Problem
Author Masking is task of rewriting the document to obfuscate the stylometric identity of original author. Given a set of documents by the same author, paraphrase the designated one so that the author cannot be verified anymore.

Evaluation
The obfuscation software would be called,
- Safe, if a forensic analysis does not reveal the original author of its obfuscated texts
- Sound, if its obfuscated texts are textually entailed with their originals
- Sensible, if its obfuscated texts are inconspicuous to human evaluators

Data
Datasets used for Author verification task at PAN 2013 to PAN 2015
- PAN13: English computer science textbooks
- PAN14 EE: English essays written by students with English as a second language
- PAN14 EN: English horror fiction novels
- PAN15: Dialogs from English plays

Approach
• Round trip translation to obfuscate the document of an author
  The idea is to introduce minor corruption (change in vocabulary, change in sentence lengths, paraphrasing, etc) while translating from one language to another.
  - 100,000 randomly selected sentences from the Europarl corpus used for training translation systems.

Key Takeaways
- In its current form this method is not useful. It can fool automatic authorship attribution systems, but so can some random junk text.
- Is it worth continuing in this direction? The results were ‘not so bad’ on training data.

Limitations
• Junk Text (until now)
  - Rate limit on use of online services like Google, Bing & Yandex.
  - Availability of generic corpus for training translation systems as compared to domain specific corpora
  - Higher computational power to handle large models

Advantages
• A text generative technique
• Length of sentences can be controlled
• Vocabulary can be controlled
• A lot of focus on translation as a tool for paraphrasing, text simplification, etc.

How can we make this usable?
• Use a different and a larger corpus which has a greater and a robust vocabulary (OpenSubtitles, paraphrase.org ?)
• Make the sentence length penalty parameter a function of the author’s stylometry rather than target language
• How much change is sufficient? Ignore low confidence translations?
• Use the word usage trends to manipulate the translations. For example, Replacing a few words that are used in recent times by those that were popular in 18th century (Genre dependent)

Results
The following were the evaluation results for safety aspect:

<table>
<thead>
<tr>
<th>Participant</th>
<th>PAN 2013</th>
<th>PAN 2014 EE</th>
<th>PAN 2014 EN</th>
<th>PAN 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mihaylova et al. [31]</td>
<td>-0.10</td>
<td>-0.13</td>
<td>-0.16</td>
<td>-0.11</td>
</tr>
<tr>
<td>Keswani et al. [20]</td>
<td>-0.09</td>
<td>-0.11</td>
<td>-0.12</td>
<td>-0.06</td>
</tr>
<tr>
<td>Mamonzadeh et al. [18]</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.04</td>
</tr>
</tbody>
</table>

In terms of the sound and the sensible aspect, our system performed the worst out of all the systems submitted.

References