is that the news article, despite all of its name and numbers, had the best per word average quality. This may be due to it being written by a professional journalist so there will be very few poor qualities to begin with. It is surprising that the academic paper is so much lower that this with an average per-word quality of just 962, compared to the news article's 1635. A possible explanation for this is that text contains a large amount of punctuation, and some technical terms as well as names and numbers, which affect the quality test as specialised terms are not common.

5.6 Results Discussion

As can be seen from the data above, there is a noticeable difference between the output of the algorithm when applying different bitstreams to the same text. In these tests, the second bitstream generally generated a lower bitrate than the first, and all (apart from the fiction text, which experienced a slight increase), resulted in a lower quality than when using the first bitstream.
The aim for the bitrate was to achieve a rate of around one bit for every 10 words. While, on average, the performance of the algorithm is just shy of that amount, it is very close. Apart from the false positives and the false negatives, there were no cases where the algorithm deobfuscated the wrong bits from a word.
On examining the output from the algorithms (which is included with the original text on the cd), there are cases where the outputted words are not suitable for the surrounding text. While generally the sentence still has meaning, the words will cause suspicion amongst human readers. This is tested in more detail though the user survey in section 6.7.

I will now look at some of the more interesting results in more detail:

5.6.1 False Positives

There is a known reason why false positives occur when deobfuscating. There are a number of words which will always appear to contain bits even if they do not. This is often a case with words that can only contain a 0 (but the bitstream required a 1), and so fail the deobfuscation test when obfuscating.
A possible fix for this would be to implement a blacklist of words which have this affect. These could be found by running the algorithm on a much larger corpus than the test data and logging the words where this occurs. These can then be ignored when obfuscating and deobfuscating. This may limit the bitrate slightly but will improve the correctness of the algorithm.

5.6.2 Improved Quality

Perhaps the most surprising result of the tests is the improvement in the quality value of the texts using the bigram frequency data. The reason that this has occurred is due to the