# Rhetorical Parallelism Detection: Original Annotation Guidelines

Stephen Bothwell

September 25th, 2023[*]

## 1   Introduction

In this document, we provide our original annotation guidelines for parallelism data. These guidelines were not officially in written form and were spoken and discussed verbally; thus, the version presented here is a close approximation of what was conveyed and considered as guidelines by our initial annotator.

## 2   Guidelines

In general, branches of parallelisms should be tagged from the first identical (or similar) word to the last identical (or similar) word; as such, each branch should be maximally represented. Branches should be detected, paired, and combined based upon whether they exhibit at least *two* of these criteria:

- They contain identical number of words (or were within approximately two words of one another).

- They have identical syntactic structure.

- They have two or more pairs of words of an identical grammatical form in identical order.

- They have two or more pairs of words that are lexically identical, are synonyms, are cognates, or are antonyms.

- They have at least two words that have a phonetically similar ending (*e.g.*, rhyme).

- They are a short distance from one another (*i.e.*, they have few intervening words). The acceptable margin is again around two words.

We use the BRAT annotation tool [1] to amass our dataset. Our annotation scheme consists of BRAT's entities and relationships. An entity can be a `ParallelArm` or one of `ChiasmA` and `ChiasmB`. A relationship can be `Parallel`, linking two `ParallelArm` entities. This relationship indicates a synchystic parallel structure. A relationship can also be `Chiasm`, linking a `ChiasmA` and a `ChiasmB` and lsabeling a chiastic structure.

## References

[1]   Pontus Stenetorp et al. "brat: A Web-Based Tool for NLP-assisted Text Annotation". In: *Proceedings of the Demonstrations Session at EACL 2012*. Avignon, France: Association for Computational Linguistics, Apr. 2012.

---

[*]The date presented here is the "last updated" date.