Results

	Bitstream 1	**Bitstream 2**
Words	1969	1969
Total Bitrate	190	182
Average bitrate (/10 words)	0.96	0.92
False Positives	23	30
False Negatives	2	2
True Positives	167	161
True Negatives	1777	1776
Error Count	0	0
Original Total Quality	2135362	2135362
Original Average Quality	1084	1084
Obfuscated Total Quality	2253878	2256190
Obfuscated Average Quality	1144	1145

Table 5.4: Results for a Piece of Fiction

Analysis

While the false positive rate seems high, it is not more than expected for a block of text of that length. The unexpected result is the false negative rate, meaning that for two words in which data was hidden no data was found. Having run the tests again there is no obvious reason for this to occur, no bits are found in the words. The only explanation that could be a possibility is in storing the output some extra whitespace has appeared, and has not been removed by the calls to the trim() function. More investigation will be required to find this problem. This is still a very small issue as it only affects 2 words in 1900 (less than 0.1% of the words). The quality of the text (in the eyes of automatic analysis using bigram frequencies) has again improved.

5.5 Comparison of Documents

There are some differences between the four different documents. The bitrate varies between them by as much as 0.2 bits for every 10 words. The lower bitrates can be found in the news article and the academic paper extract. This is due to the number of names, technical terms and numbers in these two documents, which are not contained in the WordNet dictionary. These terms also affect the quality ratings for the surrounding words, limiting the bitrate further. The USENET postings and book extract, however, contain less of these specialised terms and so there are more words that can have bits included.
All four of the documents experienced an increase in the quality after they had been processed, due to the quality testing and quality centred synonym retrieval. What was surprising