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

Prof. María José Marín
Dr. Camino Rea
Universidad de Murcia, Spain

UNIVERSIDAD DE
MURCIA

PAPER STRUCTURE

Prof. María José Marín Dr. Camino Rea. Universidad de Murcia

The 12th International
Bilkent University School of English Language Conference
EAP in UNIVERSITY SETTINGS - EAP Conference Series No.2

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.

-**T**he **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.

The 12th International
Bilkent University School of English Language Conference
EAP in UNIVERSITY SETTINGS - EAP Conference Series No.2

İhsan Doğramacı
1984
Bilkent University

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.
* http://www.coe.int/T/DG4/Linguistic/CADRE_EN.asp

The 12th International
Bilkent University School of English Language Conference
EAP in UNIVERSITY SETTINGS - EAP Conference Series No.2

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.

The 12ᵗʰ International
Bilkent University School of English Language Conference
EAP in UNIVERSITY SETTINGS - EAP Conference Series No.2

İhsan Doğramacı
Bilkent
1984
Bilkent University

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) :

Ratio = Rel F (SC)/Rel F (RC) (freq normed x 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 specialsied 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.

TYPE	FREQ UKSCC	FREQ LACELL	REL FREQ UKSCC	REL FREQ LACELL	TEXT RANGE	RATIO	KEYNESS
Conspiracy	357	12	0,13579747	0,00057098	22	237,832335	838,205322
Trafficking	280	19	0,10650782	0,00090405	14	117,811683	870,371216
Bribery	20	2	0,0076077	9,5163E-05	3	79,9436421	26,861702
Perjury	15	2	0,00570578	9,5163E-05	6	59,9577316	565,04
Fraud	447	60	0,17003212	0,0028549	28	59,5580134	1006,22614
Nuisance	57	8	0,02168195	0,00038065	13	56,959845	63,1570702
Intimidation	56	8	0,02130156	0,00038065	12	55,9605495	101,689247
Manslaughter	64	10	0,02434464	0,00047582	8	51,163931	54,9405785
Arson	4	1	0,00152154	4,7582E-05	3	31,9774569	356,98
Threats	86	34	0,03271312	0,00161778	22	20,2210389	40,4039764
Battery	165	80	0,06276354	0,00380653	4	16,4883762	204,735733
Abduction	10	5	0,00380385	0,00023791	3	15,9887284	123,98
Forgery	17	24	0,00646655	0,00114196	3	5,66267465	24,7046776
Torture	118	200	0,04488544	0,00951633	16	4,71667489	146,08847
Larceny	2	6	0,00076077	0,00028549	1	2,66478807	567,83
Mayhem	1	3	0,00038039	0,00014274	1	2,66478807	899,23
Assault	161	513	0,061242	0,02440939	30	2,50895251	87,1345749
Robbery	30	115	0,01141155	0,00547189	14	2,08548632	11,06
Theft	53	225	0,02016041	0,01070587	16	1,8831169	15,02

Prof. María José Marín Dr. Camino Rea. Universidad de Murcia

-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.

General in West's *GSL* list	Non-terms	Terms
Violence	Arson	Conspiracy
Murder	Threats	Trafficking
Robbery	Battery	Bribery
	Abduction	Perjury
	Torture	Fraud
	Assault	Nuisance

Prof. María José Marín Dr. Camino Rea. Universidad de Murcia

-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 disambiaguate 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

-**Almela, A. (2008).** *Evaluating Multiword Automatic Term Recognition Techniques on a Veterinary Medicine Corpus.* MA Thesis. Universidad de Murcia.

-**Cabré, M. T., Estopà, R., Vivaldi, J.** (**2001**). 'Automatic term detection: a review of current systems' In D. Bourigault, C. Jacquemin, M.C. L'Homme (eds.) *Recent Advances in Computational Terminology.* Amsterdam: John Benjamins, Natural Language Processing, 2, 53-87.

-**Chung, T. M. (2003)** 'A corpus comparison approach for terminology extraction'' *Terminology* 9(2), 221-246.

-**Gilquin, G., Granger, S. (2010)**. 'How can data-driven learning be used in language teaching? In O'Keefe, A. and McCarthy, M. (eds.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge.

-**Hwang, K. and Nation, P. (1995)** Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23, 1: 35-41.

-**Lemay, Ch., L'Homme, M.C., Drouin, P. (2005).** "Two methods for extracting "specific" single-word terms form specialised corpora". *International Journal of Corpus Linguistics.* 10:2, 227-255.

-**Marín, M.J., Rea, C. (2011).** "Design and compilation of a legal English corpus based on UK law reports: the process of making decisions" in Carrió Pastor, M. L. y Candel Mora, M. A. (2011). *Las tecnologías de la información y las comunicaciones: Presente y futuro en el análisis de córpora*. Actas del III Congreso Internacional de Lingüística de Corpus. Valencia: Universitat Politècnica de València. 101-110.

-**Maynard, D. and Ananiadou, S. (2000).** 'TRUCKS: A model for automatic multi-word term recognition'. *Journal of Natural Language Processing* 8(1), 101–125.

-**McEnery, T. and Wilson, A. (1996)** *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

-**Nation, P. (2001)** *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

-**Schmitt, N. (2002).** 'Using corpora to teach and assess vocabulary'. Tan, M. (ed.). *Corpus Studies in Language Education*. IELE Press.