive affinities are called such because they act as basic neurocognitive functions. We have an affinity to certain patterns of thought because these patterns are mirrored on all levels of mind-brain activity, from simple neural firing patterns to all aspects of cognition.

Similarities (a prerequisite to repetition) between phrases creates "more redundancy, and presumably efficiency, in their consecutive construal or processing in the brain" (Fahnestock, 2014). Our ability to distinguish different events or to update information is dependent on the repetition or reuse of a brain's input signals. In fact, a particular neural firing pattern will potentially be reused under "sufficiently similar circumstances" (Gregg, 1984). Hence, the repetition of words, word groups or some linguistic attribute of a text can activate a repetition of neural firing patterns. Fahnestock (2005) reports that "psychologists have recorded quicker recognition time[s] for highly constrained, predictable sentences," empirically confirming what Burke calls *collaborative expectancy* (1950). Fundamental to a predictable sentence is, as Fahnestock points out, some sort of similarity or redundancy in its form. Similarity and redundancy create familiarity, and familiarity is "a close relative of repetition" (Chien and Harris, 2010). We can predict the ending of a story if we have encountered similar ones; in other words, if we have received *repeated* exposure to a genre of story. We can predict that what comes at the end of the saying, "Readers don't need to write but writers do need to. . . " is *read* because of the repetition of the root words *read* and *write,* the suggestiveness of its overall chiastic symmetrical form (which involves the reversal—and therefore repetition—of terms), and the rhyme of *need* and *read* (a repetition of the final syllable in a word). The so-called rhyme-as-reason effect has clear empirical support. People routinely judge rhyming statements to be more accurate, more trustworthy and truer than non-rhyming paraphrases (see McGlone and Tofighbakhsh, 2000).

Familiarity and its relative repetition accomplish several things, both cognitively and functionally, allied with formal assent. In a classic passage from *Rhetoric of motive*s, Burke gives an example of a series of semantically vacant antitheses, built on proforms and deictics: "we do this, but they on the other hand do that; we stay here; but they go there; we look up, but they look down" (1950). He states that "you will find yourself swinging along with the succession of antitheses" and that:

> a yielding to the form prepares for assent to the matter identified with it. You might have no interest in that associated matter. You may even reject it. But on the level of purely formal assent you would collaborate to round out its symmetry by spontaneously willing its completion and perfection as an utterance. . . .assent on the formal level invites assent to the proposition as doctrine (Burke, 1950).

This last phrase is effectively the rhetorical equivalent of the rhyme-as-reason effect: "assent on the formal level invites assent to the proposition." Or, more precisely, rhyme-as-reason is a subtype of formal assent. And notice, too, that there is a lot of form with which to assent in Burke's passage. Burke only names antithesis, but there are several figures in his example, including alliteration, assonance, epanaphora, mesodiplosis, parison and isocolon (Harris, 2013). Take another example that makes use of repetitive structure, including parison and epanaphora:

3. Who is here so base, that he would be a bondman?

   If any, speak; for him I have offended.

   Who is here so rude, that would not be a Roman?

   If any, speak; for him I have offended.

   Who is here so vile, that will not love his country?

   If any, speak; for him I have offended. (Shakespeare, 1623)

In this case, there is a lot of repetition of word groups on top of phonological and syntactic repetition. By the third question-conditional pair, readers are familiar enough with the structure that upon hearing, "Who is here so...," that they will expect to hear shortly, "If any, speak; for him I have offended." The repetition also acts as a "piling on" of accusations, that to disagree would be to be "a bondman," not "a Roman," or to "not love his country." The ease of processing from repetition, therefore, makes us yield to the form and to want to agree that speaking out would be "base," "rude" and "vile."

1.4 Outline of study

In this study, we look at the ploce suite of figures, as well as some others. Ploce is a figure of simple lexical repetition. Under our definition, ploce can encompass single words or word groups. The following definitions are adapted from the Rhetoricon, discussed in depth in Section 2.1.1. From the ploce suite, we have:

- Epanaphora (sometimes known as anaphora), the repetition of words at the beginning of a phrase;
- Epiphora (sometimes known as epistrophe), the repetition of words at the end of a phrase;
- Symploce, a compound figure containing both epanaphora and epiphora;
- Anadiplosis, the repetition of words at the end of one phrase and at the beginning of the subsequent phrase;

- Antimetabole (sometimes known as chiasmus), the reverse repetition of words or word groups;
- Polysyndeton, the repetition of conjunctions and disjunctions; and
- Ploce, which encompasses all of the figures above, but also includes any repetition of words.

We would have liked to have included mesodiplosis, the repetition of words in the middle of a phrase, in the list, but we only found one instance among the sources we used.

We also look at:

- Asyndeton, the "opposite" of polysyndeton, or the lack of conjunctions and disjunctions separating phrases;
- Polyptoton, the repetition of word stems;
- Homoioptoton (sometimes known as homeoteleuton), the "opposite" of polyptoton, or the repetition of derivational affixes;
- Alliteration, the repetition of phonemes at the beginning of words; and
- Parison, the repetition of syntactic structure between two or more phrases.

In order to study the figures in Section 1.3, we first gather "prototypical" examples of them. We define prototypical examples as those that accompany definitions of figures in figure dictionaries and handbooks. That is, these instances clearly illustrate the form and function of the figure. For example, in Vickers, 11. appears under the definition of parison.

From the list of figures above, we then isolate three: parison, epanaphora and homoioptoton. We study these figures in some depth and refine their definitions to eliminate ambiguities in meaning. We develop detection algorithms for these three figures and run them on our data of prototypical examples gathered from the previous step, both as a proof-of-concept and to speed up the detection process. We analyze our results to see which of these three figures, if any, collocate with the figure associated with the examples, and formulate hypotheses based on our findings.

2 Background

2.1 Resources

Our detection algorithms make use of the Stanford CoreNLP library (Manning et al., 2014). The CoreNLP library provides a part-of-speech or grammatical category tagger, which splits text into sentences and provides the part-of-speech tag for each of the words in the sentence. The Stanford CoreNLP tagger uses Penn Treebank tags, available in Appendix A. The library also provides a syntax tree generator, which creates a syntax tree for a sentence alongside tagging each word.

2.1.1 The Rhetoricon

The Rhetoricon, maintained by the University of Waterloo RhetFig group, is a database containing thousands of annotated instances of rhetorical figures. Alongside these are the definitions, etymologies, forms and synonyms of hundreds of rhetorical figures. Because the history of rhetorical figures is rife with "idiosyncrasy, intellectual gallimaufry, and academic ideologies," a goal of the Rhetoricon is to provide consistency with its definitions (Harris, 2018). It attempts to provide a one-to-one mapping between rhetorical figures and definitions. Hence, these definitions are drawn from the wisdom of the tradition, but do not attempt to to accommodate all variations when a rhetorical figure has multiple definitions, solving the one-to-many problem. For example, the figure *parison* is sometimes referred to as a figure in which proximal phrases or clauses have identical structure, presumably syntactic structure, and sometimes referred to as a figure in which phrases or clauses have the same prosody. The Rhetoricon chooses the former, since there is another term for the latter (*isocolon*). Furthermore, it attempts to consolidate all synonyms under one name, typically the name under which it is most common, solving the many-to-one problem. For example, along with *isocolon* being often used interchangeably with *parison*, instances of the same prosody are also known as *compar*, so the Rhetoricon redirects users to isocolon when *compar* is searched. Etymologies are an important entry alongside definitions, because some terms are Greek, some are Latin and some are mixed neologisms developed by RhetFig to refer to never-yet-studied rhetorical figures.

Etymologies also often give insight to the form of a figure. For example, *epanaphora* means "upon, in addition" (*epi*), "up, again" (*ana*) and "to bring, carry" (*phero*, all Greek), and it refers to the repetition of the first element of a phrase or clause. The form of each figure is given in an annotation scheme developed by Harris and Di Marco (2009). Epanaphora (called *anaphora* in their paper) is given by $< [W]_a \ldots >< [W]_a \ldots >$, given phrasal and clausal boundaries, $<$ and $>$, and lexeme boundaries, [and]. (Note that, under our definition, each phrase or clause can begin with any number of words, so our revised annotation scheme would be $<< \ldots >_a \ldots ><< \ldots >_a \ldots >$.) The Rhetoricon attempts to provide a formal notation for every rhetorical figure, but especially schemes or formal figures. These notations correspond to the annotated instances, so that in an example for epanaphora such as 1a., *ask* is highlighted, the first occurrence labeled A1 and the second one labeled A2. These definitions are adopted in this study.

2.2 Collocation in other figure detection work

Unfortunately, outside of some exceptions, rhetorical figures are often neglected by rhetoricians and, outside of tropes, rhetorical figures are largely neglected by the field of natural language processing in computer science. Furthermore, the collocation of rhetorical figures is a relatively new concept in figural logic and has not yet been studied.

In his annotation tool JANTOR, Gawryjołek (2009) works with a large set of figures, including many that we will be looking at in our study. Java (2015) builds on this research, improving on the accuracy of figures more difficult to detect. Both of the tools developed by Gawryjołek and Java allow for more than one rhetorical figure to be found within a single passage, but otherwise study each figure individually. Both authors also provide examples of figures alongside their definitions, and many of their examples, like 1a., include the collocation of multiple figures without commenting on this collocation.

Hromada (2011) works on figures of repetition in isolation, and in particular builds on an antimetabole detector using regular expressions. Antimetabole is a figure of reverse repetition, such as in 1a. This antimetabole detector, in order to reduce false positive results, looks for the pattern $ABCCBA$ or the reverse repetition of three elements instead of the reverse repetition of two elements under the traditional definition. Note that, in both parts of the pattern, B is sandwiched between A and C—which is in fact mesodiplosis (medial repetition) collocating alongside antimetabole.

Strommer (2011) restricts his research to epanaphora, allowing a closer examination of one figure "instead of a shallow and superficial examination of numerous figures". He gives an example of an accidental epanaphora as one that collocates with antithesis, and surmises that antithesis is the "intended" figure, whereas the repetition in epanaphora is simply a by-product of antithesis. In his research, he measures the purposefulness or intentionality with which an author uses epanaphora, based on the number of words in the repetition, the number of repetitions that make up the epanaphora, the size of gaps between each repetition, and the grammatical category of words repeated. In his conclusion, he recommends that future studies also consider the "homogeneity of sentence length"; that is, the similarity of lengths between sentences containing epanaphora. A similarity in sentence length is prerequisite to figures of parallelism like isocolon and parison, which suggests that "intentional" epanaphora may collocate frequently with these figures.

Dubremetz (2017) studies antimetabole using a machine learning algorithm that rates examples based on "prototypicality." However, Harris (2018) has argued that her definition of prototypicality is in fact the collocation of multiple rhetorical figures. It is the collocation of antimetabole with other figures that makes it symmetric (as in 1a. and 4., seen as more "prototypical" in Dubremetz's study), whereas antimetabole in isolation need not be (as in 1b. and 5.).

4. There are only two kinds of men: the righteous who think they are sinners and the sinners who think they are righteous. (Dubremetz, 2017)

5. You hear about constitutional rights, free speech and the free press. Every time I hear these words I say to myself, 'That man is a Red, that man is a Communist!' (Dubremetz, 2017)

Dubremetz programs her algorithm to give higher rating to antimetabole that has "perfectly symmetrical switch[ing] of syntactic roles"; in 4., "righteous is a subject, sinners a complement in the first colon; then the roles switch in the second" (Dubremetz, 2017; Harris, 2018). The switching of grammatical roles means that they maintain the same syntactic structure, which is the definition of the figure parison. The algorithm also scores based on the “detection of recurrent lexical patterns,” including “the presence of negation underlying a contrast,” which again is another rhetorical figure—in this case, antithesis (Dubremetz, 2017). Harris points out that antithesis frequently collocates with antimetabole, and we can see an example of this in 1a., where *ask* and *ask not* are antithetical.

3 Method

Rhetorical figures, according to RhetFig and its Rhetoricon, can be characterised by their linguistic domain and cognitive affinities. Different linguistic domains effect different neurocognitive functions. For example, a phonological figure will involve different neurocognitive processes than a syntactic one, and both will involve different processes than a semantic one. Some figures can be constrained to relatively small domains, like morphemes in a single sentence, whereas others may be much larger, spanning an entire discursive text. In this study, we focus on three different schemes: parison, epanaphora and homoioptoton. We choose three figures that manifest in different linguistic domains: parison deals with syntactic structure, epanaphora with word groups, and homoioptoton with morphemes. Because morphemes are always embedded within words, we can say that parison and epanaphora have larger scopes than homoioptoton. While their linguistic domains differ, the cognitive affinity underlying our figures of study do not—they are all figures of repetition. With this, we hope to study schemes controlling for cognitive affinity, but varying in linguistic domain. The domains of each of the figures we study are laid out in Table 1.

We create an automatic detection algorithm to detect figures of parison, epanaphora and homoioptoton in instances of the figures listed in Section 1.4 from figure dictionaries and handbooks. In order to do this, we needed to precisely define these figures and vague notions like “phrases.” We test these algorithms on instances from the Rhetoricon.

In total, we gathered 187 instances from figure dictionaries and handbooks for figures listed in Section 1.4. We reference *Ad Herennium*, Bain, Blount, Bullinger, Burton, Christiansen, De Mille, Demetrius, Gibbons, Holmes, J.G. Smith, Johnson, Epp, Macbeth, Norwood, Peacham, Raub,

Figure Name	Number of Instances	Linguistic Domain
alliteration	6	phonological
anadiplosis	16	lexico-syntactical
antimetabole	28	lexico-syntactical
asyndeton	23	lexico-syntactical
epanalepsis	18	lexico-syntactical
epanaphora	15	lexico-syntactical
epiphora	20	lexico-syntactical
homoioptoton	6	morphological
ploce	10	lexical
polysyndeton	15	lexico-syntactical
polyptoton	8	morphological
parison	8	syntactical
symploce	12	lexico-syntactical
Total	187	

Table 1: Number of instances of each figure collected from figure dictionaries and their linguistic domains.

Vickers and Vinsauf.

Table 1 shows the distribution of figures.

3.1 The detector

The detector reads passages from a plaintext file. Each passage can be delimited by a line break (for single-sentence passages) or placed into double quotes (“ ”, for multi-sentence passages). We use the Stanford CoreNLP parser to generate syntax trees for our phrases, which uses Penn Treebank part-of-speech tags (see Appendix A). We apply the following equivalence classes for our detector, adapted from Gawryjołek (2009). This means that the following tags are treated as if they were the same part-of-speech:

- Adjectives: JJ (adjective), JJR (adjective, comparative), JJS (adjective, superlative)
- Nouns: NN (noun), NNS (noun, plural), NNP (proper noun), NNPS (proper noun, plural)

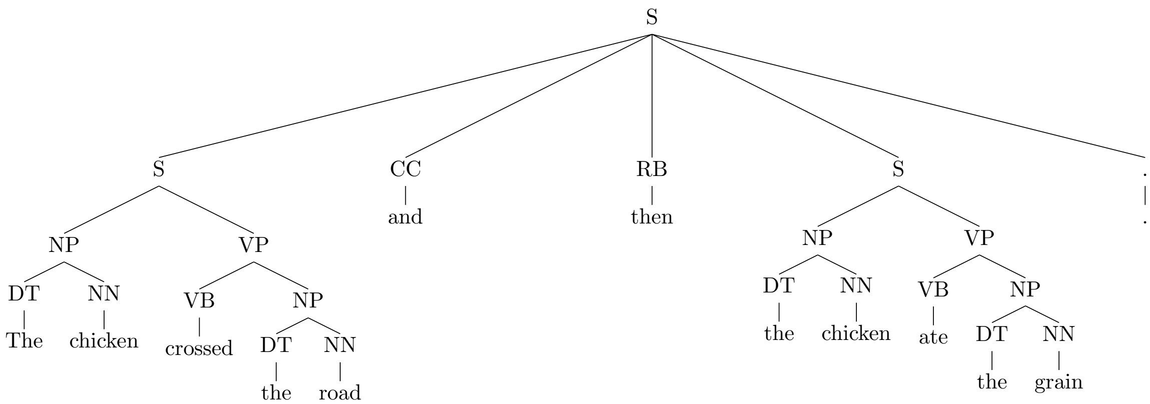

Figure 1: Syntax tree for 6a.

- Adverbs: RB (adverb), RBR (adverb, comparative), RBS (adverb, superlative), WRB (wh-adverb)
- Verbs: VB (verb), 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)
- Pronouns: WP (wh-pronoun), WP$ (wh-pronoun, possessive), PRP (pronoun), PRP$ (pronoun, possessive)

The detector uses the syntax trees to find parison and epanaphora, and the grammatical category or part-of-speech tags (without the syntactic structure) for homoioptoton. Within the parison- and epanaphora-detecting algorithms, the syntax trees' branches are compared at each height. The height is the maximum distance from the leaves (where $[\text{The}]_{DT}$, $[\text{chicken}]_{NN}$, etc. in Figure 1 are all leaves with a height of zero). All of the Noun Phrases (NPs) in Figure 1 have a height of one, since they are all one level away from the leaves. The two Verb Phrases (VPs) have a height of two, since they are one from the VB leaves and two from the NPs, and we take the maximum of one and two. This was done especially with parison in mind, since, for example, a branch with a height of two could not possibly have the same structure as a branch with a height of one.

For parison, a branch must have the exact same structure as the other branch with which it is

```
========== PARISON, EPANAPHORA, AND HOMOIOPTOTON ==========

================== PARISON AND EPANAPHORA =================
 The chicken crossed the road and then the chicken ate the grain .

================= PARISON AND HOMOIOPTOTON ================

=============== EPANAPHORA AND HOMOIOPTOTON ===============

====================== PARISON ONLY =======================
 The chicken crossed the road . It ate the grain .

==================== EPANAPHORA ONLY ======================
 The chicken crossed the road . The chicken fell .
 The chicken crossed the road . The chicken ate small grains .

=================== HOMOIOPTOTON ONLY =====================

======== NO PARISON, EPANAPHORA, OR HOMOIOPTOTON ==========
 The chicken crossed the road and then ate the grain .
 The chicken crossed the road . It fell .
```

Figure 2: Sample output with input from 6a.–6f.

being matched; for epanaphora, a branch must begin with the same word as the other branch. For example, at some point, the algorithm will compare the two S branches at height four in Figure 1. Since the left S branch has all of the same structure as the right S branch, the algorithm will have found a match for parison. Since the left S branch begins with *the* and the right S branch begins with *the,* the algorithm will also have found a match for epanaphora. The detectors then check that the branches meet the definition of a phrase (outlined in Section 3.2), and if so returns a positive result. The output is another plaintext file, divided into sections: one section for passages containing all three of parison, epanaphora and homoioptoton, three sections for combinations of two figures, three sections for only parison, only epanaphora or only homoioptoton, and one section for none of the three figures. A sample of the output (whose input is 6a.–6f.) is in Figure 2.

3.2 Phrases

Two of our figures, parison and epanaphora, use phrases in their definition. However, linguistic phrases can be as small as a single word (such as *he* in the first NP of Figure 11) in any part of a sentence, which, not being restrictive enough, does not align with our traditional notions of parison and epanaphora. We then considered the idea of clauses, which require grammatical completeness. However, clause detection is very difficult, especially due to zeugma, and perhaps too restrictive; requiring grammatical completeness would omit fragment sentences from being detected as any syntactic figure (which include both parison and epanaphora). We noticed early on in our examples that members of a list in a sentence are often parallel to one another, and since punctuation helps delimit separate clauses in a sentence, we decided to make use of it in our definition of phrases. Hence, we define phrases to either:

- contain a VP and a NP, which captures all clauses as well as VPs with direct objects; or
- be a member of a list delimited by punctuation (commas, parentheses, colons, semi-colons, full-stops and conjunctions).

In order to check for punctuation (or conjunctions), the algorithm does one of two things. First, it uses the syntax tree and looks at the leaves surrounding the branch being checked. For example, if the right S branch from Figure 1 was being checked for punctuation, it would look at the left of the right S branch and find $[\text{then}]_{RB}$ and [.]. The branch must be flanked by punctuation or conjunctions on both sides (or be the first branch in a sentence) in order to be part of a list. Second, because we recognize that the syntax tree generator is not always completely accurate, we generate a new tree based purely on punctuation (and conjunctions). An example of a punctuation-based tree is shown in Figure 3.

Like the syntax trees, the branches are checked by height, so all of the branches beginning with "0" are checked against one another; all of the branches beginning with "1" are checked against one another, etc. The parison detector would find a match for the second and fourth branches labeled "3", since *I weep for him* and *I rejoice at it* have the same parts-of-speech tags.

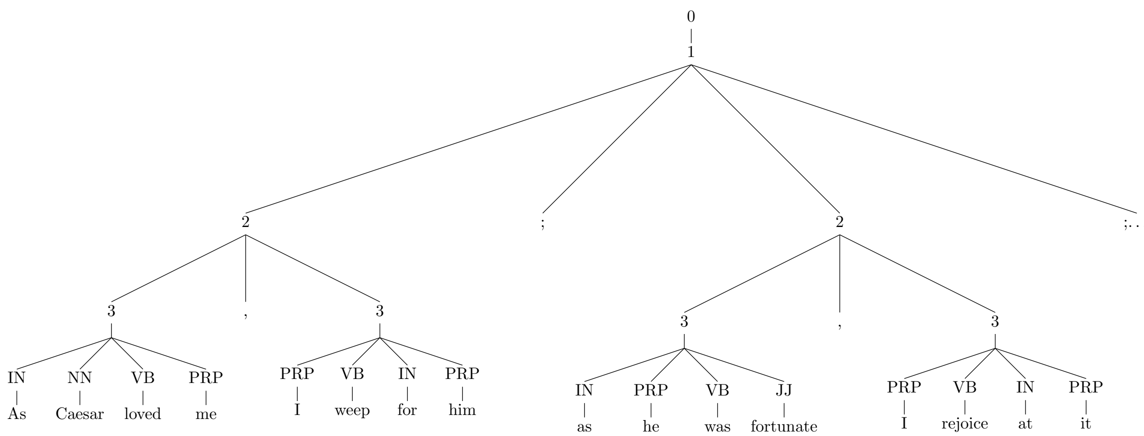

Figure 3: Punctuation-based tree for "As Caesar loved me, I weep for him; as he was fortunate, I rejoice at it;. . . " (part of 11.).

3.3 Parison

The Rhetoricon defines parison as the repetition of syntactic structure. As straightforward as this may sound at first, this definition turned out to be too vague to work with computationally. Should we, for example, consider every Subject-Verb-Object (SVO) sentence to have the same syntactic structure, no matter how the subject, verb and objects are expressed? Do "chickens make clucking noises" and "ducks and chickens are birds" then exhibit the same syntactic structure because they are both SVO sentences? The intuitive answer is, hopefully, no—but while this particular example is simple to diagnose, we can imagine there being many more examples whose answer is much less clear-cut. We therefore need to better define parison and what we mean by "repetition of syntactic structure," and we can begin by looking at relatively simple sentences like 6a.–6f. (illustrated in Figure 1 and Figure 4–7):

6a. The chicken crossed the road and then the chicken ate the grain.

6b. The chicken crossed the road and then ate the grain.

6c. The chicken crossed the road. It fell.

6d. The chicken crossed the road. The chicken fell.

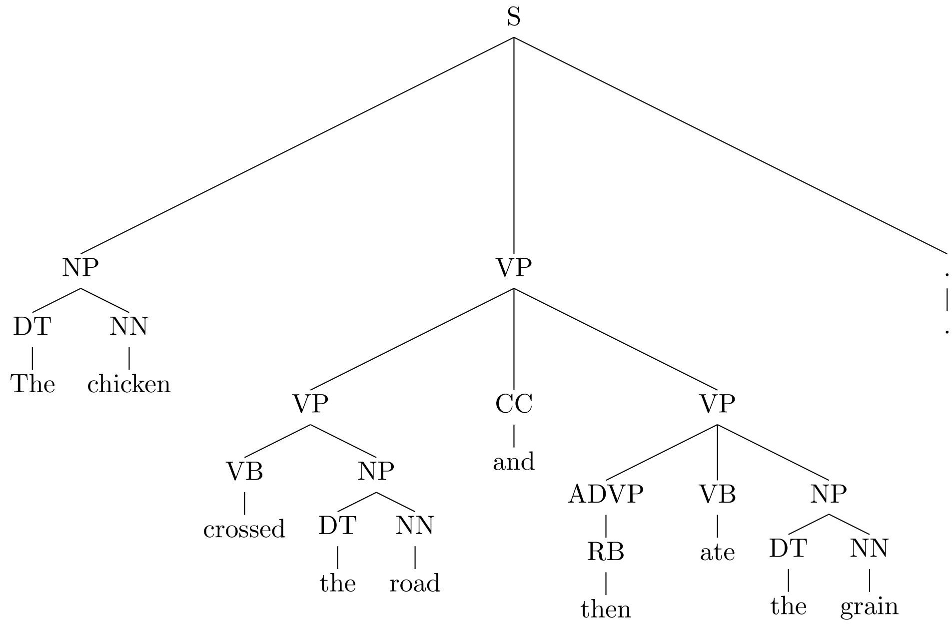

Figure 4: Syntax tree for 6b.

6e. The chicken crossed the road. It ate the grain.

6f. The chicken crossed the road. The chicken ate small grains.

We more precisely define parison to be the exact repetition of branches of syntax trees between two or more phrases. This means that both the branch structures must be the same and that the part-of-speech or grammatical category tags must be the same. That is, 1a. is not defined as a parison because $[[\text{ask}]_V]_{VP}$ and $[[\text{ask}]_V[\text{not}]_{RB}]_{VP}$ do not match, and neither do $[[\text{your}]_{PRP}[\text{country}]_N]_{NP}$ and $[[\text{you}]_{PRP}]_{NP}$. As stated in Section 3.2, phrases are defined as containing both a Noun and Verb Phrase or delimited by punctuation. Under this definition of parison and phrases, and with these equivalences classes, we can label 6a., 6b. and 6e. as parison because of the similarity in structure between *crossed the road* and *ate the grain* (which both contain both a Verb Phrase and a Noun Phrase). 6c. and 6d. do not contain parison, since, in the first sentence, the verb has a direct object, but not in the second. 6f. also does not contain parison, because *small,* an adjective, does not match grammatical category with *the,* a determiner.

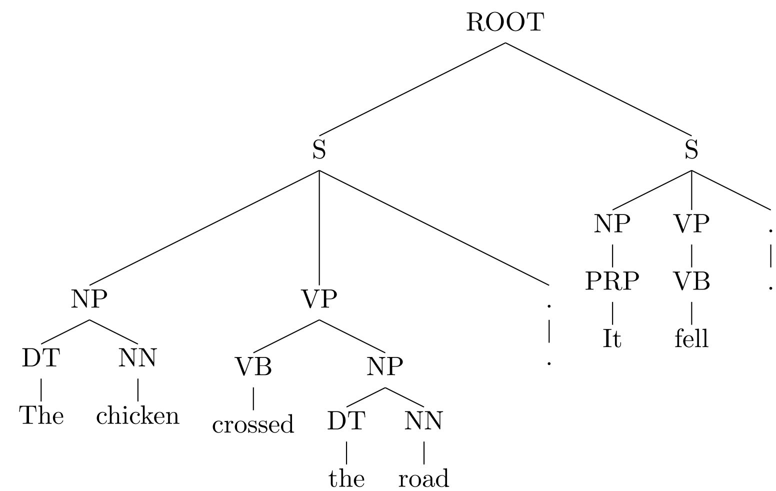

Figure 5: Syntax tree for 6c.

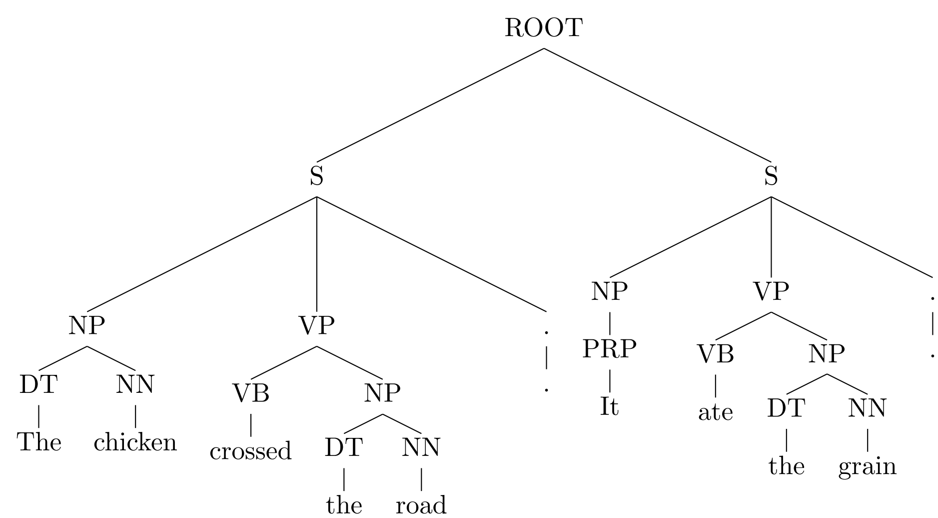

Figure 6: Syntax tree for 6d.

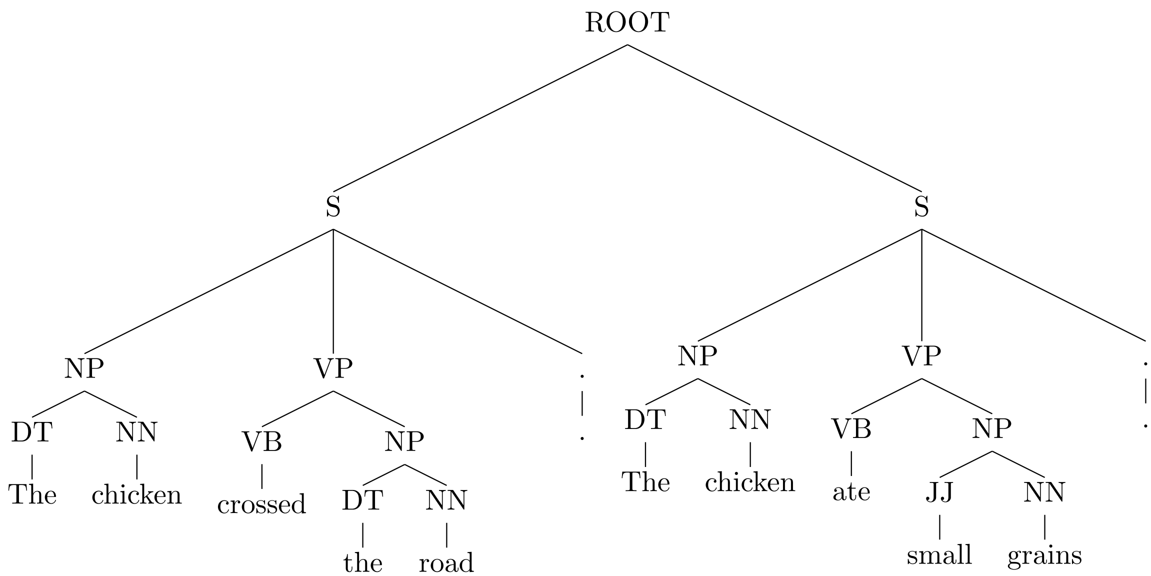

Figure 7: Syntax tree for 6e.

The following figures are positive examples of parison from our working data of prototypical instances. In Figure 8, "He is esteemed eloquent which can invent wittily, remember perfectly. . . " returns a positive result for parison because *invent wittily* and *remember perfectly* are members of the same list with the same grammatical structure. In Figure 9, "What lies behind us and what lies before us are. . . " returns a positive result for parison because *what lies behind us* and *what lies before us* are syntactically equal and grammatically complete by our standards—that is, both phrases contain both a verb and a noun.

Unfortunately, the parser does generate some mistakes, so the following syntax tree returns a false positive result for parison due to the mislabeling of some part-of-speech or grammatical category tags. In Figure 10, "Sins stain thy beautious soul; forsake thy sins" returns a false positive for parison due to the mislabelling of *thy, beautious* and *forsake.* Note that *stain* is also incorrectly labeled as a noun.

The inaccuracy of the parser can also lead to false negatives, both due to the mislabelling of part-of-speech tags and incorrect tree structure. In Figure 11, "Blessed shalt thou be in the city, and

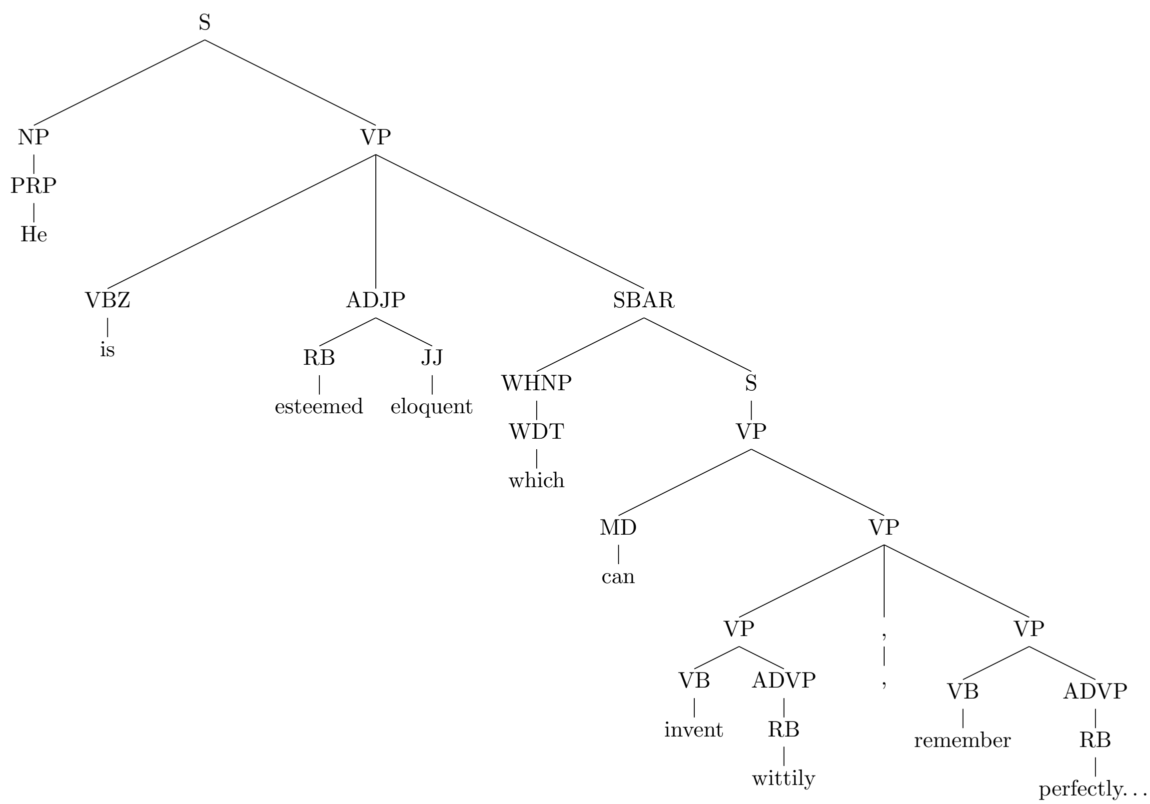

Figure 8: Syntax tree for "He is esteemed eloquent which can invent wittily, remember perfectly. . . " (Peacham qtd. in Christiansen, 2013).

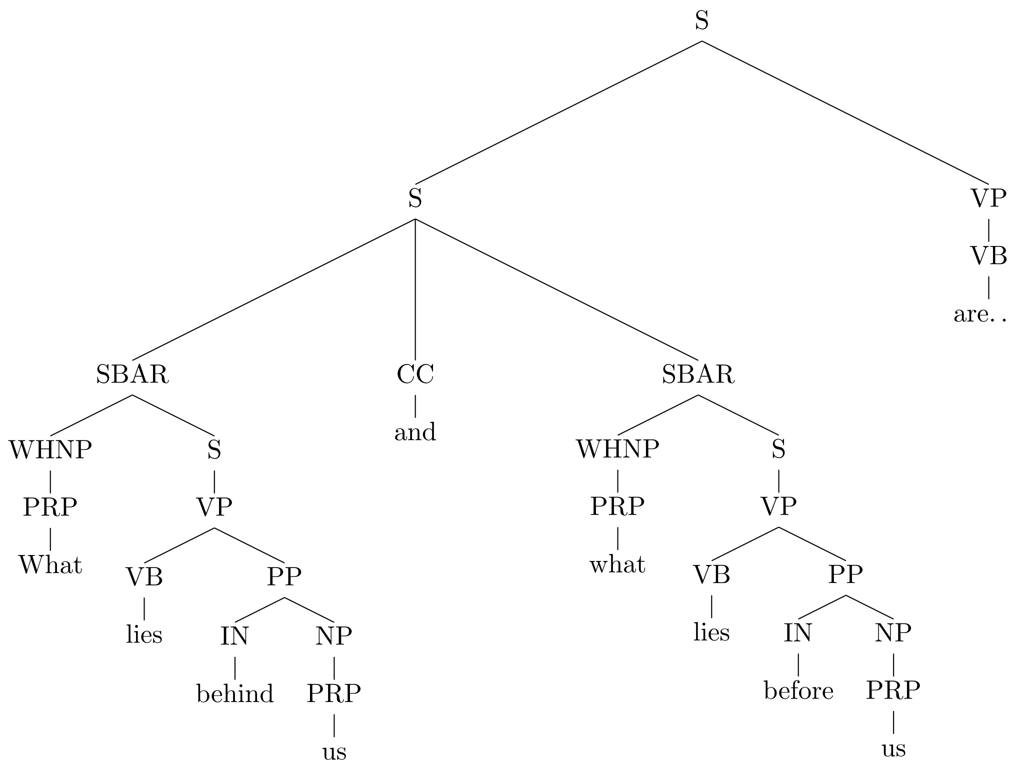

Figure 9: Syntax tree for "What lies behind us and what lies before us are. . . " (part of 10.)

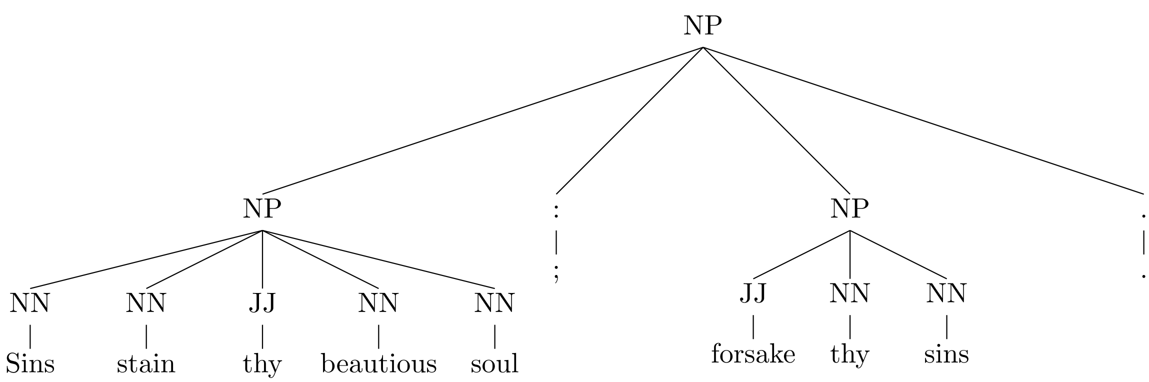

Figure 10: Syntax tree by parser for "Sins stain thy beautious soul; forsake thy sins" (Holmes, 1806).

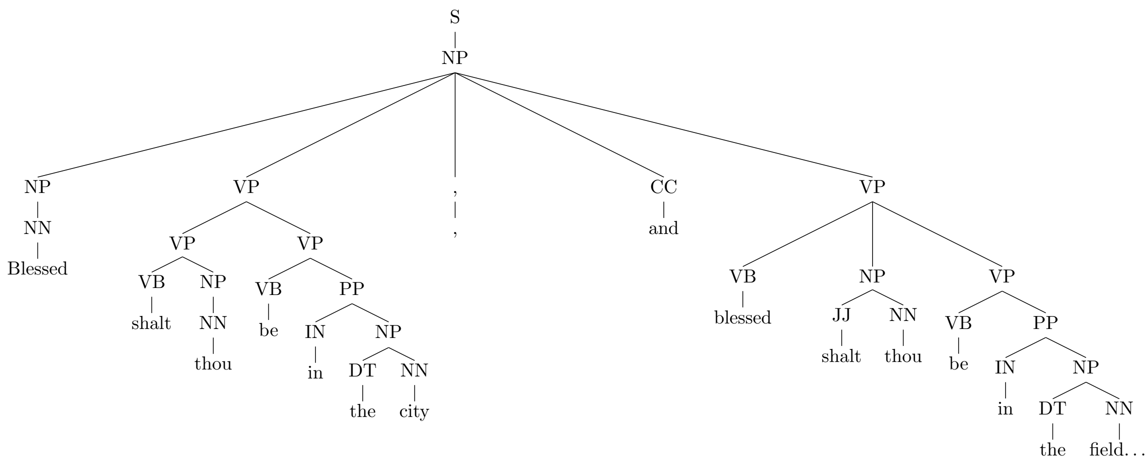

Figure 11: Syntax tree by parser for or "Blessed shalt thou be in the city, and blessed shalt thou be in the field. . . " (Deutoronomy 28:3–6 qtd. in Bullinger, 1898).

blessed shalt thou be in the field" would return a false negative for parison, due to the mislabelling of *blessed* as a noun and then as a verb, and the separation of *blessed shalt thou be in the city* into three different phrases. However, the sentence actually returns a positive result because of the presence of a parison elsewhere in the sentence.

In order to solve the false negative issue like in Figure 11, we thought about using only partial matches for part-of-speech tags instead of looking at the entire structure of a syntax tree's branches. We would then require a minimum length of tags to match in order to return a positive result, since one or two words with repeating grammatical categories would be too short to constitute a phrase (and therefore a parison). For example, we could say that, because "be in the city" and "be in the field" have matching part-of-speech tags, that is enough to warrant a positive parison result. The question then becomes, "What is the smallest length of phrase we could have to both avoid false positives and false negatives?" There is no clear-cut answer to this, and the best way to address this is by statistical methods, which is beyond the scope of this study. Since instances like those in Figure 11 happen very rarely—three times, in fact, in all of our test data—we decided to omit this minimum-length solution from our code.

3.4 Epanaphora

We define epanaphora to be the repetition of a word or word groups at the beginning of two or more phrases. Like with parison, although the definition seems straightforward at first, there are some examples for which the presence of epanaphora could be argued either way. One of the major questions was: what constitutes the "beginning" of a phrase? How many words can we have before the "beginning" of a phrase before it is no longer the beginning? Are there words that are too short to be considered part of an epanaphora? Again, we can begin by looking at examples (7a.–7f.) and decide which ones we want to include in our definition of epanaphora and which we do not:

7a. I have never wished to cater to the crowd; for what I know they do not approve, and what they approve I do not know. (Burton, 2016)

7b. I have never wished to cater to the crowd; this is because what I know they do not approve, and what they approve I do not know.

7c. I have never wished to cater to the crowd; what I know they do not approve, and what they approve I do not know.

7d. I have never wished to cater to the crowd; I know that which they do not approve, and what they approve I do not know.

7e. The scent is so sweet that it's both pleasant and unpleasant; people recoil and go nearer, recoil and go nearer; they're not sure whether to be disgusted or seduced. (Perry, 2016)

7f. When his children had departed, he took up his guitar, and played several mournful, but sweet airs, more sweet and mournful than I had ever heard him play before. (Shelley, 1818)

As with parison, we decided to follow a stricter definition of epanaphora and mandate that, aside from conjunctions and disjunctions, no words can separate the beginning of a phrase and the first word in an epanaphora. Because *what I know they do not approve* constitutes its own phrase in 7a. and 7b. (since *I know* and *they to not approve* both have NPs and VPs), which matches another phrase, *what they approve I do not know*, they both should be considered epanaphora. 7c.

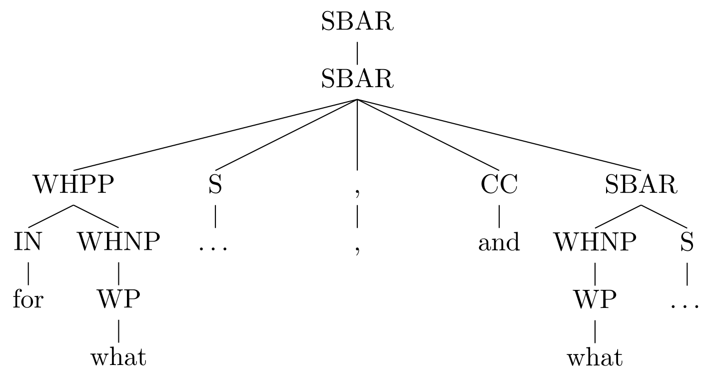

Figure 12: Partial syntax tree by parser for "...for what..., and what..." (part of 7a.).

is an unambiguous example of epanaphora under both definitions of phrases (and every definition of epanaphora). 7d. is an epanaphora because of the repetition of *I,* and we do not consider *I* to be too short of a word to be included in our definition. On the other hand, 7e. should not be an epanaphora, since *recoil and go nearer* does not meet our definition of a phrase and *people* prevents *people recoil* and *recoil* from matching as a phrase delimited by punctuation. Likewise, 7f. also should not be an epanaphora, because although the *but* in *but sweet airs* is not considered the first word of the phrase, *more* in *more sweet* is, preventing a match between the two instances of *sweet.*

Unfortunately, the inaccuracy of the syntax trees for the examples prevents them from being properly categorized. 7c. is correctly labeled enapahora by the detector, but 7a. and 7b. are not and thus produce false negatives. Figure 12, 14 and 16 show partial syntax trees for examples 7a., 7b. and 7c., respectively, as output by the parser. Figure 13 and Figure 15 show the correct partial syntax tree for 7a. and 7b., respectively, where *what I know...* and *what they approve...* are at the same height.

7e. returns a false negative, not because of an incorrect syntax tree by the parser, but because of the punctuation-based tree: *and go nearer[,]* is incorrectly labeled as a phrase because it has punctuation or conjunctions on either side, and is matched with the second *and go nearer[,]* (illustrated in Figure 17). We want to be able to match phrases like "go nearer, go farther and go anywhere" (illustrated in Figure 18), but not multiple phrases like "people recoil and go nearer,

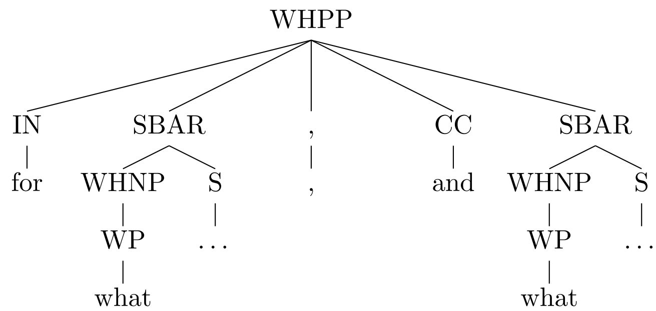

Figure 13: Correct partial syntax tree for "…for what…, and what…" (part of 7a.).

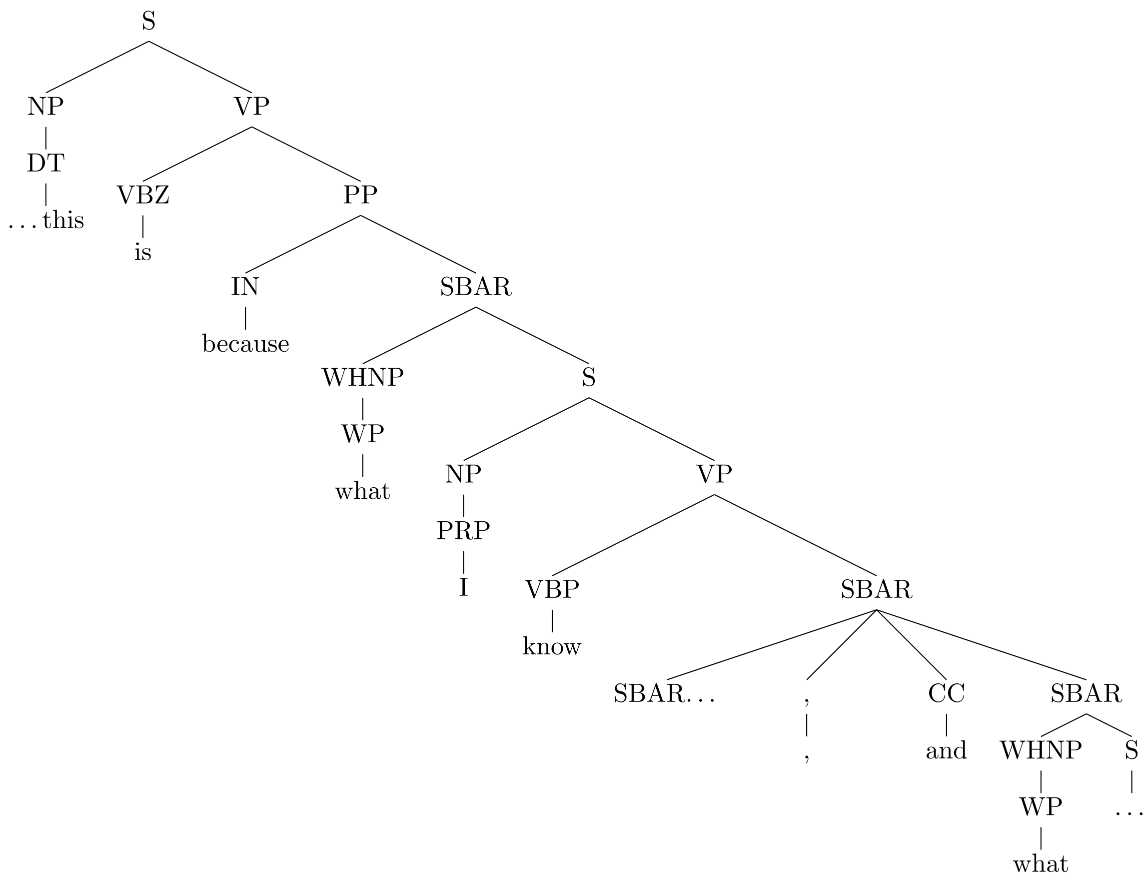

Figure 14: Partial syntax tree by parser for "…this is because what I know…, and what…" (part of 7b.).

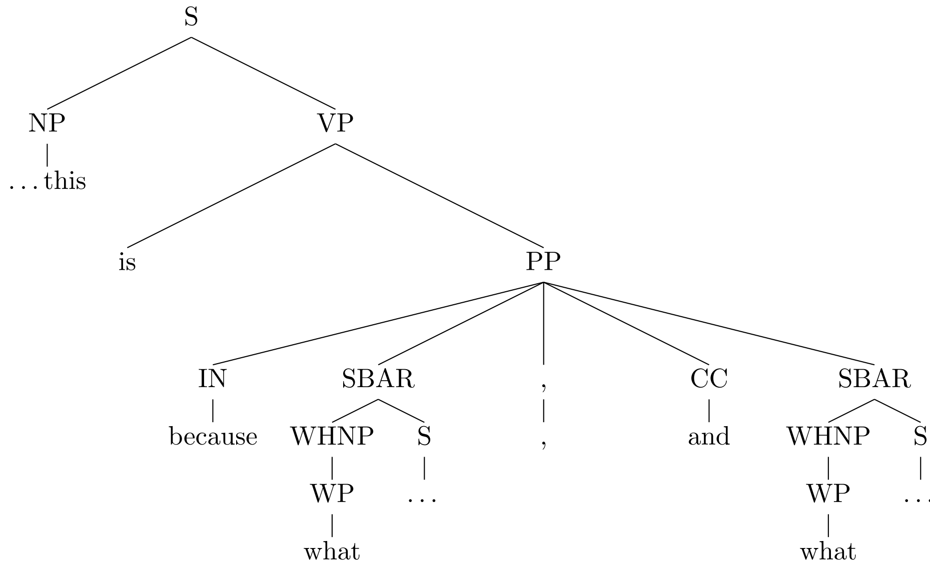

Figure 15: Correct partial syntax tree by parser for "... this is because what I know..., and what..." (part of 7b.).

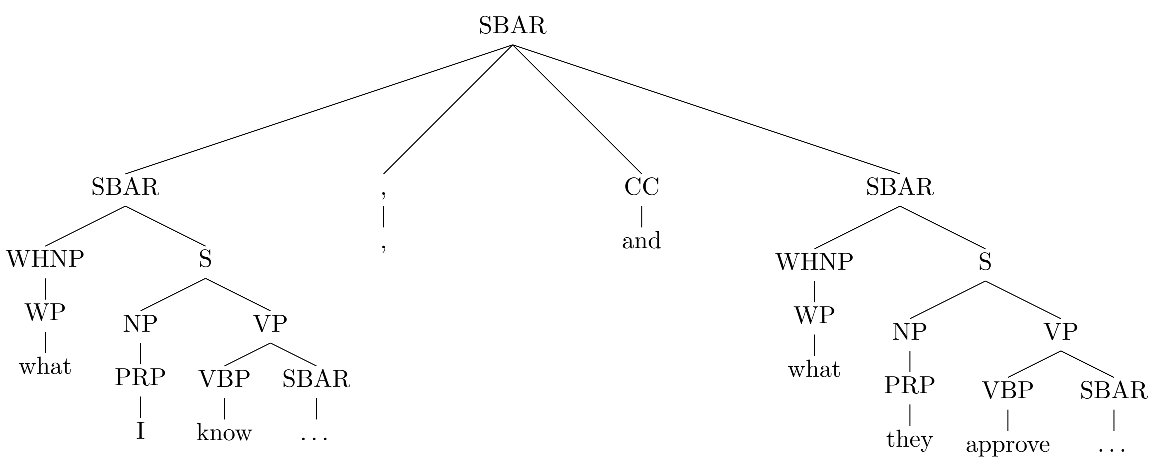

Figure 16: Partial syntax tree for "... what I know..., and what they approve...." (part of 7c.).

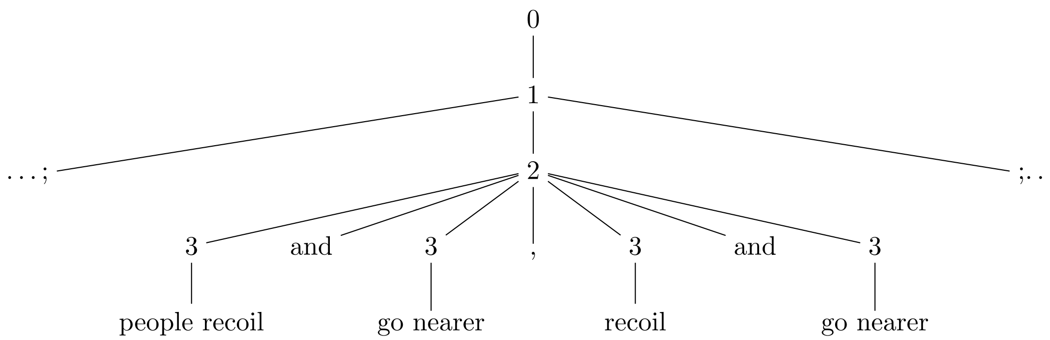

Figure 17: Partial punctuation tree by detector for "…; people recoil and go nearer, recoil and go nearer;…"

recoil and go nearer." The latter is better described by mesodiplosis (phrase-medial repetition) or epiphora (phrase-final repetition); the problem lies in the fact that *and* is more tightly-binding than the comma in latter, as illustrated in Figure 19. In "go nearer…", the *and* has the same level of binding as the comma. One possible fix could be to only include conjunctions and disjunctions (like *and*) as delimiters in a punctuation tree if it is preceded by a comma, which would fix this particular example, but may produce other false positives. We could also forgo conjunctions and disjunctions as delimiters in the punctuation tree altogether.

7f. correctly returns a negative result. We did not find any false negative examples labeled incorrectly because of the syntax tree generator.

3.5 Homoioptoton

We define homoioptoton to be the repetition of morphemes or affixes at the beginning or end of different words. The detector uses regular expressions to check the prefix and suffix of words, and then compares them to a premade list of English affixes. If the affix appears more than once in the text and the affix is in the premade list (see Appendix B), the detector checks that the words containing the affix are of the same grammatical category using the Stanford CoreNLP part-of-speech tagger. For example, for the following instance, the detector finds the repeating suffixes *-able* (from *commendable* and *profitable*) and *-ble* (from all three adjectives). Since *-able* is in our

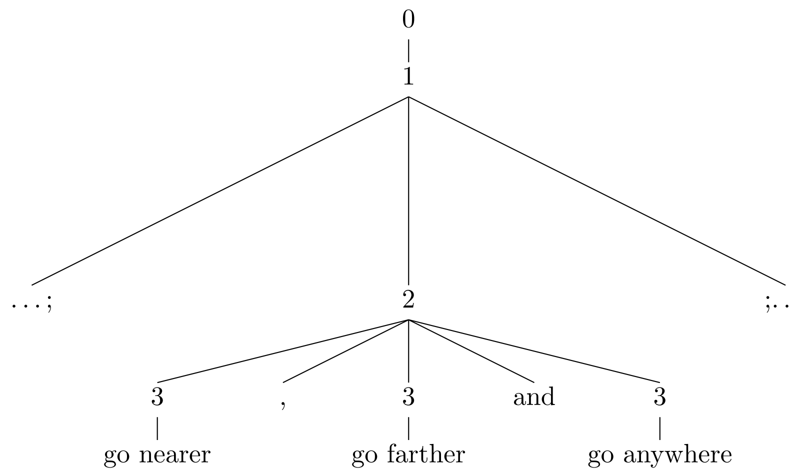

Figure 18: Partial punctuation tree for “. . . ; go nearer, go farther and go anywhere;. . . ”

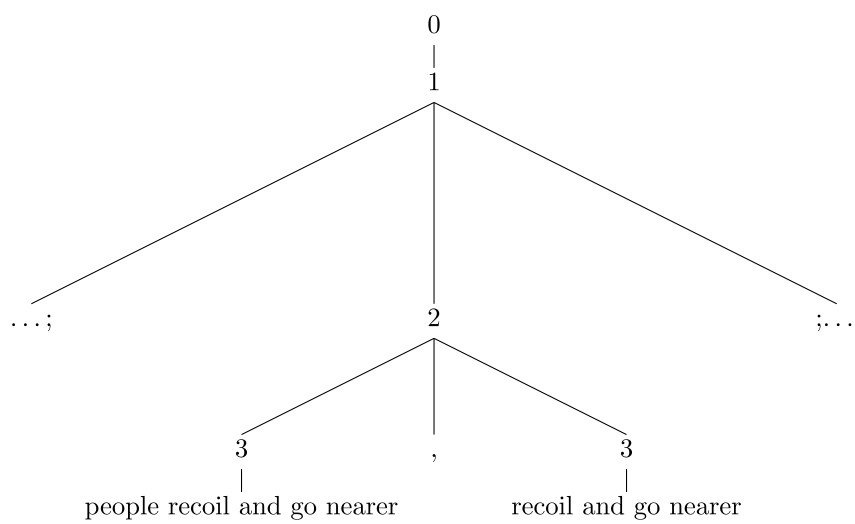

Figure 19: Correct partial punctuation tree for “. . . ; go nearer, go farther and go anywhere;. . . ”

list of affixes, the detector returns a positive result.

8. In activitie commendable, in a commonwealth profitable, and in warre terrible. (Peacham, 1954)

The homoioptoton detector is sometimes—and, in particular cases, often—inaccurate because it cannot differentiate between prefixes from simply the beginning of a word, or suffixes from the end of a word. For example, *reach* and *reapply* would return a false positive for homoioptoton, because they both begin with *re-* and are verbs. Similarly, *nothing* and *happening* could return a false positive for homoioptoton (depending on the parser's interpretation of *happening*) because they both end with *-ing* and can be nouns. A lemmatizer that could check the derivational morphology of words would solve this problem; lemmatizing *reach* and *reapply* should return *reach* and *apply,* dropping the *re-* when it is truly an affix; and lemmatizing *nothing* and *happening* should return *nothing* and *happen,* dropping the affix *-ing.* Hence, for improved homoioptoton detection, we could check that the lemmatized versions of the words do not contain the repeating affix, and that the grammatical category of the lemmas are the same. Thankfully, the detector did not return any false negatives with the instances we tested.

4 Results

Since the data set we worked with is relatively small, we cannot conclude anything definitive based on the statistics we gathered from it. However, we can use it to hypothesize about the collocation of figures and allow it to inform future studies.

4.1 Repetition and parallelism

Epanaphora and parison collocate frequently, which makes sense given that a repetition of word groups also implies a repetition of the word group's syntactic structure. In other words, a phrase that is simply repeated ("I like chickens! I like chickens!") technically contains both parison and epanaphora. However, parison requires the repetition of syntax of a phrase, as defined above in

Homer	Virgil
was the greater genius	[was] the better artist
in [him] we admire the man	in [him we admire] the work
hurries with a commanding impetuosity	leads us with an attractive majesty
scatters with a generous profusion	bestows with a careful magnificence
like the Nile, pours out his riches with a boundless overflow	like a river in its banks, [pours out his riches] with a gentle and constant stream

Table 2: Illustration of comparisons in 9.

Section 1.4, whereas epanaphora only requires the repetition of (the beginning) part of a phrase, so we can still find many instances of epanaphora without parison, and parison without epanaphora. All seven instances listed under parison also contain epanaphora somewhere in the instance, whereas parison occurs in about a third of the instances listed under epanaphora. This suggests that parison has stronger cognitive salience than epanaphora, because figure dictionary compilers tend to notice the repetition of the syntactic structure over the repetition of phrase-initial words. The following example shows parison with epanaphora, as well as other figures.

9. Homer was the greater genius, Virgil the better artist. In the one we most admire the man, in the other the work. Homer hurries us with a commanding impetuosity, Virgil leads us with an attractive majesty. Homer scatters with a generous profusion, Virgil bestows with a careful magnificence. Homer, like the Nile, pours out his riches with a boundless overflow, Virgil, like a river in its banks, with a gentle and constant stream. (Pope qtd. in De Mille, 1878)

The first sentence contains parison between *the greater genius* and *the better artist,* and the epanaphora does not appear until later in the passage. The use of parison is clearly used to illustrate a comparison between Homer and Virgil. The second sentence contains epanaphora *in the* but no parison because of the zeugma (*the man* and *the work* do not meet the requirements of a phrase in order to be marked for parison). The third sentence is where the passage-long epanaphora begins, which compares not two phrases in a sentence but whole sentences with one another, so that we arrive at a comparison as illustrated in Table 2.

Repetition, whether in syntactic structure or in words, can invoke parallelism or a similarity between two parallel phrases because of their formal similarity. In a figure of parallelism, the second parallel phrase "acts as a model with regard to the first: It is not identical to but also not isolated from the first," and because of its non-isolation, "the first member of a balanced pair suggests, or leads the memory on, to something parallel... in the second member" (Lotman qtd. in Christiansen, 2013; Robb qtd. in Gregg, 1984). The use of parison, homoioptoton (such as between *impetuosity* and *majesty*), mesodiplosis (such as in the many repetitions of *with*) and epanaphora bring each pair of comparisons into literal alignment, which then forces the reader to bring each pair of comparisons into conceptual alignment; that is, we compare *greater geniuses* with *better artists, the man* and *the work,* and so forth. Fahnestock argues that "similarity in at least one dimension... can impose a connectedness on consecutive sentences, even when their content is different" (2014). Homer and Virgil are brought together conceptually by nature of the passage's form, which contains multiple dimensions of similarity. Peacham (1954) states that the first phrase in a figure of parallelism is answered by the second phrase, like in an echo, and the repetition of *Homer* and *Virgil* resembles this. The repetition of the names brings together all of the comparisons made throughout the passage, again brought together by the nature of the passage's form (epanaphora, in this case). The bringing together of comparisons is another way of describing parison, so because epanaphora and other figures are helping illustrate parallelism in this passage, it is a figure of parallelism, parison, which stands out most in this example.

Parison and epanaphora occur at about the same rate amongst the instances listed under epiphora (so amongst approximately a third of the instances). We note that we do not have any data on the ratio of occurrences of epiphora among instances of parison and epanaphora, so we cannot definitely state which figure is more noticeable. We hypothesize, however, that epiphora, like epanaphora, occurs unnoticed under instances of parison than the other way around, given the similarity between epiphora and epanaphora. 10., though listed only under epiphora in Burton (2007), exemplifies epiphora, epanaphora and parison (as well as isocolon). It brings into parallelism the phrases *what lies behind us, what lies before us,* and *what lies within us,* and this parallelism is greater than the effects of the repetition of *what lies* and *us* alone.

10. What lies behind us and what lies before us are tiny compared to what lies within us. (Emerson

qtd. in Burton, 2007)

11. As Caesar loved me, I weep for him; as he was fortunate, I rejoice at it; as he was valiant, I honour him; but as he was ambitious, I slew him. (Shakespeare, 1623)

11. repeats description-reaction pairs, and the parallelism between each pair is more salient than any individual instance of lexical repetition, especially since the words are not consistently repeated throughout all four description-reaction pairs (the epanaphora in *as Caesar* becomes *as he* and the epiphora in *him* becomes *it* for one instance). The parallelism is used to highlight the logical cause (description) and effect (reaction) linkage in each pair, even when they switch from positive to negative descriptions (from loving, fortunate and valiant to ambitious) and reactions (from weeping, rejoicing and honouring to slaying).

4.2 Contrast and reversal

About a third of the instances listed under antimetabole, epanaphora, epiphora and asyndeton also contain parison, which means that parison can sometimes be the less "visible" figure. As we saw in Section 2.2 with 4. and 5., antimetabole and other chiastic figures are somewhat reliant on figures of parallelism to highlight the reversal. The reversal of antimetabole and other chiastic figures are based on difference, and difference can only be highlighted when it is accompanied by degrees of similarity. For example, in 10., which is listed under parallelism in Dupriez and Halsall (1991), the words *love* and *hate* stand out plainly among the repetition of words and structure around them:

12. As love, if love be perfect, casts out fear,
    So hate, if hate be perfect, casts out fear. (Tennyson qtd. in Dupriez and Halsall, 1991)

Furthermore, antimetabole and other chiastic figures are often described as figures of "symmetry" or "mirror-images." Symmetry occurs about a point, like a mirror through which images are reflected. Without this stable mid-point, there can be no symmetry. In 5., there is no stable mid-point and no symmetry, which is why the antimetabole between *hear* and *free* is so difficult to spot. In 4., the conjunction *and* serves as the stable mid-point, about which "the righteous who think they are sinners" and "the sinners who think they are righteous" are reversed. Note that 4. is not an exact parison under our definition (since *righteous* in the second phrase is an adjective, but *the righteous* and *sinners* are nouns in the rest of the sentence), but exhibits isocolon, another figure of parallelism. Stabilizing figures like those of parallelism help highlight figures of contrast, like antimetabole and other chiasmi. Although 12. is listed under parallelism in Dupriez and Halsall (1991), the antithesis between *love* and *hate* are at least as noticeable as the figures of parallelism, if not more. We note that, while we do not have data on the ratio of antimetabole occurring amongst instances labeled under parison, we hypothesize that antimetabole and other chiastic figures do indeed have more cognitive saliance and therefore "stand out" more than figures of parallelism.

4.3 Lists: Asyndeton and polysyndeton

As we saw when we defined phrases for our parison and epanaphora detectors, figures of repetition often occur within lists. In fact, 11. is a list of Caesar's traits and Brutus's reactions to them, and 9. provides a list of comparisons between Homer and Virgil. It is not surprising, therefore, that asyndeton and polysyndeton, figures that require lists, often occur with parison, epanaphora and homoioptoton. Instances under polysyndeton tend to have more of these figures than asyndeton, likely due to the additional structure (and, of course, repetition) granted by the repetition of conjunctions.

13. Her face with beauty, her head with wisedom, her eyes with Majesty, her countenance with gracefulnesse, her lips with lovelinesse... (Smith, 1721)

14. For men shall be lovers of themselves, covetous, boasters, proud, blasphemers, disobedient to parents, unthankful, unholy. (2 Timothy 3:2 qtd. in Norwood, 1792)

15. And I saw it, and I say it, and I will swear it to be true. (Puttenham qtd. in Christiansen, 2013)

16. I said, "Who killed him?" and he said, "I don't know who killed him but he's dead all right," and it was dark and there was water standing in the street and no lights and windows broke

and boats all up in the town and trees blown down and everything all blown and I got a skiff and went out and found my boat where I had her inside Mango Key and she was all right only she was full of water. (Hemingway qtd. in Burton, 2007)

13. and 14. are examples of asyndeton. 13. has plenty of structure, granted by using figures like parison, epanaphora (*her*), mesodiplosis (*with*), and some homoioptoton (*-nesse* in *gracefulnesse* and *lovelinesse*). 14., meanwhile, gives us a list without much structure (although it has homoioptoton with *un-* in *unthankful* and *unholy*), and even its members jump from one word to multiple words and from adjectives to nouns. Asyndeton is meant to be a "piling up" of words, and often draws attention away from the content or quality of words by overwhelming readers with their sheer quantity. The figures in 13., by providing more structure, adds the quality back in, in a sense; 13. shows that the person spoken of not only has very many wonderful features, but that each of these features is significant enough that the reader has to consider each one before moving on to the next.

Likewise, the excessive conjunctions in polysyndeton often slows down the reading of lists, and makes it amenable to other figures of repetition. Note that 16., where conjunctions take the place of what would normally be sentence-ending punctuation, gives a sense of breathlessness, which *speeds up* the reading. 15., however, which exhibits repetition in the form of epanaphora (I) and epiphora (*it*), punctuation and parison between each conjunction, which not only reads slower than 16., but gives it a sense of parallelism, allowing the reader to compare the different actions of the narrator. Observing the differences between 13. and 14., and between 15. and 16., we can see that asyndeton and polysyndeton can be used with or without other figures of repetition, and in doing so produce different effects. Furthermore, because repeating words and structures between members of a list seems "natural," whereas both asyndeton and polysyndeton are marked and purposeful, we hypothesize that asyndeton and polysyndeton has more cognitive salience compared to parison, epanaphora and homoioptoton.

5 Conclusion

In this study, we took "prototypical" instances of figures of repetition from rhetorical figure handbooks and dictionaries. We built a figure detection tool for parison, epanaphora and homoioptoton, and used this tool to look for their collocation with the other figures of repetition. We found that parison collocates with many of these figures. Epanaphora shows up with almost all of the prototypical instances of parison, meaning that, while the two may collocate frequently, parison may have more cognitive salience than epanaphora. We hypothesize that the same holds true for epiphora. Parison also collocates frequently with antimetabole, but with antimetabole holding the stronger salience than parison due to the markedness of reversal and difference. Lastly, all three of our figures collocate frequently with polysyndeton and asyndeton, and the frequency of collocation depends on the desired speed of reading. We hypothesize that polysyndeton and asyndeton have stronger cognitive salience than the other figures, due to the markedness of these figures' forms.

Based on these results, the linguistic domain of a figure does not seem to affect the strength of its salience, assuming that the most salient figure is almost always the one noted by handbooks: epanaphora, epiphora, antimetabole, asyndeton and polysyndeton are all lexico-syntactic figures, but they vary in strength over parison. More significant seems to be a figure's markedness or frequency in language; an uncommon figure like asyndeton will stand out more than a common figure like parison.

6 Next steps

This study is a pilot study on the collocation of figures, and it leaves many questions about collocation left unanswered. A future study should look at the frequency of particular rhetorical figures in corpora, and see if there is any correlation between the frequency and the cognitive salience of a rhetorical figure. This would, of course, require more figure detectors, many of which are provided in previous work like those of Gawryjołek (2009) and Java (2015).

All of the detectors used in this study could be improved to eliminate both false positives and false negatives. Gawryjołek (2009) and Java (2015) use a range of node heights to match branches in syntax trees for parison, rather than mandating that the node heights are equal; this could be used to eliminate some false negatives (but could also introduce more false positives). The punctuation-based tree for parison and epanaphora could also be improved upon, especially eliminating cases where, in "X and Y[,]" and "X and Y" are considered their own separate phrases. The homoioptoton detector could use the most improvements. Gawryjołek (2009) and Java (2015) use WordNet as a foundation for their polyptoton detection, and a similar method could be used for homoioptoton. WordNet organizes itself in synsets, so each word is semantically related to others, including those which may be derivational variations of the word. Using the method outlined in Gawryjołek (2009) and Java (2015), we could find these derivational variations for our potential homoioptoton pairs, find the word or words without affixes amongst these variations, and then check that the affix-less word for each pair match in grammatical category.

References

Alliheedi, M. and C. Di Marco (2014, May). Rhetorical figuration as a metric in text summarization. Montreal, QC. Canadian Artificial Intelligence Conference.

Bullinger, E. (1898). *Figures of Speech Used in the Bible: Explained and Illustrated.* New York: E. & J.B. Young & Co.

Burke, K. (1969 [1950]). *A rhetoric of motives.* Berkeley, CA: University of California Press.

Burton, G. O. (2007). Silva rhetoricae. `http://rhetoric.byu.edu`.

Burton, N. (2016, April). Are you an epicurean? Really? *Psychology Today*.

Chien, L. and R. A. Harris (2010). Scheme trope chroma chengyu: Figuration in Chinese four-character idioms. *Cognitive Semiotics 6*, 155–178.

Christiansen, N. L. (2013). *Figuring style.* South Carolina: University of South Carolina Press.

De Mille, J. (1878). *The Elements of Rhetoric.* New York: Harper & Brothers.

Dubremetz, M. (2017). *Detecting Rhetorical Figures Based on Repetition of Words: Chiasmus, Epanaphora, Epiphora.* Ph. D. thesis, Studia Linguistica Upsaliensia 18.

Dupriez, B. and A. W. Halsall (1991). *A dictionary of literary devices: Gradus, A–Z.* Toronto: University of Toronto Press.

Fahnestock, J. (1999). *Rhetorical Figures in Scientific Argumentation.* New York: Oxford University Press.

Fahnestock, J. (2005). Rhetoric in the age of cognitive science. In R. Graff (Ed.), *The Viability of Rhetoric*, pp. 159–179. New York: State University of New York Press.

Fahnestock, J. (2014). *Rhetorical style: The uses of language in persuasion.* New York: Oxford University Press.

Gawryjołek, J. J. (2009). Automated annotation and visualization of rhetorical figures. Master's thesis, University of Waterloo.

Gregg, R. B. (1984). *Symbolic inducement and knowing: A study in the foundations of rhetoric.* Columbia, South Carolina: University of South Carolina Press.

Harris, R. A. (2013). The rhetoric of science meets the science of rhetoric. *Poroi 9.* Article 8.

Harris, R. A. (2018). Antimetabole and its friends. International Association for Cognitive Semiotics. Unpublished manuscript.

Harris, R. A. and C. Di Marco (2009). Constructing a rhetorical figuration ontology. Artificial Intelligence and Simulation of Behaviour. Unpublished manuscript.

Harris, R. A. and C. Di Marco (2017). Rhetorical figures, argumentation, computation. *Argument and Computation 8*, 211–231.

Holmes, J. (1806). *The Art of Rhetoric Made Easy*. London: A. Parker.

Hromada, D. D. (2011). Initial experiments with multilingual extraction of rhetorical figures by means of PERL-compatible regular expressions. pp. 85–90. The Second Student Research Workshop associated with RANLP 2011.

Java, J. (2015). *Characterization of Prose by Rhetorical Structure for Machine Learning Classification*. Ph. D. thesis, Nova Southeastern University.

Kennedy, J. F. and T. Sorensen (1961). Inaugural address, 1961 Jan 20, in: The American presidency project [Internet]. `http://www.presidency.ucsb.edu/ws/index.php?pid=8032`.

Manning, C. D., M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky (2014). The Stanford CoreNLP natural language processing toolkit. pp. 55–60. 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations.

McGlone, M. and J. Tofighbakhsh (2000, September). Birds of a feather flock conjointly (?): rhyme as reason in aphorisms. *Psychological Science 11*, 424–428.

McQuarrie, E. and D. G. Mick (1996). Figures of rhetoric in advertising language. *Journal of Consumer Research 22*.

Norwood, C. (1792). *Divine Eloquence, Or An Essay Upon The Tropes And Figures Contained, The Holy Scriptures*. London, United Kingdom: Printed for W. Phorson.

Peacham, H. (1954). *The Garden of Eloquence*. Gainesville, Florida: Scholar's Facsimiles.

Perelman, C. (1982). *The Realm of Rhetoric*. Notre Dame, IN: University of Notre Dame Press.

Perry, S. (2016). *The Essex Serpent*. London: Serpent's Tail.

Shakespeare, W. (1993 [1623]). Julius Caesar. In J. Hylton (Ed.), *The Complete Works of William Shakespeare*. MIT.

Shelley, M. (2008 [1818]). Frankenstein. In *Oxford World's Classics*. New York: Oxford University Press.

Smith, J. G. (1721). *The Mysterie of Rhetorique Unveil'd Wherein above 130 the Tropes and Figures are Severally Derived from the Greek into English: Together with Lively Definitions and Variety of Latin, English, Scriptural, Examples, Pertinent to Each of Them Apart.* London: R. Wilkin, B. Tooke, D. Midwinter, B. Cowse, and G. Mortlock.

Smith, T. (n.d.). CSSR president's message. `http://rhetcanada.org/presidents-message/`. Accessed 14 December 2019.

Strommer, C. W. (2011). *Using Rhetorical Figures and Shallow Attributes as a Metric of Intent in Text.* Ph. D. thesis, University of Waterloo.

A Penn Treebank Part-Of-Speech Tags

Tag	Description	Tag	Description
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	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	WRB	Wh-adverb

Table 3: The Penn Treebank part-of-speech (or grammatical category) tags used in the detectors. Adapted from `https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html`.

B Affixes used in the homoioptoton detector

dis	down	extra	hyper	il
im	in	ir	inter	mega
mid	mis	non	over	out
post	pre	pro	semi	sub
super	tele	trans	ultra	un
under	up			

Table 4: Prefixes used in the homoioptoton detector.

ability	abilities	able	ably	ablies
ac	acious	acity	acities	adelic
aholic	ally	allies	ator	ade
adic	age	al	algia	an
ance	ant	ard	arian	arium
ary	ate	ation	ative	cide
cidal	cracy	crat	cule	cy
cies	cycle	dom	dox	ectomy
ee	eer	el	emia	ence
ency	er	ern	esce	escence
ese	esque	ess	est	etic
ette	ful	fy	fies	gam
gamy	gamies	geddon	gon	gonic
gyny	gynies	hood	ia	ial
ian	iasis	iatric	ible	ibly
iblies	ic	ical	ify	ifies
ile	ily	ine	ing	ion
ious	ish	ism	ist	ite
itis	ity	ities	itude	ive
ization	isation	ize	ise	less
let	like	ling	loger	logist
log	ly	mageddon	ment	ness
ocity	ocities	oholic	oid	ology
ologies	oma	onym	opia	opsy
or	ory	osis	ostomy	ostomies
otic	otomy	otomies	ous	path
pathy	pathies	phile	phobia	phone
phyte	plegia	plegic	pnea	punk
scopy	scopies	scope	scribe	script
sect	sexual	ship	sion	sis
some	sophy	sophies	sophic	tion
tome	tomies	trophy	trophies	tude
ty	ties	ular	uous	ure
ward	ware	wise		

Table 5: Suffixes used in the homoioptoton detector.