CLAS20016 Ancient Greek 4 Epithet Classifier

1. Introduction

1.1 Purpose of report

The formulaic nature of Homer's Odyssey has been debated for many years. Scholars of the early 1900s such as Düntzer, Witte and Parry all argued for the stock repetition of epithets, lines and phrases yet more recent scholars like Visser, Paraskevaides and Bakker do not believe in such a formula, opting to explore a looser construction of lines.

Parry explores the noun-epithet formula, taking particular interest in genitive and dative plural nouns with epithets that he believes are predetermined by the syllables of the noun (Parry, 1928).

Bakker also takes interest in the dative however he bases his 'nucleus' in the verb as opposed to the noun and takes the whole dative phrase as an epithet to the verb (Bakker, 2005).

This report focuses on a subsection on the analysis done by Bakker and Parry to determine whether the epithets used by Homer in the first half of the Odyssey for ships follow a formula. To do this, a classification model or classifier to predict the presence of an epithet and, if present, what the accompanying epithet would be was created and trained using features surrounding the instances. The more accurate the model, the more the epithets of 'ship' follow a pattern.

1.2 Limitations and assumptions

The biggest limitations of this research were the time period in which it was due and the inaccessibility of coding libraries to Ancient Greek.

With more time, a wider range of features could be explored, and models refined. More time would also allow for creating coding modules that exist for English texts but not Ancient Greek such as the removal of stop words and lemmatisation.

Assumptions made for this project including that the instances from the first half of the Odyssey are reflective of the patterns in latter parts of the narrative.

2 Method

2.1 Dataset and sample size

Since there was no accessible dataset profiling each instance where the word ship was used and the text does not lend itself to be read in and profiled automatically, manual imputation was used. First finding the word ship in an English translation, the corresponding line was found in West's version of the Odyssey. The book number, line number, word for ship, epithet in both Greek and English, number, case clause and scansion was recorded for each instance. If two epithets were present, two different instances were recorded for the separate epithet. A frequency of each epithet can be seen in Figure 1. and Figure 2.

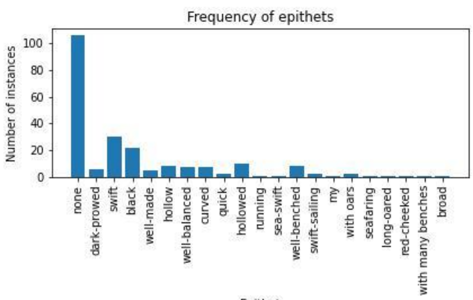

Figure 1-The distribution of class labels in the training dataset.

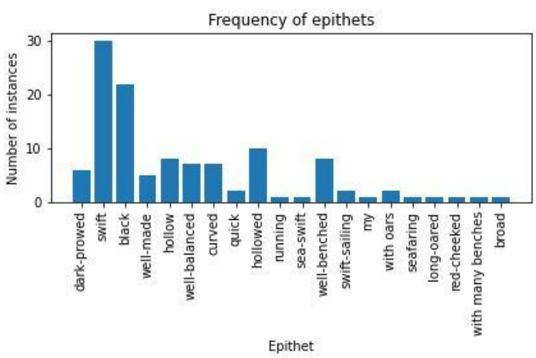

Figure 2-The distribution of class labels in the training dataset.

Further features were calculated from the attributes: the number of lines per each clause, the position in the line where the word ship occurred (start, middle or end), the position of the epithet relative to the word 'ship', the percentage of feet that were dactyls in the line and the trigrams of the clause. The line number was standardised to be relative to the number of lines in the book.

2.3 Cleaning and feature engineering

With the attributes found, feature analysis was performed. As can be seen in Figure 3., there was little correlation between the attributes, with the strips of maroon-red appearing in epithet_gr and epithet_en which are the labels. Thus, no attributes can be removed through this analysis. The most correlated features with the English epithet were the word for ship, the difference in position and the case of the noun-epithet with correlations of 0.7683, 0.7213 and 0.5129 respectively. Correlation analysis was performed on the full dataset as well as a filtered dataset excluding instances with no epithet.

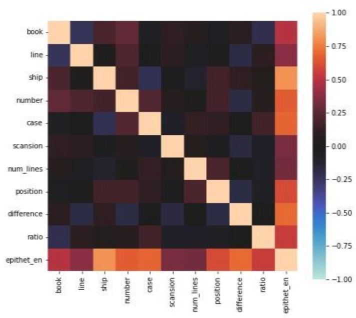

Figure 3-The correlation between all attributes

Next, chi-squared analysis was performed on the categorical variables to select the 3 best features. The two most significant numeric variables (line number, position, number of lines for the clause and the percentage of dactyls) were selected using analysis of variance (ANOVA). Due to the strong correlation between a difference of zero and the 'none' label, difference was not included in the modelling of the dataset that included the 'none' instances.

To gain a sense of the most common words that appeared in tandem with the mention of ships, WordCloud was used to visualise the distribution.

Figure 4 – Wordcloud of the raw training data

From this wordcloud of the initial data in Figure 4, there is a lot of apparent noise as exemplified by the large 'και' as well as the appearance of 'δ' and 'δε'. As previously mentioned, it is unrealistic to properly clean the data in the timeframe however basic stop word removal was performed to better visualise the data.

Figure 5 – Wordcloud with rudimentary stop word removal.

2.4 Training models

The models tested were a Decision Tree with a maximum depth of five and a Logistic Regression. Due to the skew in 'none' type epithets, the Multinomial Logistic Regression (MLR) was used. Similarly, the f1 score was used to determine accuracy since it takes the micro-average. These models were tested on the raw data, the features selected by χ-squared, the features selected by ANOVA and a combination of the two feature selection methods.

A separate model was created to take the trigrams of the data. As mentioned earlier, Ancient Greek has a limited corpus in the coding library thus trigrams were taken to account for not being able to remove stop words (such as και and αρα) and the inability to lemmatise. The aforementioned models were fit using the trigrams after being featured with term frequency-inverse document frequency (TFIDF) and using TFIDF combined with χ-squared selecting the top 100 trigrams.

Following these models, a stacker model was created with the three best performing models – all of which performed better than the baseline model – but this did not perform as well as the highest performing model: Decision Tree Classifier using ANOVA selection. This model was then fit to the data and a tested.

2.5 Predicting the test dataset

Upon implementing the models on the training dataset and finding the accuracy of each model, the highest performing combination of modelling and preprocessing was performed on the test dataset and the accuracy was measured.

3 Results

In the dataset of the scraped features, the accuracy of the models ranged from 0.356 to 0.533 as can be seen in Table 1. None of the classifiers performed better than the baseline model.

Model	Accuracy
Zero-R	0.533
One-R	0.444
Decision Tree	0.356
MLR	0.467

Table 1- Accuracy of the models on the train data using the scraped features

The best performing combination of feature selection for the scraped features were the models created using ANOVA. Here, the MLR model performed equally as well as the baseline model while the other two models had the highest accuracy compared to their performance with other preprocessing.

Model	Accuracy
Zero-R	0.533
One-R	0.511
Decision Tree	0.489
MLR	0.533

Table 2- Accuracy of the models on the train data using ANOVA feature engineering

For the trigrams, the models performed solely on the TFIDF vectorised trigrams performed best. The One-R model surpassed the baseline model while MLR was on par with it.

Model	Accuracy
Zero-R	0.533
One-R	0.578
Decision Tree	0.489
MLR	0.533

Table 3- Accuracy of the models on the train data using TFIDF

Three of the top performing models – MLR with ANOVA featurisation, One-R with TFIDF feature engineering and MLR with TFIDF and chi-square engineering – were used in a stacker model. The stacker had an accuracy of 0.511 hence a One-R model with TFIDF featurisation was used on the final test set.

4 Discussion

4.1 Analysis of results

Initially, all the models were performing with an accuracy above 0.5 however, upon analysis of the false negatives and false positives of the final classification, it was decided that difference was an inappropriate feature – mostly due to its high correlation with a difference of 0 and the absence of an epithet.

To combat this, difference was removed from modelling features with the dataset that included instances with no epithets. Additionally, modelling was performed on a dataset that excluded instances without epithets.

It was disappointing to discover that these models were even less accurate, averaging around 0.25. It can be inferred that the presence of the instances with no epithet skewed the accuracy of the data however this is a slight deviation from the main hypothesis.

Upon removing the difference feature, the classifier predicted no epithet for all final test instances. Further investigation should be done to determine if this is occurrence for all instances or if it was coincidence.

The large number of instances with no epithet may also affect the reliability of results since classifiers are mostly designed to be trained on balanced datasets. To account for this in the modelling stage, cross-validation was utilised to pick the highest performing model.

4.3 Recommendations and further experimentation

As previously mentioned, there were many limitations in the research due to the time restraint and nature of the research. More could be done to further Parry's investigation of scansion both through scraping more features as well as increasing the sample size.

More can also be done for the effect of the verb which later scholars like Bakker have put emphasis on. Additionally, the sentiment of the trigrams as opposed to the 'scansion' and syllables of the clause could prove important, despite previous research believing epithets are all semantically neutral.

5 Conclusions

From the accuracy of the final prediction model, it can be determined that none of the features explored were significant in the 'formula' of Homer. This does not however, disprove the notion of a 'formula', rather it introduces a new way to explore the verses of Homer as well as a gap in the computing corpus that would open many texts to new analysis.

6 References

E. Bakker. Pointing at the Past: From Formula to Performance in Homeric Poetics. Washington, DC: Center for Hellenic Studies, 2005.

Greek and Roman Verse. https : / / hypotactic . com / my - reading - of - homer - work - in - progress/. Accessed: 2022-10-10.

Homer, Odyssey. https : / / www . perseus . tufts . edu / hopper / text ? doc = Perseus % 3Atext % 3A1999.01.0136%3Abook%3D1%3Acard%3D1. Accessed: 2022-10-15.

M. Parry. The Making of the Homeric Verse. Oxford, 1971.

F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: Journal of Machine Learning Research 12 (2011), pp. 2825–2830.

M. West. Homerus. Odyssea. Bibliothexa scriptorum Graecorum et Romanorum Teubneriana. De Gruyer, 2017.