Analysing Style - Readability

Paul Clough, October 2000 [cloughie@dcs.shef.ac.uk]

CONTENTS

1 INTRODUCTION................................................................................................................................................. 2
  1.1 AVERAGE SENTENCE LENGTH .................................................................................................................. 2
  1.2 READABILITY SCORES ............................................................................................................................... 2
  1.3 FLESCH READING EASE SCORE .................................................................................................................. 2
  1.4 FLESCH-KINCAID GRADE LEVEL SCORE ................................................................................................. 3
2 COMPARING THE RESULTS ........................................................................................................................................ 3
  2.1 INTRODUCTION ........................................................................................................................................... 3
  2.2 CALCULATING THE Z-SCORE .................................................................................................................... 4
  2.3 CALCULATING A SINGLE SCORE FROM THE THREE ATTRIBUTE VALUES .......................................... 5
3 APPLYING HIERARCHICAL CLUSTERING TO THE RESULTS ................................................................................. 6
  3.1 DENDROGRAMS .......................................................................................................................................... 6
  3.2 CREATING THE DENDROGRAMS USING SPSS ....................................................................................... 6
  3.3 RESULTS – WITH ALL FOUR ATTRIBUTES ON THE COURTS DOMAIN ..................................................... 7
  3.4 RESULTS – WITH ALL FOUR ATTRIBUTES ON THE SHOWBIZ DOMAIN .................................................. 8
  3.5 RESULTS - WITHOUT THE PASSIVE SENTENCE SCORES ON THE COURTS DOMAIN ........................... 9
  3.6 RESULTS - WITHOUT THE PASSIVE SENTENCE SCORES ON THE SHOWBIZ DOMAIN ........................... 9
4 CONCLUSIONS .................................................................................................................................................... 10
  4.1 INTRODUCTION .......................................................................................................................................... 10
  4.2 FROM TABLE 1 AND 2 ................................................................................................................................. 10
  4.3 FROM THE DENDROGRAMS ..................................................................................................................... 10
  4.4 PROBLEMS WITH READABILITY ............................................................................................................. 11
  4.5 FURTHER WORK ....................................................................................................................................... 11
  4.6 BENEFITS OF MEASURING READABILITY ............................................................................................... 12
5 BIBLIOGRAPHY ................................................................................................................................................... 15

1 APPENDIX A – DIFFERENT VERSIONS OF THE HAMILTON COURT CASE FROM THE METER CORPUS .............. 16
  1.1 THE PA VERSION ....................................................................................................................................... 16
  1.2 THE SUN VERSION ..................................................................................................................................... 16
  1.3 THE MIRROR VERSION ............................................................................................................................... 17
  1.4 THE STAR VERSION .................................................................................................................................... 17
  1.5 THE MAIL VERSION .................................................................................................................................... 18
  1.6 THE EXPRESS VERSION .............................................................................................................................. 19
  1.7 THE TIMES VERSION ................................................................................................................................... 19
  1.8 THE TELEGRAPH VERSION ....................................................................................................................... 20
  1.9 THE GUARDIAN VERSION .......................................................................................................................... 21
  1.10 THE INDEPENDENT VERSION .................................................................................................................. 22
1 Introduction

The aim of this short experiment is to investigate four simple stylometric measures for classifying documents and apply this to the METER corpus. The four measures used for each newspaper and the PA are: average sentence length, the number of passive sentences, the Flesch readability measure and the Flesch-Kincaid grade level. The last two metrics readability measures and interrelated with average sentence length to determine a score.

Using the results from these measures, a hierarchical clustering technique is applied to determine:

1) Which newspapers are more similar to others based upon the four measures (or a combination of measures).
2) Which newspapers are “closest” to the PA according to these readability measures.
3) Whether there are differences in the clustering patterns for the courts and showbiz domains.

The results are given in Tables 1 and 2.

1.1 Average sentence length

The average sentence length is expressed in words/sentence. This is computed for each newspaper and the PA using my own rule-based sentence splitting algorithm (called split.pl). More information can be found in (Clough2000). The program also extracts text for each newspaper and the PA automatically from the METER corpus and saves these as separate files used to calculate readability scores.

1.2 Readability scores

Readability scores are supposed to be a simple method for indicating the reading level of a document (Stephens2000). The reading level describes the ease with which a document can be read (not how comprehensible a piece of text is to the reader). Examples of popular measures include the Flesch readability metric and the Gunning FOG index. Two readability measures can be extracted using the grammar checking facilities in Microsoft Word 97 - the Flesch Reading Ease and the Flesch-Kincaid Grade Level. Each readability score bases its rating on the average number of syllables per word and words per sentence. These are described in Sections 1.3 and 1.4.

1.3 Flesch Reading Ease score

Rates text on a 100-point scale; the higher the score, the easier it is to understand the document. For most standard documents, aim for a score of approximately 60 to 70.

The approach to calculating the Flesch score is:

1. Calculate the average sentence length, L.
2. Calculate the average number of syllables per word, N.
3. Calculate score (between 0-100%).
1.4 Flesch-Kincaid Grade Level score

Rates text on a U.S. grade-school level. For example, a score of 8.0 means that an eighth grader can understand the document. For most standard documents, aim for a score of approximately 7.0 to 8.0. Grade Level = (L x 0.39) + (N x 11.8) –15.59 years.

2 Comparing the results

2.1 Introduction

To allow comparison between the results, the values are normalised using a Z score (Oakes1988). This allows values on a different measurement scale (for example percentages and numeric values between a set range) to be compared. The aim of the comparison is to enable the application of a hierarchical clustering algorithm (in this case a dendrogram) to find which newspapers are most “similar” to others based upon the attributes: sentence length, number of passive sentences used, Flesch reading ease and Flesch-Kincaid grade level.

<table>
<thead>
<tr>
<th>Source</th>
<th>Using split.pl</th>
<th>Using Microsoft Word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average W/S (num. of sentences)</td>
<td>Passive sentences</td>
</tr>
<tr>
<td>pa</td>
<td>23.45 (8719)</td>
<td>6%</td>
</tr>
<tr>
<td>other</td>
<td>21.93 (58)</td>
<td>21%</td>
</tr>
<tr>
<td>times</td>
<td>21.16 (1636)</td>
<td>20%</td>
</tr>
<tr>
<td>star</td>
<td>18.25 (788)</td>
<td>21%</td>
</tr>
<tr>
<td>express</td>
<td>20.43 (1076)</td>
<td>20%</td>
</tr>
<tr>
<td>mirror</td>
<td>17.31 (1002)</td>
<td>20%</td>
</tr>
<tr>
<td>mail</td>
<td>20.33 (1556)</td>
<td>20%</td>
</tr>
<tr>
<td>guardian</td>
<td>22.04 (1741)</td>
<td>21%</td>
</tr>
<tr>
<td>telegraph</td>
<td>21.78 (2103)</td>
<td>24%</td>
</tr>
<tr>
<td>sun</td>
<td>17.15 (1083)</td>
<td>18%</td>
</tr>
<tr>
<td>independent</td>
<td>21.03 (1245)</td>
<td>25%</td>
</tr>
</tbody>
</table>

Table 1 Results for the courts domain

<table>
<thead>
<tr>
<th>Source</th>
<th>Using split.pl</th>
<th>Using Microsoft Word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average W/S (num. of sentences)</td>
<td>Passive sentences</td>
</tr>
<tr>
<td>pa</td>
<td>22.34 (1484)</td>
<td>3%</td>
</tr>
<tr>
<td>other</td>
<td>0 (0)</td>
<td>0</td>
</tr>
<tr>
<td>times</td>
<td>20.36 (144)</td>
<td>13%</td>
</tr>
<tr>
<td>star</td>
<td>18.53 (412)</td>
<td>10%</td>
</tr>
<tr>
<td>express</td>
<td>20.00 (256)</td>
<td>10%</td>
</tr>
<tr>
<td>mirror</td>
<td>18.80 (251)</td>
<td>10%</td>
</tr>
<tr>
<td>mail</td>
<td>20.67 (309)</td>
<td>11%</td>
</tr>
<tr>
<td>guardian</td>
<td>20.64 (184)</td>
<td>13%</td>
</tr>
<tr>
<td>telegraph</td>
<td>19.39 (154)</td>
<td>11%</td>
</tr>
<tr>
<td>sun</td>
<td>17.16 (353)</td>
<td>12%</td>
</tr>
<tr>
<td>independent</td>
<td>22.33 (160)</td>
<td>16%</td>
</tr>
</tbody>
</table>

Table 2 Results for the showbiz domain
2.2 Calculating the Z-Score

The Z-score is the number of standard deviations from the mean and is calculated using (1). In this equation, Z is the Z score, x is the value for an attribute (e.g. the times has on average 21.16 words per sentence in the courts domain) and \( \bar{x} \) the average value for this attribute (e.g. the average number of words per sentence for all the newspapers and pa).

\[
Z = \frac{x - \bar{x}}{\sigma^2}
\]  

(1)

The mean is calculated for an attribute such as the mean number of words per sentence and is calculated from equation (2) where N is the number of samples (in this case the number of source for the METER corpus – 11).

\[
x = \frac{\sum_{i=1}^{N} x_i}{N}
\]  

(2)

The standard deviation \( \sigma^2 \) is calculated for each value from equation (3).

\[
\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}
\]  

(3)

The Z scores are shown in Tables 3 and 4.

<table>
<thead>
<tr>
<th>Source (# sentences)</th>
<th>Average W/S</th>
<th>Passive sentences</th>
<th>Flesch Reading Ease</th>
<th>Flesch-Kincaid Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>Z score</td>
<td>Value</td>
<td>Z score</td>
</tr>
<tr>
<td>pa</td>
<td>23.45</td>
<td>1.47</td>
<td>6%</td>
<td>-2.77</td>
</tr>
<tr>
<td>other</td>
<td>21.93</td>
<td>0.73</td>
<td>21%</td>
<td>0.28</td>
</tr>
<tr>
<td>times</td>
<td>21.16</td>
<td>0.35</td>
<td>20%</td>
<td>0.07</td>
</tr>
<tr>
<td>star</td>
<td>18.25</td>
<td>-1.07</td>
<td>21%</td>
<td>0.28</td>
</tr>
<tr>
<td>express</td>
<td>20.43</td>
<td>-0.005</td>
<td>20%</td>
<td>0.07</td>
</tr>
<tr>
<td>mirror</td>
<td>17.31</td>
<td>-1.53</td>
<td>20%</td>
<td>0.07</td>
</tr>
<tr>
<td>mail</td>
<td>20.33</td>
<td>-0.22</td>
<td>20%</td>
<td>0.07</td>
</tr>
<tr>
<td>guardian</td>
<td>22.04</td>
<td>-0.20</td>
<td>21%</td>
<td>0.28</td>
</tr>
<tr>
<td>telegraph</td>
<td>21.78</td>
<td>0.65</td>
<td>24%</td>
<td>0.89</td>
</tr>
<tr>
<td>sun</td>
<td>17.15</td>
<td>-1.60</td>
<td>18%</td>
<td>-0.33</td>
</tr>
<tr>
<td>independent</td>
<td>21.03</td>
<td>0.29</td>
<td>25%</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>20.44</td>
<td>20%</td>
<td>58%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.05</td>
<td>0.05</td>
<td>0.06</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Table 3  Z scores for the courts domain
2.3 Calculating a single score from the three attribute values

Once the individual Z scores are calculate, the next stage is to compute a single value from the three Z scores for each newspaper. This is then used to compare with other Z score values. The single value is found by computing the square of the difference between the Z score for each attribute, then summing the values to create a single score – equation (4). The sum is the comparison of each Z score for each attribute (i = 1 to 4). This must be computed for all combinations of scores and the result is entered into a symmetrical 2x2 matrix (Tables 5 and 6).

\[ Z = \sum_{i=1}^{4} (Z(x_i) - Z(y_i))^2 \]  \hspace{1cm} (4)

<table>
<thead>
<tr>
<th>Source</th>
<th>Average W/S</th>
<th>Passive sentences</th>
<th>Flesch Reading Ease</th>
<th>Flesch-Kincaid Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>Z score</td>
<td>Value</td>
<td>Z score</td>
</tr>
<tr>
<td>pa</td>
<td>22.34</td>
<td>0.69</td>
<td>3%</td>
<td>-1.54</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
<td>-2.89</td>
<td>0</td>
<td>-2.18</td>
</tr>
<tr>
<td>times</td>
<td>20.36</td>
<td>0.37</td>
<td>13%</td>
<td>0.60</td>
</tr>
<tr>
<td>star</td>
<td>18.53</td>
<td>0.08</td>
<td>10%</td>
<td>-0.04</td>
</tr>
<tr>
<td>express</td>
<td>20.00</td>
<td>0.32</td>
<td>13%</td>
<td>0.60</td>
</tr>
<tr>
<td>mirror</td>
<td>16.80</td>
<td>-0.20</td>
<td>10%</td>
<td>-0.04</td>
</tr>
<tr>
<td>mail</td>
<td>20.76</td>
<td>0.44</td>
<td>11%</td>
<td>0.18</td>
</tr>
<tr>
<td>guardian</td>
<td>20.64</td>
<td>0.42</td>
<td>13%</td>
<td>0.60</td>
</tr>
<tr>
<td>telegraph</td>
<td>19.39</td>
<td>0.22</td>
<td>11%</td>
<td>0.18</td>
</tr>
<tr>
<td>sun</td>
<td>17.16</td>
<td>-0.14</td>
<td>12%</td>
<td>0.40</td>
</tr>
<tr>
<td>independent</td>
<td>22.33</td>
<td>0.69</td>
<td>16%</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Mean \( (\bar{x}) \): 18.03 10% 57% 8.12

Std Dev \( (s^2) \): 6.25 0.05 0.19 3.13

Table 4 Z scores for the showbiz domain

<table>
<thead>
<tr>
<th>pa</th>
<th>other</th>
<th>times</th>
<th>star</th>
<th>express</th>
<th>mirror</th>
<th>mail</th>
<th>guardian</th>
<th>telegraph</th>
<th>sun</th>
<th>independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>33.87</td>
<td>21.50</td>
<td>21.19</td>
<td>19.15</td>
<td>20.52</td>
<td>20.13</td>
<td>27.17</td>
<td>29.40</td>
<td>18.79</td>
<td>35.47</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
<td>2.32</td>
<td>10.27</td>
<td>4.74</td>
<td>14.52</td>
<td>4.80</td>
<td>2.17</td>
<td>1.48</td>
<td>16.10</td>
<td>1.24</td>
</tr>
<tr>
<td>times</td>
<td>0</td>
<td>3.53</td>
<td>0.47</td>
<td>6.16</td>
<td>0.59</td>
<td>0.48</td>
<td>0.93</td>
<td>7.08</td>
<td>1.07</td>
<td>1.80</td>
</tr>
<tr>
<td>star</td>
<td>0</td>
<td>1.59</td>
<td>0.42</td>
<td>1.26</td>
<td>3.23</td>
<td>5.93</td>
<td>0.96</td>
<td>6.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>express</td>
<td>0</td>
<td>3.42</td>
<td>0.05</td>
<td>0.98</td>
<td>2.18</td>
<td>4.12</td>
<td>3.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>0</td>
<td>2.95</td>
<td>5.76</td>
<td>9.52</td>
<td>0.19</td>
<td>10.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mail</td>
<td>0</td>
<td>0.81</td>
<td>2.27</td>
<td>3.64</td>
<td>3.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guardian</td>
<td>0</td>
<td>1.10</td>
<td>6.88</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>telegraph</td>
<td>0</td>
<td>11.26</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>0</td>
<td></td>
<td>12.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>independent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 2x2 Matrix of single Z scores for the courts domain (symmetric)
3 Applying hierarchical clustering to the results

3.1 Dendrograms

The clustering approach used to compare the results is the dendrogram (Oakes1988). This is a form of hierarchical clustering, the result being a tree-like categorisation where small clusters of similar documents are included within larger clusters of less similar documents. The dendrogram uses an agglomerative strategy to form the clusters which means individual documents are compared and fused to the most similar to form a cluster. This continues until all documents have been processed. This is bottom-up approach where individual documents are grouped first. This compares to a divisive clustering technique where the overall cluster is started with and continually divided until left with individual documents. This is a top-down approach.

The simple dendrogram algorithm is:

1. FOREACH document pair find \( \text{sim}(i, j) \) – the entry in the similarity matrix DO
2. Search the similarity matrix to identify the most similar remaining pair of clusters.
3. Fuse this pair K and L to form a new cluster KL.
4. Update \( \text{sim} \) by calculating the similarity between the new cluster and each of the remaining clusters.
5. UNTIL only one cluster left.

I experimented with two similarity metrics: Average Linkage and the Ward Method. I also used a Between Groups and Nearest Neighbour interval measure.

3.2 Creating the dendrograms using SPSS

To create the dendrograms, SPSS for Windows version 10.0.7 was used. This tool removes the need to calculate all the Z transform scores due to a normalisation process applied to the original data before clustering. The SPSS tool allows the input of Table 1 and 2 and applies all necessary transformations to allow comparison of the data sets.

To create the dendrogram, select Analyze > Classify… > Hierarchical Cluster. Different similarity metrics, transformations and interval measures can be selected to produce the clustering.
3.3 Results – with all four attributes on the courts domain

* * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S *

Dendrogram using Average Linkage (Between Groups) for COURTS

<table>
<thead>
<tr>
<th>CASE</th>
<th>Label</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>express</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>mail</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>times</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>guardian</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>telegraph</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>independen</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>mirror</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>sun</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>star</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>pa</td>
<td>1</td>
</tr>
</tbody>
</table>

* * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S *

Dendrogram using Ward Method for COURTS

<table>
<thead>
<tr>
<th>CASE</th>
<th>Label</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>express</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>mail</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>times</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>guardian</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>star</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>mirror</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>sun</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>telegraph</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>independen</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>pa</td>
<td>1</td>
</tr>
</tbody>
</table>
3.4 Results – with all four attributes on the showbiz domain

The other category was omitted because the 0 scores distorted the clustering.

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S *

Dendrogram using Average Linkage (Between Groups) for SHOWBIZ

<table>
<thead>
<tr>
<th>Label</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td>times</td>
<td>2</td>
</tr>
<tr>
<td>guardian</td>
<td>7</td>
</tr>
<tr>
<td>express</td>
<td>4</td>
</tr>
<tr>
<td>star</td>
<td>3</td>
</tr>
<tr>
<td>telegraph</td>
<td>8</td>
</tr>
<tr>
<td>mail</td>
<td>6</td>
</tr>
<tr>
<td>mirror</td>
<td>5</td>
</tr>
<tr>
<td>sun</td>
<td>9</td>
</tr>
<tr>
<td>indepen</td>
<td>10</td>
</tr>
<tr>
<td>pa</td>
<td>1</td>
</tr>
</tbody>
</table>

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S *

Dendrogram using Ward Method for SHOWBIZ

<table>
<thead>
<tr>
<th>Label</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td>times</td>
<td>2</td>
</tr>
<tr>
<td>guardian</td>
<td>7</td>
</tr>
<tr>
<td>express</td>
<td>4</td>
</tr>
<tr>
<td>star</td>
<td>3</td>
</tr>
<tr>
<td>telegraph</td>
<td>8</td>
</tr>
<tr>
<td>mail</td>
<td>6</td>
</tr>
<tr>
<td>mirror</td>
<td>5</td>
</tr>
<tr>
<td>sun</td>
<td>9</td>
</tr>
<tr>
<td>indepen</td>
<td>10</td>
</tr>
<tr>
<td>pa</td>
<td>1</td>
</tr>
</tbody>
</table>
3.5 Results - without the passive sentence scores on the courts domain

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * *

Dendrogram using Ward Method for COURTS domain

Rescaled Distance Cluster Combine

<table>
<thead>
<tr>
<th>CASE</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Num</td>
<td>+------------------+-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>express</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mail</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>times</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>indepen</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guardian</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>telegraph</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pa</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.6 Results - without the passive sentence scores on the showbiz domain

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * *

Dendrogram using Ward Method for SHOWBIZ

Rescaled Distance Cluster Combine

<table>
<thead>
<tr>
<th>CASE</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Num</td>
<td>+------------------+-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>times</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mail</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guardian</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>express</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>telegraph</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pa</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>indepen</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4 Conclusions

4.1 Introduction

In this short experiment, I have applied a simple readability score to the newspapers and PA text of the METER corpus. I have made use of off-the-shelf word processing and statistical tools to provide the score and cluster the results. Although I used my own sentence splitter to calculate average sentence length, I selected some documents at random and the results from both Microsoft Word and my own sentence splitter were close enough not to cause a large degree of error. This could have affected the results and perhaps I should have relied only on Microsoft Word to provide all the scores for consistency. My aim for this experiment was to simply use the resulting readability scores as the basis on which to apply a hierarchical clustering algorithm.

The results are briefly described in the following sections.

4.2 From Table 1 and 2

The PA has the greatest average sentence length, but comes highest in readability. This must be down to the use of words with few syllables. This is perhaps not surprising as the PA reports are very factual and do not add much journalese; this is performed by the newspaper.

The Sun comes next as the most readable, followed by the Mirror then Star. These are essentially the tabloids that express the news in a manner readable by a wide audience of varying reading ability. The most difficult newspaper to read according to these results is the other category that consists of the Financial Times among others. This is also not surprising as this paper assumes an audience of higher reading ability. The next hardest is the Independent.

The results are similar for both the courts and showbiz domain. The easiest paper to read in the showbiz domain is the Mirror, this time followed by the Star and then Sun. Again the tabloids are easier to read according to these figures. The hardest newspaper to read is the Independent, followed by the Times. The PA again comes out as the most readable source.

What is perhaps revealing is the differences between the readability scores for the two domains. This is perhaps not surprising as a law and court reporting domain will use perhaps more polysyllabic words and perhaps longer sentences in form of quotes.

There is a significant difference between the use of passive sentences between the PA and the newspapers. I think this will be due to the PA making more use of the active voice to describe events as they are occurring, real-time.

Appendix A contains different newspaper and PA versions of the same news story (except the other category). Make your own judgement on which you think is the most readable.

4.3 From the dendrograms

The biggest difference between the results for the dendrograms is the use of the passive sentence scores within the clustering. Due to the large difference between the PA and the newspapers, this always leaves the PA separate that only clusters at the highest level. Therefore made several clusters using the passive sentence scores and not.

A general consistent cluster that does run throughout the dendrograms is the (Sun, Mirror, Star) cluster. This is not surprising as these tabloids do use similar language that is different from the broadsheets and the Mail and Express that used to be broadsheets but are now somewhere in
between a tabloid and broadsheet. I am surprised that using a simple metric such as readability has been able to detect the differences between the papers so well and perhaps helps to validate the use of these metrics for classifying the newspapers.

If the passive sentence scores are not used, the PA seems to group with the other category in the courts domain, and the independent in the showbiz domain. In both cases, the PA is most dissimilar with the tabloids that again affirms that the tabloids will radically alter the PA supplied text to create stories that fit their audience. Again, I am not surprised that the PA groups with the broadsheets as they are more likely to simply cut-and-paste text from the PA.

What is perhaps surprising is that if the PA is the most readable and is used verbatim or with little change mostly by the tabloids, why do they have such a low readability score? I can only explain this as either an error in my experiments (perhaps all formatting from the PA was not successfully removed) or because of the large differences in sentences obtained for the PA and other category samples.

4.4 Problems with readability

Although this experiment has been useful, I would still doubt the use of readability figures as being conclusive evidence for being able to rank the newspapers in an order of readability. For a start, the measures do not account from comprehensibility. For example, the sentence “Humpty Dumpty sat on the wall” has the same Flesch scores as “Dumpty sat the Humpty wall on” (Flesch – 87.9%, Flesch-Kincaid – 2.4) which is not very helpful. Also, the scores are only based upon one semantic factor – the number of syllables per word. No account is taken of the correctness of the sentence, or the vocabulary used.

I suggest the use of the grammatical structure and vocabulary to make the scores more useful for accurate measurement of readability. For example, the parse depth or the type of sentence (simple, complex, compound etc.). However, these scores are widely used by education and industry and therefore it is useful that these are used to rank the newspapers and PA as they can be calculated using most common word processing and statistical tools.

Also, I compared the newspapers and PA with the first five books of the Bible (NIV version) to compare the results. I found that the tabloids seemed to again cluster with the readability scores obtained from the NIV (New International Version – one of the most popular versions) Bible and this is perhaps not surprising as the sentences in these books are fairly short and the language relatively simple. Moses uses words that can be understood by most people of the time. However the actual message these books convey is perhaps more difficult to understand than the newspapers. An example excerpt from the Sun and Genesis is shown in Figures 1 and 2.

Again there is a difference between the dendrograms using passive sentences and not as with passive sentences, the PA is also clustered with four of the Bible books and again this is because the variance of passive sentences is very large and affects the other scores.

4.5 Further work

This has only been a very quick study of newspaper and PA text with readability scores, but there are several issues raised that would be good to investigate further. The first is the varying sizes of newspaper and PA samples. It would be interesting to select randomly a set number of sentences from each newspaper and PA, e.g. 100 sentences and measure the readability of these to see whether the figures change. This would apply to texts selected from each domain.

It would also be interesting to calculate the number of syllables per word for each newspaper and PA and see whether this fits in with the readability scores obtained.
Another test would be to calculate the distribution of style measures used to see how the
distribution changes throughout the documents. Of course, the figures obtained could have been
caused by some exceptional cases and may not be representative of the source and domain as a
whole.

Using on-line web sites, it would be interesting to see how close the style figures from articles in
the two domains vary with the results obtained from the METER corpus. On-line newspapers
could provide a useful source for exploring different domains.

Only two domains have been analysed. It would be good to compare a range of domains to see
which appear more readable. Again, having examples from on-line newspapers could aide with
this work.

Two readability scores have been used in this experiment. It would be interesting to compare
these with other scores such as the Gunning FOG index, Fry Readability Graph, Powers-Sumner-
Kearl Formula, McLaughlin’s SMOG formula and the FORCAST formula. Details of these can be
found in (Johnson) and (Green).

I have mentioned the use of further linguistic techniques to further improve the readability scores
to measure grammatical form of the text (i.e. that it makes grammatical sense) and use
techniques such as sentence depth and sentence type to determine sentence complexity.

Finally, readability scores are all very well, but they need some form of evaluation. One of the
most popular methods is to apply the Cloze test (Green) to a text that will allow a human to judge
the scores that are obtained from the readability metrics.

4.6 Benefits of measuring readability

Obviously the experiments in this study have been interesting, but they could also benefit a media
source in the following ways:

- Knowing an average readability score could help control the articles that are written for
  that source. Akin to some form of quality control.
- Finding whether newspapers had used the PA as a source may be possible through
  readability scores (although perhaps unlikely due to journalesse).
- Media sources would probably like to know how readable their texts are and how they
  compare with other sources.
- For the PA, knowing that they produce readable material may be important. They can be
  more sure that users of their text will be able to easily read the news (although not
  necessarily understand it!). This could perhaps have an impact on charging for use of the
  source material.
**HIERARCHICAL CLUSTER ANALYSIS**

Dendrogram using Ward Method

<table>
<thead>
<tr>
<th>CASE</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Label</td>
<td>Num</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>express</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mail</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>times</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>guardian</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leviticus</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>telegraph</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>independent</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exodus</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbers</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genesis</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deut</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pa</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deuteronomy</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leviticus</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sun</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>exodus</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirror</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>numbers</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>genesis</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NOEL Edmonds yesterday announced he is quitting the BBC – because he no longer finds their shows "stimulating."

The ex-Noel's House Party host, who is walking out after 30 years, said he did not rule out a TV comeback if the right offer came along.

His spokesman said: "He is only 50 and has a lot of life left in him."

CORONATION Street temptress Tracy Shaw was committed to a mental hospital in her battle with anorexia, she revealed yesterday.

Stunning Tracy, 26, who plays crimper Maxine in the show, also considered suicide as her weight plummeted to just six stone.

In a frank TV interview she said: 'I had to go to hospital in the end. I was sectioned, and once I was in there I was in there. Being put into a psychiatric ward with people with long-term mental illness and who are shaking with the drugs they are taking, there's no way that you can go in and feel normal and be OK with yourself.'

This is the written account of Adam's line. When God created man, he made him in the likeness of God.

He created them male and female and blessed them. And when they were created, he called them "man."

When Adam had lived 130 years, he had a son in his own likeness, in his own image; and he named him Seth.

After Seth was born, Adam lived 800 years and had other sons and daughters.

Altogether, Adam lived 930 years, and then he died.

When Seth had lived 105 years, he became the father of Enosh.

And after he became the father of Enosh, Seth lived 807 years and had other sons and daughters.

Altogether, Seth lived 912 years, and then he died.

When Enosh had lived 90 years, he became the father of Kenan.

And after he became the father of Kenan, Enosh lived 815 years and had other sons and daughters.
5 Bibliography


1. APPENDIX A – Different versions of the Hamilton court case from the METER corpus

1.1 The PA version

FAYED BIGGEST CROOK IN TOWN, SAYS EX-SECURITY CHIEF<
By Jan Colley and Cathy Gordon, PA News<

Mohamed Al Fayed is "the biggest crook in town", his former head of security told the High Court today.<

Brian Dodd, a one-time SAS man and retired professional bodyguard, told Mr Justice Morland and the jury in the "cash for questions" libel action that he had been "tainted" with the same brush as former Tatton MP Neil Hamilton.<

Mr Dodd, 62, a holder of the Military Medal, worked for Mr Al Fayed from March 1984 until August 1986 when, according to the Harrods boss, he was sacked for gross misconduct.<

That was angrily denied today in a high-volume performance, which ended with the judge ordering Mr Dodd to leave the witness box and then abruptly rising from the bench.<

As the jury was swiftly ushered out of Court 13, a defiant Mr Dodd shouted out: "I apologise Mr Hamilton I do not think I have done your case any good."<

Just a few minutes earlier, Mr Dodd had told the court that he had sat on his bed last night and watched Mrs Hamilton's wife, Christine, on the news.<

``God knows how I felt ... I am tainted with the same brush as Neil Hamilton.''<

Adding that he was not sacked but that the system was changed, he went on: ``I have got to leave this court a tainted man by Mr Fayed, who's the biggest bloody crook in this town.''<

1.2 The SUN version:

MOHAMED Fayed was branded "the biggest crook in town" yesterday by his former security chief.

Ex-SAS hero Brian Dodd, 62, said the Harrods owner ordered him to carry out a string of "dirty tricks".

He claimed they included THREATENING Mr Fayed's tenants in posh London appartments and DESTROYING documents. Mr Dodd, awarded the Military Medal for his SAS service, was giving evidence for ex-Tory minister Neil Hamilton, 50, in the High Court cash-for-questions libel trial.

He said he used to be totally loyal to Mr Fayed and "would have done anything he asked."

Under questioning Mr Dodd claimed envelopes containing money were often collected from reception at the apartment block.

He said they were picked up by Tory and Labour MPs and members of the Lords.

Mr Dodd said Mr Fayed was furious when told that the names of the recipients were noted in a book by two security men.

Quizzed by Mr Fayed's QC, George Carman, Mr Dodd denied being sacked by the millionaire for gross misconduct.

He said: "I came here willingly and now I've got to leave this court tainted by Mr Fayed who is the biggest bloody crook in this town."

Mr Hamilton is suing Mr Fayed, 66, for saying he accepted cash and gifts for asking parliamentary questions.

The case resumes tomorrow with the final speeches.

The jury will retire to consider its verdict on Monday.
1.3 The Mirror version:

Mohamed Al Fayed's former bodyguard caused uproar at the Cash for Questions libel trial yesterday and was finally told to get out of the witness box by the judge.

In astonishing scenes, Brian Dodds frequently shouted, interrupted the judge and lawyers and refused to shut up.

Finally Mr Justice Morland ordered him to leave - and the judge abruptly left the court and the jury were led out.

Minutes earlier the former SAS hero and Para angrily denounced Al Fayed as "the biggest crook in town".

Mr Dodds said he had bullied terrified tenants into quitting flats owned by the Harrods boss, lied for him and burned sacks of documents about things "he wanted to hide".

And he said Al Fayed told him to kill one tenant and put him 'in a skip' if he had to.

'I would have gone to hell for Mr Fayed and done anything he asked. It's his belief he can do what he likes.'

Mr Dodds, 62, angrily denied Al Fayed's claim that he was sacked as head of security for gross misconduct.

'I am tainted with the same brush as Mr Hamilton - I came here willingly and I have to leave a tainted man by Mr Al Fayed, who is the biggest bloody crook in this town.'

Former Tory MP Neil Hamilton is suing Al Fayed for claiming he was paid thousands of pounds by the Harrods boss to ask Commons questions.

Mr Dodds said he made security officers at Al Fayed's office 'cover themselves' by keeping a book with the names and times when envelopes of cash were collected.

There were up to 15 people a week 'he was paying off'.

A note was made of everyone who arrived at the office - 'Labour MPs, Tory MPs and a whole load of lords'.

Quizzed by Al Fayed's QC about the way tenants were driven out, Mr Dodds shouted: 'I am stigmatised in the same way as Mr Hamilton' - and continued shouting after the lawyer sat down.

The case continues.

1.4 The Star version:

HARRODS tycoon Mohamed Al Fayed was branded 'the biggest bloody crook in town' by his former head of security yesterday.

High Court judge Mr Justice Morland ordered ex-SAS soldier and professional bodyguard Brian Dodds out of the witness box as he angrily screamed abuse about Mr Al Fayed in the fifth week of the 'cash for questions' libel trial.

The 62-year-old Military Medal holder worked for Mr Al Fayed from March 1984 until August 1986 when it is alleged he was sacked for 'gross misconduct'.

But Mr Dodd said he had been a loyal employee who would have 'gone to hell and back' for his boss.

He told how he burned FORTY sacks of documents which Mr Al Fayed wanted to hide.

And he claimed he was engaged in dirty tricks, keeping a book of people who used to collect envelopes of cash.

He apologised to Mr Hamilton, whose lawyers called him as a witness, for not helping his cause after he was ordered out of the witness box.

He said it was a nightmare working as a bodyguard for Mr Al Fayed, who frequently invited people he met in the store to stay at the Paris Ritz.

Mr Dodd said: 'I can understand Mr Hamilton's problem - one of the rooms in 1986 cost £3,000 a night without breakfast.

The average person could not afford to stay there.'

As George Carman QC for Mr Al Fayed, who was not in court, tried to stop him rambling he continued: 'If they said they couldn't afford to stay there he would say, 'You must come as my guest'.
Mr Hamilton seeks libel damages over the allegations on a Channel 4 Dispatches programme that he took cash for asking Parliamentary questions for Al Fayed.

The jury is expected to go out on Monday.

1.5 The Mail version

The case of Hamilton versus Al Fayed has seen various explosions since it began last month.

But nothing quite like what happened yesterday after Brian Dodd, Mohamed Al Fayed's bodyguard and head of security at Harrods, took the stand in Court 13.

The 62-year-old ex-SAS man, who was awarded the Military Medal by the Queen, was left high and dry in the witness box after the judge, Mr Justice Morland, walked out, fed up with his unsolicited outbursts.

As the jury was ushered out too, Mr Dodd called to the 50-year-old former Tory MP, on whose behalf he was there to give evidence: 'I apologise, Mr Hamilton. I don't think I have done your case any good.'

A few minutes earlier, he told the High Court how he watched TV reports of Mr Hamilton's wife Christine, who gave tearful evidence on Monday; 'God knows how I felt,' he said.

'I am tainted with the same brush as Neil Hamilton.'

Mr Dodd worked for Mr Al Fayed from March 1984 until August 1986 - a period in which Mr Hamilton is said to have been in the Harrods owner's pocket.

He angrily denied he was sacked for gross misconduct, saying: 'I have got to leave this court a tainted man by Mr Fayed, who's the biggest bloody crook in this town.'

The court is hearing Mr Hamilton's claims that Mr Al Fayed libelled him by falsely alleging he took cash and hospitality in return for asking questions in Parliament.

Mr Dodd ignored the convention of cross-examination by carrying on talking even when Mr Al Fayed's counsel, George Carman QC, had sat down.

Asked about Mr Hamilton and his wife's all-expenses-paid stay at Mr Al Fayed's Ritz Hotel in Paris, he said: 'I can understand Mr Hamilton's problem. One of the rooms in 1986 cost £3,000 a night without breakfast... the average person could not afford to stay there. If they said they couldn't afford to stay there... Mr Carman intervened: 'My Lord, will you stop this witness please?'

Mr Dodd carried on: 'So he would say: "You must come as my guest."'

Refusing to answer questions Yes or No unless he could add lengthy explanations, Mr Dodd agreed he made 'fictitious' entries in a book about comings and goings at Mr Al Fayed's offices in Park Lane, where Mr Hamilton is said to have received cash in brown envelopes.

He said staff kept a record of when envelopes of cash were collected and by whom so as to 'cover themselves'.

'There were up to 15 a week to people he (Mr Al Fayed) was paying off,' he said. 'There was also a book for anyone who went to see Mr Al Fayed - 'Labour MPs, Tory MPs and a whole load of Lords.'

Mr Dodd said Mr Al Fayed's secretaries shredded pages from the book. He did not mention this to anyone before yesterday because it was 'unimportant, petty'.

He recalled how Mr Al Fayed told him to throw a tenant of one of his Park Lane apartments 'on an effing skip' for alleged non-payment of rent.

'He said, "You kill him. You do what you effing like with him. Get him out of that flat".

He gave me the keys to go in and do it.'

He and another man went to the empty flat and left a note on a cereal packet, saying: 'We have missed you this time.'
Mr Dodd insisted: 'It was not criminal.
It was Mr Fayed's apartment - on his instructions that I went in there.
It was a damn sight less than what he wanted me to do to him.'

While working for Mr Al Fayed, he said, he would have gone 'to Hell and back' for Mr Al Fayed and would have done anything he asked.

With evidence now complete, the jury is expected to be sent out to consider its verdict next Monday.

1.6 The Express version

HARRODS tycoon Mohamed Al Fayed is 'the biggest crook in town,' his former security chief told the High Court yesterday.

Former-SAS soldier Brian Dodd said his boss 'paid off' about 15 people a week with cash in brown envelopes.

Ex-Tory minister Nell Hamilton denies he was on the secret payroll and pocketed thousands.

Mr Dodd, 62, claimed that he performed many 'dirty tricks' for Mr Al Fayed. One of them was breaking into the apartment of a troublesome tenant.

He said the Egyptian-born businessman told him to throw the tenant in a skip and kill him.

Mr Dodd, giving evidence on behalf of Mr Hamilton, became loud and heated as he said he felt he was now tarred with the same brush as the ex-MP.

He angrily denied being sacked for gross misconduct by Mr Al Fayed, adding: 'I have got to leave this court a tainted man by Mr Al Fayed who's the biggest bloody crook in this town.'

Mr Dodd ignored a request by Mr Justice Morland to leave the witness box after completing his evidence and continued with his tirade which ended only when the judge rose abruptly from the bench.

As the jury was ushered out of Court 13, defiant Mr Dodd shouted: 'I apologise, Mr Hamilton.

I don't think I've done your case any good.'

Mr Dodd, a holder of the Military Medal, said he would have gone 'to hell and back' for Al Fayed while working for him at his offices and apartments in London's Park Lane.

He recalled that in 1984 his boss asked him to get rid of a resident so his son Dodi's apartment could be extended.

Mr Dodd said Al Fayed told him: 'Get that bastard out of here.
Throw him in a f***ing skip if need be. I want him out.'

He told the jury: 'Al Fayed said 'You kill him.
Do what you effing like with him.
Get him out of that flat.'

He gave me the keys to go in.'

Mr Dodd said he and another man left a note in the flat, saying 'We missed you this time. We won't next.'

Questioned by Desmond Browne QC, for Mr Hamilton, he agreed he took part in dirty tricks at Al Fayed's bidding but denied any criminal activity.

He said he made doorman Philip Broomfield keep a visitors' book at Park Lane, logging the names and times of people collecting cash.

Mr Broomfield has told the court he handed Mr Hamilton two envelopes.

Mr Dodd said there was also a book for anyone who went to see Al Fayed - 'Labour MPs, Tory MPs and a load of Lords'.

Al Fayed walked around Harrods two or three times a week and frequently invited people to the Ritz in Paris.
Mr Dodd, to protests from George Carman QC, for Al Fayed, went on: 'I can understand Mr Hamilton's problem one room in 1986 cost £3,000 a night without breakfast.

If they said they couldn't afford to stay there he would say 'Come as my guest.'

Mr Dodd repeatedly shouted over Mr Carman and refused to answer the QC's questions 'yes' or 'no' unless he could add lengthy explanations.

He said Al Fayed's secretaries shredded pages from the book Mr Dodd had kept.

Mr Dodd worked for Al Fayed from 1984 until 1986 when he was allegedly sacked for gross misconduct.

Mr Dodd said he was not sacked the system just changed.

Mr Hamilton is suing Al Fayed over allegations on a 1997 Channel 4 Dispatches programme that he corruptly demanded and accepted cash payments, gift vouchers and a free holiday at the Paris Ritz in return for asking Parliamentary questions on behalf of Harrods.

Al Fayed denies libel and pleads justification.

Both QC5 will make their closing speeches tomorrow.

The jury is expected to consider its verdict on Monday.

1.7 The Times version

MOHAMED AL FAYED is the 'biggest crook in town', the Harrods owner's former head of security told the High Court yesterday.

During a highly charged day in the 'cash for questions' libel trial, Brian Dodd, a former SAS soldier, said that he had been 'tainted' with the same brush as Neil Hamilton, the former Tory minister who is suing Mr Al Fayed.

As he finished giving evidence, Mr Dodd, 62, became so angry that Mr Justice Morland ordered him to leave the witness box.

As the jury was ushered out of court, Mr Dodd shouted: 'I apologise, Mr Hamilton. I do not think I have done your case any good.'

Earlier the court was told that Mr Dodd worked for Mr Al Fayed from March 1984 until August 1986 when, according to the Harrods chairman, he was sacked for gross misconduct, a charge Mr Dodd denied.

He told the court that on Monday he had watched Christine Hamilton, the former MP's wife who had given evidence that day, on the evening news.

'God knows how I felt... I am tainted with the same brush as Neil HaMilton.'

He added: 'I have got to leave this court tainted by Mr Fayed, who's the biggest bloody crook in town.'

Mr Hamilton is suing Mr Al Fayed over allegations that he received cash and gifts for asking questions in the Commons on Harrods's behalf.

Mr Al Fayed denies libel and pleads justification.

Mr Dodd, who said he 'would have gone to hell and back' for Mr Al Fayed while he worked for him, was in charge of security at the Harrods chairman's Park Lane offices and apartments.

Mr Dodd said that he made the staff there record the names and times when envelopes of cash were collected.

'There were up to 15 a week to people he [Mr Al Fayed] was paying off,' he said.

There was also another book for anyone who went to see Mr Al Fayed - 'Labour MPs, Tory MPs and a whole lot of Lords'.

Mr Dodd said that Mr Al Fayed's secretaries shredded pages from the book he had kept.

'It's my knowledge, because each day the page was ripped out, it was not allowed to be seen,' he said.

'In 1986, I burnt 40 sacks of documents from Mr Fayed's office that he wanted to hide.'
Mr Dodd, who admitted under questioning from George Carman, QC, for Mr Al Fayed, to three speeding tickets and having been involved with the police over a domestic incident with his brother, denied that he was guilty of 'criminal' behaviour while working at Harrods.

Mr Carman asked him about breaking into a flat at the Park Lane apartment block.

Mr Dodd answered that Mr Al Fayed had told him to throw one of the Park Lane tenants 'on an effing skip'.

As the evidence concluded, the jury was sent away until Thursday, when counsels will make their closing speeches.

1.8 The Telegraph version

MOHAMED Fayed's former SAS bodyguard called him the 'biggest bloody crook in town' yesterday as the cash for questions libel trial reached its closing stages.

Brian Dodd, 62, who worked for the Harrods owner in the Eighties, told the High Court how he spent four hours burning 40 sacks of documents that Mr Fayed 'wanted to hide'.

He also admitted to breaking into flats at Mr Fayed's Park Lane complex on his instructions in order to intimidate and evict unwanted tenants.

Mr Dodd, holder of the Military Medal for action in Aden with 22 SAS, said there was a time when he would have 'gone to Hell and back' for Mr Fayed.

But he had been angered by his former employer's claim during the trial that he had sacked Mr Dodd for gross misconduct.

This was not true, he said.

'I am tainted with the same brush as Neil Hamilton. I was not sacked for gross misconduct.

I have got to leave this court a tainted man by Mr Fayed who is the biggest bloody crook in this town.'

Mr Dodd clashed repeatedly with George Carman, QC, for Mr Fayed, as he accused the Harrods owners of 'ordering me to do many dirty tricks'.

Eventually the judge told him to leave the witness box and abruptly adjourned the sitting when Mr Dodd insisted upon having a final say.

As the jury was ushered out, Mr Dodd shouted: 'I apologise, Mr Hamilton, I do not think I have done your case any good.'

Mr Hamilton, the former MP for Tatton, is seeking damages over Mr Fayed's allegations on Channel 4's Despatches in 1997 that he had corruptly demanded and accepted cash payments, gift vouchers and a free holiday at the Paris Ritz in return for asking parliamentary questions on behalf of Harrods.

Mr Dodd was head of security at Mr Fayed's Park Lane offices and apartments, where staff included Philip Bromfield, a doorman who has alleged that he handed two envelopes to Mr Hamilton.

Mr Dodd said he made Mr Bromfield and others keep a book in which to enter the names and times when envelopes - including some containing cash - were collected, so as to 'cover themselves'.

'We knew there was money in the envelopes. There were up to 15 a week to people he was paying off,' Mr Dodd said.

But when Mr Fayed's brother Ali found out about the book 'he went mad' and told him there should be no record of visitors.

It was agreed that the names of anyone who went upstairs would go in the book, and in the evening the book itself would go up.

The page for each day would be ripped out and destroyed by Mr Fayed's secretaries, Alison Bozek and Iris Bond - both of whom have given evidence that they saw their employer put cash in envelopes for Mr Hamilton.

'Each day the page was ripped out, it was not allowed to be seen,' he said.

'In 1986, I burned 40 sacks of documents from Mr Fayed's office that he wanted to hide - 40.'
It took me four hours, on a bonfire.'

Mr Dodd said Mr Fayed was a ‘nightmare’ for a bodyguard because he regularly walked around his Knightsbridge store, stopping to talk to friends and acquaintances.

Frequently, he would invite them to go and stay in the Ritz if they were in Paris.

"I can understand Mr Hamilton’s problem - one of the rooms in 1986 cost £3,000 a night without breakfast. The average person could not afford to stay there."

If they said they couldn’t afford to stay there, he would say, ‘You must come as my guest’.

Under heated cross-examination by Mr Carman, Mr Dodd acknowledged that he had written to Mr Fayed in 1995 asking for a job.

His letter read: ‘I would find no difficulty in giving you and your family my complete loyalty as I always have.’

He said he still admired Mr Fayed and missed the ‘good years’ working for him.

Mr Dodd said the letter had been written when he could not find security work where he lived in North Wales and his praise of Mr Fayed may have ‘been a bit of bullshit’.

Mr Dodd said that ‘at Mr Fayed’s bidding’ he had engaged in ‘many dirty tricks’.

He had once ordered his number two to go into the Park Lane offices of Mr Fayed’s former brother-in-law, Adnan Khashoggi.

Mr Fayed had also told him to evict one of the Park Lane tenants.

He said ‘throw him on an effing skip if needs be’, claimed Mr Dodd.

“He said ‘you kill him, you do what you effing like with him, get him out of that flat.”

He and another man went into the flat and took away documents.

They wrote on a cornflakes box: ‘We have missed you this time.

We won’t miss you next time.’

The jury was sent away until tomorrow, when closing speeches will be heard, and is expected to consider its verdict on Monday.

1.9 The Guardian version

Mohamed Al Fayed’s former security chief branded his ex-employer the ‘biggest crook in town’ yesterday, at the libel action brought by the former MP, Neil Hamilton.

Brian Dodd told how the Harrods owner ‘paid off’ up to 15 people a week, and deployed strong-arm tactics against those he wanted to evict from his central London apartment block.

Amid extraordinary scenes at the high court in London, Mr Dodd repeatedly spoke over barristers’ questions, claiming to have been insulted by Mr Fayed’s assertion in evidence that he had been sacked for gross misconduct.

‘I am tainted with the same brush as Neil Hamilton.

I have got to leave this court a tainted man by Mr Fayed, who’s the biggest bloody crook in this town.’

His comments prompted angry exchanges with George Carman QC, for Mr Fayed.

When the barrister intervened, Mr Dodd replied: ‘Hang on, your honour, or sir, or whatever I call you, Mr Carman.’

Later, he told the QC: ‘Don’t throw everything at me, Mr Carman ... you assume lots of things, I’ve heard you.’

The performance ended when Mr Justice Morland - who had joined the rest of the court in laughing at the outbursts - ordered the witness to leave, before abruptly adjourning the hearing.
As the jury was led out of court 13, Mr Dodd shouted across to an ashen-faced Mr Hamilton: ‘I apologise, I do not think I have done your case any good.’

Mr Hamilton is suing Mr Fayed over allegations in a 1997 Channel 4 documentary that he had corruptly demanded and accepted cash and hospitality in return for performing parliamentary services for Harrods.

Mr Fayed, who was not in court yesterday, denies libel and maintains his allegations were true.

Mr Dodd, 62, a former SAS officer, worked for Mr Fayed from March, 1984, until August, 1986, when, he claims, he left of his own accord.

As the security manager of Mr Fayed's Park Lane apartments, he ordered his staff to keep a book in which names and times were recorded of people who arrived to pick up envelopes and packages.

Knowing some of the envelopes contained cash, he told the jury, he wanted his staff to ‘cover themselves.’

‘There were up to 15 a week, to people Fayed was paying off.

There was also a book for anyone who went to see Mr Al Fayed - Labour MPs, Tory MPs and a whole load of lords.’

Mr Fayed's brother, Ali, ‘went mad' when he found out about the book.

'I said that I needed to know who is in the building in case there is a fire.

I said you can call them Donald Duck One and Donald Duck Two, but I will have the names in my book or I will not do my job.'

It was agreed that names would be recorded, and the book 'taken upstairs' at the end of the day.

Mr Dodd alleged that Mr Fayed's secretaries, Alison Bozek and Iris Bond - both of whom earlier told how they saw Mr Fayed stuff envelopes with cash for Mr Hamilton - shredded pages from the book.

'Each day the page was ripped out, it was not allowed to be seen.’

Those were not the only records destroyed on the orders of Mr Fayed.

‘In 1986, I burned 40 sacks of documents from Mr Fayed's office that he wanted to hide - 40, it took me four hours, on a bonfire’

Mr Dodd, who fiercely denied that he was guilty of criminal behaviour when he worked for Mr Fayed, told how the-Harrods owner had wanted one of the Park Lane tenants removed from his flat. 'Mr Fayed told me to throw his body on an effing skip.

He said: “You kill him, you do what you effing like with him, get him out of that flat”’

Under questioning from Mr Carman, he agreed that he went into the flat with another man, and filled four suitcases with documents which they later burned.

They left the tenant a note on a cornflakes box, which read: ‘We've missed you this time. We won't miss you next time.’

Mr Dodd: 'It was a damn sight less than what he [Fayed] wanted me to do to him.'

Mr Dodd accepted he had kept fictitious records about movements at Park Lane.

On one occasion, he recorded that two American tenants, whom he suspected of being prostitutes, brought five or six men home at night ‘as if they were running a brothel’. Admitting the entries were false, he said: ‘There were many fictitious books at Park Lane.’

The case was adjourned until tomorrow, and the jury will retire to consider its verdict on Monday.

1.10 The Independent version

THE LIBEL action between Neil Hamilton and Mohamed Al Fayed was halted amid uproar at the High Court yesterday after the multi-millionaire owner of Harrods was branded ‘the biggest crook in town’ by his former head of security.

The jury was hastily ushered out and the judge, Mr Justice Morland, also left after failing to prevent Brian Dodd, a former paratrooper and Special Air Service soldier, from shouting from the witness box, where he had been called to give evidence on behalf of the former MP for Tatton.
Mr Dodd, in a dark grey suit and regimental tie, plaintively cried out in the emptying courtroom: 'I'm sorry Mr Hamilton, I don't think I've done your cause much good.'

Earlier, he said that after watching Mr Hamilton's wife, Christine, on television news he felt '... tainted by the same brush as Neil Hamilton'.

He went on: 'I have got to leave this court a tainted man by Mr Fayed, who's the biggest bloody crook in this town.'

The latest day of evidence lived up to all the colourful theatre that had gone before.

Mr Dodd, sitting ramrod straight, repeatedly interrupted, not just Mr Fayed's QC, George Carman, but Desmond Browne, the counsel for Mr Hamilton.

Mr Hamilton, a former minister for corporate affairs under John Major, is suing Mr Fayed over allegations that he accepted cash, gifts and holidays in return for asking questions in the Commons favourable to the owner of Harrods.

Mr Dodd, the holder of the Military Medal, was in charge of security at Mr Fayed's offices and apartments in Park Lane, central London.

The staff included Philip Bromfield, a doorman, who claimed that he had handed over two envelopes to Mr Hamilton.

Mr Dodd told the jury that he had made Mr Bromfield and others keep a book in which to enter the names and times when these envelopes were collected, so as to 'cover themselves'.

He said: 'There were up to 15 people a week who he [Mr Fayed] was paying off.'

There was also a separate book for anyone who went to see Mr Fayed and this included the names of 'Labour MPs, Tory MPs and a whole load of lords'.

Mr Dodd claimed that Alison Bozek and Iris Bond, Mr Fayed's secretaries who appeared as his witnesses, had shredded pages from the book the security staff kept.

The jury is expected to retire on Monday at the end of legal arguments, closing speeches and summing up by the judge, to consider its verdict.