A DESCRIPTIVE ANALYSIS OF CRIME NOUNS IN UKSCC, A LEGAL CORPUS

Prof. María José Marín
Dr. Camino Rea
Universidad de Murcia, Spain
1. Introduction
2. Specialised corpora and the EAP class
3. UKSCC: description
4. Methodology and data analysis:
   4.1. Methodology
   4.2. Analysis of data
5. Conclusions
6. References
1. Introduction

• **Motivation:** Introduction of legal English as part of the Law degree curriculum after the Bologna Reform.

• **Main features of EAP course at law degree:**
  - Both **English for General Academic Purposes (EGAP)** and **English for Specialised Academic Purposes (ESAP)** are taught.
  - **Aims/objectives:** enabling students as **efficient communicators** in **formal** settings and as autonomous learners.
  - Possibility of dealing with **English-speaking clients** as legal practitioners.
  - **The UK legal systems** and the terms of the art must be known by Law students.
  - Students will progress from **CEFR (Common European Framework of Reference for Languages) B1** to **B2** level.
As established by the Council of Europe (2001) for B2 level: B2 English users “can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation” *

Nation and Hwang (1995) recommend that students moving onto the specialised field must focus on its lexicon as a way of improving text coverage.

This paper focuses on the most common crime nouns in legal English as an example of legal terms which crystallise technical knowledge and present different levels of specialisation, appearing in most legal English textbooks at this level (Fernández, 1994; Rice, 2007; Krois-Linder and Firth, 2008; Frost, 2009; Callanan, 2010 and Orts, 2010) Establishing the level of specialisation of these items can help to grade the materials based on them according to the students’ needs.

2. Specialised corpora and the EAP class

- Specialised corpora are a useful source of information for the elaboration of teaching materials:

  - McEnery and Wilson (1996): “such corpora can be used to provide many kinds of domain-specific material for language learning”

  - Schmitt (2002) highlights their use as an interesting resource for the ESL class as well as a tool to assess vocabulary acquisition.

  - Gilquin and Granger (2010) affirm that they provide “a large number of authentic instances of a particular linguistic item”.

- However, the scarceness of legal corpora is manifest, finding just few of them, in most cases not satisfying our needs (see Marín and Rea (2011) for a comprehensive review). So UKSCC was designed to bridge this gap.
3. UKSCC: description

-The United Kingdom Supreme Court (UKSCC) was created with the aim of bridging the methodological gap existing in the area.

-Monolingual, synchronic 2.6 million-word corpus of 193 judicial decisions issued by the Supreme Court of the United Kingdom between 2008 and 2010.

-Judicial decisions were selected due to the pivotal role they play in common law legal systems acting as the major source of information for legal practitioners.

-The Supreme Court of the United Kingdom was chosen as a source to obtain the texts owing to its position at the top of the judicial pyramid and the fact that its decisions always set precedent being most often cited by legal practitioners.

-It also deals with all branches of law providing rich and varied texts as far as their lexicon is concerned.
4. Methodology and data analysis

4.1. Methodology

ATR methods traditionally focus on multi-word terms neglecting single-word units to a certain extent (Maynard and Ananiadou 2000; Cabré et al. 2001; Lemay et al. 2005; Chung 2003; Almela, 2008).

Chung’s (2003) method presents a straightforward though effective technique to calculate a word’s level of specialisation by comparing a specialised/study corpus (SC) against a general one, the reference corpus (RC):

\[
\text{Ratio} = \frac{\text{Rel F (SC)}}{\text{Rel F (RC)}} \text{ (freq normed } x \text{ 1,000 words)}
\]

Words displaying >50 ratio value and those not in the RC are considered to be terms.

86% precision on average.
4.2. Data analysis
-Chung’s technique is applied on UKSCC by comparison with LACELL, a general English corpus of 21 million words.

-Elaboration of “to-be-taught” list of crime nouns graded to suit students’ needs. Less specialised items should be taught first, ranked according to different parameters following Nation’s (2001) advice on the introduction of specialised vocabulary.

-Frequency by itself cannot account for a word’s relevance or representativeness, indicated by their keyness value. According to Scott (2008), “a word is considered key if it is unusually frequent (or unusually infrequent) in comparison with what one would expect on the basis of the larger word-lists”.

Only two types, fraud and conspiracy are above the average frequency for UKSCC, 353.16, the rest display relatively low values.
4.2. Data analysis

- **Text range** provides information on the distribution of a word within a corpus. Choosing items which display high text range counts will contribute to the better understanding of specialised texts due to text coverage. Crime nouns appear in 10.81 texts on average against 32.31, the mean value for the whole corpus. Their text coverage is relatively low, 0.0010% of the running words in all the corpus texts against 0.24% displayed by the first 2,000 specialised terms in UKSCC. (Short sample list of nouns)

Their keyness, however, also pointing at the representativeness of these types, is considerably high: 185.29 v. 116.08 for UKSCC.

In spite of the lower frequency and distribution counts, they appear to be reasonably representative of the genre.
<table>
<thead>
<tr>
<th>TYPE</th>
<th>FREQ UKSCC</th>
<th>FREQ LACELL</th>
<th>REL FREQ UKSCC</th>
<th>REL FREQ LACELL</th>
<th>TEXT RANGE</th>
<th>RATIO</th>
<th>KEYNESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conspiracy</td>
<td>357</td>
<td>12</td>
<td>0.13579747</td>
<td>0.00057098</td>
<td>22</td>
<td>237.832335</td>
<td>838,205322</td>
</tr>
<tr>
<td>Trafficking</td>
<td>280</td>
<td>19</td>
<td>0.10650782</td>
<td>0.00090405</td>
<td>14</td>
<td>117.811683</td>
<td>870,371216</td>
</tr>
<tr>
<td>Bribery</td>
<td>20</td>
<td>2</td>
<td>0.0076077</td>
<td>9.5163E-05</td>
<td>3</td>
<td>79.9436421</td>
<td>26,861702</td>
</tr>
<tr>
<td>Perjury</td>
<td>15</td>
<td>2</td>
<td>0.00570578</td>
<td>9.5163E-05</td>
<td>6</td>
<td>59.9577316</td>
<td>565.04</td>
</tr>
<tr>
<td>Fraud</td>
<td>447</td>
<td>60</td>
<td>0.17003212</td>
<td>0.0028549</td>
<td>28</td>
<td>59.5580134</td>
<td>1006,22614</td>
</tr>
<tr>
<td>Nuisance</td>
<td>57</td>
<td>8</td>
<td>0.02168195</td>
<td>0.00038065</td>
<td>13</td>
<td>56.959845</td>
<td>63,1570702</td>
</tr>
<tr>
<td>Intimidation</td>
<td>56</td>
<td>8</td>
<td>0.02130156</td>
<td>0.00038065</td>
<td>12</td>
<td>55.9605495</td>
<td>101,689247</td>
</tr>
<tr>
<td>Manslaughter</td>
<td>64</td>
<td>10</td>
<td>0.02434464</td>
<td>0.00047582</td>
<td>8</td>
<td>51.163931</td>
<td>54,9405785</td>
</tr>
<tr>
<td>Arson</td>
<td>4</td>
<td>1</td>
<td>0.00152154</td>
<td>4.7582E-05</td>
<td>3</td>
<td>31.9774569</td>
<td>356.98</td>
</tr>
<tr>
<td>Threats</td>
<td>86</td>
<td>34</td>
<td>0.03271312</td>
<td>0.00161778</td>
<td>22</td>
<td>20.2210389</td>
<td>40,4039764</td>
</tr>
<tr>
<td>Battery</td>
<td>165</td>
<td>80</td>
<td>0.06276354</td>
<td>0.00380653</td>
<td>4</td>
<td>16.483762</td>
<td>204,735733</td>
</tr>
<tr>
<td>Abduction</td>
<td>10</td>
<td>5</td>
<td>0.00380385</td>
<td>0.00023791</td>
<td>3</td>
<td>15.9887284</td>
<td>123.98</td>
</tr>
<tr>
<td>Forgery</td>
<td>17</td>
<td>24</td>
<td>0.00646655</td>
<td>0.00114196</td>
<td>3</td>
<td>5.66267465</td>
<td>24,7046776</td>
</tr>
<tr>
<td>Torture</td>
<td>118</td>
<td>200</td>
<td>0.04488544</td>
<td>0.00951633</td>
<td>16</td>
<td>4.71667489</td>
<td>146,08847</td>
</tr>
<tr>
<td>Larceny</td>
<td>2</td>
<td>6</td>
<td>0.00076077</td>
<td>0.00028549</td>
<td>1</td>
<td>2.66478807</td>
<td>567.83</td>
</tr>
<tr>
<td>Mayhem</td>
<td>1</td>
<td>3</td>
<td>0.00038039</td>
<td>0.00014274</td>
<td>1</td>
<td>2.66478807</td>
<td>899.23</td>
</tr>
<tr>
<td>Assault</td>
<td>161</td>
<td>513</td>
<td>0.061242</td>
<td>0.02440939</td>
<td>30</td>
<td>2.50895251</td>
<td>87,1345749</td>
</tr>
<tr>
<td>Robbery</td>
<td>30</td>
<td>115</td>
<td>0.01141155</td>
<td>0.00547189</td>
<td>14</td>
<td>2.08548632</td>
<td>11.06</td>
</tr>
<tr>
<td>Theft</td>
<td>53</td>
<td>225</td>
<td>0.02016041</td>
<td>0.01070587</td>
<td>16</td>
<td>1.8831169</td>
<td>15.02</td>
</tr>
</tbody>
</table>
-Applying Chung’s method to crime nouns, it appears that 24.24% of them are technical as they either display a >50 ratio value or simply do not occur in the RC while 75.76% are non-terms or general vocabulary.

<table>
<thead>
<tr>
<th>General in West’s GSL list</th>
<th>Non-terms</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violence</td>
<td>Arson</td>
<td>Conspiracy</td>
</tr>
<tr>
<td>Murder</td>
<td>Threats</td>
<td>Trafficking</td>
</tr>
<tr>
<td>Robbery</td>
<td>Battery</td>
<td>Bribery</td>
</tr>
<tr>
<td></td>
<td>Abduction</td>
<td>Perjury</td>
</tr>
<tr>
<td></td>
<td>Torture</td>
<td>Fraud</td>
</tr>
<tr>
<td></td>
<td>Assault</td>
<td>Nuisance</td>
</tr>
</tbody>
</table>
Non-terms appear specially interesting for the EAP lecturer as they are shared by both the specialised and general fields (semi-technical vocabulary) and specialised corpora can offer authentic contextual information helping to disambiguate meaning:

**Battery:**
EXAMPLES FROM REFERENCE CORPUS, LACELL
-.... all rechargeable battery packs should be returned to a battery recycling ...
-....in a battery of public appearances to plug the new CD ...

EXAMPLES FROM UKSCC:
-.... the Ashleys say that they should be entitled to seek to establish their claim in battery ...
-....from the age of 17 for, amongst other things, battery, common assault, ...
5. Conclusions

- Results evidence the relevance of crime nouns in legal English.

- The different parameters studied might help us grade vocabulary to elaborate to-be-taught lists ranked according to the specialist’s criterion.

- Focusing on semi-technical items might be of interest for the EAP class: use of general v. specialised contexts to disambiguate meaning.

- Further research: classification of all the terms of the art in our corpus and use of these inventories for the elaboration of didactic materials.
6. References


