Ancient Greek 4 CLAS20016

The Formula of Ship Epithets [1129 Words]

Jia Williams Murnane
Student ID: 1168850
Github repository

October 16, 2022

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, 1971)[4]. 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)[1].

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[1]. The more accurate the model, the more the epithets of 'ship' follow a pattern. Assumptions made for this project include that the instances from the first half of the Odyssey are reflective of the patterns in latter parts of the narrative. The biggest limitations of this research were the time period in which it was due and the inaccessibility of coding libraries to Ancient Greek. Due to this, manual imputation was used to create the dataset. After finding the word ship in an English translation[3], the corresponding line was found in West's version of the Odyssey[6]. The book number, line number, word for ship, epithet in both Greek and English, number, case, clause and scansion[2] was recorded for each instance[2]. If two epithets were present, two different instances were recorded for the separate epithets. The frequency of each epithet can be seen in Figure 1. and Figure 2. 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[3]. The line number was standardised to be relative to the number of lines in the book.

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. From this wordcloud of the initial data in Figure 3., there is a lot of apparent noise as exemplified by the large '$\kappa\alpha\iota$' as well as the appearance of 'δ' and

[1]A classifier takes features of an instance and predicts a label, in this case an epithet (or lack thereof) for the word ship.

[2]The word for ship combines the effect of number and case as well as the dialect.

[3]A trigram being a trio of words in row.

'$\delta\epsilon$'. Thus, Figure 4. visualises the clauses with rudimentary cleaning[4]. The most correlated features to the 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. Chi-squared analysis and analysis of variance (ANOVA) were also used on the categorical and numerical features respectively. 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. Term frequency-Inverse Document Frequency (TFIDF) was performed on the trigrams[5].

The models[5] tested were Decision Trees[6] and Multinomial Logistic Regression (MLR)[7]. These models were tested on variations of engineered datasets. A separate model was created to take the trigrams of the data with a similar process for model training as the scraped features[8]. 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 of depth one (One-R) on TFIDF trigrams[9]. The baseline model of the scraped features with no engineering had an accuracy of 0.533[10]. The best performing model with the scraped features was MLR with ANOVA which had an accuracy of 0.533 as well. The best performing model, One-R with TFIDF which had an accuracy of 0.578. The stacker had an accuracy of 0.511 hence a One-R model with TFIDF featurisation was used on the final test set.

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[11]. 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. As previously mentioned, there were many limitations in the research due to the time restraint and nature of the research[12]. 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. Furthermore, 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.

[4]This cleaning was done on the most common words from Figure 3. and included removed words and lengthening/lemmatising words.

[5]As the name suggests, this feature engineering weights the importance of the trigram in the classification against its frequency in the document.

[6]One with a maximum depth of 1 and one with a maximum depth of 5.

[7]A multinomial regression was used due to the unbalanced distribution of the labels.

[8]Scraped refers to all features that were manually found and not with the Natural Language ToolKit (ie. all features excluding trigrams).

[9]A Stacker takes the labels from other models as its features to then predict a label.

[10]The accuracy was tested using f1 score which takes the micro-average - the average of the accuracy for each label.

[11]A false negative is defined as the times for each epithet where the true label is not identified while a false positive refers to when an epithet is predicted for an instance and is wrong.

[12]Cross-validation is where the data is split and tested then split again in a different denomination to randomise the data and decrease bias.

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. For now, it appears that current scholarship is correct in their analysis that the rigid formula in a mathematical sense proposed by Parry does not exist. In saying this, with more time and the use of statistical analysis through computing, their theories can be put to the test with a greater sample size and the analysis of more features.

Figures

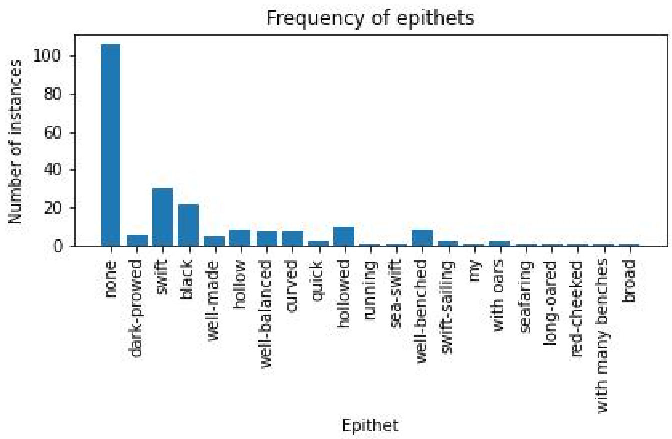

Figure 1: The distribution of all epithets in the first 12 books of the Odyssey

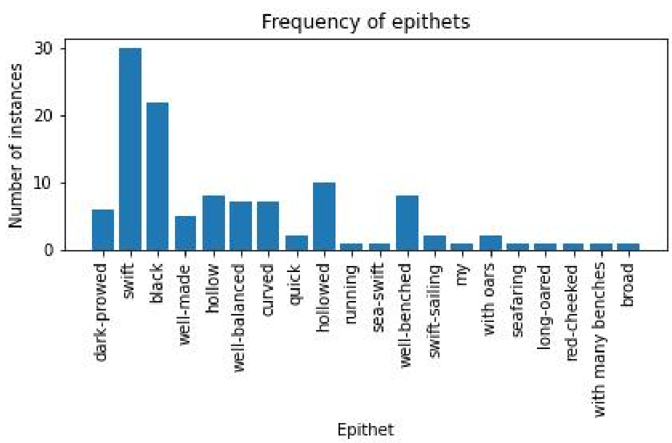

Figure 2: The distribution of epithets excluding instances without epithets

Figure 3: WordCloud visualisation of the clauses which hold the word ship

Figure 4: WordCloud visualisation with rudimentary stopword removal

References

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

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

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

[4] M. Parry. *The Making of the Homeric Verse*. Oxford, 1971.

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

[6] M. West. *Homerus. Odyssea. Bibliothexa scriptorum Graecorum et Romanorum Teubneriana*. De Gruyer, 2017.