AMSI VACATION RESEARCH SCHOLARSHIPS 2019–20

EXPLORE THE MATHEMATICAL SCIENCES THIS SUMMER

Information Entropy of Constructed Languages

Megan Crossing

Supervised by Professor Matthew Roughan

The University of Adelaide

Vacation Research Scholarships are funded jointly by the Department of Education and Training and the Australian Mathematical Sciences Institute.

Contents

Abstract

This project investigates entropic differences between natural and constructed languages, with a focus on the Neo-Quenya translation of the *New Testament*. Using the Natural Language Processing Toolkit in Python, we found the Shannon entropies and Zipf-Mandelbrot distributions of four translations of the New Testament. The translations used were the *King James Version*, the *New International Version*, the *Biblia Sacra juxta Vulgatam Clementinam* (Latin), and the *Neo-Quenya*. The study found that the entropies of the Latin and Neo-Quenya translations were generally lower than the two English translations, that the Zipf-Mandelbrot distributions of the English translations were near identical, and that the Zipf-Mandelbrot distributions of the Latin and Neo-Quenya were, whilst similar, less similar than the two English translations.

1 Introduction

There are three main types of constructed languages:

- Auxiliary languages
- Engineered languages
- Fictional languages

Auxiliary languages, such as Esperanto, are used to aid global communication and engineered languages, such as Lobjang, are created to perform particular experiments. The third, fictional languages, are created to provide verisimilitude in fantasy books (such as *Lord of the Rings*) and TV series (such as *Game of Thrones* or Star Trek). They are the focus of this paper.

Unlike engineered or auxiliary languages, fictional languages are created to be believable. Presumably, a good fictional language will be statistically similar to natural languages. We can assess this using the lens of information theory.

Information entropy is a topic that has been often studied since its initiation in 1948 by Claude Shannon (Gleick 2011). Since this time, the entropy of many natural languages has been analysed (Barnard 1955; Bentz 2017), but there has been little investigation into the entropy of constructed languages.

In this paper, we analyse the Quenya translation of the *New Testament*.

1.1 Statement of Authorship

Megan Crossing processed the texts; developed the Python code; produced, reported and interpreted the results; and wrote this report.

Professor Matthew Roughan initiated and supervised the project, and proofread this report.

2 Quenya: Its Brief History

Quenya was one of 9 Elvish dialects developed by J.R.R. Tolkien for use in his *Lord of the Rings* universe. He began working on it in 1915, and it is generally agreed on that the language was in its final form by 1954, when he published *The Fellowship of the Ring*. (Fauskanger n.d.)

As a linguist, Tolkien put a lot of time and effort into making his languages as realistic as possible, giving many of his words complete etymologies. For example, Tolkien writes that the word *Quenya* is probably derived from the same stem as *Quendi* (Elves), and thus means Elvish. However, he also writes that the stem may instead be *'quet'*, and not *'quen'*, meaning *Quendi* translates as 'those who speak with voices', and *Quenya* as 'speech'. (Tolkien 1994) Not only did Tolkien give his languages a history, he also incorporated the uncertainty he observed in natural languages.

Since Tolkien's creation of Quenya, various people have attempted to translate works into it (Derdzinski n.d.; *I Vinya Vere: The New Testament in Neo-Quenya* 2015). Despite the years Tolkien put into developing Quenya, its vocabulary stretches only so far. Translators have developed neologisms to deal with this, creating Neo-Quenya. The work investigated here is the Neo-Quenya translation of the *New Testament* (*I Vinya Vere: The New Testament in Neo-Quenya* 2015).

So, natural languages have evolved over hundreds of thousands of years by hundreds of thousands of people. Tolkien's Quenya, whilst having a mimicked evolution, has been constructed over the last hundred with only one main contributor. With this in mind, it will be interesting to see if any hidden differences remain between Quenya and natural languages.

3 Information Entropy

3.1 Shannon Entropy

Here, information entropy was found using Shannon's entropy (Shannon 1948), which is given by the following equation:

$$H(X) = -\sum_{k=1}^{N} p(k) \log_2 p(k).$$

This gives the information entropy ($H(X)$, measured in bits per token) of a discrete random variable, X, that exists over a distribution of 1 to N. In our case, we are finding the word entropies for four

translations of the *New Testament*. Thus, for a given translation, X is the set of all words in the *New Testament*, N is the number of unique words, and $p(k)$ is the probability of the $k^(th)$ word occuring.

3.2 Zipf-Mandelbrot Distribution

The Zipf-Mandelbrot distribution (Zipf 1949) is a generalisation of the Zipf distribution. The Zipf distribution was designed to model observed patterns in the probability distributions of natural languages:

$$f(k|N,s) = \frac{1}{k^s \sum_{n=1}^{N} n^{-s}},$$

where f is the probability distribution of words in a natural language, N is the vocabulary size of the language, k is the rank of a word, $k = 1$ being the most frequent word, and s is some parameter.
This paper uses the later modification by Benoit Mandelbrot (Mandelbrot 1966), as it better models the lower ranked words. It has the following probability mass function:

$$f(k|N,s,q) = \frac{1}{(k+q)^s \sum_{n=1}^{N} (n+q)^{-s}}.$$

This is similar to the Zipf distribution, with the only difference being the additional parameter q. In the case where $q = 0$, the Zipf and Zipf-Mandelbrot distributions are identical.

4 The Process

4.1 Text selection

The longest accessible piece of any artistic constructed language text is, as far as we are aware, Fauskanger's (2015) translation of the *New Testament* into Neo-Quenya.

For comparison, two English translations (the *New International* Version (1978) and the *King James Version* (2004)) were also analysed, along with the Latin *Biblia Sacra juxta Vulgatam Clemantinam (1592)*. The *New International Version* is standard and commonly used. The *King James Version* was revised in 2004, but is based off the *Authorised King James Version*, which was translated in 1611. The *King James* was chosen as its writing style is far more archaic than the *New International Version*.

The *Biblia Sacra* is a revised version of the original AD 382 translation into Latin. This Latin translation was chosen for two reasons. Firstly, Quenya grammar is partially based off Latin grammar.

Secondly, Tolkien created Quenya to be the 'Latin' of Elvish – used more for scholarship and law than for everyday writing and conversation. Perhaps Quenya and 16th century Ecclesiastical Latin lend themselves to similar writing styles.

4.2 Pre-Processing

The Neo-Quenya translation was downloaded as Word documents of the individual chapters (*I Vinya Vere: The New Testament in Neo-Quenya* 2015). These included alternating paragraphs of both the Quenya and its back translation. A find and replace search in Word was used to remove the back translation paragraphs, along with the translation notes.

The three natural language translations were downloaded as pdf documents (*Biblia Sacra juxta Vulgatam Clementinam* 1592; *King James Version* 2004; *New International Version* 1978), and copied into word documents for pre-processing, where the *Old Testament*, *Psalms* and *Proverbs* were removed. Headers were also removed.

In the Neo-Quenya translation, the *Gospel of Mark* Chapter 16, Verses 9-20 were not included. This is not due to an incomplete translation, but simply to the fact that these verses were not contained in the earlier manuscripts (*The Holy Bible: New International Version with concise bible encyclopaedia* 2007, Mark 16:8-9). Consequently, these verses were removed from the *New International Version*, the *King James Version* and the *Biblia Sacra*.

The translations were then combined and split into separate text documents containing 10 sections for individual analysis:

- New Testament
  - Gospels
    - Gospel of Matthew
    - Gospel of Mark
    - Gospel of Luke
    - Gospel of John
  - Acts of the Apostles
  - Letters
    - Paul's Letters
  - Revelation

The *New International* and *King James* translations did not attribute the authorship of the *Letters to the Hebrews* to Paul, whereas the *Biblia Sacra* did. The Neo-Quenya translation did not specify the authorship of any of its letters. Due to the variation in writing style ("Non-Pauline Letters" 2007), the *Letters to the Hebrews* were not included in segment containing *Paul's Letters*.

4.3 Generating Results

Each section was uploaded into Python (Van Rossum and Drake 2009) as four text files (one for each translation), and the following processing was performed:

1. The text was tokenised into single words. Special characters (such as accented letters) were converted to normal characters (i.e. non-accented). Upper case letters were converted to lower case. Non-alphabetic characters were then removed.

2. The total number of words was counted and recorded. [appendix B, table 5]

3. Word frequencies were calculated and ordered from most to least common. The total number of unique words was recorded. [appendix B, table 6]

4. A list of ranks of the same length as the vocabulary size was created, with 1 being the most common word.

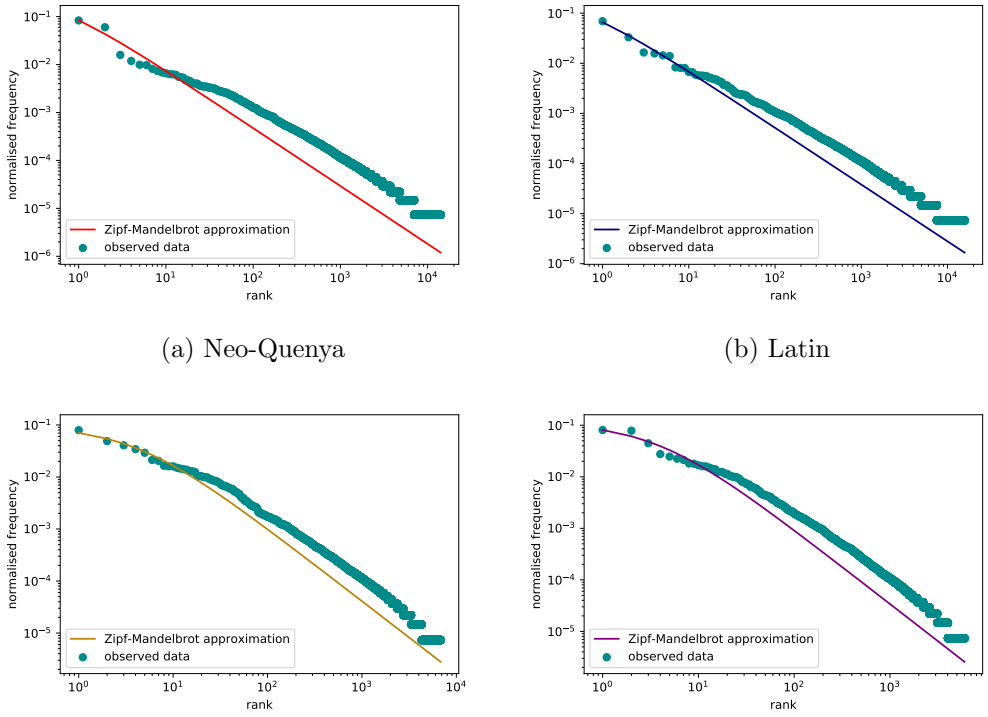
5. The frequencies were stored in an independent list and normalised.

6. Zipf-Mandelbrot parameters were found and a curve was fitted using the Python's *curve_fit* package (Virtanen 2020).

7. The Shannon entropy was calculated for each translation.

5 Results

5.1 Zipf-Mandelbrot Approximations



(a) Neo-Quenya

(b) Latin

(c) New International

(d) King James

Figure 1: Rank against frequency probability distributions with fitted Zipf-Mandelbrot curves for four *New Testament* translations. The Zipf-Mandelbrot curve fits the *New International* translation well. The fit is moderate for the *King James* and Latin translations, and very poor for the Neo-Quenya.

As shown in figure 1, the Zipf-Mandelbrot curve fits nicely for the *New International* translation, less so for the Latin and the *King James*, and very poorly for the Neo-Quenya. This problem was consistent across most subsections. There were some segments where the Zipf-Mandelbrot curve fit the Latin distribution well, but none for the Neo-Quenya translation. An implication of this is that the Zipf-Mandelbrot curve is a poor representation of the probability distribution of Neo-Quenya word frequencies, and so the entropy of the Zipf-Mandelbrot curve was not considered.

The ill-fit of the Zipf-Mandelbrot curve for the Neo-Quenya word frequencies appears to be primarily caused by the two most frequent words. In all four of the translations, these words seem to deviate slightly from the curve, but it appears that this is more evident in the Latin and Neo-Quenya translations. In order to allow for easy comparison between probability distributions, the first two

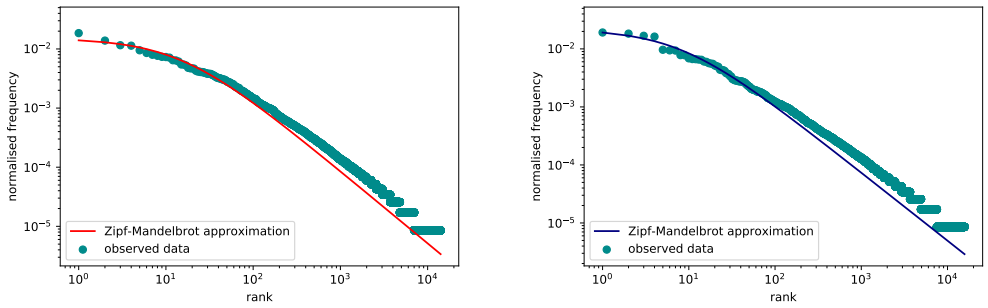
most frequent words of each language were removed, and the Zipf-Mandelbrot curves refitted. For most segments, the words removed are given by table 1.

Table 1: Two most common words in each translation of the *NewTestament*. These words were removed for further comparisons using the Zipf-Mandelbrot distribution.

Translation	Excluded Word	Word Type
Neo-Quenya	*i* (the; that)	definite article; relative pronoun/conjunction
	ar (and; day)	conjunction; noun
Latin	*et* (and)	conjunction
	in (in; on)	preposition
New International	*the*	definite article
	and	conjunction
King James	*and*	conjunction
	the	definite article

As disclosed by table 1, the two most common words in Neo-Quenya have at least two uses each. In Neo-Quenya, *'i'* represents not only the definite article (the), but also a common pronoun and conjunction (that). Likewise, *'ar'* is both a common conjunction (and), and a noun (day) that may be assumed to have fairly regular use in the *New Testament*. Latin, perhaps, deviates because it is a language lacking a definite article in ordinary use.

The words excluded were identical across most segments, with some exceptions. A pronoun, *'you'* replaces *'and'* in the *New International Gospel of John*. In *Letters* and *Paul's Letters*, Latin and Neo-Quenya are consistent, but the removed words for the *New International* and *King James* translations are *'the'*, *'to'* and *'the'*, *'of'* respectively. On a lesser note, the ranks of *'ar'* and *'i'* are inverted in the Neo-Quenya *Gospel of Mark*, as are those of *'and'* and *'the'* in the *King James Gospel of John*.



(a) Neo-Quenya (b) Latin

Figure 2: Rank against frequency probability distributions with fitted Zipf-Mandelbrot Curves for Neo-Quenya and Latin *New Testament* translations, with the two most frequent words removed from the data.

Figure 2 shows the Zipf-Mandelbrot curves fitted to the rank-frequency distributions of Latin and Neo-Quenya with the two most common words removed. Evidently, figure 2 demonstrates a far better fit for Latin and Neo-Quenya than figure 1. This allows for the comparison between Zipf-Mandelbrot curves in figure 3.

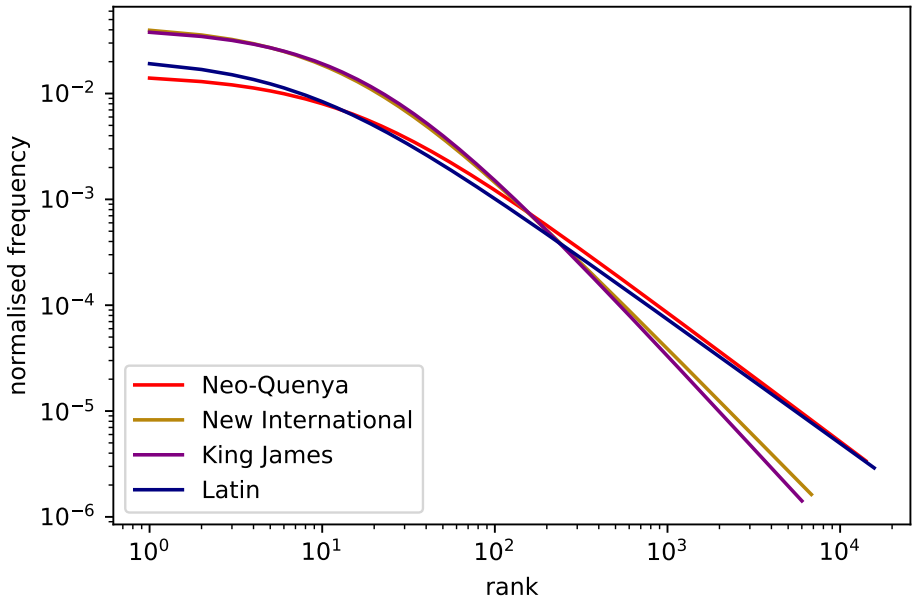


Figure 3: Zipf-Mandelbrot approximations of rank-frequency probability distributions for four *New Testament* translations with the two most common words removed from the data. There is a distinct split between the English and non-English translations.

Figure 3 presents an immediate dichotomy into English and non-English translations. The English distributions are almost identical. The Latin and Neo-Quenya distributions are not quite so similar for higher frequency words, but are very similar for lower frequency words. This dichotomy is consistent across the Zipf-Mandelbrot comparisons for all sections [appendix A, figure 6].

Table 2: Parameters s and q for the Zipf-Mandelbrot probability distributions of four *New Testament* translations with the two most common words removed from each

Translation	s	q
Neo-Quenya	1.22	14.62
Biblia Sacra	1.17	7.84
King James	1.77	18.14
New International	1.66	14.61

The parameters, s and q, for each of the Zipf-Mandelbrot distributions in figure 3 are given in table 2. Again, the English distributions have similar parameters, as do the Latin and the Neo-Quenya

5.2 Shannon Entropy

Table 3: Shannon entropies of each segment for the four *New Testament* translations.

Segment	**Neo-Quenya**	**Biblia Sacra**	**King James**	**New International**
New Testament	10.13	9.78	10.79	10.82
Gospels	9.76	9.57	10.45	10.71
Matthew	9.52	9.35	10.43	10.69
Mark	9.22	8.79	9.94	10.35
Luke	9.56	9.32	10.32	10.65
John	9.01	8.99	10.13	10.13
Acts	9.52	9.16	10.22	10.42
Letters	9.86	9.11	10.66	10.31
Paul's Letters	9.80	7.86	9.36	9.00
Revelation	8.81	8.77	10.01	10.08

Table 3 displays the Shannon entropies for each of the ten segments. Considering all translations, the entropies sit between 7.86 bits per word (*Biblia Sacra Paul's Letters*) and 10.82 bits per word

(*New International New Testament*). Generally, the English translations have similar entropies, as do the Latin and the Elvish. Although the *King James Version* usually has the highest entropies, there is some overlap: the *New International Version* has the highest entropy for *Letters* (10.66 bits per word) and *Paul's Letters* (9.36 bits per word). The Neo-Quenya entropies remain sandwiched between those of the *Biblia Sacra* and the English translations, except in Paul's Letters, where it is higher than the *King James* entropy. The Latin translation always has the lowest entropy.

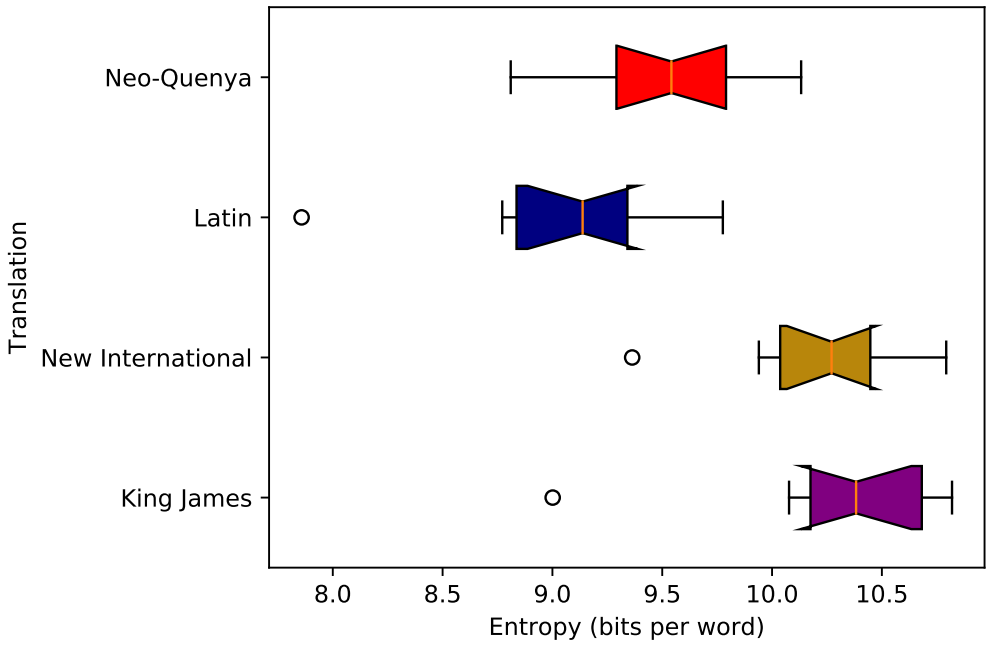


Figure 4: Box plots of Shannon entropies for the segments of four *New Testament* translations. The box plots for the *King James* and *New International* translations are similar. Although closer to each other than to the English translations, the box plots for the Neo-Quenya and Latin translations are not similar. The Neo-Quenya translation has no outlier.

The notched box plots were generated in order to give some indication of the spread of entropies across different writing styles. In figure 4, the notches represent the 95% confidence interval of the median. As in other areas, the dichotomy into English and non-English translations is evident. Upon closer investigation, differences emerge. The medians of the English translations clearly fall within each other's 95% confidence intervals, and so these distributions (and thus the translations) can be said to be statistically similar in terms of entropy. The same cannot be said of the Neo-Quenya and Latin translations. Whilst their entropies and Zipf-Mandelbrot distributions are usually similar, the Neo-Quenya and *Biblia Sacra* translations of the *New Testament* are statistically different. Further-

more, the Neo-Quenya entropies have the smallest range when outliers are included. This suggests that as a fictional language, Quenya lends itself to less diversity in the way it can be used.

As demonstrated by Bentz et al. (2017), vocabulary size has a strong impact on entropy. Potentially, the variation observed amoungst the box plots is merely due to fluctuations in the number of unique words in each segment. Consquently, we investigated the possibility that variation in entropy was due to vocabulary size, rather than writing style.

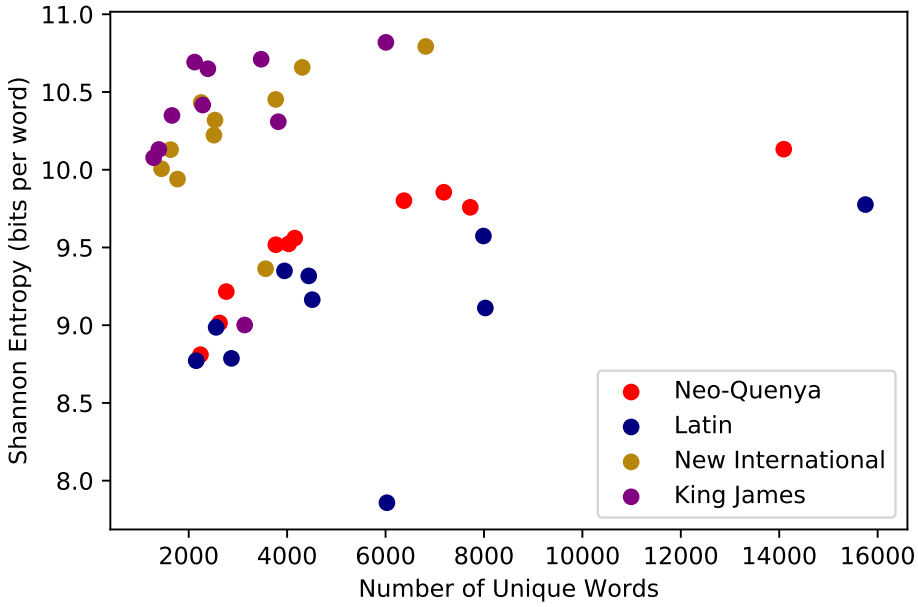


Figure 5: Scatter plot of entropy against vocabulary size for each *New Testament* translation. Only weak correlations are present for the natural language translations. Again, the dichotomy between English and non-English translations may be observed in the two clusters.

Table 4: Pearson's correlation coefficient for relationship between entropy and vocabulary size in four *New Testament* translations.

Translation	Pearson's Correlation Coefficient
Neo-Quenya	0.853
Biblia Sacra	0.471
King James	0.232
New International	0.509

To confirm the low impact of vocabulary size, the vocabulary size of each segment was plotted against its Shannon entropy (figure 5). No strong correlations between entropy and vocabulary size are evident. However, the Pearson's correlation coefficients in table 4 give deeper insight. Entropy and vocabulary size are poorly correlated for the *New International*, *King James* and Latin translations, with coefficients of 0.51, 0.23 and 0.47 respectively. Whilst increasing vocabulary size leads to a general increase in entropy, this trend is not reliable for any of the natural language translations. Conversely, the Neo-Quenya translation has a correlation coefficient of 0.85, suggesting that vocabulary size has a strong impact on the Neo-Quenya entropies. This suggests an entropic divide between Quenya and the natural languages.

6 Discussion

6.1 A Dichotomy in Languages

Generally, the results demonstrated that the Neo-Quenya *New Testament* is not at all similar to either of the English translations, and that it is closer to the Latin translation. Both the entropy box plots (figure 4) and the Zipf-Mandelbrot comparisons (figure 3) demonstrated this split.

The segment entropy box plots (figure 4) displayed an evident dichotomy. The similar distributions of the *King James* and *New International* translations strongly indicate that their entropies are statistically comparable despite great variety in writing style. Whilst the distributions of the Neo-Quenya and Latin segment entropies were both lower than the English distributions, their box plots were statistically dissimilar. The median for the Latin translation lay outside the 95% confidence interval of the Neo-Quenya median and vice versa. The Neo-Quenya and the Latin may be dissimilar due to the fact that they are different languages, or it may be because Latin is a natural language, and Neo-Quenya is constructed. Comparisons with a greater variety of translations would be required to support or negate either conclusion.

As has been observed above, the Zipf-Mandelbrot comparisons after the removal of the two most common words displayed a clear dichotomy between the English and the non-English *New Testament* translations (figure 3). The Zipf-Mandelbrot distributions for the non-English translations started lower, but ended higher than the English. This was reflected in the differences between the total number of words and the number of unique words in each section [appendix B]. For each section, the English translations had by far the higher word count [appendix B, table 5], but when the number

of unique words was tabulated [appendix B, table 6], it was found that they had fewer unique words than the non-English translations. Hence, this leads to higher frequencies in the more common words and a steep drop-off as the word rank is increased. The most obvious explanation for this is inflection.

Inflection in language is essentially a categorisation of grammar structure and the importance of word order. English, as a weakly inflected language, is heavily dependent on word order, articles, prepositions, pronouns and auxiliary verbs in order to correctly convey the information in a sentence. On the other hand, heavily inflected languages (such as Quenya and Latin), usually indicate this information by changing noun and verb endings. In an extreme example:

English: The dog gives a dog to a dog.

Latin: canis canem cani dat.

Here, the different endings to *'canis'* (dog) indicate which dog it is. The *'-is'* ending denotes the dog doing the action, the *'-em'* ending denotes the dog having the action done to it, and the *'-i'* ending denotes the dog having something given to it. A similar thing occurs with verb person, number, mood, tense and aspect. Consequently, inflected languages use fewer words of a greater variety to convey the same information.

It is also worth noting that, as demonstrated by the entropy table (figure 3) and box plot (figure 4), heavily inflected languages appear to have lower Shannon entropies than weakly inflected ones. However, in order to draw a confident conclusion, more translations would need to be analysed.

6.2 Potential Extentions

The possible extensions of this project are many and varied. Ideally, comparisons would be performed between the Neo-Quenya and at least 30 other natural languages, to allow for some indication of statistical significance. A range of inflected and non-inflected languages, as well as non-Indo-European ones could be included, to investigate the relationship between Quenya, languages it is based off (e.g. Latin, Finish and Ancient Greek) (Fauskanger n.d.) and languages that have no deliberate connection (e.g. Mandarin or Wiradjuri). Ideally, the entropies would be calculated in such a way as to make strongly and weakly inflected languages comparable.

Whilst better than letter entropy, word entropy is far from perfect in measuring the entropy of a language. The selection of a word as the token of measure, causes a direct split between heavily and weakly inflected languages. Potentially, the ideal measure for entropy in language would be the morpheme. This would designate the set of tokens as being the set of smallest units of meaning within a language, allowing easier comparison between Quenya and a variety of languages.

A similar investigation could be performed upon the Klingon translation of *Hamlet*, or the *Bible*, once it is complete. The *Bible* would be ideal, as it would allow for comparison between Quenya and Klingon. It would be interesting to see if the two constructed languages displayed any consistent differences when compared with natural languages, as they were constructed in very different ways.

7 Conclusion

In summary, Quenya differs significantly from English in terms of information entropy, but is similar to Latin. Potentially, this is due to the similar levels of inflection between Latin and Quenya. The section entropies of Neo-Quenya are more clustered than those of both the English and the Latin translations, suggesting that the entropy of fictional languages differs less across different writing styles. Additionally, the notched box plots demonstrate that the entropies of the two English tranlsations are statistically similar, and that those of the Neo-Quenya and Latin translations are not. However, a greater number and variety of translations would be required in order to add confidence to these conclusions.

Bibliography

Barnard, G.A. (1955). "Statistical Calculation of Word Entropies for Four Western Languages". In: *IRE Transactions - Information Theory* 1, pp. 49–53.

Bentz, C. et al. (2017). "The entropy of words - learnability and expressivity across more than 1000 languages". In: *Entropy* 19, p. 275.

Biblia Sacra juxta Vulgatam Clementinam (1592). Trans. by Jerome. `https://www.wilbourhall.org/pdfs/vulgate.pdf`. first translated AD 382.

Derdzinski, R, ed. (n.d.). *Quenta Silmarillion Eldalambenen*. http://www.elvish.org/gwaith/silmarillion_project.htm.

Fauskanger, H.K. (n.d.). *Quenya - the ancient tongue*.

Gleick, J. (2011). *The information: a history, a theory, a flood*. London: Fourth Estate.

I Vinya Vere: The New Testament in Neo-Quenya (2015). Trans. by H.K. Fauskanger. `https://folk.uib.no/hnohf/nqnt.htm`.

King James Version (2004). `www.turnbacktogod.com/king-james-bible-kjv-bible-as-pdf/`. first translated 1611.

Mandelbrot, B. (1966). "Information theory and psycholinguistics: a theory of words and frequencies". In: *Readings in Mathematical Social Science*. Ed. by P Lazafeld N Henry. Cambridge MA: MIT Press.

New International Version (1978). `www.turnbacktogod.com/wp-content/uploads/2011/02/NIV-Bible-PDF.pdf`.

"Non-Pauline Letters" (2007). In: *Journal for the Study of the New Testament* 29 (5), pp. 107–112.

Shannon, Claude (1948). "A mathematical theory of communication". In: *The Bell System Technical Journal* 27 (3), pp. 379–423.

The Holy Bible: New International Version with concise bible encyclopaedia (2007). Minto, N.S.W: Bible Society in Australia.

Tolkien, J.R.R. (1994). "Quendi and Eldar". In: *The War of the Jewels*. Ed. by C Tolkien. London: Harper Collins.

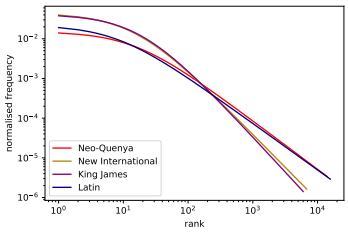
Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods*.

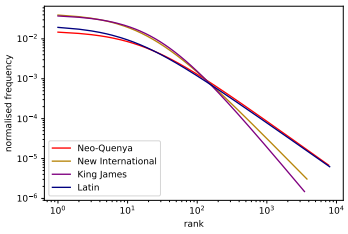
Zipf, G.K. (1949). *Human behaviour and the principle of least effore*. Addison-Wesley Press.

Appendices

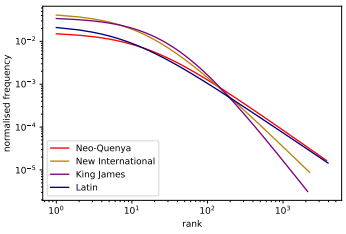
A Zipf-Mandelbrot Comparisons



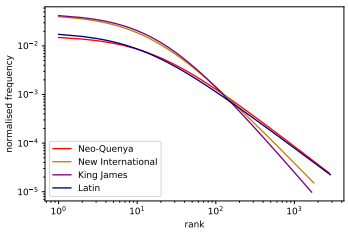
(a) New Testament



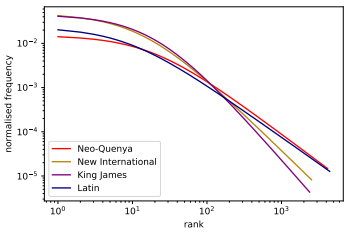
(b) Gospels



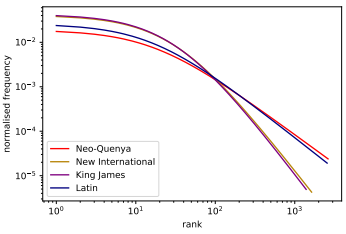
(c) Gospel of Matthew



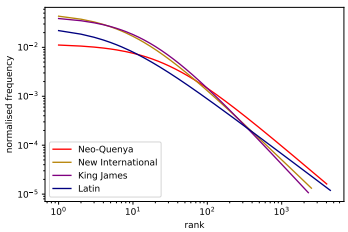
(d) Gospel of Mark



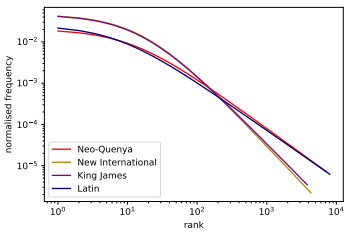
(e) Gospel of Luke



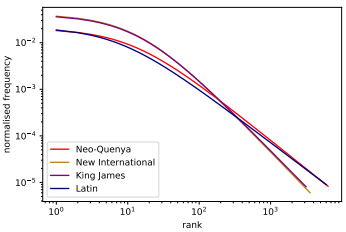
(f) Gospel of John



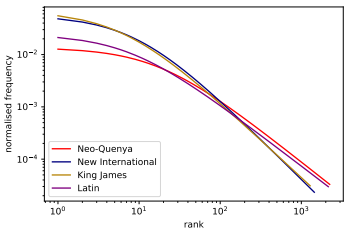
(g) Acts of the Apostles



(h) Letters



(i) Paul's Letters



(j) Revelation

Figure 6: Zipf-Mandelbrot approximations for the frequency agaisnt rank distributions of the segments of four *New Testament* translations. The two most common words have been removed from each segment and translation.

B Total and Unique Word Counts for Four *New Testament* Translations

Table 5: Total number of words in each segment for four *New Testament* translations.

Segment	Neo-Quenya	Biblia Sacra	King James	New International
New Testament	136 384	126 575	174 422	180 279
Gospels	61 523	58 892	78 804	83 652
Matthew	17 364	16 553	22 569	23 690
Mark	11 006	10 151	13 537	14 917
Luke	18 990	18 090	24 128	25 945
John	14 163	14 098	18 570	19 100
Acts	18 492	16 785	22 917	24 249
Letters	47 649	42 409	61 411	60 388
Paul's Letters	39 810	30 648	44 424	43 405
Revelation	8720	8489	11 290	11 990

Table 6: Number of unique words in each segment for four *New Testament* translations.

Segment	Neo-Quenya	Biblia Sacra	King James	New International
New Testament	14 092	15 752	6817	6007
Gospels	7721	7989	3768	3472
Matthew	3770	3946	2253	2120
Mark	2763	2868	1772	1659
Luke	4155	4439	2534	2389
John	2626	2557	1632	1393
Acts	4031	4510	2512	2284
Letters	7183	8028	4307	3820
Paul's Letters	6374	6028	3562	3137
Revelation	2238	2152	1447	1287