way in which the synonym sets are built. The lists are ordered according to the frequency of each possible word and the word before (in the text). This means that whenever a word is substituted, the word will be one of the top three in this list (depending on which bit is hidden). Quite often, these will be of higher quality that the original word. This, combined with the test performed during obfuscation (which does not hide bits in words if the resulting quality is too low) is the reason that the quality generally increases.

5.6.3 False Negatives

After more inspection of the output of the program, this appears to be due to some bug in the implementation. The example from the usenet output is the word "short", preceded by "a". The program is not able to generate a synset for this word when deobfuscating. The synsets are generated in exactly the same way when deobfuscating as when obfuscating, and the words where this issue occurs all generate synsets while obfuscating. This is a low priority issue as the program and algorithm are only a prototype and this issue appears to be related to the implementation rather than the design of the algorithm. It is also worth noting that Chang and Clark [4] experience a small number of false negatives also.

5.7 User Survey

To test the robustness to human analysis aim of the algorithm a short survey was prepared which contains output sentences from the documents used for the statistical testing. The subject is presented with 10 sentences, split into two parts. For the first part they do not know anything about the algorithm, and for the second part they receive a brief explanation. The subject is asked to first give the sentence a brief glance and assign the sentence a score of 1-10, where 1 means that the sentence seems perfectly normal and 10 means the sentence does not make any sense at all or seems completely wrong. They are then asked to study the sentence in greater detail, and provide the score again. They are then free to make any comments on the sentence, and are encouraged to circle any words that they feel are incorrect or out of place.
As a control, in each section one of the sentences (number 3) are the original text without the algorithm applied and no data hidden. This is control data to test if the algorithm does truly impact the text. There is at least one sentence in each section from each of the four documents processed.
The survey is visible as Appendix A.

5.7.1 Results

The survey was completed by six people. Two of these were non-English native speakers. The average score was taken for each sentence and is presented in the table below. It is