Dialogue Data Evaluation

Zoltán Csik

2023-09-08

Data Preparation

This R Markdown serves the purpose of evaluating acquired dialogue data. The collected data represents several linguistic dialogue features.

```
##                      Feature    HH1    HH2    HH3    HC1    HC2    HC3
## 1            Sentence Length 52.558 55.209 73.681 33.372 22.452 38.393
## 2         Words per Sentence 13.767 13.478 20.931  8.721  5.710  9.714
## 3 Unique words per Sentence 12.093 11.791 17.056  8.093  5.613  8.857
## 4          Lexical Diversity  9.287 11.006  9.816  9.618  6.915  8.277
```

Divided Dialogues:

To compare human-human data with human-robot data, the dataframe has to be split into two:

```
##                      Feature    HH1    HH2    HH3
## 1            Sentence Length 52.558 55.209 73.681
## 2         Words per Sentence 13.767 13.478 20.931
## 3 Unique words per Sentence 12.093 11.791 17.056
## 4          Lexical Diversity  9.287 11.006  9.816
```

```
##                      Feature    HC1    HC2    HC3
## 1            Sentence Length 33.372 22.452 38.393
## 2         Words per Sentence  8.721  5.710  9.714
## 3 Unique words per Sentence  8.093  5.613  8.857
## 4          Lexical Diversity  9.618  6.915  8.277
```

Human-Human Diaogue Data

Count the mean value for each linguistic feature data:

```
##                      Feature    HH1    HH2    HH3     Mean
## 1            Sentence Length 52.558 55.209 73.681 60.48267
## 2         Words per Sentence 13.767 13.478 20.931 16.05867
## 3 Unique words per Sentence 12.093 11.791 17.056 13.64667
## 4          Lexical Diversity  9.287 11.006  9.816 10.03633
```

Human-Robot Diaogue Data

Count the mean value for each linguistic feature data:

```
##                     Feature    HC1    HC2    HC3      Mean
## 1           Sentence Length 33.372 22.452 38.393 31.405667
## 2        Words per Sentence  8.721  5.710  9.714  8.048333
## 3 Unique words per Sentence  8.093  5.613  8.857  7.521000
## 4         Lexical Diversity  9.618  6.915  8.277  8.270000
```

Combined Mean Values

Combine mean values from both Data Frames

```
##                     Feature       HH       HC
## 1           Sentence Length 60.48267 31.405667
## 2        Words per Sentence 16.05867  8.048333
## 3 Unique words per Sentence 13.64667  7.521000
## 4         Lexical Diversity 10.03633  8.270000
```

Visual Comparison