## 5.4 Statistical Evaluation Results

### 5.4.1 News Article

The news article used was taken from the BBC website, on the subject of the London 2012 Paralympic Games [24].

#### Results

| | **Bitstream 1** | **Bitstream 2** |
|---|---|---|
| **Words** | 857 | 857 |
| **Total Bitrate** | 64 | 60 |
| **Average bitrate** (/10 words) | 0.7 | 0.7 |
| **False Positives** | 10 | 13 |
| **False Negatives** | 0 | 0 |
| **True Positives** | 57 | 55 |
| **True Negatives** | 790 | 789 |
| **Error Count** | 0 | 0 |
| **Original Total Quality** | 1340403 | 1340403 |
| **Original Average Quality** | 1564 | 1564 |
| **Obfuscated Total Quality** | 1401203 | 13977089 |
| **Obfuscated Average Quality** | 1635 | 1630 |

Table 5.1: Results for News Article

#### Analysis

This piece of data provided a slightly disappointing bitrate of 0.7 bits for every 10 words. A possible explanation of this is that the text contains a large number of names and numerical data, such as times. These are generally not in the WordNet database and so will not be able to contain data, and also may cause words that are suitable for obfuscation to fail the quality checks if the follow or proceed a name or time.
There were no incorrect bits found, which is a good sign, although there were 10 cases where words were found to have bits where there were none.
A surprising result is in the quality tests. In this piece of text, the quality was not decreased, in fact there is a noticeable improvement in the quality score! This will help prevent against automatic detection. There are, however, some cases in the output where the replacement word will arise suspicion amongst humans who carefully read the text, such as "**fourth** gold medal" being replaced with "**quarter** gold medal", and "four **million**" replaced by "four **billion**".